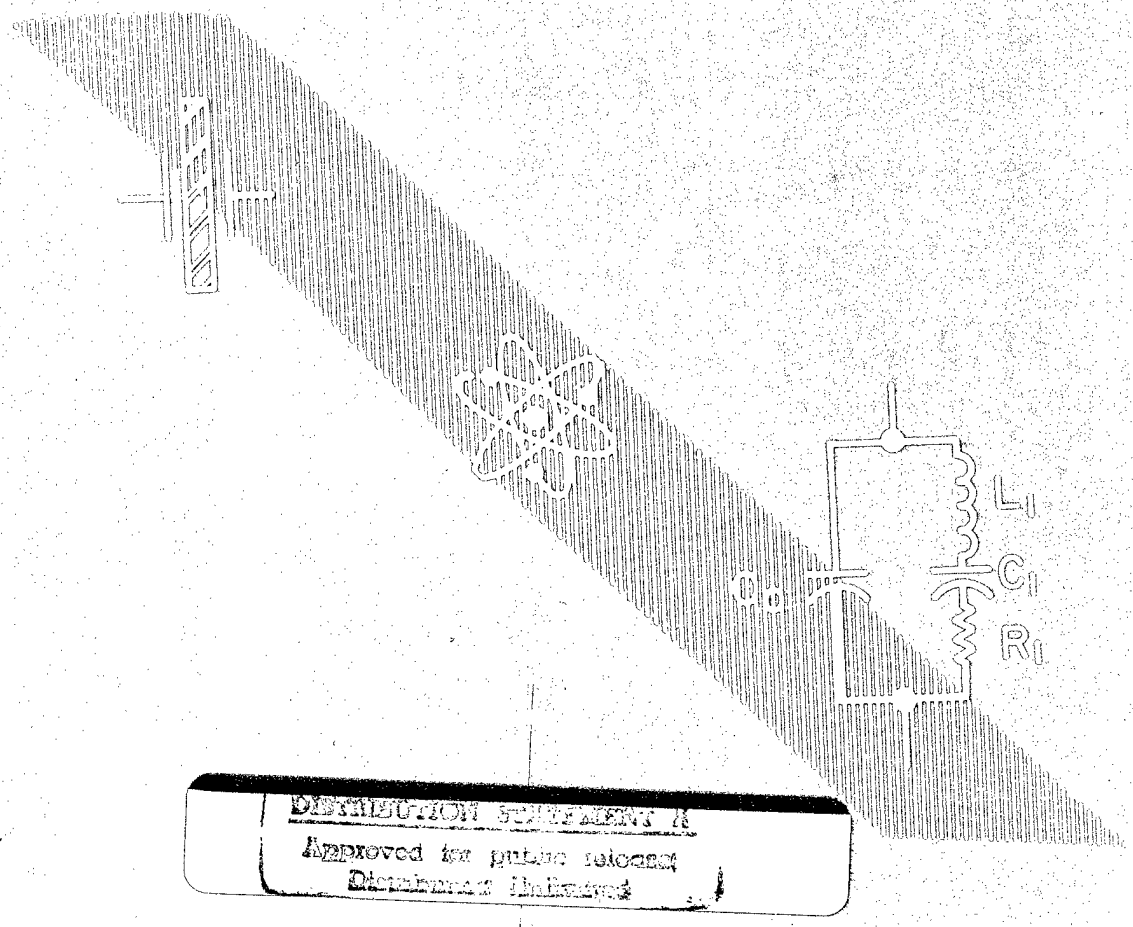
Proceedings of the

1993 IEEE International Frequency Control Symposium



DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

2-4 June 1993

93CH3244-1
87-654207

Proceedings of the

1993 IEEE International
Frequency Control Symposium

(The 47th Annual Symposium)

Sponsored by
The Institute of Electrical and
Electronics Engineers, Inc.

Ultrasonics, Ferroelectrics, and
Frequency Control Society

with the participation of personnel of the
Army Research Laboratory, Ft. Monmouth, NJ, U.S.A.

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

IEEE Catalog No. 93CH3244-1

Library of Congress No. 87-654207

19970623 244

2-4 June 1993

Salt Lake City Marriott Hotel
Salt Lake City, Utah, U.S.A.

Copyright and Reprint Permission: Abstracting is permitted with the credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 27 Congress Street, Salem, MA 01970. Instructors are permitted to photocopy isolated articles for non-commercial classroom use without fee. For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. All rights reserved. Copyright © 1993 by the Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Number: 93CH3244-1
ISBN 0-7803-0905-7 (Softbound)
ISBN 0-7803-0906-5 (Casebound)
ISBN 0-7803-0907-3 (Microfiche)
Library of Congress Number: 87-645207

1993 IEEE International Frequency Control Symposium

SYMPOSIUM CHAIRMEN

General Chairman.....Gary R. Johnson

Technical Program Chairman.....Jack Kusters

Local Arrangements Chairman.....Errol P. EerNisse

Finance Chairman.....Thomas E. Parker

Tutorials Chairman.....David W. Allan

Editorial and Awards Chairman.....John R. Vig

Travel Assistance Chairman.....Raymond L. Filler

TECHNICAL PROGRAM COMMITTEE VICE CHAIRMEN

Materials & Resonators.....Dan Stevens
Oscillators & Circuits.....Michael Driscoll
Atomic Standards & Timing.....Fred Walls

TECHNICAL PROGRAM COMMITTEE

David Allan, NIST
Ivan Avramov, Inst. Solid State Physics, Bulgaria
Arthur Ballato, U.S. Army Research Laboratory
Martin Bloch, Frequency Electronics, Inc.
Jean-Simon Boulanger, NRC, Canada
Jan Brown
Leonard Cutler, Hewlett-Packard Company
Michael Driscoll, Westinghouse Systems Center
Errol EerNisse, Quartztronics, Inc.
Raymond Filler, U.S. Army Research Laboratory
Marvin Frerking, Rockwell International
Jean-Jacques Gagnepain, CNRS-LPMO, France
Edward Garber, TRW
Michael Garvey, Frequency & Time Systems, Inc.
Helmut Hellwig, Air Force Office of Scientific Research
James Pon-Wei Hou, Quartz Freq. Tech. Ltd., Taiwan
Charles Jensik, Piezo Crystal Company
Gary Johnson, Sawyer Research Products, Inc.
Hirofumi Kawashima, Seiko Electronics, Japan
Robert Kinsman, Motorola, Inc.
William Klepczynski, U.S. Naval Observatory
Jack Kusters, Hewlett-Packard Company
Theodore Lukaszek, U.S. Army Research Laboratory
Lute Maleki, Jet Propulsion Laboratory
Donald Malocha, University of Central Florida
Thrygve Meeker, Consultant
Thomas Parker, Raytheon Research Division
Victor Reinhardt, Hughes Aircraft Company
Gerald Roberts, General Electric Company
Tadashi Shiosaki, Kyoto University, Japan
Robert Smythe, Piezo Technology, Inc.
Samuel Stein, Timing Solutions Corporation
Dan Stevens, AT&T Bell Laboratories
Joseph Suter, Johns Hopkins University
Richard Sydnor, Jet Propulsion Laboratory
John Vetelino, University of Maine
John Vig, U.S. Army Research Laboratory
Fred Walls, NIST
S. Clark Wardrip, Consultant
Rolf Weglein, Consultant
Werner Weidemann, Ball Efratom Division
Joseph White, Naval Research Laboratory
Roger Williamson, BRN Europe, Ltd., UK
Gernot Winkler, U.S. Naval Observatory
Nicholas Yannoni, U.S. Air Force/RADC
Lidia Zhourkina, Soviet Committee IEC, Russia

DTIC QUALITY INSPECTED 3

TECHNICAL SESSION CHAIRMEN

PLENARY SESSION
Jack Kusters, Hewlett-Packard Company

BULK ACOUSTIC RESONATOR OSCILLATOR DESIGN AND ANALYSIS
John R. Vig, U.S. Army Research Laboratory

OPTICALLY PUMPED FREQUENCY STANDARDS
John D. Prestage, Jet Propulsion Laboratory

RESONATOR PROCESSING
Roger W. Ward, Quartzdyne, Inc.

FREQUENCY SYNTHESIZERS
Marvin E. Frerking, Rockwell International (Collins)

QUARTZ PROCESSING & EVALUATION
Charles A. Adams, Hewlett-Packard Company

FREQUENCY STABILITY MEASUREMENTS TIME DOMAIN-FREQUENCY DOMAIN
R. Michael Garvey, Frequency and Time Systems, Inc.

RESONATOR DESIGN THEORY
Daniel S. Stevens, AT&T Bell Laboratories

TRANSDUCERS AND SENSORS
Errol P. EerNisse, Quartztronics, Inc.

RESONATOR DESIGN AND PERFORMANCE
William P. Hanson, Piezo Crystal Company

FREQUENCY & TIME COORDINATION USING NAVIGATIONAL SATELLITES
David W. Allan, Allan's TIME

POSTER SESSION
Cryogenic Oscillators and Dielectric Resonators: Michael M. Driscoll, Westinghouse
Oscillators: Fred L. Walls, National Institute of Standards and Technology
Estimators for Long-Term Stability: Samuel R. Stein, Timing Solutions Corp.
Atomic Frequency Standards: Lute Maleki, Jet Propulsion Laboratory
Special Topics in Piezoelectric Devices: Jack Kusters, Hewlett-Packard Company
Filters: Robert C. Smythe, Piezo Technology, Inc.

SAW'S -- ANALYSIS & DESIGN
Thomas E. Parker, Raytheon Research Division

MATERIALS
Tadashi Shiosaki, Kyoto University, Japan

CRYSTAL OSCILLATORS AND AMPLIFIERS
Robert G. Kinsman, Motorola, Inc.

ION STANDARDS
Leonard S. Cutler, Hewlett-Packard Laboratories

TABLE OF CONTENTS

PAGE

AWARD PRESENTATIONS 1

PLENARY PAPERS

Relativity in the Future of Engineering 2
N. Ashby, University of Colorado

Precise Timing in Electric Power Systems 15
K.E. Martin, Bonneville Power Administration

FREQUENCY - TIME - NAVIGATION

Analysis of the Frequency Stability of On-Orbit GPS Navstar Clocks 23
T.B. McCaskill, W.G. Reid, J.A. Buisson, US Naval Research Laboratory &
H.E. Warren, Sachs Freeman Associates, Inc.

Using a New GPS Frequency Reference in Frequency Calibration Operations 33
T.N. Osterdock, Stellar GPS Corporation & J.A. Kusters, Hewlett-Packard Co.

Use of the Navstar Global Positioning System (GPS) in the Defense Communications System (DCS) 40
Q.D. Hua, Stellar GPS Corporation, M.S. Williams, Quantic Industries, Inc.,
J.Evans, Defense Information Systems Agency & M.A. Trippy, US Air Force

3S Navigation R100 Preliminary Implementation of GPS/GLONASS Time Transfer 45
J. Danaher & A. Balendra, 3S Navigation

Precise Time Dissemination Using the INMARSAT Geostationary Overlay 55
A. Brown, NAVSYS Corporation, D.W. Allan, Allan's TIME & R. Walton,
COMSAT

Comparative Analysis of Parameters of GLONASS Spaceborne Frequency Standards When Used Onboard and on Service Life Tests 65
Y.G. Gouzhva, A.G. Gevorkyan, A.B. Bassevich, P.P. Bogdanov &
A.Y. Tyulyakov, Russian Institute of Radionavigation and Time, Russia

ATOMIC FREQUENCY STANDARDS

NIST-7, The New US Primary Frequency Standard 71
R.E. Drullinger, J.P. Lowe, D.J. Glaze & J. Shirley, National Institute of
Standards and Technology

An Analysis of Major Frequency Shifts in the LPTF Optically Pumped Primary Frequency Standard 75
G.D. Rovera, E. deClercq & A. Clairon, BNM-Observatoire de Paris, France

Reducing the Effect of Local Oscillator Phase Noise on the Frequency Stability of Passive Frequency Standards 81
C. Szekely, F.L. Walls, J.P. Lowe, R.E. Drullinger & A. Novick, National
Institute of Standards and Technology

Potential for Improving the Rubidium Frequency Standard with a Novel Optical Pumping Scheme Using Diode Lasers 87
N.D. Bhaskar, The Aerospace Corporation

Prospects for Developing GLONASS Spaceboard Atomic Beam Frequency Standard with Laser Pumping 97
A.G. Gevorkyan, A.N. Bessedina & A.B. Bassevich, Russian Institute of Radionavigation and Time, Russia

A Dual Frequency Synthesis Scheme for a High C-Field Cesium Resonator 105
E. Rubiola, A. DelCasale & A. DeMarchi, Politecnico di Torino, Italy

Prospects for the Use of Multilambda Rabi Cavities in Atomic Beam Frequency Standards 109
E. Bava, Politecnico di Milano, A. DeMarchi, Politecnico di Torino & P. Tavella, Istituto Elettrotecnico Nazionale G. Ferraris, Italy

Atomic Phase Delay in a Rubidium Atomic Clock 114
J.C. Camparo & R.P. Frueholz, The Aerospace Corporation

Life, Line Q, Spin-Exchange Tuning and Stability of Atomic Hydrogen Maser Frequency Standards 120
H.B. Owings, P.A. Koppang, C.C. MacMillan & H.E. Peters, Sigma Tau Standards Corporation

Recent Investigations with the Harvard-Smithsonian Cryogenic Hydrogen Maser 129
R.L. Walsworth, E.M. Mattison, R.F.C. Vessot, Harvard-Smithsonian Center for Astrophysics & I.F. Silvera, Harvard University

Long Term Stability of Hg^+ Trapped Ion Frequency Standards *(INVITED)* 132
R.L. Tjoelker, J.D. Prestage, G.J. Dick & L. Maleki, Jet Propulsion Laboratory

Progress at CSIRO Australia Towards a Microwave Frequency Standard Based on Trapped, Laser-Cooled $^{171}Yb^+$ Ions 139
P.T.H. Fisk, M.A. Lawn & C. Coles, CSIRO Div. of Applied Physics, Australia

Improved Linear Ion Trap Physics Package 144
J.D. Prestage, R.L. Tjoelker, G.J. Dick & L. Maleki, Jet Propulsion Laboratory

Doppler Sideband Spectra for Ions in a Linear Trap 148
J.D. Prestage, R.L. Tjoelker, G.J. Dick & L. Maleki, Jet Propulsion Laboratory

Traceability in Atomic Frequency Standards 155
J.A. Kusters & C.P. Little, Hewlett-Packard Company

Subminiature Rubidium Frequency Standard for Commercial Applications 164
M. Bloch, I. Pascaru, C. Stone & T. McClelland, Frequency Electronics, Inc.

SHORT AND LONG TERM STABILITY, SYNTHESIZERS AND OTHER CIRCUITRY

Dependence of SAW Resonator *1/f* Noise on Device Size 178
T.E. Parker & D. Andres, Raytheon Company

Measured vs. Volume Model-Predicted Flicker-of-Frequency Instability in VHF Quartz Crystal Resonators 186
M.M. Driscoll, Westinghouse Electronic Systems Group & W.P. Hanson, Piezo Crystal Company

A Thermodynamic Resonance in Piezoelectric Crystal Plates of Thickness-Shear Vibrations 193
Y.I. Yevdokimenko & Y.S. Shmaly, Kharkov Military University, Ukraine

Prediction of Acoustic Resonator-Stabilized Oscillator Sustaining Stage Phase Noise Characteristics From Low Q Resonator Oscillator Noise Measurements 202
M.M. Driscoll, Westinghouse Electronic Systems Group

A High Performance VHF Oscillator with Optimized Crystal Drive Power 209
V.H. Estrick, Hughes Aircraft Company

A High Isolation Low Noise Amplifier with Near Unity Gain Up to 100 MHz 216
A. DeMarchi, F. Mussino & M. Siccardi, Politecnico di Torino, Italy

A New High-Isolation, Environmentally Insensitive, Wideband Distribution Amplifier 220
R.T. Lee & K. Ho, Hewlett-Packard Company

Spur Reduction Techniques in Direct Digital Synthesizers 230
V.S. Reinhardt, Hughes Space & Communications Company

Discrete Spurious Signals and Background Noise in Direct Digital Frequency Synthesizers 242
V.F. Kroupa, Academy of Sciences of the Czech Republic

A New Technique of Frequency Synthesis 251
W. Zhou, Xidian University, P.R. China

The Composite DDS - A New Direct Digital Synthesizer Architecture 255
L.J. Kushner, Lincoln Laboratory, MIT

Induced End-of-Life Errors in a Fast Settling PLL 261
T.J. Endres, G.T. Calvetti, Hughes Aircraft Company & J.B. Kirkpatrick, Nippondenso Technical Center

A Digital Equivalent of an Analog Spectrum Analyzer 270
J.A. Barnes, Austron, Inc.

Measurement of the Phase Noise Characteristics of an Unlocked Communications Channel Identifier 283
W.K. Lam, K.W. Wan, J. Austin, E. Vilar, University of Portsmouth & D.F. Bacon, Radiocommunications Agency, United Kingdom

Reducing Errors, Complexity, and Measurement Time of PM Noise Measurements 289
F.L. Walls, National Institute of Standards and Technology

Relationship of AM to PM Noise in Selected RF Oscillators 298
L.M. Nelson, C.W. Nelson & F.L. Walls, National Institute of Standards and Technology

Investigations of AM and PM Noise in X-Band Devices 303
F.G. Ascarrunz, E.S. Ferre & F.L. Walls, National Institute of Standards and Technology

Cross Correlation Analysis Improves Time Domain Measurements 312
A. Lepek, Hebrew University, Israel & F.L. Walls, National Institute of Standards and Technology

Flicker Noise Process Analysis 321
W. Guo, Shaanxi Astronomical Observatory, P.R. China

Simultaneous Measurement of Drifts and Noise Coefficients of Oscillators: Application to the Analysis of the Time Stability of the Millisecond Pulsars 326
F. Vernotte, F. Meyer, F. Naraghi, Observatoire de Besancon & E. Lantz, CNRS, France

A New Approach to Clock Modeling and Kalman Filter Time and Frequency Prediction 331
W. Su, GEO-Centers, Inc. & R.L. Filler, U.S. Army Research Laboratory

A Simple Precision Frequency Standard Comparator 335
W. Zhou & Z. Xuan, Xidian University, P.R. China

PIEZOELECTRIC MATERIALS, RESONATORS, FILTERS AND TRANSDUCERS

Peculiarities of Technology, Physical Properties and Applications of New Piezo-Electric Material Langasite ($La_3Ga_5SiO_{14}$) 339
A.N. Gotalskaya, D.I. Drezin, V.V. Bezdelkin & V.N. Stassevich, Scientific Research Institute "Fonon", Russia

Present Stage of $La_3Ga_5SiO_{14}$ - Research 348
I.M. Silvestrova, P.A. Senyushenkov, Y.V. Pisarevsky, Institute of Crystallography of the Russian Academy of Science & V.V. Bezdelkin, Scientific Research Institute "Fonon", Russia

New Piezoelectric Materials 351
I.M. Silvestrova, Y.V. Pisarevsky, P.A. Senyushenkov, Institute of Crystallography of the Russian Academy of Science & V.V. Bezdelkin, Scientific Research Institute "Fonon", Russia

Properties, Production and Application of New Piezoelectric Crystal Lithium Tetraborate $Li_2B_4O_7$ *(INVITED)* 353
S-J. Fan, Chinese Academy of Sciences, P.R. China

Pure-Mode Measurements of Dilithium Tetraborate Material Properties 359
J.A. Kosinski, A. Ballato, U.S. Army Research Laboratory & Y. Lu, Rutgers, The State University of New Jersey

A Study of the Time-Dependence of Electrodiffusion in Quartz 371
J.E. Ferris, D.W. Hart & J.J. Martin, Oklahoma State University

Quartz Crystals for Monolithic Piezoelectric Filters 377
I.V. Kabanovich, Scientific Research Institute "Fonon", Russia

Controlled Dissolution Applied to Berlinite and Quartz Materials *(INVITED)* 381
M. Deleuze, A. Goiffon, A. Ibanez, E. Philippot, CNRS, & O. Cambon, CEPE, France

Etching of Quartz Crystal Spheres 390
R.W. Ward, Quartzdyne, Inc.

Bi- and Three-Dimensional Prediction of Etching Shapes in Quartz Micromachined Structures 397
C.R. Tellier & T.G. Leblois, LCEP/ENSMM, France

Defects in Surface Layers of Lapped Crystal Plates and their Influence on Q-Factors of Crystal Units 407
A.S. Moshkovsky, Scientific Research Institute "Fonon", Russia

The Use of an X-Ray Three Beam Technique to Improve the Adjustment of Bars for Cutting 416
H. Bradaczek, G. Hildebrandt & W. Uebach, Freie Universität, Germany

Thermoluminescence (TL) Related to $[SiO_4/M]^\circ$ Defects (M = Li or Na) in Quartz 420
A. Halperin, The Hebrew University, Israel

The Determination of Internal Stress in the Quartz Crystals 427
A. Pogrebnyak & S. Abdrafikov, The Plant "Kristall", Russia

An Analysis of Transversely Varying Thickness Modes in Quartz Resonators with Bevelled Cylindrical Edges 431
H.F. Tiersten & Y.S. Zhou, Rensselaer Polytechnic Institute

Calculated Orientations and Aspect Ratios of Stiffened Rectangular Support Systems of SC-Cut Quartz Resonators Which Minimize the Influence of Fabrication Imperfections on Acceleration Sensitivity 442
Y.S. Zhou & H.F. Tiersten, Rensselaer Polytechnic Institute

A Perturbation Analysis on Contoured Crystal Plates 448
W. Zhang, Motorola, Inc. & J. Yang, Princeton University

Effect of Stress on Guided EM Waves in Anisotropic Dielectric Plates 461
P.C.Y. Lee, J.D. Yu, Princeton University & A. Ballato, U.S. Army Research Laboratory

Calculation of Radiated Electromagnetic Power From Bulk Acoustic Wave Resonators 472
C.F. Campbell & R.J. Weber, Iowa State University

Exact Analysis of the Propagation of Acoustic Waves in Multilayered Anisotropic Piezoelectric Plates 476
J.T. Stewart & Y-K. Yong, Rutgers University

Numerical Analysis of Two Dimensional Thin Film Resonators 502
K.M. Lakin, TFR Technologies, Inc.

Self-Sustained Acoustoelectric and Photoelectric Oscillations in Systems With Relaxation 509
V.A. Vyun, Institute of Semiconductor Physics, Russia

Rigorous Modeling of Corrugated Surfaces in Microacoustics 514
A.R. Baghai-Wadji, Vienna University of Technology, Austria &
A.A. Maradudin, University of California

Frequency Temperature Characteristics of the x-Length Strip Resonators of AT-Cut Quartz 523
J. Zelenka, Technical University of Liberec, Czech Republik & P.C.Y. Lee, Princeton University

Forced Vibrations of KT-Cut Width-Extensional Mode Quartz Crystal Resonators 527
H. Kawashima, Seiko Electronic Components Co., Ltd., Japan

Application of TC-Cut Quartz Resonators Excited by Lateral Electrical Field..... 535
I.V. Abramson, A.N. Dikidzhi, A.I. Kutalev & S.V. Plachotin, Omsk Research Instrumentation Institute, Russia

Thin Rotated Y-Cut Quartz Resonators Vibrating in B-Mode Over a Wide Temperature Range 541
M. Nakazawa, M. Takeuchi, Shinshu University, H. Yamaguchi, Miyota Co. Ltd., Japan & A. Ballato, U.S. Army Research Laboratory

Research Aimed at Designing a Miniature Ruggedized Configuration of BVA-Type Crystal Unit..... 548
A.E. Karaulnik & V.B. Grouzinenko, Scientific Research Institute "Fonon", Russia

The Effects of Metal Depositions Upon Acceleration Induced Frequency Shifts for Quartz Crystal Units 560
L.D. Clayton & E.P. EerNisse, Quartztronics, Inc.

Investigations of Gold Films on Quartz Crystals 574
M. Hendrickson, J.R. Vig, U.S. Army Research Laboratory & S. Laffey, Vitronics, Inc.

Precision Frequency Trimming of SAW and STW Resonators Using Xe^+ Heavy Ion Bombardment 582
V.S. Aliev, Institute of Semiconductor Physics, Russia & I. Avramov, Bulgarian Academy of Sciences, Bulgaria

Application of Quartz Micromachining to the Realization of a Pressure Sensor *(INVITED)* 587
J.S. Danel, M. Dufour & F. Michel, LETI, France

Enhancement of Quartz Electrical Conductivity by Ion Implantation..... 597
J.S. Danel, P. Martin, M. Dufour, A. Ermolieff, S. Marthon, F. Pierre & M. Dupuy, LETI, France

Measuring Liquid Properties with Smooth- and Textured-Surface Resonators..... 603
S.J. Martin, K.O. Wessendorf, C.T. Gebert, G.C. Frye, R.W. Cernosek, L. Casaus & M.A. Mitchell, Sandia National Laboratories

Tactical BVA Quartz Resonator Performance 609
J.R. Norton, The Johns Hopkins University Applied Physics Laboratory & R.J. Besson, École Nationale Superieure de Mecanique et des Microtechniques, France

Results of the Quartz Crystal Measurement Data Obtained at the Leadless Resonator Measurement Workshop Held in Japan..... 614
Y. Oomura, Y. Watanabe, Tokyo Metropolitan University & M. Koyama, Nihon Dempa Kogyo Co. Ltd., Japan

Eight-Pole Monolithic Filters Using Lithium Tantalate 620
I. Mateescu, Institute of Physics and Technology of Materials, Romania & J. Kosinski, U.S. Army Research Laboratory

Modelization of Thickness Shear Integrated Filters 626
J. Detaint, J. Schwartzel, C. Joly, France Telecom CNET, A. Zarka & B. Capelle, L.M.C.P., France

Observations of the Crystal Perfection and of Surface Acoustic Waves in Lithium Niobate 632
A. Zarka, B. Capelle, L.M.C.P., J. Detaint, J. Schwartzel, CNET & J.M. Hode, Thomson-Sintra, France

Delta Function Model Analysis of SSBW Spurious Response in SAW Devices 639
K. Hashimoto & M. Yamaguchi, Chiba University, Japan

Design, Fabrication and Performance of 2 GHz Surface Transverse Wave Resonators 645
S. Denisenko, Institute of Semiconductor Physics, Russia & I.D. Avramov, Institute of Solid State Physics, Bulgaria

An Analysis of Self-Supported Circular Structures for Reduced Stress Sensitivity in SAW Devices 650
S. Ballandras & E. Bigler, CNRS, France

Withdrawal Weighted Fan-Shaped SAW Transducers 656
E.V. Bausk & I.B. Yakovkin, Institute of Semiconductor Physics, Russia

Deformation-Sensitive Cuts for Surface Acoustic Waves in α-Quartz 660
R.M. Taziev, E.A. Kolosovsky & A.S. Kozlov, Institute of Semiconductor Physics, Russia

Precision SAW Pressure Sensors 665
Y.N. Vlassov, A.S. Kozlov, N.S. Pashchin & I.B. Yakovkin, Institute of Semiconductor Physics, Russia

OSCILLATORS - QUARTZ, MICROWAVE AND CRYOGENIC

Voltage Controlled S-TCXO's Employing NS-GT Cut Quartz Crystal Resonators *(INVITED)* 670
H. Kawashima & K. Sunaga, Seiko Electronic Components Co. Ltd., Japan

Manufacturing Hybrid TCXOs 679
B. Achille, T. Collins & D. Marvin, Motorola, Inc.

The New Method of Statistic Piesewise-Linear Interpolation and its Application to DTCXO Creation 687
A. Kosykh, V. Bagaev, B. Ionov, A. Lepetaev, S. Zavialov, Omsk Politechnic Institute & A. Vasiliev, Microsystem SSI, Russia

Universal, Computer Facilitated, Steady State Oscillator, Closed Loop Analysis Theory 698
B. Parzen, Consulting Engineer

Dynamic Analysis of Modulated Oscillators 706
R.L. Clark, McCoy Electronics & D.T. Comer, Penn State University

The Lever Oscillator for Use in High Resistance Resonator Applications 711
K.O. Wessendorf, Sandia National Laboratories

Overmoded High Q Resonators for Microwave Oscillators 718
G.R. Kline, K.M. Lakin & K.T. McCarron, TFR Technologies, Inc.

A Study of a MSW Device Using a Low-Loss Bi,Ga Substituted YIG Film Having a Proper Growth Induced Anisotropy Magnetic Field Characterized by Ultra-Low Frequency-Temperature Dependencies 722
M. Tanno, Shin-Etsu Chemical Co., Ltd. & M. Nakazawa, Shinshu University, Japan

Extremely High-Loop Power GHz Range Surface Transverse Wave Oscillators Using AB-Class Amplifiers 728
I.D. Avramov, Institute of Solid State Physics, Bulgaria

Phase Noise and Frequency Stability of Ka-Band Harmonic Dielectric Resonator Oscillators 733
M. Mizan, T. Higgins & D. Sturzebecher, U.S. Army Research Laboratory

A Temperature Compensated 1 GHz STW Based Multifrequency Oscillator 740
M.A. Taslakov, Institute of Electronics, Bulgaria

High-Overtone, Bulk Acoustic Resonator Frequency Stability Improvements 744
S.P. Caldwell, M.M. Driscoll, D.S. Bailey, Westinghouse Electronic Systems Group, S.D. Stansberry, Design and Producibility Engineering Division, Westinghouse & H.L. Salvo, Advanced Technology Division, Westinghouse

High-Q TE Stabilized Sapphire Microwave Resonators for Low Noise Applications 749
M.E. Tobar, A.J. Giles, University of Western Australia, S. Edwards & J. Searls, Poseiden Scientific Instruments Pty., Ltd., Australia

Improved Sapphire Dielectric Resonators for Ultrastable Oscillators 757
A.N. Luiten, A.G. Mann & D.G. Blair, University of Western Australia, Australia

Microwave Oscillators Incorporating Cryogenic Sapphire Dielectric Resonators 763
C.A. Flory & R.C. Taber, Hewlett-Packard Laboratories

Closed Loop Tests of the NASA Sapphire Phase Stabilizer 774
D.G. Santiago & G.J. Dick, Jet Propulsion Laboratory

An Uncooled Microwave Oscillator with 1-Million Effective Q-Factor 779
D.P. Tsarapkin, Moscow Power Engineering Institute, Russia

A Thallium-Based Superconducting Cavity for Microwave Oscillator Applications 784
S.C. Han, B.L. Zhou, C.P. McCarroll, Xsirius Superconductivity, Inc., T.G. Hammersley, T. Apgar, Communications Techniques, Inc., P.N. Arendt & D.W. Cooke, Los Alamos National Laboratory

The Flight Performance of the Galileo Orbiter USO 788
D. Morabito, T. Krisher & S. Asmar, Jet Propulsion Laboratory

POST DEADLINE PAPERS

Factorizational Synthesis of SAW Bandpass Filters 806
A.S. Rukhlenko, Belorussian State University, Republic of Belarus, CIS

X Ray Topography Study of Gallium Phosphate Crystals and Resonators 813
B. Capelle, A. Zarka, Y. Zheng, L.M.C.P. Universités P VI et P VII, CNRS,
J. Schwartzel, J. Detaint, CNET Laboratoire de Bagneux, A. Ibanez &
E. Philippot, L.P.C.M.S. Université de Montpellier, CNRS-URA, France

Application of Low Cost Frequency Standards for Commercial and Military GPS 821
L.D. Vittorini, Rockwell International & B. Robinson, Rakon Ltd., New Zealand

AUTHOR INDEX 835

PROCEEDINGS ORDERING INFORMATION 837

SPECIFICATIONS AND STANDARDS RELATED TO FREQUENCY CONTROL 838

1993 AWARD WINNERS

Cady Award

The Cady Award was presented to Hirofumi Kawashima, Seiko Electronics Components, Ltd., "for outstanding contributions to photolithographic miniature quartz crystal units with excellent temperature stability." The award was presented by Gary R. Johnson, Sawyer Research Products, Inc.

Rabi Award

The Rabi Award was presented to Robert F. C. Vessot, Harvard-Smithsonian Center for Astrophysics, "for contributions to hydrogen maser technology and applications." The award was presented by Leonard S. Cutler, Hewlett Packard Co.

Sawyer Award

The Sawyer Award was presented to Jack L. Saunders, Saunders and Associates, "for leadership in the development and manufacture of quartz resonator measurement equipment used throughout the industry." The award was presented by Charles Adams, Hewlett Packard Co.



Robert F. C. Vessot, Rabi Award winner; Hirofumi Kawashima, Cady Award winner; and Jack L. Saunders, Sawyer Award winner.

Distinguished Lecturer Award

Also presented during the Symposium award ceremony was the Distinguished Lecturer Award to John R. Vig, U.S. Army Research Laboratory. The award was presented by UFFC Society President James F. Greenleaf. Vig was elected by the UFFC-Soc Ad Com to be the Society's Distinguished Lecturer for 1992-93. The title of his lecture was "High-Accuracy Oscillators and Clocks."

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

RELATIVITY IN THE FUTURE OF ENGINEERING

Neil Ashby
Department of Physics, Campus Box 390
University of Colorado
Boulder, CO 80309-0390

Abstract

Improvements in clock technology make it possible to develop extremely accurate timing, ranging, navigation, and communications systems. Three relativistic effects–time dilation, the Sagnac effect, and gravitational frequency shifts–must be accounted for in order for modern systems to work properly. These effects will be related in a non-mathematical way to fundamental relativity principles: constancy of the speed of light, and the principle of equivalence. Examples of current and future engineering applications will be discussed, such as in the Global Positioning System, in time synchronization systems, geodesy, and communications.

Atomic Clocks

Relativistic effects become important in applications requiring very accurate timing, time transfer, or synchronization. Many engineering systems are beginning to rely on modern atomic clocks which have fractional frequency stabilities of the order of 10^{-12} or 10^{-13}. An excellent example is the Global Positioning System (GPS), in which about a dozen relativistic effects must be accounted for in order for the system to work properly. Atomic clock technology not only provides the basis for the definition of the second as the unit of time, this technology is expected to improve rapidly in the future. Vessot *et at.*[1] have summarized potential future performance improvements in several promising devices including cryogenic H-masers, Cs fountains, and trapped Hg ions; these predictions are summarized in Fig. 1. These analyses show there is some hope that fractional frequency stabilities in the range 10^{-16} to 10^{-17} can be achieved. For this paper I shall however adopt a conservative fractional frequency stability figure of 10^{-15} as a guideline for determining what relativistic effects might be important in the future.

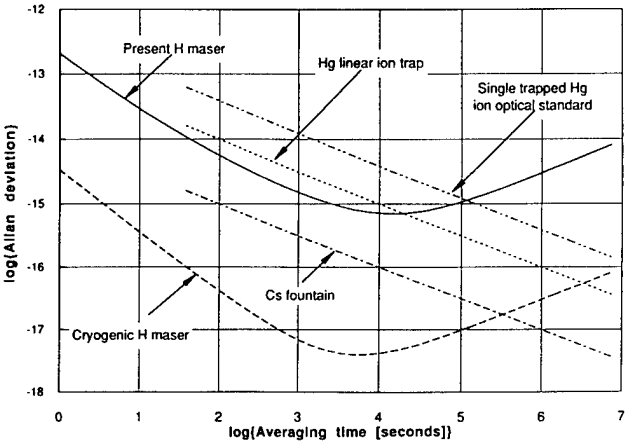


Fig. 1. Predicted Allan variance for future frequency standards. (This assumes no systematic effects in Cs and Hg devices.)

Constancy of the Speed of Light

Relativity enters metrology in a most fundamental way through the so-called 'Second Postulate' of the special theory of relativity, the principle of the constancy of the speed of light, c. This now widely accepted principle states that the speed of light in free space has the same value in all inertial systems, independent of the motion of the source. The speed of light is also independent of the motion of the observer. The numerical value of c has been defined by convention:

$$c = 299\ 792\ 498 \text{ meters/second.} \qquad (1)$$

In conjunction with the adopted unit of time, this value for c defines the SI unit of length, the meter. In thinking about the speed of light, a convenient alternative rule of thumb is that c is approximately equal to one foot per nanosecond (1 nanosecond = 1 ns = 10^{-9} second).

In an inertial frame of reference, the principle of the constancy of c provides a means of synchronizing

0-7803-0905-7/93 $3.00 © 1993 IEEE

remotely placed clocks. Consider two standard clocks, A and B, placed at rest a distance L (meters) apart. (The distance L could be found by measuring the time on clock A required for a light signal to propagate from A to B and back, and multiplying by $c/2$. This would not depend on the presence of a clock at B.) Now suppose a signal originates at clock A at time t_A. The time required for the signal to propagate in one direction from A to B is L/c. The clock at B will then be synchronized with that at A if the signal arrives at the time t_B given by

$$t_B = t_A + L/c. \tag{2}$$

This procedure is called the 'Einstein Synchronization Procedure' and clocks distributed at rest in any inertial frame will be presumed to be synchronized by this or an equivalent procedure.

Clearly in discussing electromagnetic signals as I have done above, I am ignoring quite a few practical difficulties. Signals must have sufficient spectral bandwidth that it is possible to reconstruct well-defined pulses in time. Noise in real clocks, frequency drifts due to environmental factors, etc., are not a concern here. Also I'm usually going to ignore effects on propagation speed which might arise because the signals propagate through a medium rather than through a vacuum.

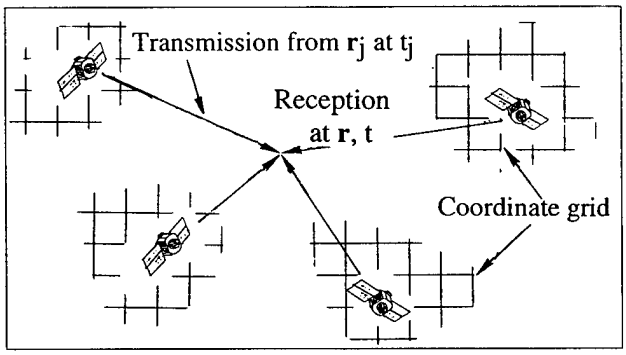


Fig. 2. Idealized conception of a navigation and time transfer system.

Navigation

Keeping these caveats in mind, the constancy of c leads to the following idealized conception of a navigational system. Referring to Fig. 2, suppose four transmitters, each with its own standard clock, are placed at known locations $\mathbf{r}_j$. Assume the clocks are synchronized by the Einstein procedure. There is a receiver at unknown position $\mathbf{r}$ carrying a standard clock which has not been synchronized. Let these transmitters rapidly transmit synchronized pulses which are tagged with the transmitter's position and time, so that a receiver can determine the time t_j and the location $\mathbf{r}_j$ of the pulse from transmitter j. The receiver's position $\mathbf{r}$ and clock time t can then be determined by solving four simultaneous propagation delay equations:

$$(\mathbf{r} - \mathbf{r}_j)^2 = c^2(t - t_j)^2; \qquad j = 1, 2, 3, 4, \tag{3}$$

for the unknowns $\mathbf{r}$ and t. These equations just express the principle of the constancy of the speed of light in an inertial frame. Clearly a timing error of one nanosecond would lead to a an error of about a foot in position determination.

Event Detection

There is a kind of reciprocity in this situation which can be used for event detection: suppose that instead of transmitters at the locations $\mathbf{r}_j$ there are receivers, tied to synchronized standard clocks. Suppose that an event occurs at the position $\mathbf{r}$ at time t causing a signal to be transmitted, which is received at the four receivers at the respective known positions $\mathbf{r}_j$ at the measured times t_j. Then by solving four propagation equations of the form of Eqs. (3), the position of the event and the time at which it occurs can be determined. If some information about the position of the event is available, it may be possible to locate the event by solving fewer than four propagation delay equations.

Fault Location

As an example of event detection using only two synchronized clocks, consider the problem of determining the location and time of a fault that occurs in a power line stretching between two detectors a distance L apart. In Fig. 3, clocks at the ends of the line are synchronized from some independent primary reference clock. A fault occurring at distances L_1, L_2 from the respective detectors at the ends of the lines sends out a signal at time t which is received at times t_1, t_2 at the respective ends of the line. A previous survey would give

$$L = L_1 + L_2, \tag{4}$$

whereas from the constancy of c, the times t_1, t_2 are related to the time t by propagation delay equations:

$$t_1 = t + L_1/c, \quad t_2 = t + L_2/c. \tag{5}$$

Solution of only two propagation delay equations, in conjunction with Eq. (4), gives the time and position of the fault. To locate the fault to within one foot requires synchronization to better than a nanosecond.

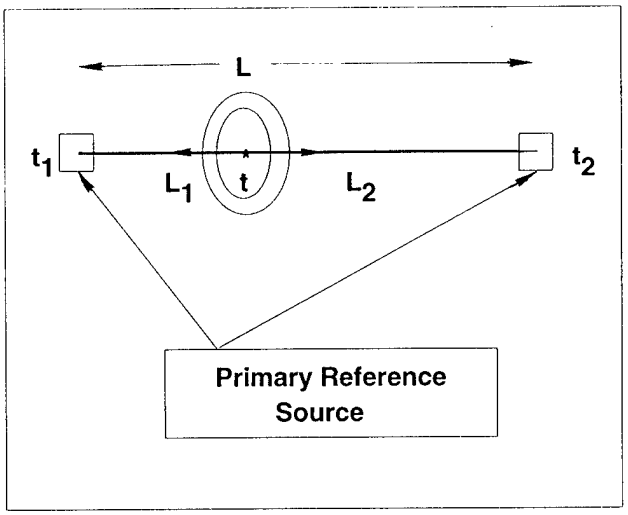


Fig. 3. Fault location using constancy of c.

Breakdown of Simultaneity

The discussion above assumes that the clocks are at rest in some inertial reference frame. Usually however clocks are in motion; for example in Fig. 2 the transmitters could be orbiting the earth. Relative motion introduces subtle new effects; perhaps the most profound of these is the breakdown of the concept of simultaneity. Events which appear to occur simultaneously in one inertial frame may not appear simultaneous to observers in some other inertial frame, which is moving with respect to the first. This is a direct consequence of the principle of the constancy of c.

In discussing measurements made by observers in two different, relatively moving inertial frames, one always imagines that each observer is equipped with his/her own measuring rods and standard clocks, that the clocks used by observers in one frame are at rest, and that they are synchronized by the Einstein procedure. In each of the inertial frames, any particular electromagnetic signal propagates with speed c.

Consider then as in Fig. 4. two events consisting of two lighting strokes which hit the two ends of a train of length $L = 2x$ simultaneously *as seen by observers on the ground*. The train is assumed to be moving to the right at speed v relative to the ground. For ease of discussion, I'll refer to the ground as the 'rest' frame, and the train as the 'moving' frame. Observers on the ground (in the rest frame) can determine the midpoint between the two lightning strokes, a distance x from either end of the initial position of the train. They will then find that light signals from the two events will propagate along the tracks and collide at the midpoint. This has nothing to do with the motion of the train.

Now look at the sequence of events involving a moving observer, sitting at the midpoint of the moving train. As the train moves forward, this observer approaches the oncoming light emitted from the event at the front of the train, and recedes from the light signal emitted from the event at the back of the train. Therefore the moving observer will encounter light from the front event first, and will have to conclude that the event at the front of the train occurred first. By the principle of the constancy of c, light must travel with speed c no matter what the value of the relative speed v is. So if light from event A arrives before that from event B, which is the same distance away, then event A must occur first.

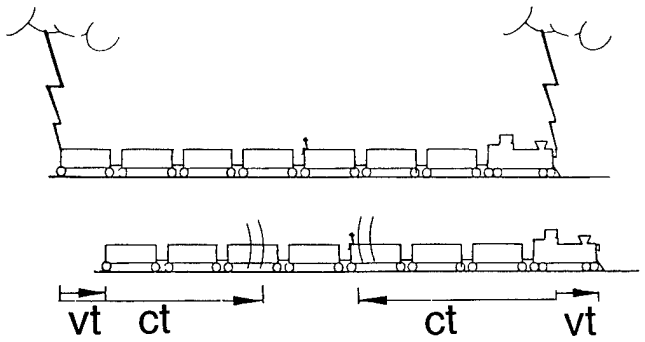


Fig. 4. Thought experiment illustrating relativity of simultaneity.

To analyze this approximately is not difficult. Suppose the zero of time for observers in both the rest and the moving frames is set to occur at the instant the midpoint of the train encounters the signal from the lightning stroke at the front of the train. I'll use primes to denote quantities measured by the moving observer. Then to the moving observer, the time t' of the stroke at the front of the train is

$$t' = -\frac{x}{c}. \tag{6}$$

To observers in the rest frame, however, the midpoint of the train is approaching the signal at the relative speed $c + v$, so to first order in v,

$$t = -\frac{x}{c+v} \approx -\frac{x}{c} + \frac{vx}{c^2}. \tag{6}$$

Therefore

$$t' = t - \frac{vx}{c^2}. \tag{8}$$

The term $-vx/c^2$ is a relativistic correction for breakdown of simultaneity. The effect is proportional to the relative velocity and proportional to the distance x.

Putting in some numbers, suppose $v = 1000$ km per hour (typical for a jet aircraft) and $x = 3500$ km. Then the correction is 108 ns. The negative sign in Eq.

(8) means that of two events simultaneous in the rest frame, to the moving observer the event farther out in front, at the more positive x, occurs earlier.

Sagnac Effect

The above discussion of the breakdown of simultaneity can be used to understand the peculiar physics on the edge of a slowly rotating disc. The prime engineering application is to time transfer and synchronization on the surface of the rotating earth. For purposes of illustration, therefore, I'll use the angular velocity of rotation of the earth, $\omega = 7.29 \times 10^{-5}$ rad/sec, and for the equatorial radius of the earth, $R = 6.378 \times 10^6$ meters.

In this case the rest frame is a local non-rotating frame, with axes pointing toward the fixed stars, but with origin at the center of the earth. The moving frame is a reference frame extending over a small portion of the rotating earth's surface, having velocity $v = \omega r$ relative to the rest frame, where r is the distance of the clocks from the rotation axis.

Now imagine two clocks fixed a small east-west distance x apart on the equator of the earth. Viewed from the nonrotating frame they will be moving with approximately equal speeds $v = \omega r$. If a clock synchronization process involving electromagnetic signals were carried out by two earth-fixed observers using Einstein synchronization in the moving frame, then the two clocks would not be synchronous when viewed from the nonrotating frame. The magnitude of the discrepancy is $vx/c^2 = \omega rx/c^2 = (2\omega/c^2)(rx/2)$. If this synchronization process is performed successively all the way around the circle, then effectively the distance x is $x = 2\pi r$, and the time discrepancy is thus

$$\Delta t = 2\omega/c^2 \times \pi r^2, \tag{9}$$

where πr^2 is the area enclosed by the path followed during the synchronization process. For example, synchronization around the earth's equator involves a discrepancy

$$\Delta t = \frac{2\omega}{c^2}\pi R^2 \approx 208 \text{ ns.} \tag{10}$$

upon arriving back at the starting point.

This effect is known as the *Sagnac effect*. If the synchronization path were westward around the earth rather than eastward, then the discrepancy would be of opposite sign. This means that Einstein synchronization in a rotating reference frame is not self-consistent: If A is synchronized with B and B is synchronized with C, then A is not necessary synchronized with C. In order to avoid difficulties with such non-transitivity it is best to adopt time in the *non-rotating* frame as the measure of time in the rotating frame. Thus one discards Einstein synchronization in the rotating frame.

To put it another way, if Einstein synchronization is used in the earth-fixed rotating frame, then it is necessary to apply a 'Sagnac correction' to the readings of clocks on the rotating earth, in order the obtain an internally consistent 'coordinate time' on earth's surface.

This is illustrated in Fig. 5, where there is a sketch of a flattened rotating earth. For a sequence of synchronization processes forming a closed circuit on the rotating earth, upon projecting the path onto the equatorial plane of the earth one can determine the projected area A_E. The Consultative Committee for the Definition of the Second and the International Radio Consultative Committee have agreed that, in order to obtain consistently synchronized clocks on the earth's surface at the subnanosecond level, the correction term to be applied is of the form

$$\Delta t = 2\omega/c^2 \times A_E, \tag{11}$$

where A_E is the projected area on the earth's equatorial plane swept out by the vector whose tail is at the center of the earth and whose head is at the position of the electromagnetic signal pulse. The area A_E is taken as positive if the head of the vector moves in the eastward direction. If two clocks located on the earth's surface are compared by using electromagnetic signals in the rotating frame of the earth, then Δt must be subtracted from the measured time difference (east clock minus west clock) in order to synchronize the clocks so they will measure coordinate time on the rotating earth. They will effectively measure time in the local non-rotating frame attached to the earth's center.

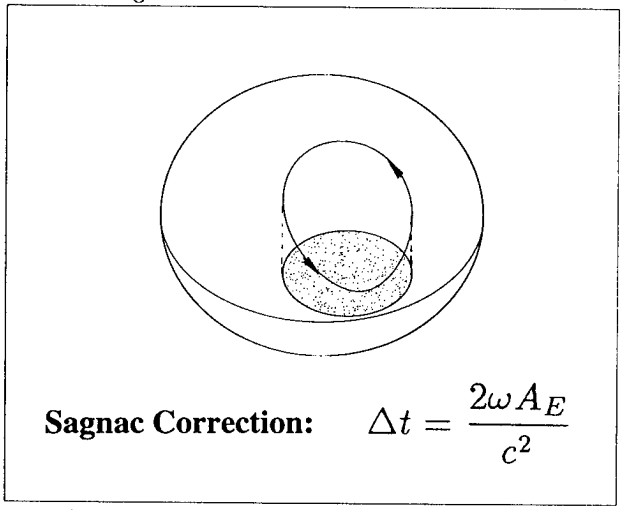


Fig. 5. Projected area for a sequence of Einstein synchronization processed forming a closed circuit on the rotating earth's surface.

Lack of transitivity in synchronization has implications for devices which rely on accurate synchronization. Suppose a communications network distributes synchronization through a series of nodes, along two different paths, to the ends of a communication link as in Fig. 6. If the area enclosed by the path, projected onto the earth's equatorial plane, is not zero, then problems with inconsistent synchronization can arise. For example, suppose one synchronization link goes from San Francisco directly to New York, while a second link goes from San Francisco to Miami and then to New York. The discrepancy in synchronization between these two paths due to the Sagnac effect is about 11 ns. While this is not significant if the signal is 60 Hz as in a power grid, in an optical communications network operating at 10^{15} Hz the discrepancy amounts to 10^7 cycles of oscillation. Depending on the design of the system this may become significant in the future.

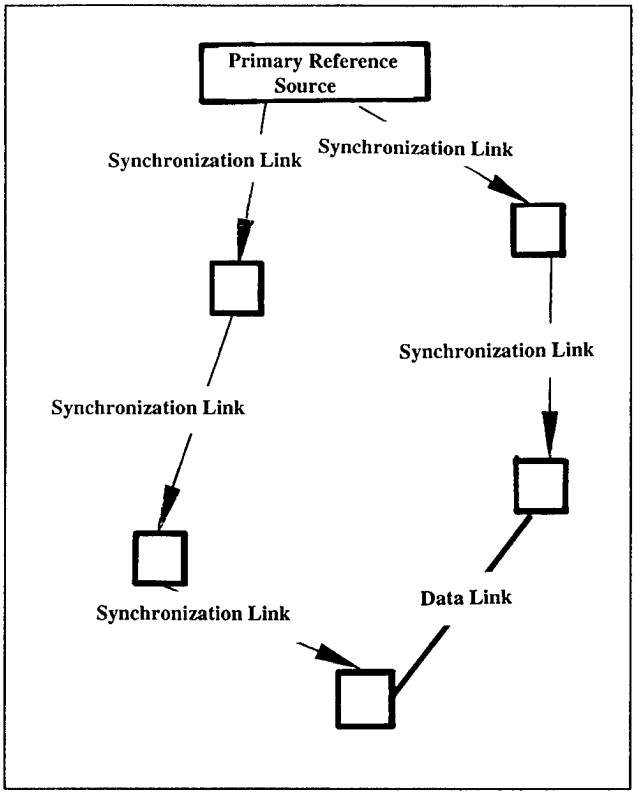


Fig. 6. Distribution of synchronization for a communications network.

Furthermore, if the trouble is taken to incorporate hardware delays to compensate for the Sagnac effect while sending in one direction, then when the communications are sent back the other way over the same link the effect will become twice as big. The effect is asymmetric. The same effect will occur in optical fiber communications networks where the speed of signal propagation may be significantly less than c. In the rotating reference frame the Sagnac effect is a property of space and time, not dependent on signal propagation speed.

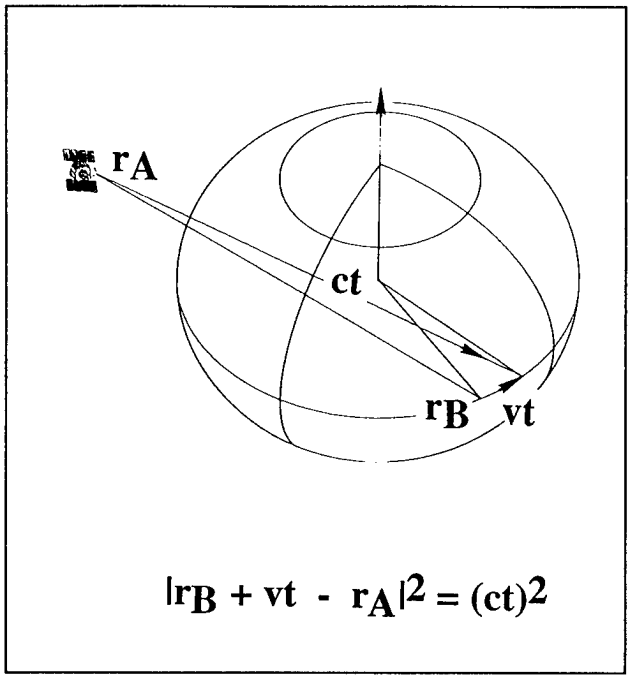


Fig. 7. The Sagnac effect will be automatically included if receiver motion due to earth rotation during signal propagation is accounted for.

An equivalent way of looking at this phenomenon is diagrammed in Fig. 7, which shows a signal transmitted from a satellite to a ground-based receiver. From the point of view of the nonrotating frame, the signal goes in a straight line with speed c, from the initial transmitter position $\mathbf{r}_A$ to the final receiver position. If in this frame one accounts for the motion of the receiver during the propagation of the signal, then the Sagnac effect will be automatically accounted for. This if the initial position of the the receiver is $\mathbf{r}_B$, the velocity of the receiver is $\mathbf{v}$, and the signal propagation time is t, constancy of c requires

$$|\mathbf{r}_B + \mathbf{v}t - \mathbf{r}_A| = (ct)^2. \qquad (12)$$

Iterative solution of Eq. (12) for t is equivalent to calculating the Sagnac correction.

Time Dilation

In the previous section I discussed two effects which are of first order in the velocity–the breakdown of simultaneity, and the Sagnac effect. In this section I shall discuss another famous effect–time dilation–which is of second order in the velocity. Imagine two

inertial frames, a 'rest' frame or laboratory frame, and a moving frame. A clock in the moving frame beats more slowly than clocks in the rest frame with which it is successively compared. The following thought experiment will readily convince anyone that the principle of the constancy of the speed of light requires the 'moving' clock to beat more slowly. A prime denotes quantities measured in the moving frame.

Suppose that observers in the two inertial frames each possess a set of rectangular Cartesian coordinate axes which they orient so that the x, x' and y, y' axes are parallel. The direction of relative motion is parallel to the x, x' axes. The moving observer orients a rod of length L' along the y' axis, and sends a light signal up along this rod from one end to the other. The situation is diagrammed in Fig. 8. To simplify the discussion one assumes that the light starts out at the instant the origins of the two reference frames pass by each other.

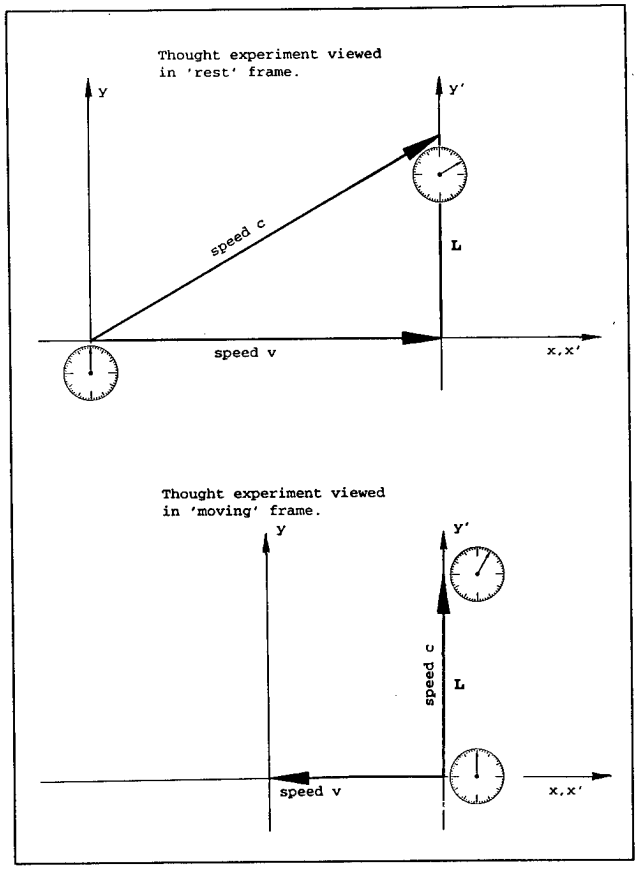


Fig. 8. Thought experiment showing that 'moving' clocks beat more slowly than clocks that remain 'at rest'.

The lower part of Fig. 8 shows the situation from the point of view of observers in the moving frame. The time t' required for light to travel along the rod is simply

$$t' = L/c. \tag{13}$$

The clock faces on the lower part of Fig. 8 indicate time at the beginning and end of the experiment.

The upper part of Fig. 8 shows the experiment from the point of view of observers in the rest frame. Breakdown of simultaneity would create difficulties for measurements of lengths oriented parallel to the relative velocity. But since this rod is oriented perpendicular to the relative velocity, by symmetry it is not possible for the rod to appear changed in length. So this rod has length $L = L'$ as it moves through the rest frame. The rod is moving to the right with speed v and the light travels along the rod, so there has to be a horizontal component of velocity of the light equal to v. The vertical component of the velocity of the light certainly has to be less than c; therefore the time required for the light to reach to upper end of the rod certainly has to be greater than L/c. This argument shows qualitatively that the clocks in the moving frame will beat more slowly than the sequence of clocks with which they are compared in the rest frame.

The top part of Fig. 8 actually gives the right answer, for by the principle of the constancy of the speed of light, the vertical component of the light velocity in the rest frame is just $\sqrt{c^2 - v^2}$. Thus for observers in the rest frame, the time t required for the light to reach the upper end of the rod is just

$$t = L/\sqrt{c^2 - v^2}, \tag{14}$$

so the relationship between t' and t obtained by eliminating L from Eqs. (13) and (14) and $L = L'$ is:

$$t' = \sqrt{1 - v^2/c^2}\, t. \tag{15}$$

Usually the ratio v/c is small, so the square root can be expanded, giving approximately

$$t' \approx \left(1 - \frac{1}{2}\frac{v^2}{c^2}\right) t. \tag{16}$$

The fractional slowing is given by the correction $v^2/2c^2$ in the above equation. This correction is also commonly called the second-order Doppler shift, or transverse Doppler shift.

Some examples of the size of this effect are as follows. For a clock at rest on the earth's equator, and viewed from the nonrotating frame,

$$-\frac{1}{2}\frac{v^2}{c^2} \approx -1.2 \times 10^{-12}; \tag{17}$$

this would accumulate to about 104 ns in one day. For a clock in a satellite orbiting the earth at 100 km altitude,

$$-\frac{1}{2}\frac{v^2}{c^2} \approx -3.4 \times 10^{-10}. \qquad (18)$$

For a clock in a GPS satellite,

$$-\frac{1}{2}\frac{v^2}{c^2} \approx -8.34 \times 10^{-11}. \qquad (19)$$

Keeping in mind that in the future the fractional frequency stability of orbiting clocks may approach a part in 10^{15}, these are very large effects. Even for clocks of frequency stability 1×10^{-13}, as in the GPS Block II satellites, the second-order Doppler shift for an earth-fixed clock is significant.

Gravitational Frequency Shifts

The Sagnac effect and the second-order Doppler shift are effects which can be understood on the basis of the Special Theory of Relativity. A third effect–the gravitational frequency shift–occurs when signals are sent from one location to another having a different gravitational potential. The effect can be understood in an elementary way using the fundamental assumption of the General Theory of Relativity–Einstein's Principle of Equivalence.[2]

The Principle of Equivalence

Einstein's Equivalence Principle states that over a small region of space and time, a fictitious gravitational field induced by acceleration cannot be distinguished from a gravitational field produced by mass. Thus the fictitious centrifugal force one feels in turning a corner in a vehicle has the same physical effects as a real gravitational field. An immediate consequence of the Equivalence Principle is that gravitational fields can be reduced to zero by transforming to a freely falling reference frame. The fictitious gravitational field due to the acceleration then exactly cancels the real gravitational field.

All experiments performed in a real gravitational field, such as in a laboratory on the surface of the earth where there is a gravitational field $\mathbf{g}$, will have the same results at experiments performed in a laboratory in free space which is accelerated in the opposite direction with acceleration $\mathbf{a} = -\mathbf{g}$. In Fig. 9a are sketched some experiments performed in a laboratory fixed on the earth's surface. For example two objects of different compositions are observed to fall downward with equal accelerations g. (This is related to the deep experimental fact of the strict proportionality of inertial and gravitational mass, a subject we shall not go into here.[3]) In Fig. 9b, a similar experiment is performed in a laboratory in free space which is being pulled upward with acceleration g. A non-accelerated observer sees that the apple and the lead ball have no forces exerted on them so remain at rest with respect to each other and the laboratory is accelerated past the objects, whereas the observer in the accelerated frame sees the objects 'fall' downward with identical accelerations g.

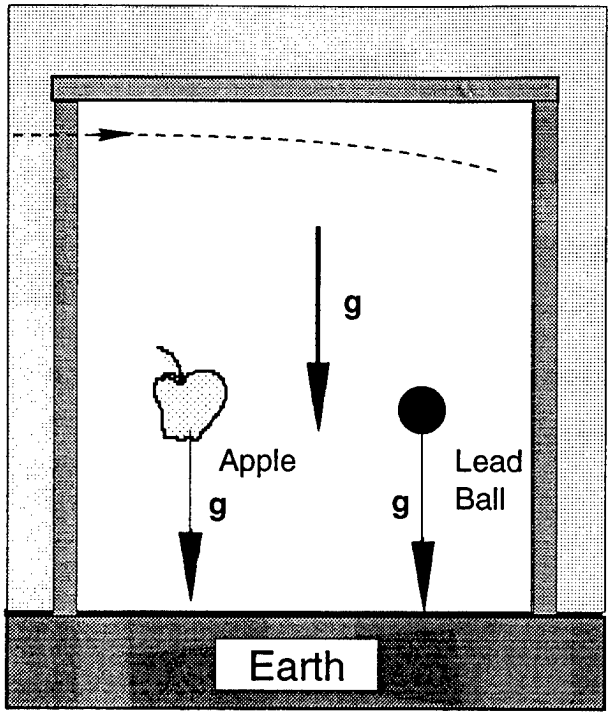


Fig. 9a. All objects fall with equal accelerations in a laboratory near the earth's surface.

The equivalence of the two laboratories implies that a beam of light is deflected toward the source of the gravitational field. Let a beam of light–which travels in a straight line in free space–enter the side of the accelerated laboratory (near the top, in Fig. 9b). The observer in this laboratory is accelerated past the light, so it must appear to fall down just as do the massive objects. The experiment must have the same outcome in the non-accelerated laboratory on earth, so to an observer in a real gravitational field light must fall down. A beam of light passing near any massive body will be deflected towards the body.

Time Delay

If one imagines the wavefronts in a beam of light as the beam is deflected toward the massive source of a gravitational field, then one can picture the portions of the wavefront nearest the mass being slowed down

slightly with respect to the portions of the wavefront farther away from the source. The wavefront then tilts over and the beam is thereby deflected. This means that of two beams of light passing near a massive source, the one which passes closer will take longer to get by. Thus not only is light deflected, it is slowed down by a gravitational field.

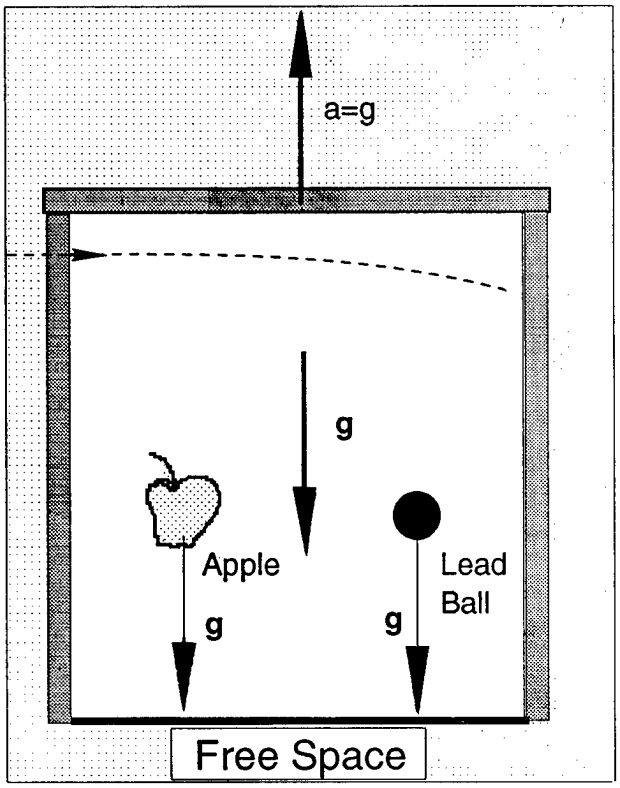


Fig. 9b. By the Equivalence Principle, experiments performed in an acclerated lab in free space have the same outcomes.

Time delays of signals in the neighborhood of the earth can be a few tenths of a nanosecond. Such time delays are determined by a complicated logarithmic function of signal path parameters, times the quantity $4GM_E/c^3$, where G is the Newtonian gravitational constant and M_E the earth's mass. For earth $GM_E/c^2 = 0.443$ cm, so the scale of such effects near earth is

$$\frac{4}{c}\frac{GM_E}{c^2} \approx \frac{1.77 \text{ cm}}{c} = .06 \text{ ns.} \qquad (20$$

This is not enough to worry about at the present time but could be significant in the future–a timing error of .1 ns in a navigational system would give rise to a 3 cm error in position.

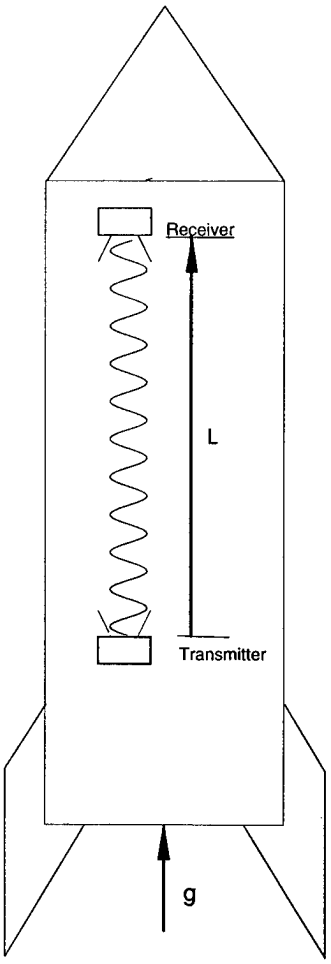


Fig. 10. A signal travelling upwards in a gravitational field is shifted towards lower frequencies.

Gravitational Frequency Shifts

It follows from the Equivalence Principle that an electromagnetic signal passing upwards in a gravitational field will be redshifted. In Fig. 10 is a sketch of an experiment performed in an equivalent laboratory, a rocket having acceleration g upwards in free space. Imagine the situation from the point of view of a non-accelerated frame. Suppose a signal leaves the accelerated transmitter at the initial instant, when the transmitter velocity is still zero. The signal upwards a distance L, and is received by the accelerated receiver. The time required for the signal to propagate from transmitter to receiver is:

$$t = L/c. \qquad (21)$$

During this time, the receiver has picked up a velocity

$$v = gt = gL/c. \qquad (22)$$

To the receiver, the signal appears to come from a receding source and is Doppler shifted. To a first approximation the fractional frequency shift is $\Delta f/f = -v/c$; therefore the fractional frequency shift in the 'effective' gravitational field g is

$$\frac{\Delta f}{f} = -\frac{v}{c} = -\frac{gL}{c^2}. \tag{23}$$

The quantity gL can be interpreted as the change in gravitational potential, $\Delta\phi$, of the signal.

At the surface of the earth,

$$g/c^2 = 1.09 \times 10^{-13} \text{ per km}, \tag{24}$$

which is very important for today's time standards. For example a signal of definite frequency originating at mean sea level would be redshifted by 1.79 parts in 10^{13} upon arriving at the altitude of the NIST frequency standards laboratory in Boulder, CO. Consequently the contribution of the NIST time standard to Universal Coordinated Time (UTC) requires that a paper correction of -15.5 ns/day be applied to the NIST clock before it can be compared to time standards at mean sea level.

For a clock in a satellite orbiting the earth at 100 km altitude compared to one on the geoid,

$$\frac{\Delta\phi}{c^2} = 1.08 \times 10^{-11}. \tag{25}$$

Not only will these effects be large in the future when clock stabilities approach a part in 10^{15} or better, it will be necessary to compute them quite accurately. This will mean, for example, that there will be a need for improved precision of the ephemerides of clock-carrying satellites.

The Global Positioning System

The best existing example of an engineering system in which relativity plays an essential role is the GPS. This consists of a constellation of perhaps 24 earth-orbiting satellites carrying atomic clocks which synchronously transmit navigation signals, much as described in the discussion of Fig. 2. The satellite orbits are at approximately 20,200 km altitude. Therefore clocks in the satellites will be significantly blueshifted in rate, compared to clocks on the ground. The second-order Doppler shift of such clocks was given in Eq. (20). Also, if the orbits are not perfectly circular (and they almost never are), the clocks' yo-yo motions towards and away from the earth will generate periodic additional gravitational frequency shifts, and second-order Doppler shifts. Further, observers on the ground who wish to make use of the navigational signals will experience the Sagnac effect due to earth's rotation.

A complete discussion of all the significant relativistic effects, with analytical expressions for the necessary corrections, can be found elsewhere.[4] Here I shall just indicate roughly the magnitudes of some of the corrections.

First, consider ground-based clocks in receivers which are at rest on the earth's surface. Standard clocks on the geoid are used to define the unit of time; however, from the point of view of a local, nonrotating frame, there is a frequency shift due to earth's mass; the fractional frequency shift is about -7×10^{-10}. The earth's oblateness is associated with a quadrupole contribution to the gravitational potential which cannot be neglected; the fractional frequency shift is about -4×10^{-13}. If earth-based clocks are not on the geoid they suffer a gravitational frequency shift (see Eqs. (23-24)). Finally there is a second-order Doppler shift due to the earth's rotation; the fractional frequency shift from this effect can be as large as -1.2×10^{-12} (see Eq. 17).

For GPS receivers in motion relative to the earth's surface, there is an additional second-order Doppler shift due to their speed with respect to the ground; this can be of the order of 10^{-12} depending on the ground speed. Also, the Sagnac effect–or motion of the receiver during propagation of the navigation signal–may give rise to effects of several hundred nanoseconds magnitude.

The transmitters themselves suffer a frequency shift due to the earth's gravitational potential, and a second-order Doppler shift due to orbital motion; these effects are several parts in 10^{10}. The additional frequency shifts due to orbital eccentricities can be tens of nanoseconds; for a GPS satellite of eccentricity $e = .01$, the maximum size of the effect is about 23 ns.

Propagation of signals from transmitter to receivers are subject to the Sagnac effect, involving relativistic corrections of up to several hundred nanoseconds. Relativistic time delay of signals or relativistic deflection of signals is a few tenths of nanoseconds and is currently neglected in the GPS.

The Concept of Coordinate Time

With so many significant relativistic effects occurring on earth-fixed and earth-orbiting clocks, the problem of synchronization of the clocks becomes an acute one. Rates are affected by motional and gravitational effects; synchronization on the spinning earth is inconsistent if the Einstein procedure is used. How is it possible to synchronize a network of distributed, rapidly

moving clocks so that a navigational system will work as conceived in Fig. 2? What has been found to work extremely well in the GPS is to use the time in a hypothetical underlying local inertial frame, with origin attached to the earth but not spinning, as the measure of time. This time is not time on any standard clock orbiting the earth; instead one makes use of general relativity to correct the readings of such clocks so they would agree with hypothetical clocks at rest in the local inertial frame. The time obtained by so correcting all the clocks in the system, is an example of *coordinate time*.

Thus, imagine an underlying nonrotating frame, or local inertial frame, unattached to the spinning earth, with with its origin at the center of the earth. This frame is sometimes called the "Earth-Centered Inertial" frame, or ECI frame. In this frame, introduce a fictitious set of standard clocks available anywhere, all synchronized via the Einstein procedure, and running at agreed upon rates such that synchronization is maintained. Gravitional effects are incorporated by choosing one clock as a Master Clock and requiring that all other clocks be syntonized to the Master clock by simple transmission of signals without any frequency shift corrections. The resulting time scale is called coordinate time.

Now introduce a set of standard clocks distributed around the surface of the rotating earth, or orbiting the earth. To each one of these standard clocks a set of systematic corrections is applied, so that at each instant the standard clock as corrected agrees with the time on a fictitious standard clock, at rest in the ECI frame, with which it instantaneously coincides. The set of corrected standard clocks will therefore be keeping coordinate time. In other words, coordinate time is equivalent to time measured by standard clocks in the ECI frame.

Time measured on coordinate clocks has two highly desirable properties. First, synchronization is reflexive: if A is synchronized with B, then B is synchronized with A. Second, synchronization is transitive: if A is synchronized with B, and B is synchronized with C, then A is synchronized with C. Internal inconsistencies are therely eliminated.

GPS time is an example of coordinate time. To an observer on the earth's geoid, a standard clock in a GPS satellite in a nominally circular orbit would appear to be blueshifted by .4465 parts per billion, or about 39,000 ns per day; this is a net effect of gravitational frequency shifts and motional Doppler shifts of satellite clocks relative to reference clocks fixed on the ground. To compensate for this, the 10.23 MHz reference frequency of satellite clocks is adjusted downward to 10.229 999 995 43 MHz. The adjustment is accomplished on the ground before the satellites are launched.

Also, if the orbit of the satellite clock is not perfectly circular, there will be additional gravitational and motional rate shifts which have to be accounted for. The additional correction required to achieve synchronization when the orbit eccentricity is not zero is given by the expression[4]

$$\Delta t = +4.428 \times 10^{-10} e\sqrt{a}\,\sin E \text{ sec}, \qquad (26)$$

where a is the semi-major axis in meters and E is the eccentric anomaly. Usually the software in the user's receiver makes this correction.

Application of Satellite Navigation in Geodesy

The motivation to obtain accurate measurements of movements of the earth's crustal plates is intense. Knowledge of these very slow motions is crucial to the development of improved earthquake prediction capability; the potential impact on construction codes, building restrictions, etc., is considerable. In recent years the GPS has been successfully used to measure very long baselines between fiducial points on different crustal plates by a method described as "carrier phase double difference." Two receivers are placed at the ends of a baseline of interest, and signals from two satellites are then "double differenced" in a manner to be described below. Differenceing removes the need for some systematic corrections but as will be seen, there are residual relativistic effects which must be accounted for.

Referring to Fig. 6 and the propagation time t given in Eq. (12), let the satellite position at the instant of transmission t_S be denoted by $\mathbf{r}_S$ and the receiver or observer position at the same instant be denoted by $\mathbf{r}_O$. Let the coordinate time of arrival of the signal at the observer be denoted by t_O. Then solving Eq. (12) for the propagation time gives

$$t_O = t_S + K + \frac{|\mathbf{r}_S - \mathbf{r}_O|}{c} + \frac{2\omega}{c^2} \cdot \left[\frac{1}{2}\mathbf{r}_S \times \mathbf{r}_O\right]. \qquad (27)$$

The last term is the Sagnac correction and K represents a possible time offset or error of the receiver's clock. The rate adjustment applied to satellite clocks means that the quantity t_S will have the correct scale when received on the geoid. There is a further correction, from the non-circular motion of the satellite, given by Eq. (26). Thus when all relativistic effects

are incorporated,

$$t_O = t_S + K + \frac{|\mathbf{r}_S - \mathbf{r}_O|}{c} + \frac{2\omega}{c^2} \cdot \left[\frac{1}{2}\mathbf{r}_S \times \mathbf{r}_O\right] + 4.428 \times 10^{-10} e\sqrt{a}\,\sin E. \tag{28}$$

Let subscripts 1 and 2 denote the two different satellites and the two different observers. Suppose there are receivers at two different positions which receive a time signal originating from a single satellite. Upon taking the first differece of the arrival times, it is immediately seen that the eccentricity term cancels out, leaving the expression:

$$t_{O2} - t_{O1} = K_2 - K_1 + \frac{|\mathbf{r}_S - \mathbf{r}_{O2}|}{c} - \frac{|\mathbf{r}_S - \mathbf{r}_{O1}|}{c} + \frac{2\omega}{c^2} \cdot \left[\frac{1}{2}\mathbf{r}_S \times (\mathbf{r}_{O2} - \mathbf{r}_{O1})\right]. \tag{29}$$

The Sagnac correction is still needed. The time of transmission of the signal, t_S, cancels out which lessens the impact of selective availability.

Now the same set of measurements is taken, at essentially the same time, using a second satellite. Writing another equation similar to Eq. (29) for the second satellite and taking the difference, it can immediately be seen that even the clock offsets in the receivers cancel out, leaving only the usual propagation delay terms with relativistic corrections due to the rotation of the earth:

$$\begin{aligned}(t_{O2} - t_{O1})_{S2} - (t_{O2} - t_{O1})_{S1} &= \frac{|\mathbf{r}_{S2} - \mathbf{r}_{O2}|}{c} - \frac{|\mathbf{r}_{S2} - \mathbf{r}_{O1}|}{c} \\ &- \frac{|\mathbf{r}_{S1} - \mathbf{r}_{O2}|}{c} + \frac{|\mathbf{r}_{S1} - \mathbf{r}_{O1}|}{c} \\ &+ \frac{2\omega}{c^2} \cdot \left[\frac{1}{2}(\mathbf{r}_{S2} - \mathbf{r}_{S1}) \times (\mathbf{r}_{O2} - \mathbf{r}_{O1})\right].\end{aligned} \tag{30}$$

The Sagnac correction is still necessary. In this application the correction is largest when the baseline is at right angles to the line between the satellites; it can be several hundred nanoseconds.

In Fig. 11 are plotted some baseline measurement data taken repeatedly on baselines in the Southwest Pacific, of lengths up to 2500 km.[5] Only the length of the baseline is shown here. The vertical scatter in the plotted points gives a measure of the errors involved. For the 2500 km baseline the spread is only a few cm.

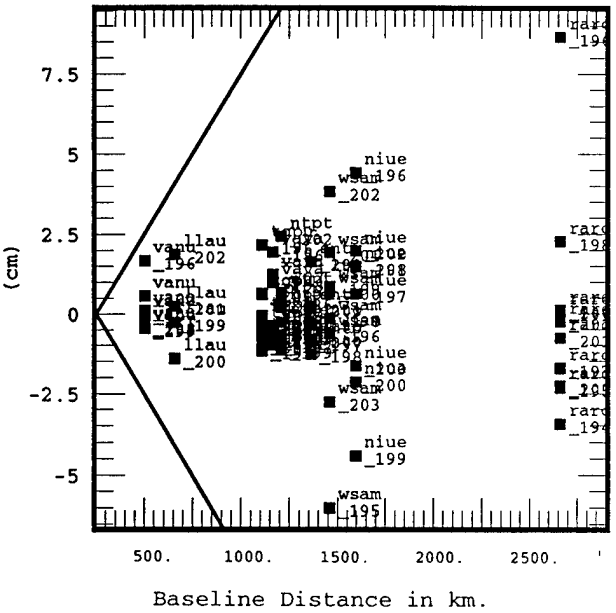


Fig. 11. Scatter in baseline lengths for several different baselines measured during the Southwest Pacific 1992 GPS Campaign. The data were provided by UNAVCO.

Impact on Fundamental Metrology

The previous sections have been devoted almost exclusively to the impact of relativity on the measurements of time, with distance derived by multiplying by c. At the level of a centimeter or less, there are additional effects on the measurement of position which arise because space in the neighborhood of a massive body is distorted. Consider as in Fig. 12 an attempt to establish a system of spatial coordinates in the neighborhood of earth, against which to measure the positions of the earth's crustal plates. Suppose that we wish to measure angles in the usual Euclidean way, so that a circle of coordinate radius r centered on the earth would have a circumference $2\pi r$, measured with standard rods or with the help of the constancy of c. Two such circles, of coordinate radii r_1 and r_2 are indicated in Fig. 12. The standard distance from the inner circle straight out along a radius to the outer circle is not $r_2 - r_1$; instead one finds the standard distance d is

$$d = r_2 - r_2 - \frac{2GM_E}{c^2}\ln(r_2/r_1). \tag{31}$$

The correction due to space curvature is of the order of 1 cm.

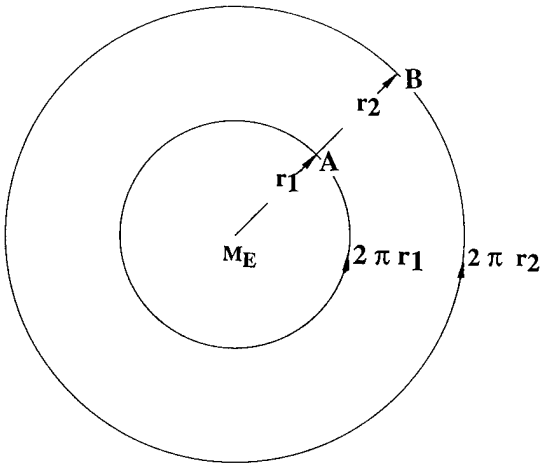
More generally, the fact that c has a defined nu-

merical value means that the physical unit of length depends on the clock used to define the unit of time. For example, in Barycentric Dynamical Time (TDB), the unit of time is the same as that of clocks on earth, in orbit around the sun, and the point of view taken is that of an observer in a reference frame at rest with respect to the solar system barycenter. The clocks on earth beat more slowly than clocks at rest at infinity in this system by the factor[6]

$$1 - L = 1 - 1.55 \times 10^{-8}. \qquad (32)$$

Therefore, the meter is physically longer, so the length of a physical object is numerically smaller by this factor. The mass of the earth can be used to construct a quantity having the physical dimensions of a length, namely GM_E/c^2. However c has a defined value; this means that in TDB coordinates, GM_E is numerically smaller than in SI units:

$$(GM_E)_{TDB} = (1 - L)(GM_E)_{SI}. \qquad (33)$$



Standard distance from A to B:

$$r_2 - r_1 - (.888\ \text{cm}) \ln(r_2/r_1)$$

Fig. 12. Effect of spatial curvature on standard distance measurements.

Some Remarkable Cancellations

So far in this discussion I have ignored the fact that the earth is actually an oblate ellipsoid; clocks near one pole will be closer to the center of the earth and will therefore be subject to a gravitational redshift; on the other hand in the ECI frame such clocks are moving more slowly than clocks near the equator and are subject to less second-order Doppler shift. This is diagrammed in Fig. 13. Over the ages the earth's surface has assumed the approximate shape of a hydrostatic equipotential in the rotating frame–the average shape of the ocean's surface defines the geoid. it is a remarkable fact that on the geoid, there is a very precise cancellation of gravitational frequency shifts and motional Doppler shifts, so that all clocks at rest on the geoid beat at the same rate! Therefore it is possible to construct a network of standard clocks on the earth's geoid, all beating at the same rate. However, to synchronize these clocks consistently it is necessary to correct for the Sagnac effect, due to the earth's rotation.

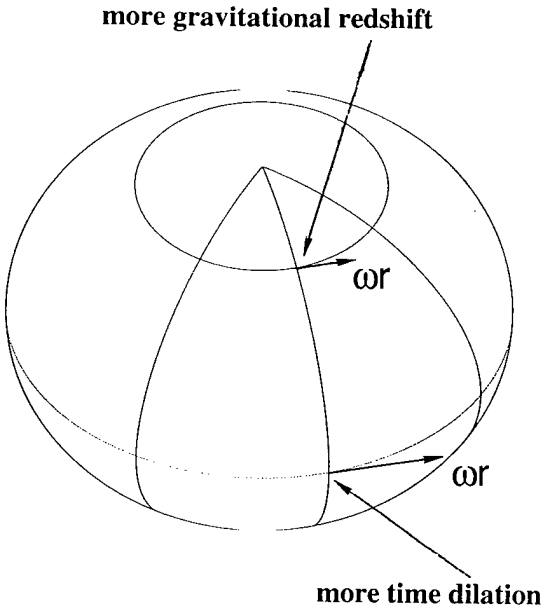


Fig. 13. On the oblate rotating earth's geoid, changes in gravitation frequency shift are precisely compensated by second-order Doppler shifts.

So far I have also ignored the possibility that the sun, moon, or other planets might contribute to gravitational frequency shifts. Also, the earth's orbit is not perfectly circular so one might expect a yo-yo effect on the rates of earth-orbiting clocks somewhat analagous to the correction given in Eq. (25) for GPS clocks. For example, when a satellite is in earth's shadow its clock should be gravitationally blueshifted as compared to a satellite-borne clock between the sun and earth. For such a configuration, the fractional frequency shift be-

tween clocks in the two satellites, due to the sun, is about three parts in a trillion, which in an hour would cause a 12 ns timing error to build up. Fortunately we do not have to worry about this! This effect is cancelled to high precision by other relativistic effects arising because the entire system of earth plus satellites is in free fall around the sun. By the principle of equivalence, we should not be surprised that for a system in free fall, the gravitational effects of the sun are transformed away. Detailed analysis of this situation is rather delicate; when comparing clocks in the ECI frame–which is falling around the sun–with clocks in the solar system center-of-mass frame, there is disagreement about the meaning of simultaneity in the two frames. Using coordinate time in the ECI frame, with clocks synchronized by the Einstein procedure (modified by gravitational effects), the gravitational effects due to other solar system bodies will cancel to high accuracy. The residual gravitational effects are due to tidal potentials only, and are less than a part in 10^{16}.

Conclusion

In this paper, numerous examples of relativistic effects which are important for current and future navigation, timing, and communications systems have been discussed. Relativistic effects are always systematic, but depend on knowledge of the positions and velocities of the various clocks in the given reference frame. These effects are not noise; they are well-understood, and can be corrected for to a high level of accuracy. As clock stability and accuracy continues to improve it will become increasingly important for system designers and practitioners to become familiar with these effects so they will be accounted for properly. I hope this paper helps in a small way to educate those for whom the mathematical apparatus of general relativity is excessively cumbersome.

References

[1] R. F. C. Vessot, E. M. Mattison, M. W. Levine, and R. L. Walsworth, "Status of local oscillators for operating ultra-high resolution frequency discriminators as frequency standards," 24^{th} Annual PTTI Applications and Planning Meeting, McLean, VA, December 1-3, 1992.

[2] A. Einstein, *The Meaning of Relativity*, 3^{rd} ed., Princeton Univ. Press, Princeton, NJ, (1950).

[3] C. W. Misner, K. S. Thorne, J. A. Wheeler, *Gravitation*, W. H. Freeman & Co., San Francisco, CA (1993), p. 13 ff.

[4] N. Ashby, "A Tutorial on Relativistic Effects in the Global Positioning System," NIST Contract No. 40 RANB9B8112, Final Report, Feb. 1990. Copies of this report are available from the author on request.

[5] John Braun, UNAVCO, private communication.

[6] R. W. Hellings, "Relativistic Effects in Astonomical Timing Measurements," *Astron J.* **91**, 650 (1986).

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PRECISE TIMING IN ELECTRIC POWER SYSTEMS

Kenneth E. Martin

Bonneville Power Administration
PO Box 491 m/s ELIP
Vancouver, WA 98666

Abstract

Electric power systems in North America are interconnected into several massive grids that span the continent. AC synchronism within the grids and 60 Hz frequency is maintained by matching power generation levels to transmission capability and load patterns. Controls are primarily protection relays and set point stability monitors.

Cost and environmental concerns discourage building new facilities. This requires using existing facilities more efficiently including heavier power transfer over existing lines, better scheduling of both generation and load, and integration of diverse power sources into the network. A new generation of controls is needed to maintain stability as these systems become more highly stressed.

Precise timing offers a new approach to system wide controls. It enables measuring the system wide AC phase as well as timing events and synchronizing controls. The Global Position System (GPS) is the first system to provide timing with enough precision to make this approach practical. GPS also provides the wide area of coverage, continuous availability, and high reliability required for power system control.

A number of utilities, including the Bonneville Power Administration, have R & D programs for GPS timing systems or advanced protection and control systems based on precise timing. Several utilities have operational phase angle measuring systems and more will go into service in 1993. At least two utilities will start using GPS based fault location in 1993. The use of precise time will continue to grow as more equipment becomes available, applications are developed, and confidence in time based control develops.

Introduction

Time and frequency systems are an integral part of operations and maintenance of electric power systems. They are used for scheduling, billing, coordination, and analysis. They are used to control power system frequency, timetag disturbances, and locate power line faults. In this paper I will survey power system operations that require time and frequency, discuss some of the related research projects at the Bonneville Power Administration (BPA), and finally indicate some potential directions of future developments.

BPA is a Federal power marketing agency whose service area includes Washington, Oregon, Idaho, and Western Montana. It operates the largest Northwest (NW) power grid and controls 23,000 megawatts (MW) of generation. It is a member of the Western States Coordinating Council (WSCC), the western U.S. power grid. Examples in the paper will be drawn from this system.

Power System Basics

The simplest power system is a generator connected to a resistive load. There is no need for reactive compensation, no need for synchronization with other generators, and no concern for power signal frequency. Indeed, the first power systems were used primarily for lighting and were that simple. When motors and transformers are introduced, some frequency control is required. Long transmission lines introduce reactance so voltage control can become a problem. Increasing power demand is satisfied with more generators which require additional synchronization.

Today electric power systems are interconnected into four massive grids that span North America.[1] WSCC is the largest of those grids, covering the US and Canada from the West coast to just east of the Rocky Mountains. There are many benefits to interconnecting the individual systems. System reliability is enhanced by multiple sources and many transmission paths. Sharing generation can result in significant savings in meeting peak loads, staggering maintenance, and using the most economical sources. However interconnection also requires

U.S. Government work not protected by U.S. copyright

synchronization and controls to prevent a problem in one part of the system from causing problems elsewhere.

There are two distinct synchronizing issues found in power systems. First, electric energy must be used as it is generated. No one has created a successful electricity reservoir or battery suitable for power system use. Load must be constantly in balance with generation. Second, power is produced and transmitted primarily as a 60 Hz (50 Hz in Europe) alternating current. Synchronous devices such as generators and motors must be kept in phase with each other; they must constantly track through frequency changes and disturbances.

Electric power is produced primarily by turning an alternator with a steam or water turbine. Power P transferred from the turbine is expressed by $P = T\omega$ where T is the torque and ω is the angular velocity. A decrease of electric load on the generator reduces the resisting torque so the speed of rotation will increase to absorb the turbine power. Conversely, an increase in electric load will cause the machine to slow down. This characteristic inherently stabilizes and synchronizes a group of generators. Each generator will find an operating point and the constant small load changes will automatically redistribute among the group. Similarly, if mechanical power input to one of a group increases, it will accelerate and take more load. The increased load will control its acceleration and the decreased load on the other units will allow them to speed up, keeping the whole group synchronized.

Power transfer between two substations in an AC system is defined by

$$P = \frac{V_1 V_2 \sin\emptyset}{Z} \qquad [1]$$

where V1 and V2 are the terminal voltages, Z is the line impedance between them, and $\emptyset$ is the included voltage phase angle. Terminal voltages are controlled to be nearly constant with inductors and capacitors. Since the line impedance is fixed, the power transfer is determined primarily by the phase angle between stations.[2]

Present Systems

Generation Controls

Each power grid is divided into a number of interconnected control areas (usually individual utility service areas). Each area controller is responsible for maintaining generation-load balance and consequently frequency within that area. The control areas are connected with tie lines (Figure 1). Power is scheduled over the lines to maintain load balance within each area, and errors show up as inadvertent interchange over those lines. An error in one area will cause an error in another area. The control error (E) is computed by[3]

$$E = (T_1 - T_0) - 10B(F_1 - F_0) \qquad [2]$$

where T1 = area interchange (MW)
T0 = interchange schedule (MW)
F1 = actual system frequency (Hz)
F0 = scheduled frequency (Hz)
B = area frequency bias (MW/0.1 Hz).

The area controller measures power interchange and frequency and computes interchange error. An Automatic Generation Control (AGC) is usually employed to minimize the control error by continuously adjusting generation. Errors in frequency or power measurement will show up as inadvertent interchange or a system frequency offset. For example, BPA with B = 250 MW/0.1 Hz, would compute 25 MW in error if the frequency measurement were off by 0.01 Hz. AGC would minimize E and consequently cause a 25 MW inadvertent interchange. This is less than 1 % of a typical 3000 MW area interchange schedule, but illustrates the need for accurate frequency measurements.

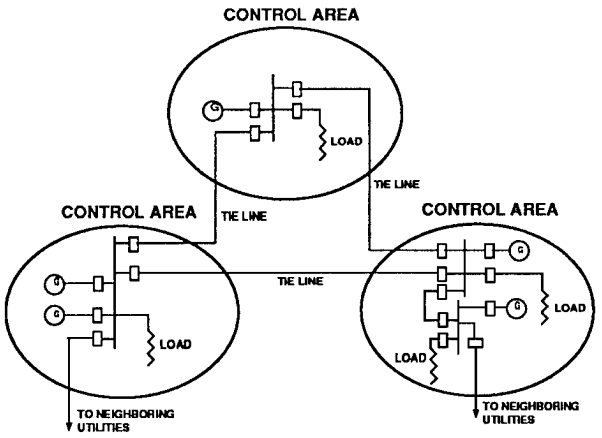


Figure 1. Diagram of Control Areas with generation and load interconnected with Tie Lines.

As a further complication, each utility must estimate and schedule generation to serve its load. They may generate their own power, but most buy some from other producers. Load may vary by 300% throughout the day so generation is scheduled in hourly increments. A generation dispatcher can make schedule adjustments throughout the day to maintain balance, although a gross imbalance such as a generator failure may require power purchase on the spot market, possibly at great expense. Advance scheduling is essential since startup can take a whole day for a large thermal plant. That also means some generation must be kept on line and running well below capacity for emergency backup. If generation is lost this "spinning

reserve" replaces the power and prevents system collapse. Since spinning reserve is expensive, only a minimal amount is kept running to serve the entire grid.

Close operational coordination throughout a grid is essential to maintain proper power flows. A major utility is designated the system timekeeper. They integrate the 60 Hz power signal and compare it with an accurate clock to compute accumulated time error ΔT. If the accumulated time error exceeds the allowable limit, a new frequency is scheduled to bring the ΔT back to zero. WSCC allows up to a 2 second error before rescheduling.[1]

System Protection

A catastrophic event such as generator failure or transmission line outage requires instantaneous response. Protective devices called relays (which originally were electromechanical relays) detect power line faults and loss of synchronism. They operate circuit breakers and send control signals for remedial action schemes. Their purpose is to take appropriate action fast enough to protect equipment and minimize system disturbances. Relaying is predicated upon analysis of what will happen in various scenarios under various operating conditions. Power systems have become increasingly complex making traditional relaying more difficult to do successfully. New concepts of adaptive relaying where the relay adapts to current operating conditions are being developed. Advanced measurement methods discussed later in this paper are essential for adaptive relaying.

Some protection equipment may never be required to operate during its lifetime of 10-20 years; others will operate constantly. It must reliably sit for long periods of time in a ready mode. Each operation is a test of the system that no one wants to repeat, but every engineer wants to thoroughly analyze. Consequently power systems are monitored extensively with automatic recording equipment. Two of the most common are the Digital Fault Recorder (DFR) and the Sequence of Events Recorder (SER). The DFR records the actual power line voltage and current waveforms that occur during a disturbance. With this the protection engineer can determine the actual conditions to which the control system responded. The SER records the relay and switch responses. From this the engineer can determine how the control system responded. Sometimes the DFR and the SER are combined into one instrument which makes correlation of the data easy. When they are separate, timetags are required for data correlation. Recordings made on separate instruments in the same station can be hard wired to a common source so it does not matter particularly if the device is "on time." But if we wish to compare recordings between distant stations, "on time" is essential. Utility wide synchronization at the millisecond level is required to adequately timetag DFR and SER records. The WSCC coordination goal is a minimal 8 ms across the system and 1 ms within each station.

Fault Location

Another important monitoring device is a fault locator. A short circuit or fault usually can be cleared by momentarily disconnecting the line. Occasionally equipment is damaged and repair is required. Automatic fault location is much faster and cheaper than patrolling the entire line.

When a fault occurs on a transmission line, the current increases, voltage decreases, and a high frequency wave propagates in both directions from the point of incidence at nearly the speed of light. Fault location has relied on computing the distance from the apparent line impedance based on voltage and current. This technique does not work well with series compensation (capacitors) and can be thrown off by load current and mutual coupling. However, location can also be computed by comparing the arrival times of the traveling wave at substations on either side of the fault (Figure 2). BPA designed and built the Fault Location Acquisition Reporter (FLAR) based on this principle.[4,5]

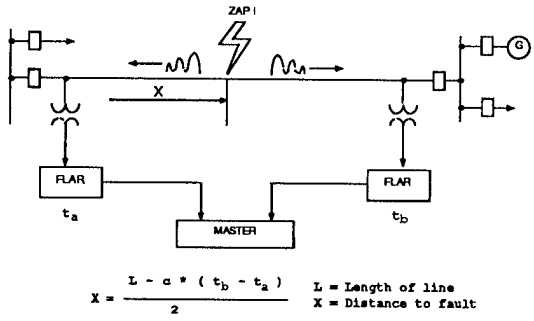


Figure 2. FLAR traveling wave fault locator diagram.

The FLAR system has microcomputer units installed at 24 key substations. Each has a precision clock synchronized by a high frequency pulse sent over the microwave system. The traveling waves are timetagged and reported to a master computer at the control center. The master correlates the timetags, computes the fault location and reports it to system dispatching.

The system has proven to be accurate and reliable. The drawback is that the synchronization pulse sent over the microwave system uses 60 Khz of bandwidth, requires an analog microwave, and is only available to major stations within BPA's service area. The system is currently being synchronized and extended with GPS.[6]

Power Industry Requirements

Industrialized nations are increasingly dependent on continuous, reliable electric power. Outages are costly for both power producers and power users. Power systems are heavily capitalized and must operate near capacity to remain economical. The equipment is expensive; mistakes and failures can cause millions of dollars of damage in seconds. Consequently protection and control systems must be highly reliable and always available.

Likewise, time and frequency systems for power system controls must be reliable, available, and accurate. They must operate continuously over large geographical areas through all weather and seasonal conditions. They must operate unattended but reliably for long periods of time. They need to have low periodic maintenance requirements and alarm for failures. They have to be economical enough to be paid for by the benefit they bring over the life of the system (which can be 15-20 years). Above all they must be reliable.

System Function	Measurement	Optimum Accuracy
Fault Locator	300 meters*	1 microsecond
Relaying	1000 meters	3 microseconds
Stability Control	±1 degree	46 microseconds
State Estimator	±1 degree	46 microseconds
Oscillograph		1 millisecond
Event Recorder		1 millisecond

*300 meters is the distance between two HV towers.

Figure 3. Power System Timing Requirements Summary.

Present timing accuracy requirements are summarized in Figure 3. A system that can deliver microsecond accuracy and meet the other industry requirements would be an ideal source. GPS now meets those requirements. It provides a highly redundant, multiple access time service. With multiple satellites there is little problem with signal blocking. Power system noise does not interfere with GPS's L-Band spread spectrum signal. The basic 200 ns accuracy meets accuracy requirements very well. Even the ±1 μs maximum S/A error is within requirements. The antennas are simple, easy to install and don't need adjustments. Finally, the receivers themselves are mostly digital, requiring little tuning or adjustment. Eventually receivers will be made on a single chip or two increasing reliability and reducing cost. Although other timing systems have been proposed and may be developed in the future[7], GPS with its present quality of service will satisfy power system needs for the foreseeable future.

Applications R & D

Electric power utilities have historically been involved in Research and Development of power system controls. Many system developments are new advances in the state of the art and there are no off the shelf solutions. Two areas of research BPA has been involved with, precise timing and phase angle measurement, will be discussed.

Time Dissemination and FLAR Extension

BPA requirements called for fault location on a line that extended outside of its service area. The synchronizing pulse could not be reliably sent over a digital microwave. The FLAR remote was finally synchronized with GPS receivers. Following that the FLAR functions were built into a GPS receiver. GPS-FLAR receivers could be the basis for a complete closed loop time system (Figure 4).

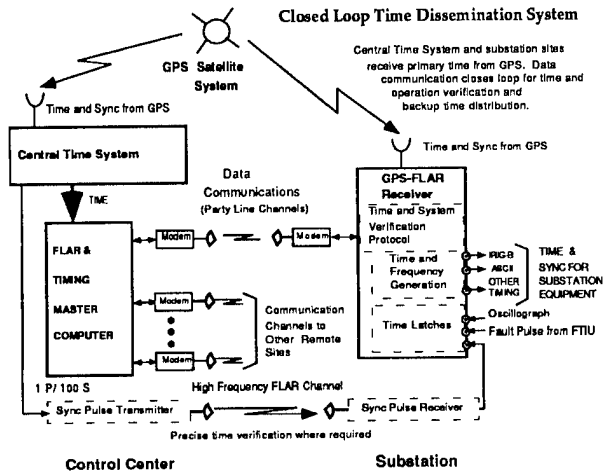


Figure 4. Closed Loop System. GPS provides primary time with time and alarm monitor on data channel; backup sync uses timetags.

With this proposed system GPS-FLAR receivers could be installed at all substations requiring time or fault location. Data communications closes the loop with time verification and alarm reporting as well as timetag reporting for fault location, where required. For applications where precise time verification is required, an additional timetag input channel can verify time from an external source (such as the sync pulse) to within 100 ns.

Presently the FLAR system does not disseminate time of day, only precise synchronization. Time for all other applications is primarily broadcast as an IRIG-B signal over the microwave. A few stations that do not have microwave use GOES receivers with moderate success. IRIG-B over the microwave has never worked very well and is marginal with the automatic decoding equipment presently used. BPA is considering replacement of its

present time dissemination system with this GPS based alternative. Cost is still high compared with the present system. Simpler, independent GPS receivers are also being considered. In the mean time more GPS-FLAR remotes are being installed for tie line fault location. Laboratory receiver testing is continuing, particularly to find better ways to diagnose receiver degradation before failure. The new central time system at the power system control center will have GPS synchronization.

Synchronous Phase Angle Measurement

A phasor is a vector representation of a sinusoidal quantity which shows both magnitude and relative phase angle. A power signal is a 60 Hz sinusoid and is commonly analyzed in phasor format. Dr. Arun Phadke at Virginia Polytechnic Institute and State University developed a microcomputer system that will accurately compute power system phasors in real time. The Phasor Measurement Unit (PMU) digitizes the three phase waveforms at a 720 sample/second rate. After each sample the 60 Hz component is derived by a Fourier transform using the last 12 samples. The three 60 Hz components are combined into a single positive sequence phasor which represents the magnitude and phase of a balanced three phase system. It is a good representation of the state of a real power system in all but extreme fault conditions. By precisely timing the sampling clock with a GPS receiver, phase angle can be accurately computed between any two measurement points.[8]

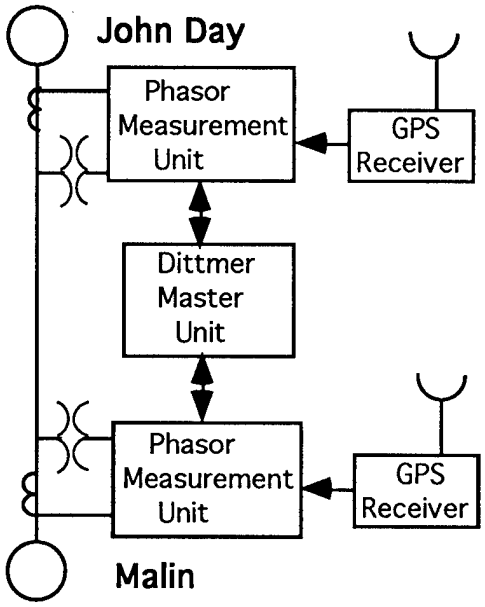


Figure 5. Phasor Measurement System Test Diagram

BPA tested two prototype PMU's on BPA's 500 kVAC Pacific Northwest-Southwest Intertie, the main transmission link between the two regions. PMU's were installed about 250 miles apart at John Day Substation along the Columbia River and Malin Substation at the California border. A master terminal at the BPA Laboratories in Vancouver, Washington recorded the data. The two PMU's were synchronized by GPS receivers (Figure 5).

Figures 6 and 7 are phasor data recordings of a disturbance caused by a bus fault which tripped off an 1100 MW nuclear power plant near Richland, Washington. A sharp voltage dip is seen at both John Day and Malin during the fault (about 30 milliseconds duration). It is followed by a voltage rise and decrease in phase angle between John Day and Malin as the power flow into California drops off. The system gradually approaches its old operating point during the 2 minutes following the fault.

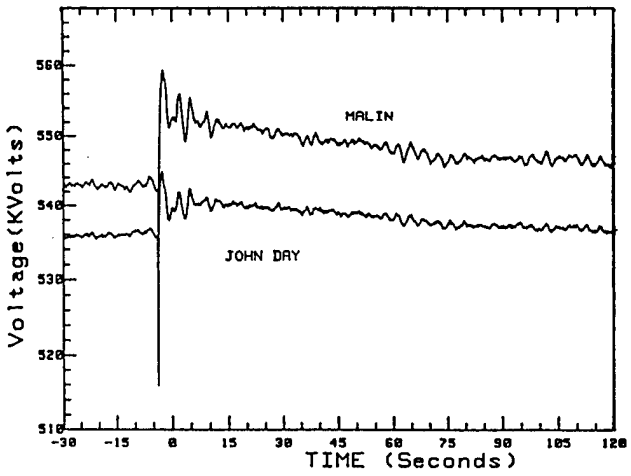


Figure 6. Voltage Magnitude at John Day and Malin Substations after loss of 1094 MW of generation.

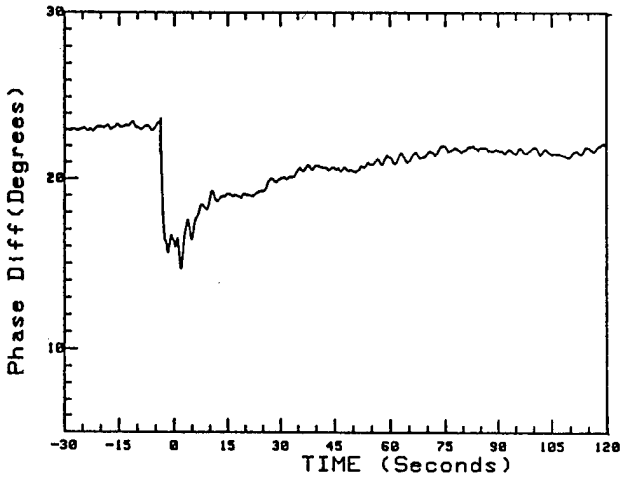


Figure 7. Voltage Phase Angle between John Day and Malin Substations for same disturbance as Figure 6.

The purpose of the test program was both to evaluate the phasor measurement system and to provide operational information on the GPS receivers used for the precise time source. The overall results were excellent. In 4 years of field deployment the only hardware failure was a chip in a GPS receiver.[9] Phasor data responded with greater accuracy and less noise than comparable analog telemetered data. Proposals for expansion of the system into an operational data network are now being considered.

Future Applications

The future will see more power system development and integration. Lines will be loaded more heavily and generation will be used more at capacity. Power will be supplied by more diverse sources from more units. New systems will be invented for generation, transmission, and energy storage. All these advances will require controls. Precise timing is one of the key technologies that will enable the development of new control systems and the monitoring required to maintain them. Some of these areas of potential development are described in the following paragraphs.

System Timing

GPS is the key to making time domain fault location practical. The dedicated resource required for the system BPA has used is too expensive for common use and not available for all utilities. Two utilities, BC Hydro and Hydro Quebec, are deploying time domain fault locators of their own using GPS synchronization. BC Hydro is additionally considering low cost GPS receivers at all substations for system timing.

Most utilities use some kind of time reception at their control centers. Centers with AGC require a steady input which is usually WWVB. These centers must coordinate frequency and time error estimation. Better time and frequency reference will improve control. A common source for all utilities would also improve coordination. GPS covers all areas in North America with a more accurate and reliable signal than any other time service. BPA is replacing its central time system with a GPS based system which will also have a rubidium local reference in the event time signals are lost for any period of time.

The larger utilities have proprietary communication systems. Both microwave and fiber optic systems are digital and require precise frequency sources to prevent frame slips. Fiber has been found to be particularly compatible with power systems. It can be used in the same right of way and is unaffected by electric fields. Consequently several utilities have considered entering the data communication business. Optics are also being used to measure power system voltage and current using special glass that will distort in electric and magnetic fields. Precise frequency sources will be required to maintain the accuracy of these systems.

Stability Controls

Stability control schemes are designed to prevent unnecessary generator shutdown, loss of load, and separation of the power grid. Most equipment is protected by local controls and relays that detect overload conditions and remove the equipment from service. Stability control schemes extend over larger areas to maintain system integrity. When a disturbance is great enough to threaten the system, stability controls respond with system control actions including dynamic braking, generator dropping, load shedding, DC or static var ramping, or controlled islanding (See Appendix for brief description).

Existing control schemes are developed and tested on a computer model. They are based largely on switch position and set point arming. They are tested under every power system condition that is anticipated. They are biased toward security to be sure they will operate for any condition where action is necessary.

Generally they work very well. Most customers experience little loss of service. However they tend to overreact causing more equipment outage and loss of revenue than is necessary. Equipment cannot be fully loaded because a large safety margin is required to allow for unplanned conditions and inexact measurements.

Controls based on real time phasor measurements could improve operation. The controls could be programmed to respond to unstable conditions rather than certain switch positions and set points. The system would adapt to current conditions. The limitation of having to simulate every scenario would be eliminated. Safety margins could be set more consistently. Control actions could be fine tuned eliminate unnecessary operation.

Phase angle measurements have been explored by several utilities.[10,11] Hydro Quebec used a proprietary microwave time dissemination to achieve a 10 μs sync. Pacific Power tried GOES receivers for their system but have turned to GPS. Electricité de France, Georgia Power, Florida Power, New York Power Authority, and the Tennesee Valley Authority are installing systems that measure phase angle.

System State Estimation

System state estimation is a mathematical technique that has evolved for determining stability of a power system based on its characteristic equations. It is used not only for assessing operating point, but for estimating the effect of taking system components out of service for maintenance. The major input requirement is the complex voltages at the power system buses throughout the system.[12] Present

estimators use a least squares algorithm to compute these complex values from power and voltage magnitude measurements. The slow response time (seconds to minutes) renders the system usable only for static analysis. In addition, if data is inconsistent or missing, or the system is in an abnormal condition the solutions may not converge.

A state estimator based on phasor measurements could avoid most of these problems. The system could operate in real time since the complex voltages used in the algorithm would be measured rather than computed. Abnormal phase angles and system oscillations would not affect the process. State estimation could be incorporated into dynamic control schemes.

System Enhancements

Phasor measurements can also be useful in a number of other supplemental controls. Certainly its direct measurement of phase angle between buses as related to equation [1] is a good measure of power flow. Lines that are only lightly loaded and are not required for system stability could be taken out of service to reduce reactive losses. Too much load and line reactance without enough capacitive support can cause voltage collapse and subsequent blackout. The high speed measurement and high accuracy of the phasor process could enhance schemes that monitor reactive support for long transmission lines. It is also ideal for measuring subsynchronous resonance, where a large machine has underdamped response to system perturbations. After a system has been separated into islands following a major disturbance, phase angle measurements would be critical to re-synchronizing and connecting the system. Since most power systems are now controlled by remote control from dispatch centers, accidental islanding is not always apparent. Real time voltage phase angle shown on dispatching displays could prevent accidental connection of systems that are out of phase, an action that usually has disastrous consequences.

Conclusions

Planning, operation, and maintenance of electric power systems is such a broad subject it is difficult to do it justice in one paper. I have tried to focus on the big picture of how present systems work with emphasis on aspects that involve time and frequency. Power systems are undergoing revolutionary changes with the acceptance of electronic and computer controls. In one decade relays have changed from the electromechanical devices they had been for nearly a century to sophisticated microprocessor instruments. The trend is not over yet. New concepts and ideas are being implemented and tested constantly.

Several predictions for the power system future can be made with some certainty. Power use will grow. For economic and environmental reasons, more use will be squeezed out of existing facilities. Power will be produced from more diverse sources. Power grids will become more tightly coupled.

Dealing with these changes will require more sophisticated and reliable measurements and controls, some of which will depend on timing with ever greater precision. GPS is today the quantum leap forward; tomorrow it may be Lasers or Fiber Optic systems or better atomic clocks.

Appendix

Power System Control Techniques

Control actions include dynamic braking, generator dropping, load shedding, DC or Static Var ramping, or controlled islanding. These terms are briefly described in the following paragraphs.

A dynamic brake is a large resistor. When load is momentarily lost, the brake is switched on line briefly to absorb the output of a generator to keep it in phase with the rest of the system. Sudden generator shutdown stresses equipment and restart can take time. A dynamic brake allows a generator to remain on line during a momentary outage. With phasor measurements, brake application could be precisely controlled by phase angle rather than triggered for a fixed time as it is now.

Load shedding is simply switching off load when it becomes too great for the generation or transmission. It is usually done as a last resort when power transfer overloads key transmission lines. Generator dropping is a fast action done in response to a sudden loss of load when there is not enough dynamic braking to maintain synchronism or the loss of load persists. This is also a last resort, especially with thermal units which restart slowly. Controlled islanding is the separation of the system into smaller independent but disconnected areas. Synchronism between areas is not required. Without inter-area power transfer, some load and generation will usually be lost. However, most service will be continued which is better than overloading successive areas until the whole system blacks out. In all these cases, a key stability indicator like system phase angle could improve the reliability and security of these schemes.

DC interties are used between the AC power grids and for several long distance transmission projects. DC ties do not require synchronization and the lines are cheaper to build, but require very expensive converter stations. Since the power transfer depends only on terminal voltage and line resistance, DC ties can be controlled independently from

the AC system to enhance stability. Fast ramping or modulation can damp out AC system instabilities.

A Static Var system uses solid state devices to control current to an inductor-capacitor combination to control AC system reactance. System reactance controls power flow in an AC system. Like DC controls, static var systems can be operated quickly and independently of the rest of the AC system. They don't offer the range of control that a DC link does, but offer the benefit of AC voltage support and lower overall cost. Both systems are solid state - static - and can be operated at high speed for system stability schemes. Phasor measurements provide the high speed and accuracy needed to get the most out of these systems.

References

[1] R.E. Wilson, "Uses of Precise Time and Frequency in Power Systems," Proceedings of the IEEE, Vol 79, No. 7, pp. 1009-1018, July 1991.

[2] C. L. Wadhna, Electrical Power Systems, Wiley Eastern Limited, 1983, Ch 17, pp 514.

[3] N. Cohn, "Considerations in the Regulation of Interconnected Areas," IEEE Transactions on Power Apparatus and Systems, Vol PAS-86, No. 12, pp 1527-1538, Dec 1967.

[4] D. C. Erickson and J. Andres, "Automatic Fault Location Using MICROTIME," BPA Document No. AFL-100-01, May 25, 1982.

[5] J. Esztergalyos, D. C. Erickson and J. Andres, "The Application of Synchronous Clocks for Power System Fault Location, Control, and Protection," Western Protective Relay Conference, October 23, 1984.

[6] M. Street, "Delivery and Application of Precise timing for a Traveling Wave Fault Detector System," Proceedings of 22nd Annual Precise Time and Time Interval Applications and Planning Meeting, 1990, NASA Publication 3116 (1991).

[7] D. Wayne Hanson and B. A. Howe, "Industrial Time Service," National Bureau of Standards, BPA Contract No. DE-AI79-84BPI9476, 1985.

[8] A. G. Phadke, J. S. Thorpe and M. G. Adamiak, "A New Measurement Technique of Tracking Voltage Phasors, Local System Frequency and Rate of Change of Frequency," IEEE Transactions on Power Apparatus and Systems, Vol. PAS-102, No. 5, May 1983.

[9] K.E. Martin, "Phasors Measurement System Test," 1992 BPA Engineering Symposium, Vol 2, pp 689-704, Portland, OR, April 1992.

[10] G. Missout, J. Beland, G. Bedard and Y. Lafleur, "Dynamic Measurement of the Absolute Voltage Angle on Long Transmission Lines," IEEE Transactions on Power Apparatus and Systems, PAS-100, Nov. 1981.

[11] D. J. Hansen and C. P. Dalpiaz, "Real-time Digital Encoding and Telemetry of Key Power System Transient Quantities," Western Protective Relay Conference, October 23, 1984.

[12] A. G. Phadke, "Precise Synchronization of Phasor Measurements in Electric Power Systems," Proceedings of 22nd Annual Precise Time and Time Interval Applications and Planning Meeting, 1990, NASA Publication 3116 (1991).

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ANALYSIS OF THE FREQUENCY STABILITY OF ON-ORBIT GPS NAVSTAR CLOCKS

Thomas B. McCaskill
Wilson G. Reid
James A. Buisson

U.S. Naval Research Laboratory
Washington, D.C., 20375

Hugh E. Warren
Sachs Freeman Associates, Inc.

ABSTRACT

The Naval Research Laboratory analyzes the performance of the on-orbit Navstar clocks for the Global Positioning System (GPS) Joint Program Office† and conducts special analyses for the GPS master control station (MCS) whenever a clock exhibits anomalous behavior. This analysis will focus on the long term behavior of the cesium and rubidium clocks using sample times of one-day or more. The analysis includes frequency and aging histories, frequency-stability profiles, time-domain noise process profiles, and anomaly detection. Events that perturb the normal clock behavior are of particular interest. The semiannual eclipse seasons will be superimposed on selected plots to investigate the temperature sensitivity of the clocks. Of particular interest is the nonstationary stochastic behavior of one of the cesium clocks, illustrated by its frequency-stability history, showing that the frequency stability is not always time-invariant. The frequency stability is presented for sample times of one day to 300 days. Time-domain noise-process analysis shows the dominant noise type to be white frequency noise for sample times of one to ten days. Composite frequency-stability profiles are presented for the Block I and Block II clocks. Several clocks show a frequency stability of a few pp10^{14} for long sample times.

BACKGROUND

The Global Positioning System is a Department of Defense (DoD) space-based navigation system that provides precise position, time, and frequency to users located anywhere in the world. When the system becomes operational, 21 space vehicles with three on-orbit spares will be deployed in six planes at an inclination of fifty-five degrees. Each Navstar space vehicle broadcasts time-coded signals and a navigation message. Information in the navigation message is used to determine the position of the space vehicle at the time of observation and to relate the time of each of the space vehicle clocks to a common system time. Each user passively receives the time coded signals from four of the space vehicles and uses the information to compute three-dimensional position and velocity plus time. The system time is steered so that a precise-time user can determine the time offset of his clock from Coordinated Universal Time (UTC) by using an additional correction provided in the navigation message.

The Global Positioning System will provide real-time, continuous position determination for ships, submarines, planes, missiles, low-Earth orbiting space vehicles, and search and rescue missions. Precise time applications include synchronization for secure communications. Possible future uses include differential GPS navigation for aircraft landings and precise attitude measurement.

Operational control of the system is exercised from the master control station located at the Consolidated Space Operations Center (CSOC) at Falcon Air Force Base (FAFB) in Colorado Springs, Colorado. The control segment consists of the master control station plus five monitor stations located in Colorado Springs, Hawaii, Kwajalein, Diego Garcia and Ascension islands. *Figure 1* depicts the flow of information from a single space vehicle to each of the monitor stations. The monitor stations passively track all Navstar space vehicles in view, collecting pseudorange and pseudorange-rate data from each space vehicle. This tracking information is sent to the master control station where the ephemeris and clock parameters for each

† *This work was sponsored by the GPS Joint Program Office.*

0-7803-0905-7/93 $3.00 © 1993 IEEE

space vehicle are estimated and predicted. The master control station periodically uploads to each space vehicle the ephemeris and clock parameters which are included in the navigation message. Additional measurements referenced to the DoD master clock are collected at the U.S. Naval Observatory (USNO) precise-time station for use in determining the steering of GPS system time to UTC (USNO).

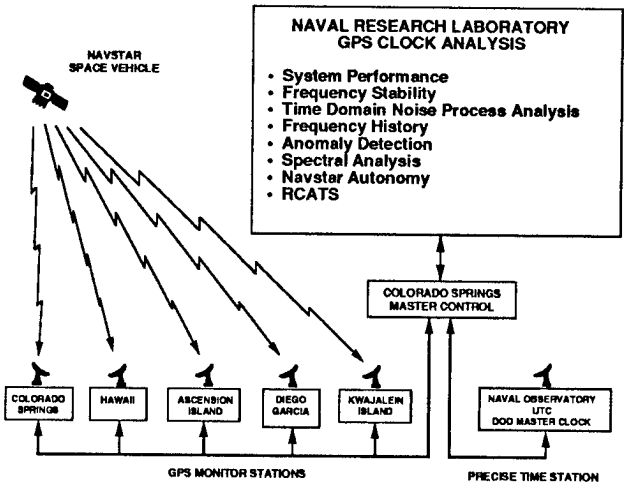


Figure 1

INTRODUCTION

Frequency stability analyses [1] are conducted by the Naval Research Laboratory (NRL) to assess the performance of the on-orbit Navstar clocks. This work is performed in support of the Global Positioning System and is sponsored by the GPS Joint Program Office. These analyses are performed for all orbiting Navstar clocks. Detailed analyses are made upon request from the master control station or from the Joint Program Office. The Naval Research Laboratory has analyzed the performance of more than 40 orbiting Navstar clocks with a total operating lifetime of more than 100 years. The results of our analyses indicate that the majority of the Navstar clocks have provided excellent performance that has exceeded the GPS performance specifications.

The phase offset between the reference clock at the tracking station and the remote clock in the space vehicle is estimated using pseudorange measurements collected as the space vehicle passes over the tracking station. The one-day measurements of the phase offset were obtained from 13-minute phase-offset measurements by taking the measurement nearest the time of closest approach, *i.e.*, the one having the highest elevation angle which minimizes the effect on the measurement of the ionosphere, the troposphere and the along-track orbit error. The 13-minute phase-offset measurements were obtained in turn from a linear least squares fit to 13 minutes of six-second phase-offset measurements evaluated at the beginning of the observation interval. The six-second phase-offset measurement was obtained from the difference between the pseudo-range measurement and the predicted range obtained from the Keplerian orbital elements broadcast in the navigation message. The frequency offset is estimated by averaging the change in the phase offset over a sidereal day.

The Block I data to be presented was collected at the Naval Observatory precise-time station using a single-frequency, time-transfer receiver with ionospheric corrections obtained from a model of the ionosphere contained in the navigation message. The broadcast signal for the Block I space vehicles does not suffer degradation due to selective availability (SA). The Block II data to be presented was collected at the Naval Observatory using dual-frequency, authorized-user receivers which measure the ionospheric delay and automatically correct for the effects of selective availability. In both cases the broadcast ephemeris was used to compute the position of the space vehicle in making the pseudorange measurement. The time-transfer receiver is driven by the DoD master clock which is characterized by a long-term frequency stability that is significantly better than any of the space vehicle clocks.

Analyses of the clock performance include frequency and aging histories, frequency-stability profiles, time-prediction uncertainty profiles, time-domain noise-process analysis, spectral analysis, and anomaly detection. Events that perturb normal space vehicle or monitor station clock operation are of particular interest. In on-orbit analysis, a variety of system effects are superimposed on the phase-offset history of the space vehicle clock. These effects may possibly enhance but probably detract from the estimate of the clock performance. Therefore, deviations from a typical frequency-stability profile are further analyzed to identify some of these effects.

CURRENT CONSTELLATION

Figure 2 presents the current GPS constellation represented by twenty Block II space vehicles in six planes and by three remaining Block I space vehicles in three planes. Each plane contains four numbered positions spaced approximately 90 degrees apart in the plane. Block I space vehicles were placed into orbits with an inclination of nominally 63 degrees while the Block II space vehicles are placed into orbits with an inclination of nominally 55 degrees which optimally balances the navigation coverage [2] for users at any

latitude. The three Block I space vehicles still in use will be deactivated after the additional Block II space vehicles are launched and positioned into the constellation. All currently operating clocks are cesium except for the three rubidium clocks that are circled. Navstar 37, which is in the C-plane, had been launched as of June 1, 1993 but had not yet been declared "healthy" so that no data is presented for this space vehicle.

Block I Navstars

Plane	Slot 1	Slot 2	Slot 3	Slot 4
A'	(10)			
B'				
C'	9			(11)

Block II Navstars

Plane	Slot 1	Slot 2	Slot 3	Slot 4
A		(25)	27	19
B	22	20	13	
C		28	31	37
D	24	15	17	
E	14	21	16	23
F	32	26	18	29

◯ Rubidium clock

Figure 2

The three rubidium clocks included in this analysis are Rockwell clocks that use the Efratom physics package. The cesium clocks were built by the Frequency and Time Systems (FTS) company except for three second-source cesium clocks. Two of the second-source cesium clocks (Navstar 31 and 32) were built by Frequency Electronics, Inc. and one (Navstar 29) was built by Kernco.

BLOCK I CLOCKS

Figure 3 presents the residuals of a linear fit to the frequency offset history of the three Block I space vehicle clocks for eighteen months from November 1992 to May 1993. The Navstar number and clock type (Cs for cesium and Rb for rubidium) are indicated immediately below the history for each of the three clocks. The ordinate in *Figure 3* represents the frequency offset with each major division representing a frequency difference of 2.5 $pp10^{12}$.

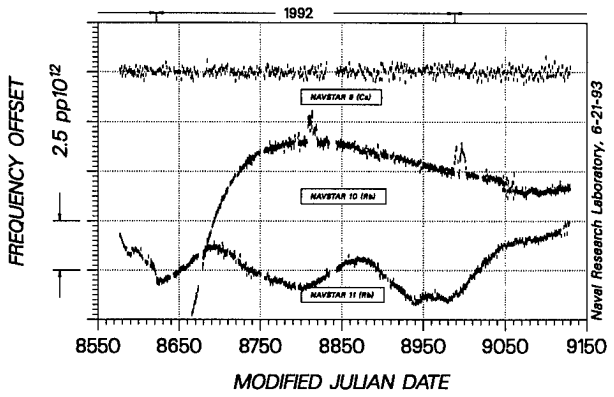


Figure 3

The frequency offset for an atomic clock is expected to appear as white frequency noise with a constant aging. The frequency offset of the Navstar 9 cesium clock shown at the top of the plot meets this expectation. This clock has been in continuous operation for almost nine years having been placed into operation 30 June 1984. The frequency offset appears as a white frequency noise process over the entire time span. Because of the longevity and excellent performance of this clock, it was selected to illustrate several of the techniques used by the Naval Research Laboratory to analyze on-orbit clocks.

The frequency offset of the Navstar 10 rubidium clock is shown in the center of *Figure 3*. This rubidium clock was placed into operation on 7 February 1992 and is the third clock to be operated on Navstar 10. The current Navstar 10 clock features additional temperature control beyond that used in earlier Block I rubidium clocks. The most significant behavior of the frequency offset is a random walk or change in the aging during the first 120 days of operation. Also notable in the behavior of the frequency offset is the occurrence of two anomalies in July 1992 and January 1993.

The frequency offset of the Navstar 11 rubidium clock is shown at the bottom of *Figure 3*. This rubidium clock was placed into operation on 4 January 1989

and is the second clock to be operated on Navstar 11. The most significant feature of the frequency offset is the presence of systematic behavior in the form of a quasi-periodic function with a period of about six months.

Figure 4 presents the residuals of a *7th* order polynomial fit to the frequency offset of the Navstar 11 clock for the entire period of operation of the clock from January 1989 through May 1993. The Navstar 11 eclipse seasons—the shaded regions—have been superimposed on the frequency offset. It can be seen that the systematic behavior of the frequency offset is highly correlated with the eclipse seasons. Specifically, it can be seen that the behavior of the frequency offset during and between alternate eclipse seasons is similar suggesting an annual periodic component in the behavior related to the period of the Earth's revolution about the sun.

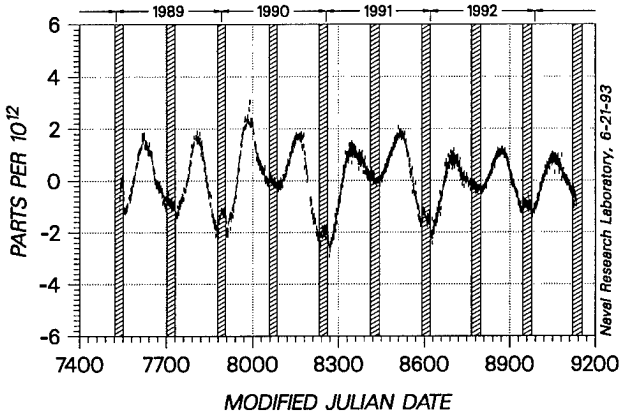


Figure 4

Each space vehicle in the constellation periodically encounters eclipse seasons. These eclipse seasons occur at a nominal semi-annual rate and have a duration of 30 to 45 days. During the eclipse season, the sun is eclipsed by the Earth for a portion of each orbit. Entrance of a space vehicle into the shadow of the Earth causes changes in the heat load from solar radiation thereby perturbing the environment of the space vehicle. The major effect of the eclipse season on the space vehicle is a change in the temperature of the space vehicle and consequently a change in the temperature of the clock.

Analysis of an earlier Navstar rubidium clock for which clock temperature measurements were avalable indicated a temperature sensitivity of approximately 2 $pp10^{12}/°C$. It is believed that the temperature sensitivity of the Navstar 11 rubidium clock is of the same order of magnitude.

Figure 5 presents the residuals of a linear fit to the frequency offset for the Navstar 10 rubidium clock. The eclipse seasons for this space vehicle have been superimposed on the frequency offset. Correlation of the anomalies of July 1992 and January 1993 with the eclipse seasons is puzzling in light of the fact that this rubidium clock has the additional temperature control that was expected to reduce its temperature sensitivity as it indeed appears to have done compared to the systematic behavior so pronounced in the Navstar 11 rubidium clock.

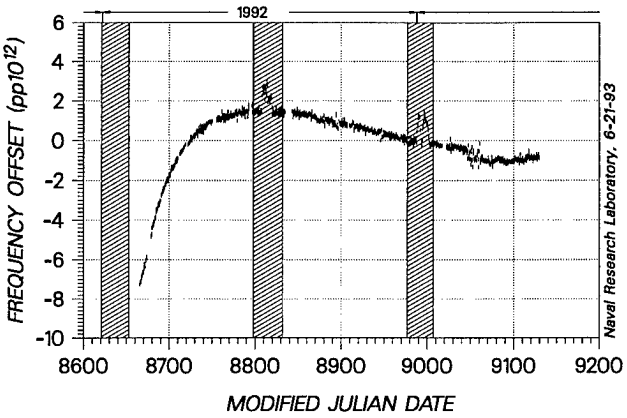


Figure 5

Figure 6 presents the frequency offset of the Navstar 9 cesium clock for a time span of almost nine years. This space vehicle is equipped with one cesium clock and three rubidium clocks. The cesium clock was the first to be activated and has been in continuous operation for the entire time.

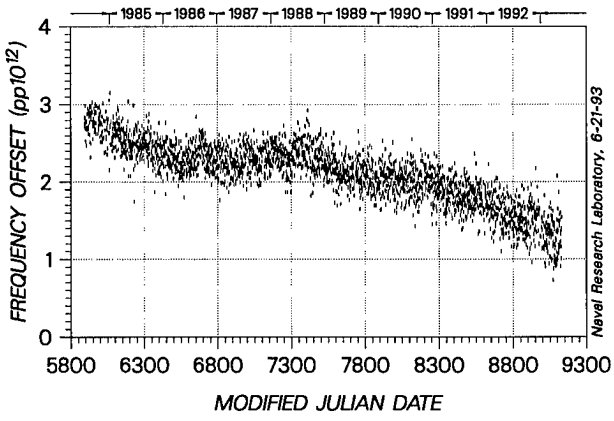


Figure 6

The initial mean frequency offset computed from the first seven days of data was 2.8 $pp10^{12}$. The mean

frequency offset after more than eight years of operation was still 1.5 $pp10^{12}$. The total change in the mean frequency offset during this time was only 1.3 $pp10^{12}$ which indicates that the mean frequency offset changed by less than 2 $pp10^{13}$ for each year of operation.

The frequency stability profile for the Navstar 9 cesium clock is presented in *Figure 7*. All available data from 13 July 1984 through 15 July 1992 was used to estimate the frequency stability. The measured one-day frequency stability was 1.7×10^{-13}. The stability was computed removing an aging of $-1.9 \times 10^{-15}/day$.

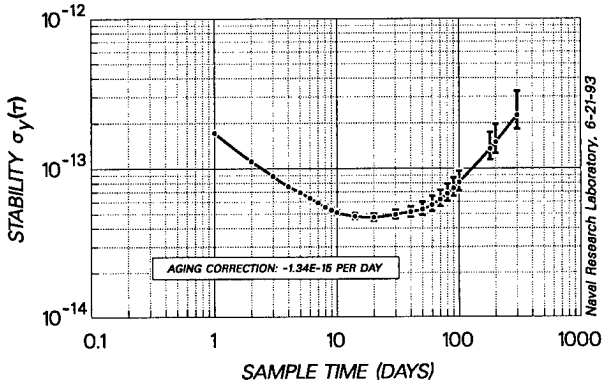


Figure 7

The current Navstar 9 cesium clock has set several records: (1) It has exceeded for eight years the expected performance level for Block I cesium clocks. (2) It has exceeded the expected operating lifetime specification for cesium clocks. And (3) it has had a useful operating lifetime that is greater than the expected lifetime of the Navstar space vehicles.

NOISE PROCESS ANALYSES

The frequency stability estimates for the entire operating lifetime of the Navstar 9 cesium clock were used to compute the type of random noise that was dominant as a function of sample time. The noise process analysis is presented in *Figure 8* which shows that the dominant noise process for sample times up to ten days was white frequency noise except for a small departure for a sample time of two days. For sample times of ten to sixty days the dominant noise process was flicker frequency noise. For sample times greater than 60-days the dominant noise process transitioned to random walk frequency noise. At sample times above 140 days aging became the dominant process affecting the frequency stability.

It was expected that white frequency noise should be dominant for sample times up to one-week with a gradual transition to flicker frequency noise followed by a transition to random walk in the frequency. Therefore for sample times up to 30-days this cesium clock has performed as expected. A single correction for aging, computed from the mean of the second differences of the phase offset, was applied to the entire nine-year span of data.

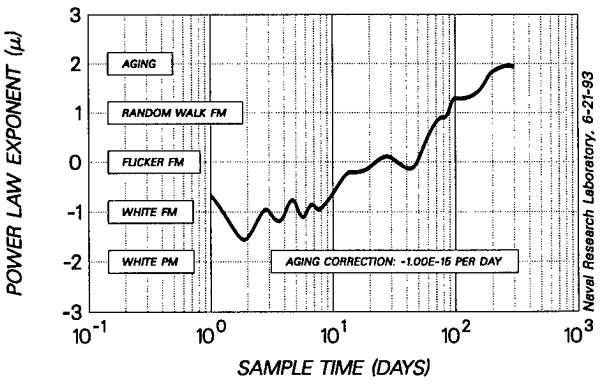


Figure 8

STABILITY INVARIANCE

The physics of atomic clocks indicates that the frequency stability of an ideal atomic clock should remain time invariant. As a test of this theory, the data for the Navstar 9 cesium clock was partitioned into six segments during the nine-year time span in an effort to determine if the frequency stability of this clock has remained constant or has changed during that time. Three of the six segments covered a time span of two years. The initial time span was 1.5 years while the last two time spans were each less than one year. Each of these segments was sufficiently long so that the uncertainty in the estimate of the stability due to the number of samples was small.

Figure 9 presents the six frequency stability profiles for the Navstar 9 cesium clock. The first three profiles are almost identical—white frequency noise being the dominant noise type for sample times of one to ten days. The fourth profile indicates a slight decrease in performance compared to the first three. The fifth profile was estimated using 6.5 months of data from 1 January 1992 to 14 July 1992. Following that date, a significant degradation was noted in the frequency stability. This analysis indicates that the frequency stability of the Navstar 9 cesium clock was essentially time invariant for the first eight years of its lifetime.

A test of the invariance of the frequency stability was devised that examines the stability of the clock for one value of the sample time as a function of elapsed

time. This time history of the frequency stability corresponds to the output of an N-day moving average filter operating on the sequence of squared first differences of the one-day frequency offset measurements.

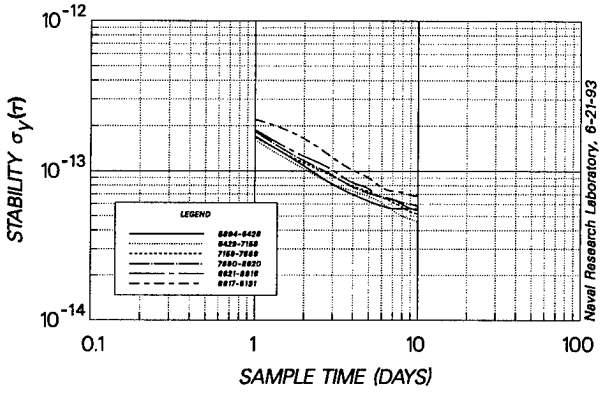


Figure 9

Computation of the frequency stability history entails a tradeoff between the number of samples used in the N-day moving average to compute one estimate of the frequency stability and the total number of samples available. The larger the number of samples averaged the greater will be the confidence in the estimate of the stability but the smaller will be the time over which the stability history can be represented. The smaller the number of samples averaged the greater will be the time over which the history can be represented but the smaller will be the confidence in the estimate of the stability.

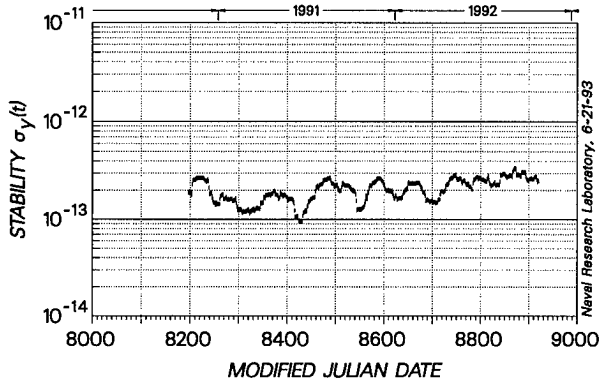


Figure 10

Figure 10 presents for the Navstar 15 cesium clock the frequency-stability history which corresponds to the output of a 30-day moving average filter operating on the sequence of squared first differences of the one-day frequency offset measurements. *Figure 10* shows that the frequency stability exceeded 2×10^{-13} on 14 April 1992 (mjd 8726) and that following that date the one-day frequency stability estimates based on a 30-day sample all exceeded 2×10^{-13}.

Figure 11 presents frequency stability profiles for the Navstar 15 cesium clock corresponding to three contiguous time spans from 16 October 1990 to 24 October 1992. For the first partition of 108 days, the one-day frequency stability was estimated to be 2.1×10^{-13} with an aging correction of $+4.4 \times 10^{-15}/day$. White frequency noise was the dominant noise type for sample times of one to ten days. For the second partition of 540 days, the one-day frequency stability was estimated to be 2.0×10^{-13} with an aging correction of $+1.6 \times 10^{-15}/day$. This second partition marked a departure from white frequency noise. For the third partition of 92 days, the one-day frequency stability was estimated to be 2.8×10^{-13} with an aging correction of $-9.5 \times 10^{-15}/day$.

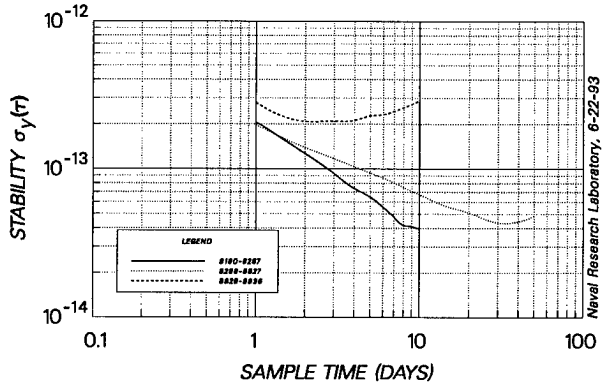


Figure 11

Comparison of the performance during the last time span with the performance during the first two time spans indicates a significant degradation in the one-day frequency stability, a change from positive to negative aging, and a departure from white frequency noise for sample times longer than two days. *Table 1* summarizes the performance of the Navstar 15 cesium clock for the three time spans.

BLOCK II CLOCKS

Figure 12 presents a comparison of the frequency-offset history for eight Block II cesium clocks: Navstar 13, 14, 15, 17, 18, 19, 20, and 22. *Figure 13* presents a comparison of the frequency-offset history for eight Block II cesium clocks: Navstar 23, 24, 26, 27, 28, 29, 31, and 32. In both cases the mean of

each of the frequency-offset histories has been removed, but the aging has been retained. The spacing between grid lines along the ordinate is 2.5 $pp10^{12}$ in both figures. The aging, measured with respect to the DoD master clock, of these sixteen cesium clocks was 5.5 $pp10^{15}/day$, or less. The frequency-offset history for each of these clocks was nominal and within specifications.

Table 1

PERFORMANCE SUMMARY
NAVSTAR 15
1990–1992

Time Span (*days*)	Frequency Stability 1 day ($pp10^{13}$)	Frequency Stability 10 days ($pp10^{14}$)	Aging ($pp10^{15}/day$)
108	2.1	4.0	+4.4
514	2.0	6.8	+1.6
92	2.8	12.9	−9.5

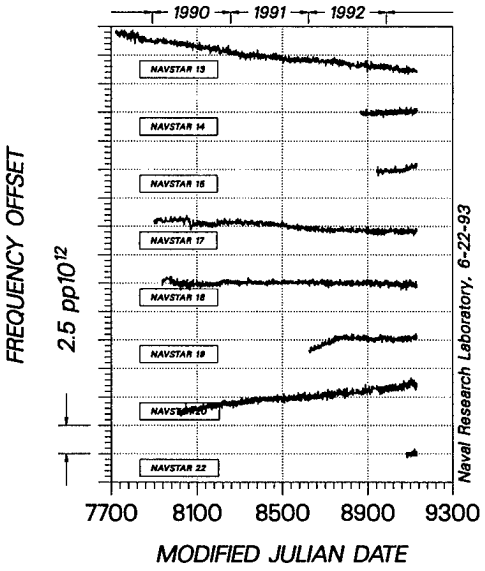


Figure 12

Figure 14 presents the frequency-offset histories for three Block II clocks with signatures that required additional analysis. Because of the large aging associated with rubidium clocks, the residuals of a linear fit to each of the histories has been plotted so that these histories could be plotted on the same scale as the preceding sixteen, *i.e.*, 2.5 $pp10^{12}$.

The Navstar 25 rubidium clock at the bottom of the figure had an aging coefficient estimated to be $-1.9\ pp10^{13}/day$ for the data span analyzed. This value of aging is about two orders of magnitude higher than that measured for the cesium clocks. This clock also exhibits random walk behavior in the frequency.

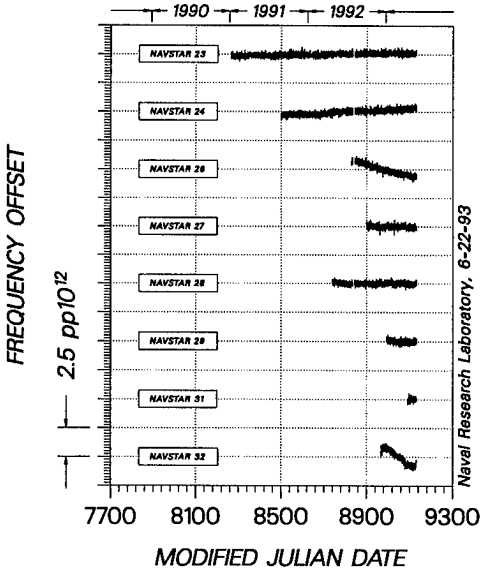


Figure 13

The frequency offset for the Navstar 21 cesium clock is shown in the middle of the figure. A decrease in the frequency offset can be observed near modified Julian day 8720 (mid 1992). This clock was reported to have experienced large variations in the cesium beam current which were found to correlate with the change in frequency. The frequency offset for the Navstar 16 cesium clock plotted at the top of the figure shows a fairly consistent periodic variation with a period of about six months.

Figure 15 shows for Navstar 16 the residuals of a linear fit to the frequency offset smoothed by a seven-point moving average filter. The seven-point average corresponds to one-week for the case of uniformly sampled data. This amount of averaging is appropriate because the white frequency noise for these cesium clocks typically extends to sample times of one-week or more. Furthermore, seven days is much smaller than the period of the behavior being studied. The eclipse seasons have been superimposed as the shaded vertical bars. With the white noise suppressed by averaging, it is not difficult to see that the frequency offset is correlated with the eclipse seasons. During the three months centered on the first eclipse season of 1992 the frequency dipped by more than 1 $pp10^{12}$.

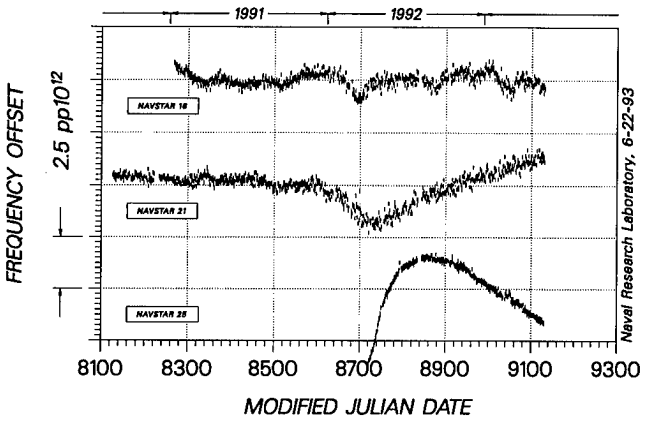


Figure 14

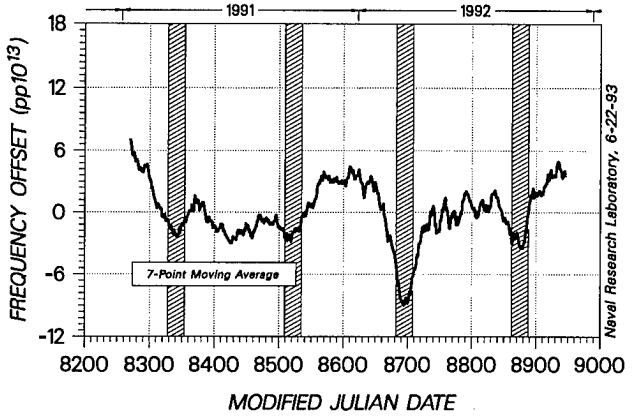


Figure 15

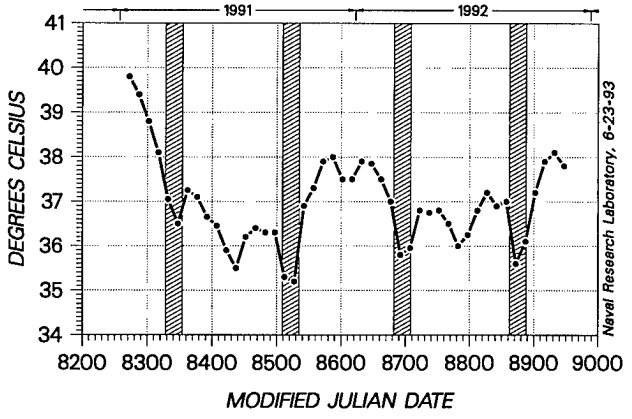


Figure 16

Figure 16 presents the daily average temperature from a sensor located near the Navstar 16 clock. Dips in the average temperature are seen to correlate with the eclipse seasons. The scales of *Figure 15* and *Figure 16* have been selected so that the correlation of the Navstar 16 frequency offset with temperature can be observed. A preliminary analysis of this data indicates that the frequency offset of the clock exhibits a temperature sensitivity on the order of 2.5 $pp10^{13}/^{o}C$. This represents a significant departure from the prelaunch testing of the cesium clocks which indicated temperature sensitivities of 1 $pp10^{13}/^{o}C$ or less.

FREQUENCY STABILITY

The frequency stabilities to be presented are all referenced to the DoD master clock located at the U.S. Naval Observatory. The space vehicle positions were determined using the broadcast ephemeris in the navigation message. The maximum sample time included in these results is at least a factor of ten less than the time spanned by the data. For example, a phase-offset history of at least 100-days duration would be required to estimate the frequency stability for a 10-day sample time. The factor of ten ensures that the confidence limits surrounding the stability estimate are reasonably close together.

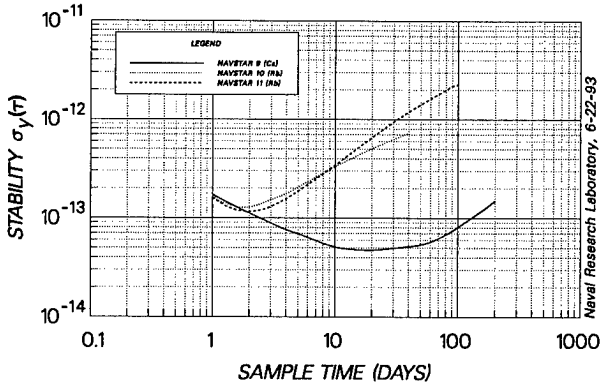


Figure 17

Figure 17 presents the frequency stability profiles for the three Block I Navstar clocks. The stability profiles for these three clocks were computed from data collected by the single-frequency GPS receiver at the Naval Observatory. Previous analysis [3] indicates that use of a single-frequency receiver could result in a small bias for the one-day frequency-stability estimates because of error in the ionospheric model. In the case of the dual-frequency receiver, which measures the ionospheric delay, this error is minimized. The difference in the frequency stability between the cesium and rubidium clocks, as expected, is most pronounced at sample times greater than a few days because of the extremely large aging present in the rubidium clocks. For a sample time of one day the frequency stability of all three

clocks is about the same. *Table 2* presents a summary of the performance of the three Block I clocks.

Table 2

PERFORMANCE SUMMARY
BLOCK I
NAVSTAR CLOCKS

Navstar Number	Time Span (*days*)	Frequency Stability 1 day ($pp10^{13}$)	Frequency Stability 10 days ($pp10^{14}$)	Aging ($pp10^{15}/day$)
9	2924	1.7	5.1	−1.3
10	466	1.6	32.9	−129.0
11	159	1.4	34.4	−106.0

Figure 18 presents the frequency stability profiles for the three Block II Navstar clocks. The stability profiles for these three clocks were obtained from data collected by the dual-frequency GPS receiver at the Naval Observatory. These three clocks have frequency stability profiles that differ from the expected profile because of anomalous behavior or, in the case of the Navstar 25 rubidium clock, because of the large component of random walk noise in the frequency and the extremely large aging. The Navstar 25 rubidium clock, in spite of these limitations, demonstrated for a sample time of one day the best frequency stability of all the Navstar clocks operated thus far. *Table 3* presents a summary of the performance of these three Block II clocks.

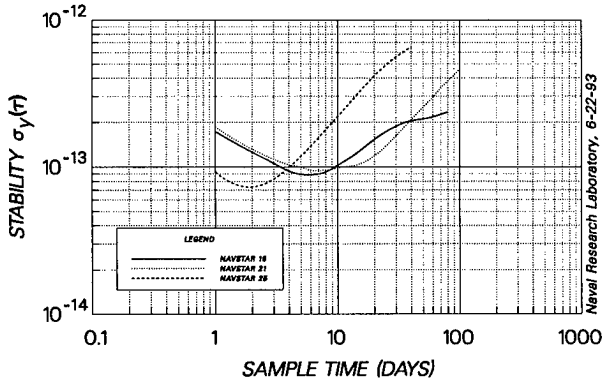


Figure 18

Figure 19 presents the frequency stability profiles for the remaining sixteen Block II Navstar clocks. The stability profiles for these clocks were also obtained from data collected by the dual-frequency GPS receiver at the Naval Observatory. *Table 4* presents a summary of the performance of these Block II Navstar clocks.

Table 3

PERFORMANCE SUMMARY
THREE BLOCK II CLOCKS

Navstar Number	Time Span (*days*)	Frequency Stability 1 day ($pp10^{13}$)	Frequency Stability 10 days ($pp10^{14}$)	Aging ($pp10^{15}/day$)
16	864	1.7	10.2	−3.4
21	1005	1.9	9.8	1.2
25	426	0.9	21.9	−183.0

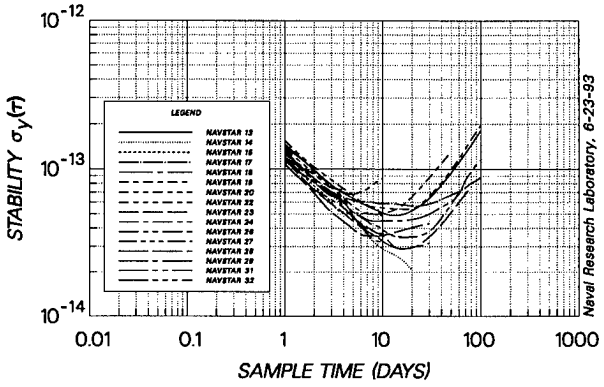


Figure 19

Table 4

PERFORMANCE SUMMARY
BLOCK II CLOCKS

Navstar Number	Time Span (*days*)	Frequency Stability 1 day ($pp10^{13}$)	Frequency Stability 10 days ($pp10^{14}$)	Aging ($pp10^{15}/day$)
13	1413	1.5	5.2	−4.7
14	267	1.2	2.1	+3.0
15	193	1.2	4.9	+1.6
17	1233	1.2	4.6	−1.4
18	1197	1.3	4.5	−1.5
19	508	1.2	5.1	+3.4
20	1124	1.6	5.8	+5.5
22	52	1.4		+2.1
23	870	1.2	3.2	−2.0
24	633	1.1	3.7	+2.4
26	304	1.2	3.8	−4.6
27	235	1.4	4.0	+0.7
28	394	1.1	3.7	−1.1
29	141	1.1	3.1	−1.2
31	42	1.4		+1.3
32	168	1.5	13.1	−2.7

These sixteen Navstar clocks are performing with an estimated one-day frequency stability of 1.6×10^{-13} or better.

REFERENCES

[1] McCaskill, T.B., Buisson, J.A., and Stebbins, S.B., "On-Orbit Frequency-Stability Analysis of the GPS Navstars 3 and 4 Rubidium Clocks and Navstars 5 and 6 Cesium Clocks", *Proceedings of the Fifteenth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, December 6-8, 1983.

[2] Buisson, J.A. and McCaskill, T.B., "TIMATION Navigation Satellite System Constellation Study", *NRL Report 7389*, 27 June 1972.

[3] McCaskill, Thomas B., Buisson, James A., Reid, Wilson G., and Warren, Hugh E., "On-Orbit Frequency Stability Analysis of the GPS Navstar Cesium and Rubidium Clocks", *Proceedings of the Twenty-Third Annual Pecise Time and Time Interval (PTTI) Applications and Planning Meeting*, 3–5 December 1991.

ACKNOWLEDGEMENT

The authors wish to express their appreciation to Mr. Richard Harding of Richard W. Harding Technical Services for his assistance in preparing the manuscript for publication.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

USING A NEW GPS FREQUENCY REFERENCE IN FREQUENCY CALIBRATION OPERATIONS

Terry N. Osterdock, Stellar GPS Corporation
800 Charcot Avenue, Suite 110
San Jose, CA 95131

John A. Kusters, Hewlett Packard Company
5301 Stevens Creek Boulevard
Santa Clara, CA 95050

Abstract

A new GPS receiver has been developed which offers the user a means of performing precise frequency calibration traceable to the USNO without all of the prior preparation and coordination required in common view measurements. This new receiver, the Model 100 GPS Clock™, has a unique architecture that offers significantly improved frequency stability and time pulse (1 PPS) jitter of less than 1 nanosecond. Use of this new low cost receiver will be discussed as it pertains to the general area of frequency calibration. Techniques for frequency calibration using the GPS that can be used in most calibration labs will be identified.

Introduction

A number of papers have been written in the past several years about the use of the NAVSTAR Global Positioning System (GPS) in time comparisons between two locations. The impact of Selective Availability (SA) on the stability of timing pulses derived from GPS reduces the accuracy of single measurements relative to UTC(USNO) to 300 nanoseconds. The jitter on such derivations is typically 20 to 60 nanoseconds which greatly affects the ability of the user to make use of the GPS, using standard GPS timing receivers, for precise frequency comparisons in the short term.

If the user wished to make ultra precise time transfer measurements between two locations, then they would have to use common-view timing measurements. Common-view measurements can be made with a few nanoseconds of measurement accuracy, but generally require advance preparation, coordination of the two sites, and tracking of specific satellites during specific time periods.

In this paper we shall describe a number of important definitions involved in frequency calibration, then several frequency calibration techniques, followed by a discussion of several traceable frequency sources and a description of the Stellar GPS Model 100 GPS Clock.

Definitions

The following definitions are provided to establish a common basis for discussion:

Accuracy: the degree of conformity of a measured or calculated value to its definition or with respect to a standard reference.

Drift: the linear (first order) component of the systematic change in frequency of an oscillator over time. Drift is due to aging plus changes in the environment and other factors external to the oscillator.

0-7803-0905-7/93 $3.00 © 1993 IEEE

Offset: the difference between the realized value and a reference value.

Precision: the degree of mutual agreement among a series of individual measurements. Precision is often expressed, but not necessarily, by the standard deviation of the measurements.

Resolution: the degree to which a measurement can be determined is called the resolution of the measurement; the smallest significant difference that can be measured with a given instrument.

Stability, or more correctly, Instability: the spontaneous and/or environmentally caused frequency change within a given time interval, or within a given range of an environmental variable.

NOTE: Generally, one distinguishes between systematic effects such as frequency drift and stochastic frequency fluctuations. Special variances have been developed for the characterization of these fluctuations. Systematic instabilities may be caused by radiation, pressure, temperature, humidity, etc. Random or stochastic instabilities are typically characterized in the time-domain and/or frequency-domain. It is typically dependent on the measurement system bandwidth and/or on the sample time or integration time.

Common Frequency Calibration Techniques

The calibration of precision frequency and time sources requires comparison between the device to be calibrated and a precision reference source. All secondary frequency sources, such as Rubidium frequency standards and quartz oscillators, require periodic calibration, usually against a local reference standard. For the highest precision, comparison against a national reference standard may be required.

Local clocks are generally those time and frequency standards which are maintained at the calibration site. In some cases, where a low degree of precision is required, the time and frequency reference signals may be transmitted from one site to another. As a result of the transmission process, the received time and frequency signals are degraded by the propagation delay and induced line noise. A more subtle effect which degrades the reference signal is the environmental performance of the transmission media.

Frequency Counters

The modern electronic frequency counter is a versatile device. Most simply it can be used to directly measure the frequency of a signal applied to its input port. The accuracy of the measurement is directly related to the internal resolution of the counter and the stability of the internal frequency source. The performance of the frequency counter can be significantly improved in both accuracy and in stability by using the reference frequency source as an external time base for the counter. However, modern counters are still limited by their internal design to resolutions on the order of 1 part in 10^9. As a result, most high precision frequency sources cannot be adequately evaluated by direct measurement with a frequency counter.

Heterodyne techniques can greatly improve the resolution of the frequency counter measurement based upon the reciprocal counting technique. [1,2] A typical "single-mixer heterodyne" technique, is shown in figure 1. In the figure, the counter's internal time base has been replaced with the external frequency reference. The frequency reference is also applied to a

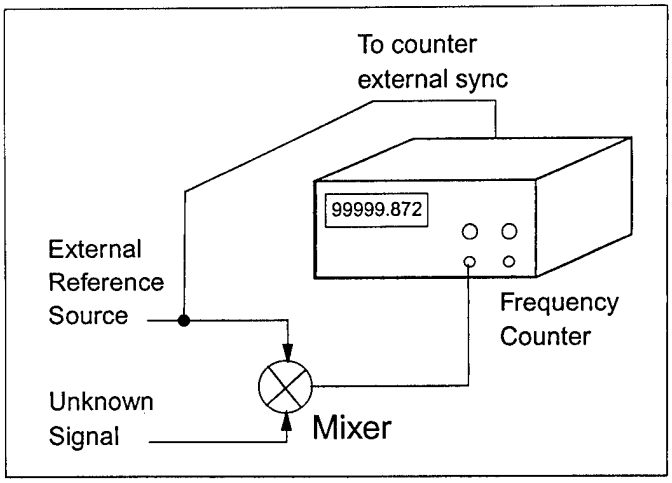


Figure 1
Single Mixer Heterodyne Frequency Measurement Technique

diode mixer, where the signal being measured is heter-

odyned against the reference. The difference frequency is then measured against the reference.

Overall accuracy and stability will be governed by the signal with the worst stability. Therefore, unless it is known that the frequency source is significantly better than that being measured, we can only conclude that the signal being measured is no worse than the measurement indicates and may be much better.

Time Interval Counters

Time interval counters, such as the HP 5370B, are now available with resolutions on the order of picoseconds. With this degree of precision, frequency measurements of very high precision sources can be made against a known frequency reference source. The degree of precision is governed by the signal with the greatest amount of noise, or instability in the signal.

The "dual-mixer time-difference" technique (figure 2) is an excellent method for comparing two signals which are essentially identical in frequency. The offset frequency source is used to heterodyne both channels and it does not need to be of exceptionally high quality as its instability is common mode to both measurement channels and will be eliminated to first order. By tracking the counter readings over a period of time, a data plot (figure 3) results. The measured time interval between the start and stop signals are plotted as a function of elapsed time. The maximum time interval that can accumulate is the period of the highest frequency applied to either the "start" or "stop" inputs of the counter. If a full period of time interval does accumulate, the data reduction becomes more complicated as proper one-period adjustments must be made to all of the data obtained after the data step. If both signals are relatively stable, a determination of the unknown frequency can be determined by computing the slope of the data. As mentioned before, the results will indicate that the unknown frequency is no worse than the measurement indicates and may be much better.

DUAL-MIXER TIME-DIFFERENCE
Frequency Comparison Technique

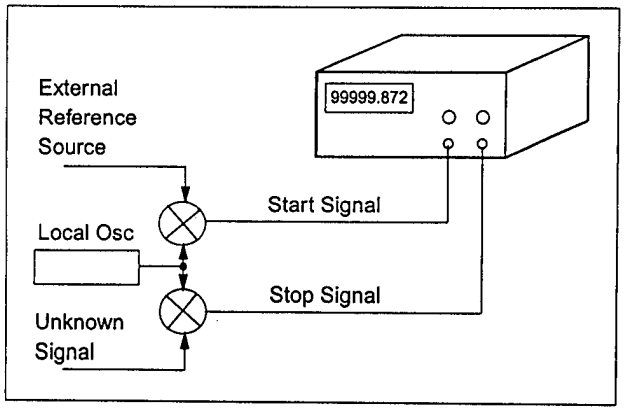


Figure 2 Dual-Mixer Time-Difference

DUAL-MIXER TIME-DIFFERENCE
Typical Data Plot

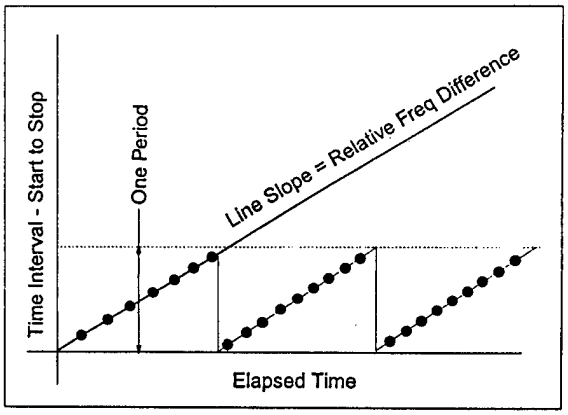


Figure 3 Typical Data Plot

The direct time-interval technique (figure 4) is a much simpler technique. Both signals are applied directly to the inputs of the time-interval counter where the start signal is the reference frequency source, and the unknown signal is very close in frequency to the reference. In this case also, a curve similar to that of figure 3 results. This method is commonly used to measure large numbers of cesium and rubidium standards essentially simultaneously through continuously sampling. [3] As an advanced time interval counter is accurate to about 100 picoseconds, a resolution of 1 part in 10^{10}, and as there is no heterodyne advantage to this method, data must be obtained over long periods of time. Analysis and experimentation indicates that if the two sources are within

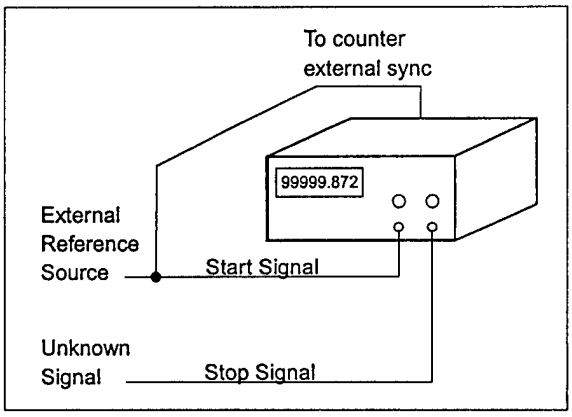


Figure 4 Direct Time Interval

1 part in 10^{10} of each other, data obtained over 10,000 seconds and properly averaged can yield accuracies of 1 part in 10^{13} with respect to the reference source.

Phase Comparison

Another technique for comparing two signals that are essentially identical involves comparing the change in phase between the two sources. Both signals are applied to a linear phase comparator and the result is then plotted using a strip chart recorder. The data plot looks something like figure 3 but is generally continuous. The slope of the plot again indicates the difference in frequency of the unknown signal versus the frequency reference.

Traceable Frequency Sources

The main concern of the calibration process is being able to determine the frequency of a given unit relative to the accepted definition. For most users, in order to relate the measurement to the accepted definition a means of obtaining "traceable" comparisons is required. There are a variety of acceptable means of making such a comparison. In the United States, there are two organizations which provide services which will allow the user to calibrate frequency references to the internationally accepted definition, the National Institute of Standards and Technology (NIST) and the U.S. Naval Observatory (USNO). A brief discussion of their services is contained in the following paragraphs. In the references, additional information is available about each of the time and frequency dissemination means. For our purposes here, we shall only briefly identify the means and the relative accuracies available.

National Institute of Standards and Technology

NIST, formerly the National Bureau of Standards, is part of the U.S. Department of Commerce. The Time and Frequency Division is located in Boulder, Colorado.

WWV and WWVH Radio Broadcast Stations

WWV in Fort Collins, Colorado and WWVH in Kekaha, Kauai, Hawaii broadcast on several standard frequencies time signals corresponding to the UT1 time scale with accuracies better than 2 parts in 10^{11} when transmitted. In transmission, the accuracy is degraded due to atmospheric propagation effects, and to Doppler shift when the transmission path includes reflection from the ionosphere. The Doppler effect can approach 1 part in 10^{7} in the worst conditions. The signals are readily available, can be received with inexpensive receivers, and are accurate enough to set clocks to within one millisecond of UT1.

WWVB Radio Broadcast Station

A more accurate service provided by NIST is the WWVB signal also broadcast from Fort Collins, Colorado. The signal is broadcast at 60KHz and is also referenced to the atomic time scale. Commercial equipment is available which performs phase comparisons between a local reference and the signal derived from WWVB. With normal signal conditions, excluding diurnal shift at sunrise and sunset, frequency accuracies in the United States of 2 to 3 parts in 10^{11} in a 24 hour period can be obtained. Time pulse modulation permits time synchronization to 100 microseconds or better provided compensation for the propagation delay is known. Generally, it is impossible to initially set remote clocks to high accuracy with WWVB alone.

Other Methods of Time Dissemination from NIST

Other services provided by NIST are (in increasing order of accuracy):

- Telephone voice messages duplicating the WWV message obtained by dialing (303) 499-7711; accuracy is about 30 milliseconds,
- Computer modem time transfer programs capable of synchronizing computer internal clocks to one time tick, or several milliseconds,
- Remote synchronization of time bases to about 1 part in 10^9,
- Common view of Loran-C stations with 24 hour continuous monitoring using NIST provided measuring equipment. Over a long period, overall frequency accuracy approaches parts in 10^{12},
- Common view of GPS satellites providing time transfer accuracies on the order of 30-50 nanoseconds.

United States Naval Observatory

The USNO is part of the U.S. Department of Defense and is located in Washington, D.C. The USNO provides a number of services to both civilian and military users of time and frequency for the purpose of calibration, setting and controlling clocks, timing systems, or secondary frequency standards.

LORAN-C

Loran-C (LOng RAnge Navigation) is the federally provided land-based radio navigation system for civil marine use in the U.S. coastal waters. Although primarily used for navigation, Loran-C transmissions are also used for time dissemination and frequency reference purposes. The frequency of each of the broadcast stations, operating on 100 KHz are controlled by Cesium Standards and comparison data is provided by the USNO as well as NIST.

Measurements of time difference are made using receivers which achieve high accuracy by comparing zero crossings of a specified rf cycle within the transmitted pulses. Typical Loran-C receivers can achieve frequency accuracies of 1 part in 10^{12} and time accuracies of better than 1 microsecond after monitoring for 24 hours. The Loran signal does not contain any intelligence and therefore, to determine the exact UTC time, the user must first set the clock to about 10 milliseconds using one of the other methods discussed.

NAVSTAR Global Positioning System (GPS)

The GPS is designed and deployed by the U.S. Department of Defense as an all-conditions, worldwide, 24 hour availability, three-dimensional navigation system. The GPS enables users to determine their three dimensional position, velocity, and time with unprecedented accuracies. The system is composed of 21 satellites plus three active spares in six different orbital planes with an orbital period of 12 hours. The L1 signal is available to all civilian users and provides positional accuracies of about 100 meters rms, and timing accuracies to about 300 nanoseconds. These accuracies are the result of artificial degradation of accuracy by the Department of Defense using what is known as Selective Availability (SA).

The GPS receiver must track the spread spectrum codes from at least four satellites in view. By measurement of the time of arrival of the signals from four or more satellites in view, knowing the speed of the radio signal, when each signal is transmitted and the position of each satellite, the final solution yield an exact three-dimensional position of the antenna as well as time as determined on the UTC time scale. The position of the satellite is known as each satellite message contains a satellite ephemeris.

With the system now fully functional, at least four satellites are within sight of any ground station 24 hours a day. Under ideal conditions, frequency accuracies of parts in 10^{12} and timing accuracies of better than 300 nsec are possible (SA on).

Other Methods of Time Dissemination

Like NIST, the USNO provides a number of other alternatives methods of disseminating time and frequency information.

- Telephone voice messages provide accuracy to about 30 milliseconds,
- Computer modem time transfer programs of synchronizing computer internal clocks to one time tick, or several milliseconds,
- Common view of Loran-C stations with 24 hour continuous monitoring. Over a long period, overall frequency accuracy approaches parts in 10^{12},
- Common view of GPS satellites providing time transfer accuracies on the order of 30-50 nanoseconds.

Other National Frequency and Time Standards

National standards organizations of many countries provide similar services to their local communities.

Use of the GPS

The GPS is the first truly worldwide frequency and time dissemination system available with unprecedented accuracies. The entire system is based upon very accurate time as kept by atomic standards on board each of the satellites which are monitored and controlled by the U.S. Naval Observatory. As a result, use of GPS receivers which are locked to the satellites can provide the user with very accurate, and traceable, frequency and time very inexpensively.

The Model 100 GPS Clock™

One such GPS receiver is the Stellar GPS Corporation Model 100 GPS Clock. It is a small, lightweight, low cost, GPS receiver that has a unique frequency-based architecture* offering and easy-to-use precise, accurate, and stable frequency, as well as time, that can be used directly in calibration operations.

To operate the GPS Clock, the user need only attach the antenna and cable, plug the power supply into the power outlet and the receiver will do everything else. The unit will automatically acquire and track satellites, lock up its internal oscillator to the satellite frequency and when the "data valid" light is lit, the 10 MHz and 1PPS are locked to GPS. The frequency and time outputs have accuracies that are directly and continuously traceable to the USNO. The frequency accuracy is 1 part in 10^{11} over a one day average, and 5 parts in 10^{12} over a one week average. The time accuracy relative to UTC (USNO) is 300 nanoseconds (SA on) and 100 nsec (SA off). The stability of the frequency output (10 MHz) is 1 part in 10^{11} for averaging times from .1 to 100 seconds (with the oven oscillator option) and time stability (1 PPS) is less than 1 nsec of pulse-to-pulse jitter, rms.

The GPS Clock can have its antenna located as much as 1000 feet away from the receiver, allowing it to be installed in difficult locations.

The architecture of a typical *time-based* GPS receiver is as shown in figure 5. Typically, the manufacturer uses an OEM GPS receiver designed for navigational purposes, but provides a 1 PPS output. The 1 PPS signal is then provided as an output to the user, but because of the method by which the 1 PPS is derived in the OEM receiver, it has jitter of from 20 to 60 nanoseconds, rms. If the time-based receiver also provides a frequency output, it is generally derived from the jittery 1 PPS signal by slaving, in a Phase-Locked-Loop (PLL), an additional oscillator to the 1PPS. The output performance of the frequency output is determined by the quality of both the oscillator and the PLL.

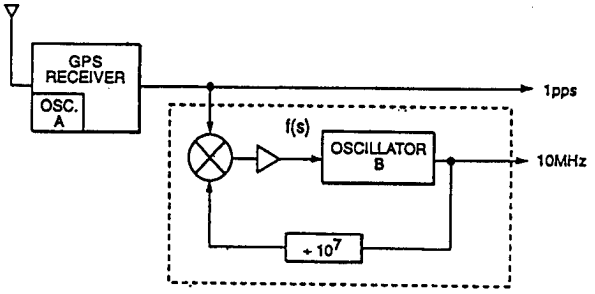


Figure 5 Typical Time-Based GPS Receiver Design

The architecture of the GPS Clock is different from the standard time-based GPS timing receivers. It is optimized for frequency applications, which improves its timing performance as well. The GPS Clock architecture (figure 6) is *frequency-based.* Instead of slaving an oscillator to the 1 PPS output of a GPS receiver, the GPS Clock slaves the oscillator of the GPS receiver to the satellites and derives the 1 PPS from the locked oscillator frequency output. The result is that the GPS Clock output is more stable. more accurate, and more precise.

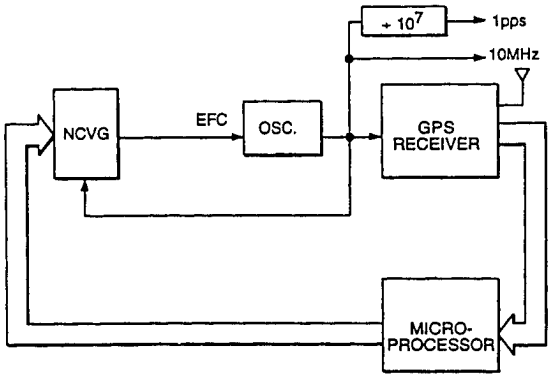


Figure 6 Model 100 GPS Clock™ Receiver Design

Use of the GPS Clock in Calibration

The GPS Clock output can be used directly in applications needing precise frequency or can be used as the reference frequency standard in any of the calibration and measurement systems previously discussed. The results provided have real-time direct traceability to the USNO and thereby ultimately to the international defined frequency and time.

References

[1] Application Note 358-12, "Simplify Frequency Stability Measurements with Built-in Allan Variance Analysis," Hewlett-Packard Co., 1989

[2] S.R. Stein, "Frequency and Time - Their Measurement and Characterization," Precision Frequency Control, Vol. 2, pp. 192-231, Academic Press, Orlando, 1985.

[3] J.A. Kusters, "Determination of the Aging Rates of High Stability Rubidium Frequency Standards," Proceedings of the 2nd International Conference on Frequency Control and Synthesis, 1989.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

USE OF THE NAVSTAR GLOBAL POSITIONING SYSTEM (GPS) IN THE DEFENSE COMMUNICATIONS SYSTEM (DCS)

Quyen D. Hua, PhD, Stellar GPS Corporation
800 Charcot Avenue, Suite 110
San Jose, CA 95131

Michael S. Williams, Quantic Industries, Inc.
990 Commercial Street
San Carlos, CA 94070

John Evans, Defense Information Systems Agency (DISA)
Reston, Virginia

Matthew A. Trippy, Capt, USAF, GPS User Equipment
NAVSTAR GPS Joint Program Office
Los Angeles, California

Abstract

Network synchronization in any telecommunications system is essential in order to achieve the required performance objectives. CCITT Recommendation G. 811 specifies that all clocks at network nodes should have a long term frequency departure of not greater than $1x10^{-11}$. This paper describes the network synchronization technique used in the Defense communications System and the equipment used to provide the stability required to meet the CCITT Recommendation.

Introduction

The Defense Communications Systems (DCS) is a worldwide communications network serving the needs of the Commanders-in-Chiefs (CINCs) and their respective areas. The network frequency reference is in the process of being switched over from the LORAN-C network to the NAVSTAR Global Positioning System (GPS). The DCS is managed by the Defense Information Systems Agency (DISA).

The DCS, like other synchronous digital communications system, relies heavily on accurate frequencies being available throughout the network. Accurate and stable frequency is essential to reduce Loss of Bit Count Integrity (LBCI) and provide mission critical communications services to field operations. With the planned phase over from LORAN-C transmissions, the DCS is in the process of installing GPS frequency references at major nodes in the network to improve and maintain network stability.

The GPS receivers to be used in the DCS are manufactured by Quantic Industries, Inc. (Model Q-5200/SM) and contain the Stellar GPS Corporation Model 100 GPS Clock™, modified to handle Selective Availability (SA) signals.

The Defense Communications System (DCS)

The Defense Communications Agency (renamed Defense Information Systems Agency in 1991) began converting the DCS from an analog system to a digital system in the early 1970s, much in the same way that AT&T was converting their system. During the transition period the DCS digital transmis-

0-7803-0905-7/93 $3.00 © 1993 IEEE

sion subsystems were timed by clocks intrinsic to the equipment (e.g., internal clock) and by use of pulse stuffing and buffering, also intrinsic to the transmission equipment. However, the introduction of digital switches and other subsystems required the use of synchronous transmission, which requires that an accurate clock signal accompany the digital data signal.

Network Synchronization in the DCS

After studying various network synchronization schemes which could be used to support a synchronous network, an independent clock approach was selected to maximize network survivability. This approach employs plesiochronous clocks located at major equipment nodes, and buffers intrinsic to each piece of equipment to accommodate any timing differences between nodes. The design objectives which DISA intended to meet are outlined in CCITT Recommendation G. 811, which specifies that all clocks at network nodes should have a long-term frequency departure of not greater than $1x10^{-11}$ (Figure 1). In order to meet these objectives, an acquisition program was initiated to acquire the timing and synchronization (T&S) subsystems needed for the new approach.

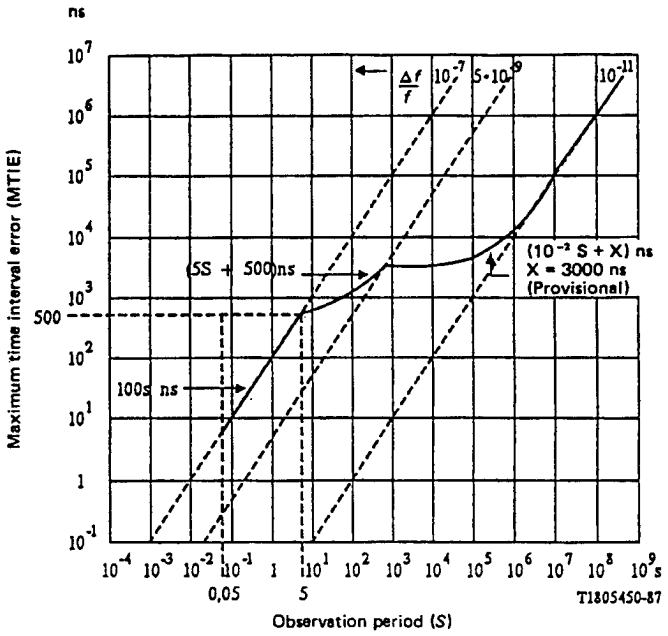


Figure 1. CCITT Recommendation G. 811

LORAN-C was selected as the original reference standard for the candidate timing and synchronization subsystem primarily because of its low cost, off-the-shelf availability, and virtual worldwide coverage necessary for a global communications system. A functional diagram of the AN/GSQ-215 Station Clock is shown in Figure 2. The LORAN-C receiver acquires and tracks selected stations and provides a corrected 1 MHz output reference for the redundant oscillators. During periods of outages, the oscillators enter a holdover mode in which they attempt to maintain the correct frequency. The output signals from the two oscillators, and a third which is directly tied back to the corrected frequency output from the LORAN-C receiver, are fed to a clock distribution system which employs a majority vote logic system. The resulting output signal is then used to generate a family of basic frequencies which are divided to generate the various clock rates needed by the equipment at the nodes.

By 1986, it became evident that the long-term goal of the DoD was to phase out LORAN-C and other navigation systems in favor of GPS. As a means toward this end, the USCG had announced their plans to turn over maintenance of all LORAN-C stations outside of the continental US to their respective host nations by the mid 1990s. Because DISA did not want to rely upon foreign operation and maintenance of LORAN-C stations for network synchronization, other primary reference sources were considered.

GPS appeared to be the most cost effective means of replacing LORAN-C as the primary reference. Prices for commercial-off-the-shelf (COTS) C/A Code receivers were falling rapidly in the early 1980's and the initial plans called for one of these receivers to replace the LORAN-C receiver in the existing timing system rack. Because both Selective Availability (SA) and the DCS are expected to be operational during wartime scenarios, it was soon decided that a Precise Positioning Service (PPS) receiver was the desired choice. However, other GPS alternatives were also examined.

Following suit with AT&T, a solution which combined computer software and GPS hardware was considered. This approach involved collecting 1PPS

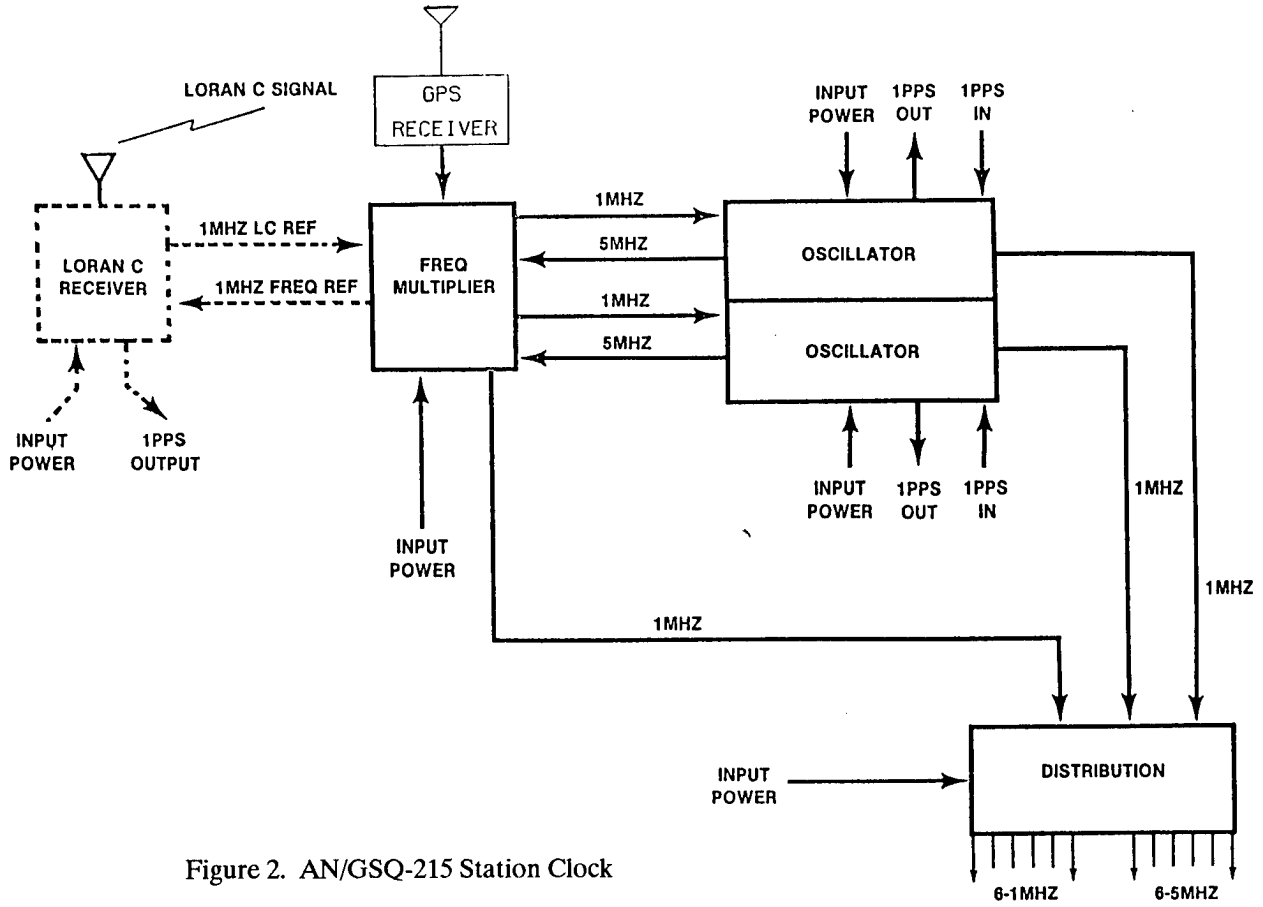


Figure 2. AN/GSQ-215 Station Clock

data from a C/A code receiver, and filtering outliers to determine daily clock corrections to a pair of rubidium oscillators. This solution was not regarded viable for the worldwide DCS because it would have required replacement of all existing timing system racks (which burdens the spare parts or "logistics system"), and because of the implications the software would place on operations (e/.g., procurement and maintenance of PCs, software distribution, debugging and maintenance).

Again, because of the geographical dispersion and sheer number of sites involved, differential correction of GPS signals did not appear viable for the same reason. Moreover, making network timing and synchronization dependent on yet another system seemed unnecessary. Because DISA was eligible as an authorized user of the Precise Positioning Service (PPS), a PPS capable receiver was the chosen solution.

Network synchronization requirements strongly point toward a frequency-based system, as opposed to a time transfer based system. P-Code PPS systems are more accurate in timing but do not offer perceptible improvements over a well designed frequency-based C/A Code receiver when frequency is the requirement. Therefore, a C/A Code receiver integrated with a PPS-Security Module (PPS-SM) was chosen.

The Quantic Industries Q-5200/SM Timing GPS Receiver (TGR) sets are being procured for DISA through the GPS Joint Program Office (JPO) at the Space and Missiles Systems Center, Los Angeles Air Force Base, California.

The Q-5200/SM Timing GPS Receiver

The Quantic Model Q-5200/SM TGR Set, consisting of the receiver and its associated antenna module, will replace the LORAN-C Receiver in the AN/GSQ-215 Station Clock used in the Defense Communications System (DCS) for network synchronization. The TGR set provides very accurate frequency as well as timing signals and information about the satellites. The stable frequency is used to feed the

AN/GSQ-215 Station Clock the reference frequency which is then used to generate frequencies required within the station equipment.

The TGR is a 3 1/2 inch high 19 inch rack mount GPS receiver containing the GPS Clock™ module (manufactured by Stellar GPS Corp.), a control and display front panel, and a power supply (Figures 3 and 4).

The GPS Clock module is a L1 C/A GPS Receiver optimized for frequency and time applications. The GPS Clock provides the user with a very stable 10 MHz reference frequency. The GPS Clock module employs a unique new architecture* that is specifically designed to obtain the ultimate in precision and stability available from the GPS. As a result of this new architecture, the one pulse per second (1PPS) timing signal exhibits an RMS pulse-to-pulse jitter less than 1 nanosecond, rather than the normal 20 to 60 nanosecond jitter observed on 1PPS outputs from time-based GPS receivers.

In the Q-5200/SM, the GPS Clock is modified to interface with a PPS-SM and incorporates the associated algorithms in order to provide PPS to the DCS.

The GPS Clock module responds to commands from either the RS-232 port or from the front panel keypad and provides output data to the front panel display or RS-232 port as required by specifications. When provided with an external 1 or 5 MHz frequency or an external 1PPS signal and when commanded to do so, the time and frequency measurement (TFM) section of the TGR continuously performs frequency and/or timing offset measurements relative to GPS. Measurement results are output by the TGR processor via the front panel display or the RS-232 port.

The 1 PPS offset is measured to a resolution of 5 nanoseconds and the frequency accuracy is measured to a resolution of $1x10^{-12}$.

The TGR is easy to operate. It has a default mode which allows the unit, when the antenna is connected and power applied, to automatically acquire and track satellites without operator intervention. When the "Loss of GPS" indicator is not illuminated, the unit is providing accurate frequencies and timing pulses that are corrected and locked to the GPS. On the other hand, the unit can be commanded from either the front panel or the RS-232C serial interface port to operate in a number of different modes. The RS-232C port allows the unit to be remotely monitored, making the unit especially suitable for unattended operations. If power were to fail and be reapplied, the unit will automatically resume operation using the last commanded mode settings supplied.

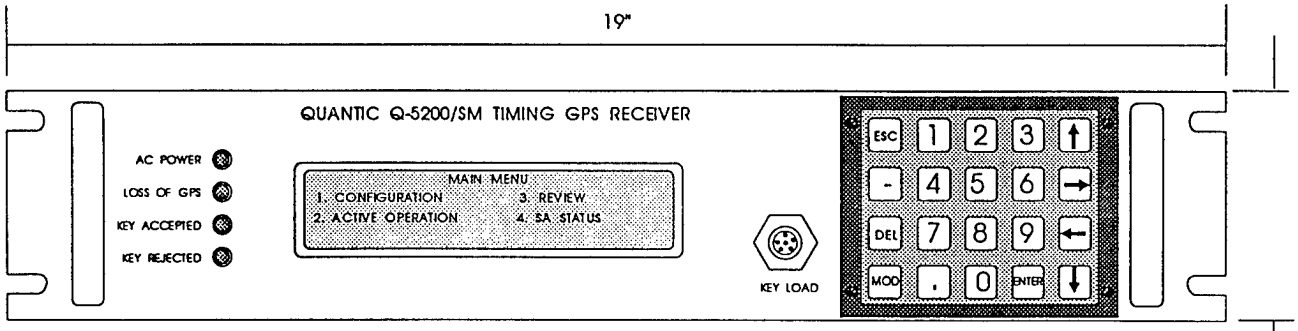
The front panel control and display subassembly contains a four line, 40 character liquid crystal display (LCD) and a 20-key pad for input of commands and data by the operator. The control and display is implemented with user friendly menus and sub-menus.

The TGR operates from either 115/230 Vac and/or -48 Vdc, if supplied. If the TGR loses AC power, it automatically switches to the DC supply, if available, without impact on the unit operation.

The TGR interfaces to the Antenna Module via a 200 foot RG-58/U cable or optional 1000 foot cable. The TGR sends both +15 Vdc power and a 11.114 MHz reference frequency up the cable to the Antenna Module. The Antenna Module receives the satellite L1 frequency from the satellites and down converts the signal to an IF of 153.028 MHz which is sent down the cable to the receiver.

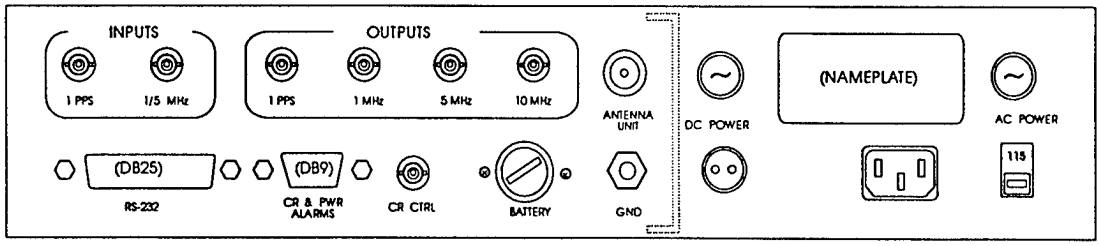
To maximize performance, time delays induced by the Antenna Module, variations in cable length, and the GPS Clock processing are measured. Provisions in the GPS Clock incorporate compensation for the system time delays to minimize the absolute timing errors.

* Patent Pending.



a) Front Panel

3.5"



b) Rear Panel

Figure 3. Q-5200/SM Timing GPS Receiver

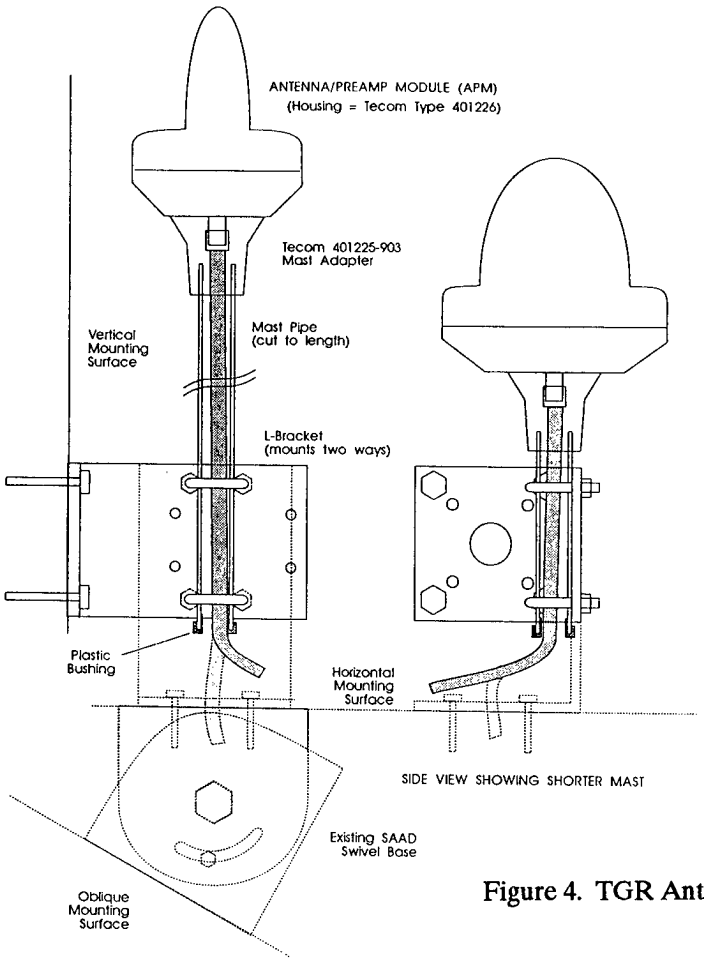


Figure 4. TGR Antenna and Mount

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

3S Navigation R100 Preliminary Implementation of GPS/GLONASS Time Transfer.

By

James Danaher and Anushia Balendra

3S Navigation
23141 Plaza Pointe Drive
Laguna Hills, CA 92653

Abstract

Introduction of the GPS common-view method has led to a most precise and accurate method for time comparison. It has been observed that GPS time transfer provides an accuracy at the level of 1-2 nanoseconds within the continent and 10-20 nanoseconds between continents. The Russian GLONASS system operates in a manner that is very similar to GPS and should be able to match this time transfer performance. The absence of Selective Availability (SA) on the GLONASS system makes it very attractive for time transfer and real-time precision frequency reference applications. GLONASS satellites also transmit P code signals on both L1 and L2 without Anti-Spoofing (AS) encryption. These signals should be usable for ionosphere density sounding, which will improve the performance of GPS/GLONASS receivers.

This paper presents the initial implementation of GLONASS time transfer results using 3S Navigation R100 Integrated GPS/GLONASS receiver. It also presents the capability of R-100 GLONASS C/A code signals to provide real time frequency reference.

1. Introduction.

Integrated GPS/GLONASS navigation and timing offers better availability, performance and integrity than GPS-only operation.[1]. The 3S Navigation R-100 family of receivers includes the R-100 and the R-101. The R-100 is a C/A code integrated GPS/GLONASS receiver. It has a twelve channel capability with each GLONASS satellite requiring one channel and each GPS satellite requiring two channels. The R-100 is therefore capable of simultaneously tracking up to twelve GLONASS satellites, up to six GPS satellites or a combination of both. The basic R-101 is a four channel dual frequency (L1 and L2) GLONASS P code receiver. Optional expansion boards allow increase of the number of channels and inclusion of the R-100 GPS/GLONASS C/A code capabilities. Both receivers are designed to meet the needs of a wide variety of users, from the navigator, requiring mainly position, velocity, and acceleration information to the researcher, requiring detailed raw data. Outputs include navigation and time solutions in several coordinate systems and based on GPS-only, GLONASS-only, and integrated GPS/GLONASS observations. Raw pseudo-range and integrated carrier phase measurements with associated error characteristics, carrier smoothed pseudo-range, C/No, satellite ephemerides and almanac are provided. 3S Navigation receivers are also well suited for operating as a differential reference station.

Since there is no SA on the GLONASS system, it can provide a very precise real-time frequency reference. In addition the GLONASS P code is currently not encrypted and has a wider bandwidth implying a better range measurement and compared to C/A code.

The second section of this paper shows GLONASS C/A Vs P code range residuals. The wave length of the carrier is much shorter than the wave length of the code. Hence precision of the carrier phase measurement is much higher than the precision of the range measurements. Integrated carrier phase residual for both GLONASS C/A and P code are also presented for comparison. GLONASS based time transfer between two R-100 digital boards using C/A code and the accuracy improvement when integrated carrier phase is used are also included. The "CCDS Group on GPS Time Transfer Standards" [2] format data is presented for code and integrated carrier phase based computations.

The third section presents the oscillator characterization using R-100 GLONASS C/A code signal. Plots are provided to show the results. Conclusions and possible improvements are given in section 4.

2. GLONASS Time Transfer

The GLONASS pseudo-range and carrier phase residuals for both C/A and P code are presented in Figures 2.1 and 2.2. The antenna location in SGS85

0-7803-0905-7/93 $3.00 © 1993 IEEE

reference is computed from surveyed WGS84 location using the WGS84 to SGS85 transformation[3] published by MIT Lincoln Laboratories. Broadcast ephemerides and modeled ionospheric and tropospheric delays are used to determine the residuals.

Figure 2.1 shows the GLONASS channels 1 range residual for the run conducted on June 2, 1993 from approximately 21:13:20 UTC to 21:30:00 UTC for C/A and P code. A Sapphire Cesium clock from the Russian Institute of Radio navigation and Time (RIRT) was provided by Norell, Inc. for this test. As expected, P code range residual is less noisier than the C/A code range residual. The higher accuracy carrier phase residuals are also shown. Figure 2.2 shows the same residual plot as in Figure 2.1 in a bigger scale.

A preliminary implementation of the CCDS time transfer data format for two R-100 digital boards are shown in Figures 2.3 and 2.4 respectively. This is a zero base-line test where two R-100 digital boards were connected to a single antenna and RF unit. The Russian Sapphire Cesium clock was used as reference. The same data computed using integrated carrier phase is given in Figures 2.5 and 2.6. The time difference between two boards A and B is shown in Figure 2.7.

From Figure 2.7, it can be concluded that time transfer was performed between two R-100 digital boards provides less than 4.0 ns RMS difference. When integrated carrier phase is used, the accuracy is improved with less than 0.1 ns RMS difference but the absolute time reference was lost. When carrier phase smoothing is applied to the pseudo-range data, the accuracy should fall between the two approaches shown in Figure 2.7.

In this test, multipath, ionospheric delay, and analog delay variations were common-mode. In field conditions these effects will be different for each receiver. Thus the time transfer result will be worse than the capability of the digital and software processing as shown in Figure 2.7.

3. Oscillator Characterization.

R-100 GLONASS C/A code signals can be used to provide real time frequency reference and for oscillator characterization and calibration. Figure 3.1 shows the carrier phase residual in meters when a TCXO (crystal), two different Ball Ephratom Rubidium oscillators, and the Russian Sapphire Cesium oscillator are used as external reference. Figure 3.1 is enlarged in Figure 3.2. As seen from these figures, cesium has the least noise, and rubidium #1, rubidium #2, and cesium are very stable compared to the TCXO. Figure 3.3 presents the Allen variance plot for the same data set. This figure shows that the R-100 can use GLONASS C/A code signals to provide a real time reference with $\sigma(\tau)$ better than 2×10^{-11} when τ is between 1 and 100 seconds.

4. Conclusions and Improvements.

The accuracy of the GLONASS C/A code time transfer can be improved by integrated carrier phase processing. In real time the performance will be degraded due to multipath, analog delay variations in R-100 and ionospheric effects. GLONASS signals can provide a real time precise frequency reference. Improvements in the performance shown in this paper can be achieved with tuning loops for static operation of R-100, processing multiple satellite measurements, improved real time GLONASS orbit prediction, differential operation, and dual frequency ionospheric correction.

5. References.

[1] J. Besser and J. Danaher, "The 3S Navigation R-100 Family of Integrated GPS/GLONASS Receivers: Description and Performance Results," presented at the ION 49th Annual Meeting, San Francisco, California, January 21-23, 1993.

[2] Report of the 2nd meeting of the CCDS Group on GPS Time Transfer Standards, Paris, France, June 11, 1992.

[3] P. Mistra, R. Abbott, and E. Baylis, "SGS85 - WGS84 Transformation: Interim Results," presented at ION 49th Annual Meeting, San Francisco, California, January 21-23, 1993.

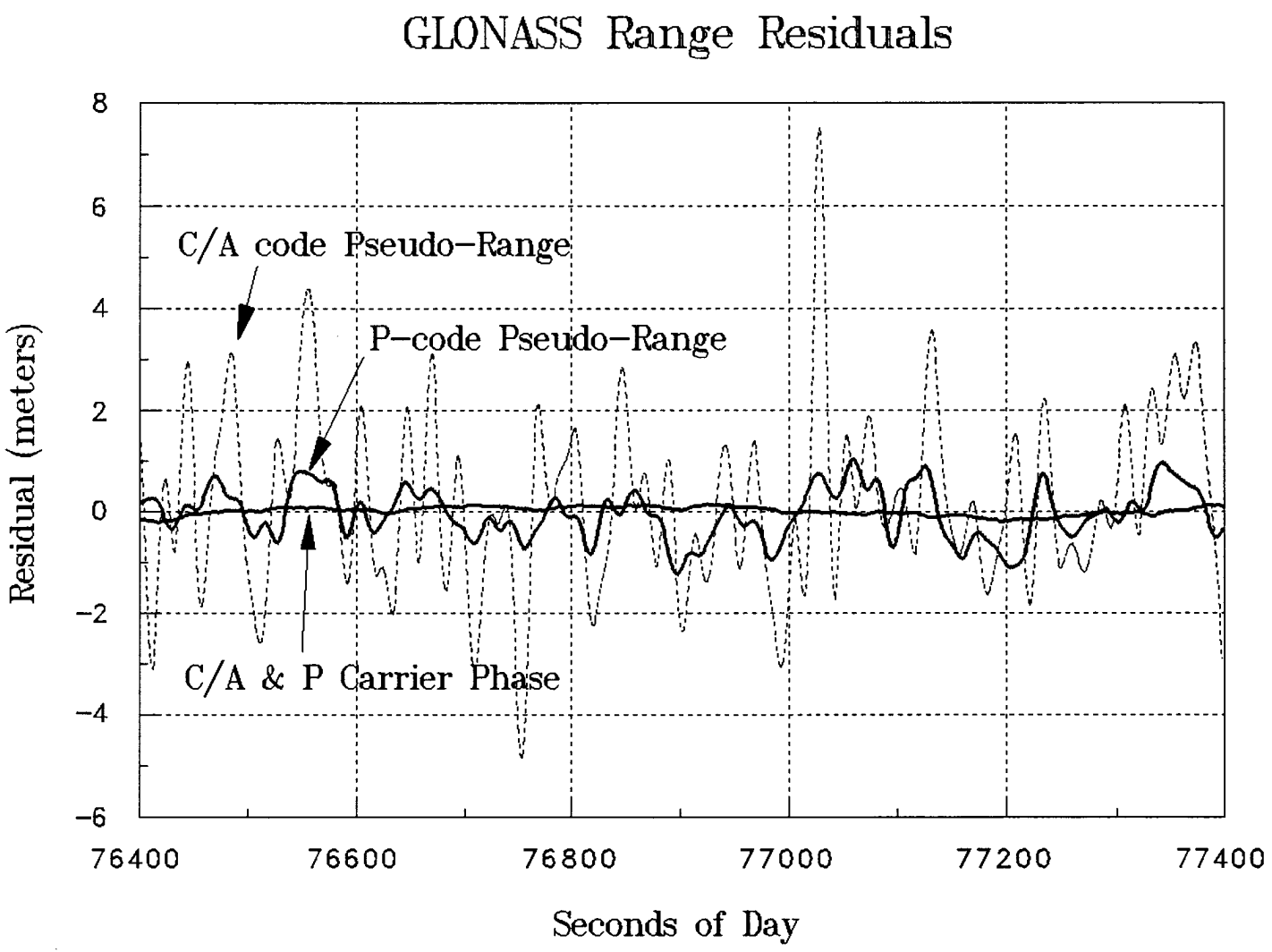


Figure 2.1 GLONASS ch. 1 (L1, Cs) R-100 & R-101.
June 2, 1993 at 3S Navigation.

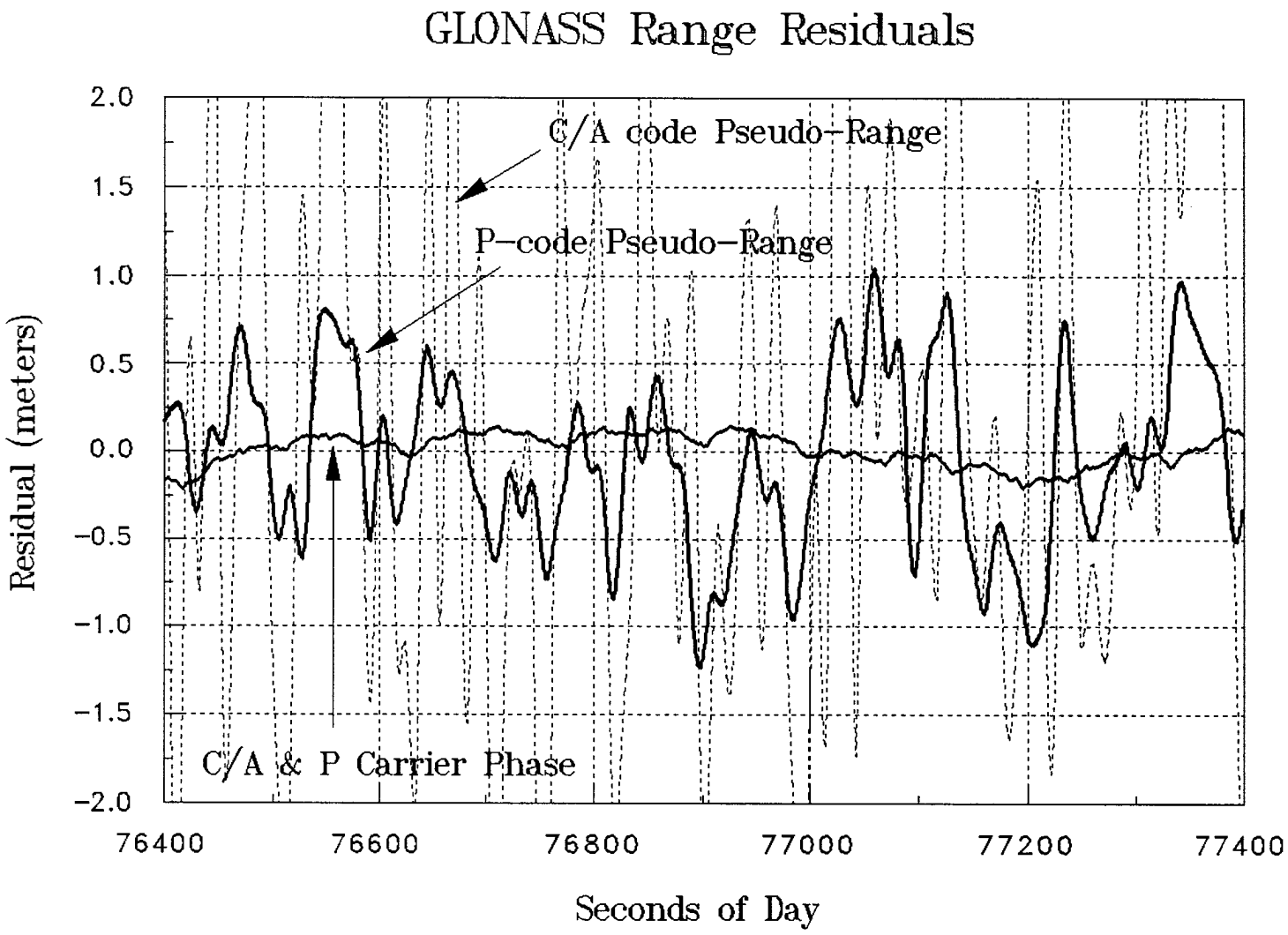


Figure 2.2 GLONASS ch. 1 (L1, Cs) R-100 & R-101.
June 2, 1993 at 3S Navigation.

```
CCDS GPS DATA FORMAT VERSION 1
3S R100 GB1/SN0002 1992 707G/NEW000
CH = 1 (GLONASS L1 C/A) INPUT = 24jun07a.mea

UTC(3SNAV) = 778525362 ns, -79.9 ps/s AT 49033 00:00:00

X = -2473150.49m
Y = -4706100.58m
Z = +3512038.98m
INT DLY = 500.0ns
CAB DLY =  41.6ns
REF DLY =   0.0ns
UTC(3SNAV)

PRN CL  MDJ  HHMMSS TRKL ELV AZTH   REFSV      SRSV     REFGPS    SRGPS  DSG IOE MDTR SMDT MDIO SMDI
             UTC    s  .1dg .1gd    .1ns     .1ps/s     .1ns    .1ps/s .1ns     .1ns.1ps/s.1ns.1ps/s
 12  1 49162 071000  780 400 1214 -4731910085  +766 -4729571218  +857   83  41  126 +16   73   +8
 12  1 49162 072300  780 349 1282 -4731909519  +693 -4729570582  +784   69  41  142 +24   80  +10
 12  1 49162 073600  780 294 1341 -4731909054  +674 -4729570034  +774   73  43  165 +35   89  +12
 12  1 49162 074900  780 238 1391 -4731908596  +543 -4729569498  +643   85  43  200 +56  100  +15
 12  1 49162 080200  780 181 1433 -4731908161  +516 -4729568998  +616  115  45  260 +100 113  +18
 12  1 49162 081500  780 124 1468 -4731907739  +623 -4729568497  +723  286  45  376 +212 128  +20
```

Figure 2.3. CCDS Format Data for R-100 Board A.

```
CCDS GPS DATA FORMAT VERSION 1
3S R100 GB1/SN0002 1992 707G/NEW000
CH = 1 (GLONASS L1 C/A) INPUT = 24jun07b.mea

UTC(3SNAV) = 778525362 ns, -79.9 ps/s AT 49033 00:00:00

X = -2473150.49m
Y = -4706100.58m
Z = +3512038.98m
INT DLY = 500.0ns
CAB DLY =  41.6ns
REF DLY =   0.0ns
UTC(3SNAV)

PRN CL  MDJ  HHMMSS TRKL ELV AZTH   REFSV      SRSV     REFGPS    SRGPS  DSG IOE MDTR SMDT MDIO SMDI
             UTC    s  .1dg .1gd    .1ns     .1ps/s     .1ns    .1ps/s .1ns     .1ns.1ps/s.1ns.1ps/s
 12  1 49162 071000  780 400 1214 -4731910074  +720 -4729571208  +811   54  41  126 +16   73   +8
 12  1 49162 072300  780 349 1282 -4731909535  +675 -4729570598  +766   67  41  142 +24   80  +10
 12  1 49162 073600  780 294 1341 -4731909045  +633 -4729570025  +733   88  43  165 +35   89  +12
 12  1 49162 074900  780 238 1391 -4731908557  +626 -4729569459  +726   80  43  200 +56  100  +15
 12  1 49162 080200  780 181 1433 -4731908127  +499 -4729568963  +599  121  45  260 +100 113  +18
 12  1 49162 081500  780 124 1468 -4731907733  +648 -4729568491  +748  282  45  376 +212 128  +20
```

Figure 2.4. CCDS Format Data for R-100 Board B.

```
CCDS GPS DATA FORMAT VERSION 1
3S R100 GB1/SN0002 1992 707G/NEW000
CH = 1 (GLONASS L1 C/A) INPUT = 24jun07a.mea ( Integrated Carrier Phase Based Computation)

UTC(3SNAV) = 778525362 ns, -79.9 ps/s AT 49033 00:00:00

X = -2473150.49m
Y = -4706100.58m
Z = +3512038.98m
INT DLY = 500.0ns
CAB DLY =  41.6ns
REF DLY =   0.0ns
UTC(3SNAV)

PRN CL   MDJ   HHMMSS TRKL ELV AZTH   REFSV      SRSV     REFGPS    SRGPS  DSG IOE MDTR SMDT MDIO SMDI
              UTC    s  .1dg .1gd    .1ns     .1ps/s     .1ns    .1ps/s .1ns    .1ns.1ps/s.1ns.1ps/s
 12  1 49162 071000  780 400 1214 +10092191031   +888 +10094529897   +979  0  41 126  +16   73   +8
 12  1 49162 072300  780 349 1282 +10092191708   +844 +10094530645   +935  1  41 142  +24   80  +10
 12  1 49162 073600  780 294 1341 +10092192307   +743 +10094531327   +843  1  43 165  +35   89  +12
 12  1 49162 074900  780 238 1391 +10092192890   +738 +10094531988   +838  1  43 200  +56  100  +15
 12  1 49162 080200  780 181 1433 +10092193452   +688 +10094532616   +788  0  45 260 +100  113  +18
 12  1 49162 081500  780 124 1468 +10092193985   +694 +10094533227   +794  2  45 376 +212  128  +20
```

Figure 2.5. CCDS Format Data for R-100 Board A Based on Integrated Carrier Phase.

```
CCDS GPS DATA FORMAT VERSION 1
3S R100 GB1/SN0002 1992 707G/NEW000
CH = 1 (GLONASS L1 C/A) INPUT = 24jun07b.mea ( Integrated Carrier Phase Based Computation)


UTC(3SNAV) = 778525362 ns, -79.9 ps/s AT 49033 00:00:00

X = -2473150.49m
Y = -4706100.58m
Z = +3512038.98m
INT DLY = 500.0ns
CAB DLY =  41.6ns
REF DLY =   0.0ns
UTC(3SNAV)

PRN CL   MDJ   HHMMSS TRKL ELV AZTH   REFSV      SRSV     REFGPS    SRGPS  DSG IOE MDTR SMDT MDIO SMDI
              UTC    s  .1dg .1gd    .1ns     .1ps/s     .1ns    .1ps/s .1ns
.1ns.1ps/s.1ns.1ps/s
 12  1 49162 071000  780 400 1214 +10109378660   +888 +10111717527   +979  0  41 126  +16   73   +8
 12  1 49162 072300  780 349 1282 +10109379338   +844 +10111718275   +935  1  41 142  +24   80  +10
 12  1 49162 073600  780 294 1341 +10109379937   +743 +10111718957   +843  1  43 165  +35   89  +12
 12  1 49162 074900  780 238 1391 +10109380519   +738 +10111719617   +838  1  43 200  +56  100  +15
 12  1 49162 080200  780 181 1433 +10109381082   +688 +10111720246   +788  0  45 260 +100  113  +18
 12  1 49162 081500  780 124 1468 +10109381615   +694 +10111720856   +794  2  45 376 +212  128  +20
```

Figure 2.6. CCDS Format Data for R-100 Board B Based on Integrated Carrier Phase.

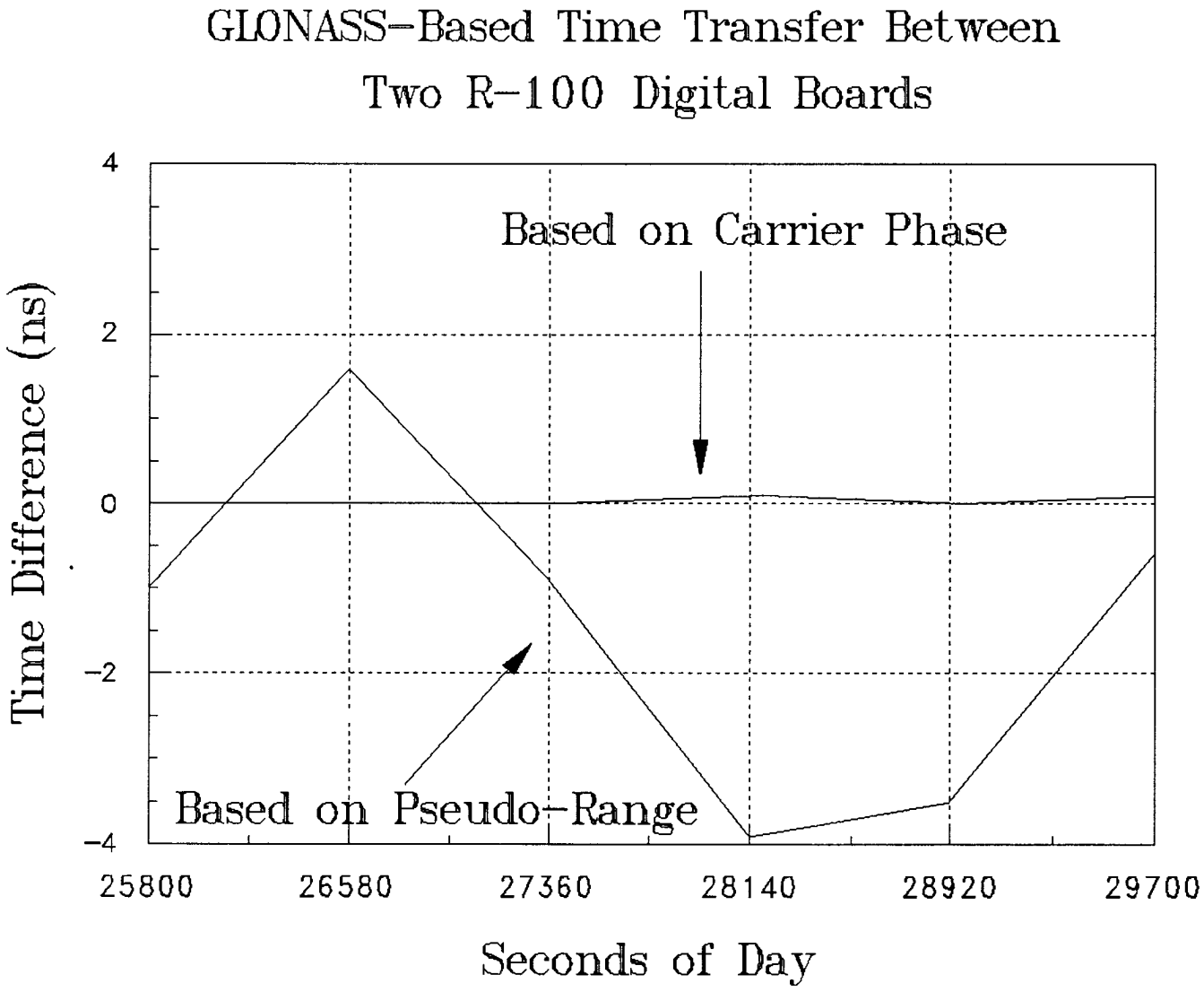


Figure 2.7. Time Transfer Between R-100 A(Cs) and R-100 B(Cs) Using GLONASS L1 C/A ch. 12. June 24, 1993 at 3S Navigation.

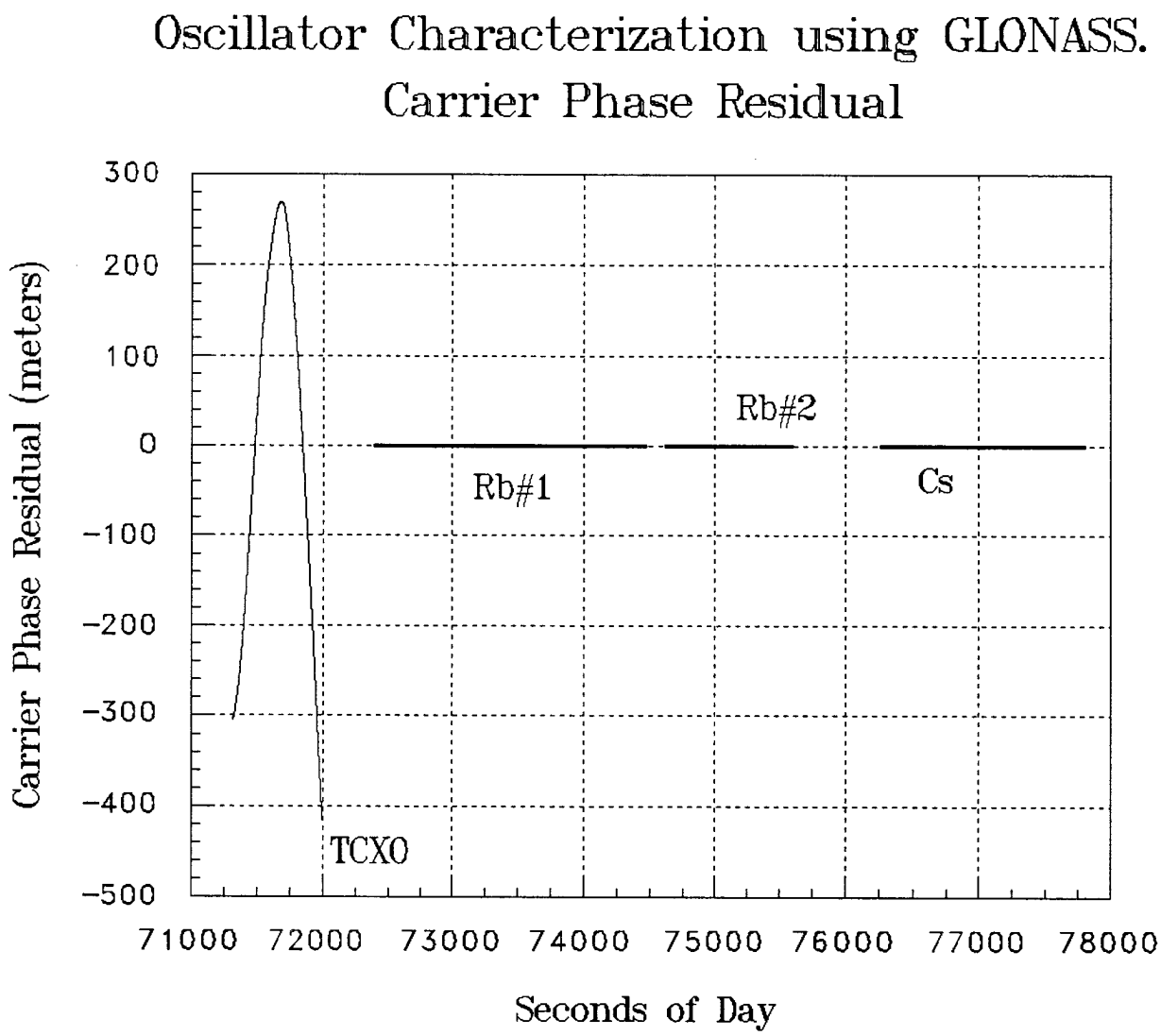


Figure 3.1. Oscillator Characterization Using GLONASS L1 C/A ch. 13. May 24, 1993 at 3S Navigation.

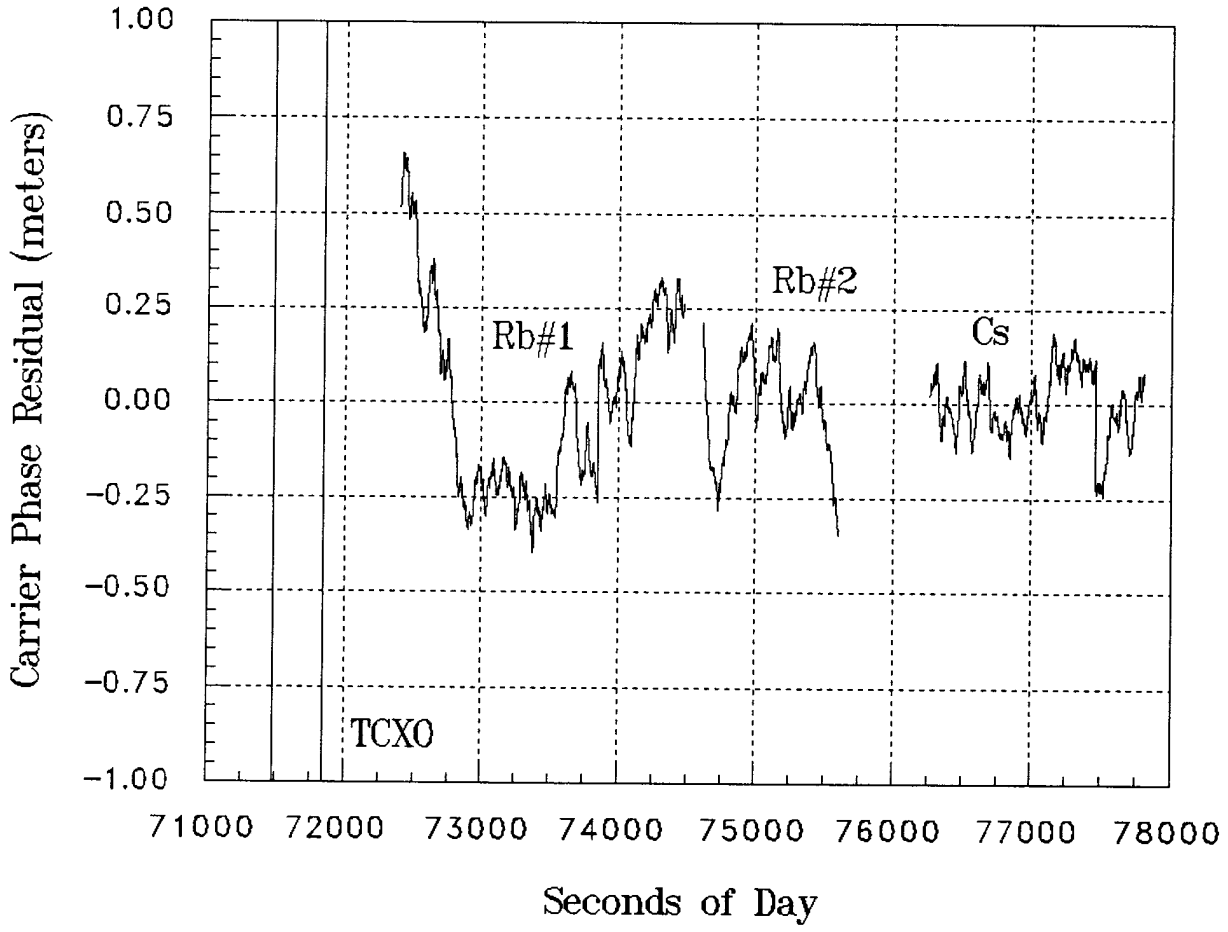


Figure 3.2. Oscillator Characterization Using GLONASS L1 C/A ch. 13. May 24, 1993 at 3S Navigation.

Oscillator Characterization using GLONASS.

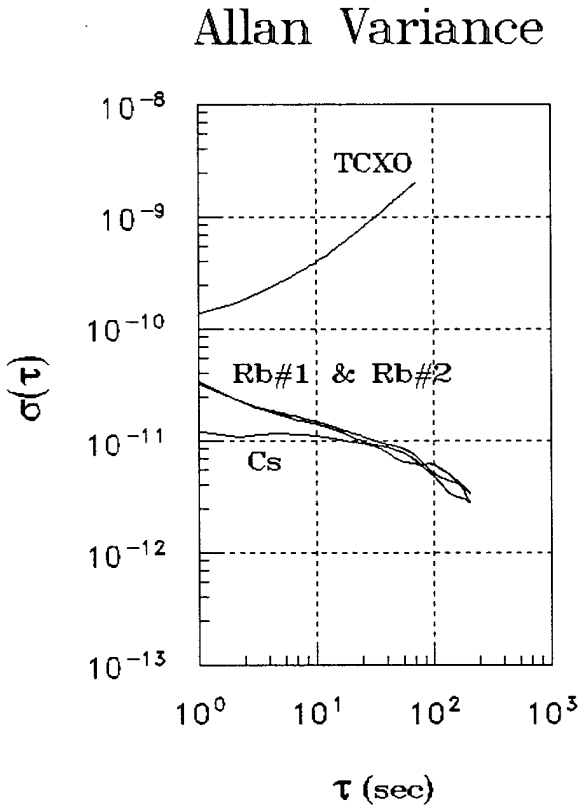


Figure 3.3. Oscillator Characterization Using GLONASS

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PRECISE TIME DISSEMINATION USING THE INMARSAT GEOSTATIONARY OVERLAY

Alison Brown, NAVSYS Corporation
14960 Woodcarver Road, Colorado Springs, CO 80921

David W. Allan, Allan's TIME, and Rick Walton, COMSAT

Abstract

INMARSAT has designed a GPS (L1) transponder that will be included in their third generation satellites. This transponder will broadcast a pseudo-GPS signal that can be used for navigation and also for disseminating integrity data and/or differential corrections for the GPS satellites. This INMARSAT Geostationary Overlay (IGO) service will be used to enhance the performance of the GPS navigation service for civil aviation and other users.

The IGO service can also be used as a method of disseminating a precise time reference. Since the IGO signal is compatible with GPS, conventional GPS timing receivers can easily be modified to utilize the proposed service. Preliminary test results taken through the IGO satellite have demonstrated a timing accuracy of 10 nanoseconds.

Introduction

The INMARSAT-III constellation of four geostationary satellites will provide redundant coverage over most of the earth, as illustrated in Figure 1. In addition to the communications payload, the INMARSAT-III satellites will also carry a navigation transponder that will be used to broadcast GPS-like signals. These signals can be used for navigation and will broadcast a GPS Integrity Broadcast (GIB) generated by the Federal Aviation Administration (FAA) that provides warnings to users when the GPS service is not operating correctly.

The INMARSAT Geostationary Overlay (IGO) signal is generated at the satellite earth station and is controlled so that the IGO signal broadcast by the satellite appears to be synchronized with the GPS satellite signals. It is also possible to use this architecture to precisely synchronize the IGO signal to a time reference. Since only a single satellite signal is required for precise time dissemination at fixed installations, the four INMARSAT-III satellites can provide redundant world-wide coverage for precise time dissemination.

NAVSYS, COMSAT, and NIST have signed a cooperative research and development agreement to perform a timing experiment using the FAA's GIB Test-Bed equipment. The architecture to be employed in this experiment is illustrated in Figure 2. The IGO signal generator design by NAVSYS is installed at the Southbury Earth Station operated by COMSAT. This signal generator is designed to synchronize the IGO signal to a precision time reference installed at the earth station. A monitor station installed at NIST will be used to measure the accuracy of the broadcast IGO signals. NIST will also provide time and frequency corrections to the earth station to steer the precision time reference and to calibrate for observed offsets in the system. This testing will continue through 1994. Preliminary results from the initial tests are included in this paper. In the long term, INMARSAT time could be that of UTC as generated at the Bureau Internationale de Poids et Mesures.

International High Accuracy Timing

The INMARSAT transponder can be used to provide an international high accuracy timing service. Usages of a service vary depending on the application and include, for example, the need for time accuracy, time stability, time predictability, frequency accuracy, and frequency stability.[1] The typical user, of course, may need high accuracy or stability at some generic site, but wishes not to have a high investment in timing equipment. At the same time, reliability and redundancy are often very important issues.

The capabilities of different time and frequency dissemination systems are summarized in Table 1. This table was prepared by the International Telecommunications Union (ITU) Radiocommunication Study Group 7A held in Geneva in April 1993.[2]

Toward the high accuracy end of time and frequency transfer systems, GPS has had a major impact on international timing. The large number of users for positioning and navigation have driven GPS receiver prices down to around a thousand dollars (US). The same is not true for GPS timing receivers, because of the smaller number of users. However, with the

0-7803-0905-7/93 $3.00 © 1993 IEEE

growing high accuracy needs within telecommunications, such as with the Synchronous Optical Network (SONET) and the Synchronous Digital Hierarchy (SDH), as well as within the power industry, the prices of receivers will decrease in natural consequence.

In response to the current and anticipated needs within the telecommunications industry, the 1993 Consultative Committee for the Definition of the Second (CCDS) wrote a recommendation encouraging the study of the technical problems associated with the goal of 100 ns worldwide synchronization. As these problems are studied, it is anticipated that many of them will have solutions within currently available resources. The resources within time and frequency have improved dramatically over the last decade, and there is good reason to believe they will continue to improve. The availability of GPS, alone, has been very significant to the telecommunications industries (telecom) as well as to the time and frequency community.

The official time and frequency reference within the US for telecommunications is UTC. [3] Yet if you ask the question of telecom leaders, "How many of you are using UTC?" the answer will almost always be no one! Everyone knows UTC is an outstanding time scale, but the problem is that it tells you what time it was several weeks after the fact, and in addition, it is not readily accessible in any direct way. Indirectly, it is accessible via GPS, which also broadcasts UTC (USNO). UTC (USNO) is now steered to within 100 ns of UTC.

GPS, being a US military system, has not been accepted by all countries as a reliable reference. In this regard, apprehensions may be greater than they need to be. An official Civil GPS Interface Service Committee (CGSIC) has been set up between the US Department of Defense and the Department of Commerce. A computer bulletin board, accessible internationally, has been set up giving current information about the status of GPS as well as a set of post-processed precise ephemerides. In addition, a Memorandum of Agreement has been signed between the Department of Defense and the Department of Transportation (effective 8 January 1993 through the year 2005) for the civil use of GPS.

The INMARSAT timing system described in this paper could provide three significant steps forward. It could provide an international civil reference time scale, which could be a real time reference to UTC. It could also provide a set of corrections providing higher accuracy usage of GPS, given the degradation caused by selective availability. In addition, as will be described in this paper, it could also provide higher accuracy at less cost and with more reliability then can be obtained with GPS.

In telecommunications, support from the time and frequency community could significantly enhance the accuracy and the rate of information flow. UTC could become their real-time, ultra-accurate reference on a global basis. Now most of the telecom servers generate their own timing, yet different servers have to communicate with each other, which often creates information flow conflicts. In practice, to avoid some of the conflicts in data flow, the servers tend to use the biggest server as a reference. Having an externally unbiased and readily available reference—always somewhat better than their needs—would mitigate many of the current problems. This external support from UTC as a real-time timing reference could significantly improve data flow efficiency. Such support may also reduce costs, since the servers would not have to use as much of their resources to supply precise timing.

INMARSAT Geostationary Overlay (IGO)

INMARSAT has ordered four INMARSAT-III satellites which will include C-to-L1 transponders and C-to-C band transponders (for atmospheric corrections). Contracts for all four spacecraft launches have been signed, with the first launch scheduled for December 1994.

The INMARSAT Council took the decision to include navigation transponders on the third generation satellites in order to accomplish three major objectives:

1) to provide real-time (within 6 seconds as established by FAA and ICAO) integrity status of each of the GPS satellites and other navigation satellites such as GLONASS;

2) to provide a geostationary overlay to existing navigation satellite systems (hence the term INMARSAT Geostationary Overlay or IGO) which would augment other satellite navigation systems with an undegraded (no selective availability) navigation signal; and

3) to provide, if possible, Wide Area Differential GPS (WADGPS) corrections to enhance air traffic safety.

In addition to providing a navigation and integrity service, the transponder signal can also be used to

disseminate a precise time reference. This would appear similar to a GPS satellite signal to a user, but would be synchronized to a precision time reference and would provide access to this reference over a large area.

INMARSAT has been a strong participant in the Radio Technical Commission for Aviation (RTCA) and other international forums which are responsible for the design and specification of the signal structure, data format, and operational characteristics of the Integrity/WADGPS broadcast. The RTCA Sub-committee 159 is responsible for writing the performance specifications (MOPS) for the Integrity broadcast that will eventually form the basis for a GPS-based sole means of navigation for commercial aircraft. The INMARSAT-III transponder will be an integral part of the Integrity broadcast to warn pilots when a GPS satellite is providing erroneous signals.

The navigation transponder on each INMARSAT-III consists of fully redundant (except for the antenna) C-to-L1 and C-to-C band translators as well as redundant transmit amplifiers (HPAs). A block diagram of the transponder is shown in Figure 3. The transponder receives GPS-like signals on the 6.4 GHz up-link to the satellite from the earth station and retransmits these signals at L1 (1575.42 MHz) on an earth-coverage antenna. The GPS-like signals use a 1.023 Mbps C/A code from the same family of codes used for GPS. The C/A code is modulated onto the carrier with 50-500 bps data which includes the Integrity message and WADGPS data.

The data rate on normal GPS signals is 50 bps, but the Integrity and WADGPS signals require a 250 bps rate with rate ½ convolutional coding which brings the total rate to 500 bps. This data rate is required to provide timely warnings for all of the GPS (and GLONASS) satellites and to meet the accuracy requirements for non-precision approaches.

IGO Signal Generator

A block diagram of the IGO signal generator is included in Figure 4. This has been designed by NAVSYS to provide precise synchronization of the IGO signal to an external time reference. The signal generator (SIGGEN) includes the following components:

1) a communication server which is used to receive the formatted Integrity and WADGPS message to be transmitted on the IGO signal;

2) a SIGGEN time and frequency reference to which the IGO signal is synchronized;

3) a SIGGEN controller which generates and controls the IGO IF signal output to the earth station for up-link to the satellite; and

4) a SIGGEN monitor which receives the IGO signal and provides the feedback data used in the SIGGEN control algorithms.

SIGGEN Communication Server

The FAA is developing a network of ground-based reference stations which will be used to continuously monitor the status of the GPS satellites and generate differential corrections for the observed range errors. This data is processed at a central facility to generate a GPS Integrity Broadcast (GIB) message for transmission by the IGO. The function of the communication server is to continuously receive the GIB message from the FAA central facility and then pass this data to the SIGGEN controller for modulation on the IGO signal.

SIGGEN Time and Frequency Reference

The SIGGEN time and frequency reference provides the time standard to which the IGO signal is synchronized. In the initial test phase, an HP 5071A primary frequency standard has been provided on loan by Hewlett Packard. The HP 5071A includes an improved cesium beam tube design that results in an accuracy of $\pm 2 \times 10^{-12}$. The HP 5071A will be operated during the test phase under remote control by NIST to adjust the reference for frequency offset and maintain it synchronized with NIST's time standard.

SIGGEN Controller

The purpose of the SIGGEN controller is to generate the IGO signal and control its timing relative to the SIGGEN precise time reference. The IGO signal is steered so that the timing of the signal (the C/A code and data epochs) appear to be synchronous with the SIGGEN time reference when they are transmitted by the INMARSAT-III satellite. In order to achieve this, the signal output by the SIGGEN controller must be advanced in time to compensate for the delays on the up-link path through the satellite transponder.

The signal output by the SIGGEN controller is characterized by the following equation.

$$S_{IF}^{XTM}(t) = D(t + \tau_C)\, C\,(t + \tau_C) \cos 2\pi(f_{IF}t + \delta f_C t + \theta) \qquad (1)$$

where S_{IF}^{XTM} is the IF signal output by the SIGGEN
C(t) is the C/A code at time t
D(t) is the integrity data modulated on the signal
τ_C is the controller time advance
f_{IF} is the nominal frequency of the IF signal (near 70 MHz)
δf_C is the frequency offset inserted by the controller

The IF signal is mixed up to C-band by the earth station, adjusted to compensate for the satellite Doppler, and broadcast up to the satellite where it is mixed to L-band. The signal broadcast by the INMARSAT-III satellite is characterized by the following equation.

$$S_{SV}^{XTM}(t) = D(t + \tau_C - \tau_D)\, C\,(t + \tau_C - \tau_D) \cos 2\pi(f_{L1}\, t + \delta f_C t - \delta f_D t + \theta') \quad (2)$$

where S_{SV}^{XTM} is the L1 signal broadcast by the INMARSAT-III satellite
τ_D is the time delay in the signal path from the SIGGEN
f_{L1} is the GPS L1 frequency (1575.42 MHz)
δf_D is the composite frequency offsets in the signal path from the SIGGEN

In order for the IGO signal to appear as a synchronous GPS-type satellite signal, the time and frequency offsets inserted by the controller must be driven to cancel out the time delay and frequency offsets in the signal path from the controller to the satellite (i.e. $\tau_C=\tau_D$ and $\delta f_C=\delta f_D$). These offsets consist of the following composite effects.

$$\tau_D = \tau_{TXES} + R/c + \tau_{TROPO} + \tau_{IONO}^{C} + \tau_{SV} \quad (3)$$
$$\delta f_D = \delta f_{ES} - R'/c + \delta f_{SV} \quad (4)$$

where τ_{TXES} is the group delay in the earth station path to the up-link antenna
R and R′ are the range and range rate to the satellite in meters and m/s
c is the speed of light (m/s)
τ_{TROPO} is the group delay from the tropospheric portion of the atmosphere
τ_{IONO}^{C} is the ionospheric group delay on the C-band up-link to the satellite
τ_{SV} is the group delay in the satellite transponder
δf_{ES} is the frequency compensation applied at the earth station
δf_{SV} is the frequency offset due to drifts in the transponder frequency reference

SIGGEN Monitor

In order to dynamically compensate for the group delays and frequency offsets, the SIGGEN monitor is used to measure the time and frequency offsets of the received signal relative to the SIGGEN time and frequency reference. The received SIGGEN signal is described by the following equation.

$$S_{ES}^{RX}(t) = D(t - \tau_R) \cos 2\pi(f_{L1} t + \delta f_R t + \theta'') \quad (5)$$

where S_{ES}^{RX} is the L1 signal received by the SIGGEN monitor
τ_R is the measured time delay from the reference
δf_R is the measured frequency offset from the reference

The measured time offset is related to the controller signal through the following equations.

$$\tau_R = \tau_D - \tau_C + R/c + \tau_{TROPO} + \tau_{IONO}^{L1} + \tau_{RXES} + n_{PR} \quad (6)$$
$$\delta f_R = \delta f_D - \delta f_C - R'/c + n_{DR} \quad (7)$$

where τ_{IONO}^{L1} is the ionospheric group delay on the L1 down-link from the satellite
τ_{RXES} is the group delay in the earth station reception path
n_{PR} is the measurement error in the SIGGEN code tracking loops
n_{DR} is the measurement error in the SIGGEN frequency tracking loops

SIGGEN Control Algorithm

The SIGGEN control algorithm uses the measurements of the received time and frequency offsets (τ_R and δf_R) and a measurement of the controller state (τ_C) to synchronize the IGO signal with the time reference. The following steps are performed by the algorithm.

1) Estimated delays are calculated to correct the observations for the up-link and down-link atmospheric, earth station, and satellite group delays τ_{est}^{U} and τ_{est}^{D}. The ionospheric delays are observed through dual frequency observations. The tropospheric delays are modeled and the earth station and satellite group delays are removed through calibration parameters.

$$\tau_{est}^{U} = \tau_{TXES} + \tau_{TROPO} + \tau_{IONO}^{C} + \tau_{SV} \quad (8)$$
$$\tau_{est}^{D} = \tau_{RXES} + \tau_{TROPO} + \tau_{IONO}^{L1} \quad (9)$$

2) The range and range-rate (R and R′) to the satellite are estimated through a Kalman filter using the following observable (Z_R). This is applied as an update

to the filter to generate estimates of the range and range-rate.

$$Z_R = \tau_R - \tau_{est}^D + \tau_C - \tau_{est}^U = 2R/c + \epsilon_{est}^U + \epsilon_{est}^D + n_{PR} \quad (10)$$

The accuracy of the final range and range-rate estimates (ϵ_R and ϵ_R') is a function of the calibration errors in the up-link and down-link to the satellites (ϵ_{est}^U and ϵ_{est}^D), the noise on the receiver code measurements (n_{PR}), and the time constant of the Kalman filter. Since the satellite is in a highly predictable geostationary orbit, the range and range-rate estimates can be smoothed to reduce their residual error to a minimal level.

3) The time and frequency of the controller are adjusted to minimize the following residuals.

$$Z_C = \tau_C - R/c + \tau_c^U = \epsilon_R + \epsilon_{est}^U \quad (11)$$

$$Z_F = \delta f_C + R'/c = \epsilon_{R'} + n_{DR} \quad (12)$$

The accuracy of the final closed loop synchronization is primarily a function of the calibration errors in the satellite up-link. The SIGGEN is designed to measure the state of the broadcast IGO signal (τ_{est}) very precisely. Because of the highly predictable nature of the satellite orbit, the range and range-rate residual error can also be reduced to a minimal level. The dominant error source then becomes the residual errors in the up-link and down-link calibration parameters (ϵ_{est}^U and ϵ_{est}^D).

Preliminary Test Results

A test program is currently being performed in conjunction with NIST using the FAA GPS Integrity Broadcast Test Bed to demonstrate the timing accuracy that can be provided through the INMARSAT Geostationary Overlay. In Figure 5 through Figure 7, preliminary test results are included showing the accuracy of the range estimation and code synchronization residuals using signals broadcast through the INMARSAT Atlantic Ocean Region-West satellite.

These preliminary tests were carried out over a period of 25 hours. In Figure 5, the estimated range-rate using the SIGGEN observations is plotted against the range-rate provided by INMARSAT TT&C. The two results show excellent agreement over the period of the testing.

In Figure 6, the residual error on the timing control loop is plotted. This error was maintained within 3 meters (10 nanoseconds) 1-sigma and had a mean offset of only 6 cm (0.2 nanoseconds). Modifications made to the SIGGEN controller since these tests are anticipated to improve on these preliminary results.

In Figure 7, the residual error on the frequency control loop is plotted. The mean frequency offset during the tests was maintained within 9 mHz which is equivalent to a frequency stability at L-band of 6×10^{-12}. Further experimentation is being performed with the SIGGEN control loops to demonstrate the ultimate performance possible with the INMARSAT signal.

Conclusion

The INMARSAT timing system described in this paper has three significant advantages over existing time and frequency dissemination systems. The global coverage provided by the INMARSAT satellites will allow this service to be provided as an international civil reference time scale. The ability to monitor and steer the INMARSAT time reference remotely from an establishment such as NIST provides the capability to generate a real-time reference to UTC. Finally, the accuracy and reliability of the system will be significantly improved over existing services, including GPS.

References

[1] D.W. Allan and A. Lepek, "Trends in International Timing," Proceedings of 1993 European Frequency and Time Forum.

[2] Draft New Recommendation, "Systems, Techniques and Services for Time and Frequency Transfer," available from ITU secretariat for documents, Geneva, Switzerland.

[3] D. Bodson, et al, "Time and Frequency Information in Telecommunications Systems Standardized by Federal Standard 1002A," IEEE Proceedings, Vol 79, No 7, July 1991.

[4] Civil GPS Service Interface Committee (CGSIC) has a GPS Information Center (GPSIC) bulletin board. Call (703) 866-3826 for information on how to access.

Table 1 Time and Frequency Systems Comparison

TYPE	TYPICAL TIME ACCURACY, CAPABILITY vs UTC	TYPICAL FREQUENCY TRANSFER CAPABILITY	COVERAGE	AVAILABILITY	EASE OF USE	APPROX RELATIVE USER COST ($US 1992)	EXAMPLE SYSTEM	COMMENTS (1992)
HF Broadcast	1 ms to 10 ms	10^{-6} to 10^{-8} (over 1 day)	Global	Continuous, but operator & location dependent	Depends on accuracy requirements	50 to 5,000	Many services worldwide	Accuracy depends on path length, time of day, receiver calibration, etc
LF Broadcast	1 ms	10^{-10} to 10^{-11}	Regional	Continuous	Automatic	3,000 to 5,000	See Recommendation 768	Depends on distance from the source & diurnal propagation (ionosphere height)
LF Navigation (pulsed)	1 μs	10^{-12}	Regional	Continuous	Automatic	12,000	Loran-C	Northern hemisphere coverage. Stability & accuracy based on ground wave reception
VLF Broadcast	10 ms	10^{-11} (over 1 day)	Global	Continuous	Automatic	4,000	OMEGA	Carrier resolution can provide better time accuracy
Television broadcast terrestrial links	10 ns for common view	10^{-12} to 10^{-14} (over 1 day)	Local	Dependent on local broadcast schedule	Automatic	5,000		Calibration required for timing
Navigation satellite, broadcast	50 to 500 ns	10^{-10} to 10^{-12}	Global	Continuous	Automatic	3,000 to 15,000	GPS & GLONASS	One day averaging necessary to meet specified frequency transfer capability. Best broadcast system available today with commercial receivers.
Navigation satellite, common view	5 to 20 ns	10^{-13} to 10^{-15} one to 50 days	Intercontinental	Continuous	Automatic data acquisition. Requires post-processing	10,000 to 20,000 per site	GPS & GLONASS	Most accurate, widely used time synchronization method available today with commercial receivers for baselines less than 8,000 km.
Meteorological satellite, broadcast	100 μs	Not recommended for frequency transfer	Regional (satellite footprint)	Continuous	Automatic	4,000 to 5,000	GOES	May not be available during satellite eclipse.

Time and Frequency Systems Comparison (continued)

TYPE	TYPICAL TIME ACCURACY, CAPABILITY vs UTC	TYPICAL FREQUENCY TRANSFER CAPABILITY	COVERAGE	AVAILABILITY	EASE OF USE	APPROX RELATIVE USER COST ($US 1992)	EXAMPLE SYSTEM	COMMENTS (1992)
Other geostationary broadcast satellites	20 μs	5×10^{-10}	Regional (satellite footprint)	Continuous	Automatic	4,000	INSAT	May not be available during satellite eclipse.
Comm satellite, two-way	1 to 10 ns	10^{-14} to 10^{-15}	Regional (satellite footprint)	Continuous (as scheduled)	Data acquisition can be automatic (depending on satellite). Post-processing required	50,000 per site	North American & European networks exist	Most accurate operational method at this time.
Telephone time code, two-way	1 to 10 ms	10^{-8} (over 1 day)	Telephone calling range	Continuous	Automatic	100	Europe & North America	Phone line must have same path in both directions. Assumes computer & software availability.
Optical fibre	10 to 50 ps	10^{-16} to 10^{-17}	Local less than 50 km	Continuous	Automatic	Transmitter & receiver $US 30k per set plus cable & underground installation costs	Dedicated to frequency transfer	Cable must be temperature stabilized (e.g. 1.5 m underground).
	100 ns	10^{-13} to 10^{-14} (10 to 100 days)	Long distance 2,000 km	Continuous	Automatic	N/A. Equipment is part of a specific communication system	SDH	Part of a digital communication system.
Microwave link	1 to 10 ns	10^{-14} to 10^{-15}	Local	Continuous	Automatic	50,000 to 75,000		Sensitive to atmosph conditions & multipath effects. Must be two-way to achieve stated accuracy & stability
Coaxial cable	1 to 10 ns	10^{-14} to 10^{-15}	Local	Continuous	Automatic	5 to 30 per meter		Sensitive to temperature, VSWR, humidity, barometric pressure.

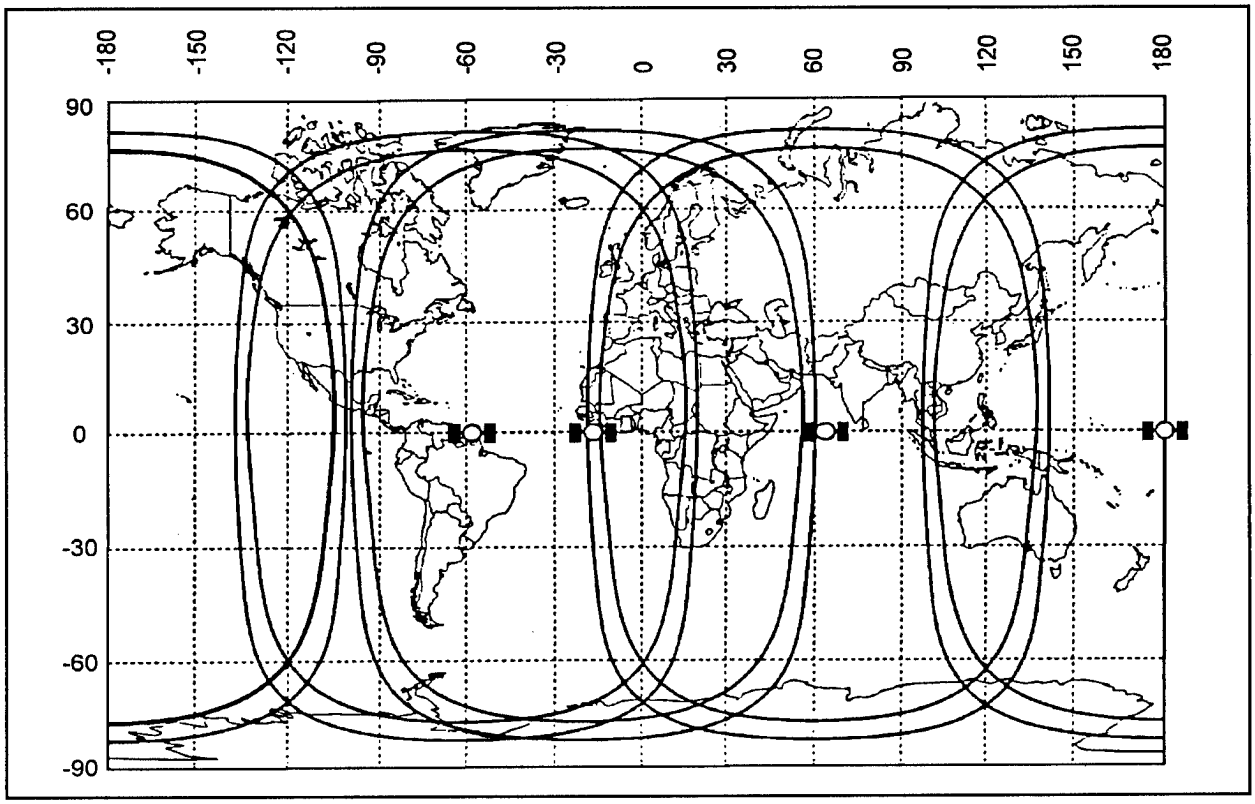


Figure 1 INMARSAT Satellite Coverage

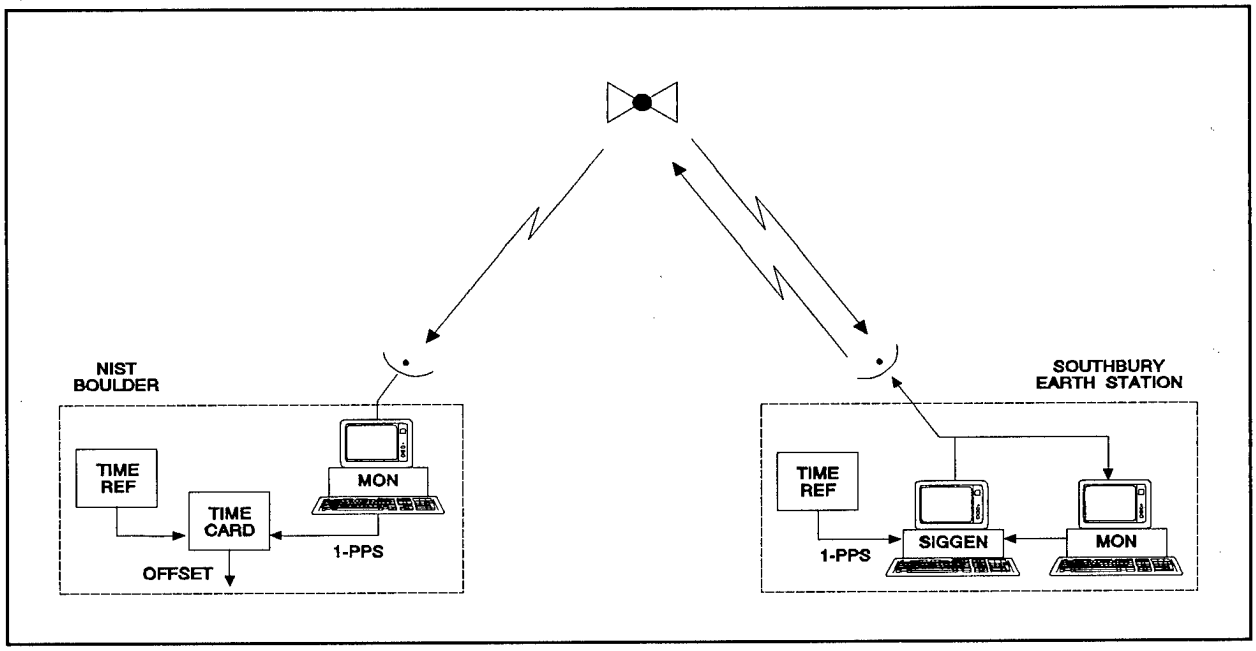


Figure 2 Experimental Test Architecture

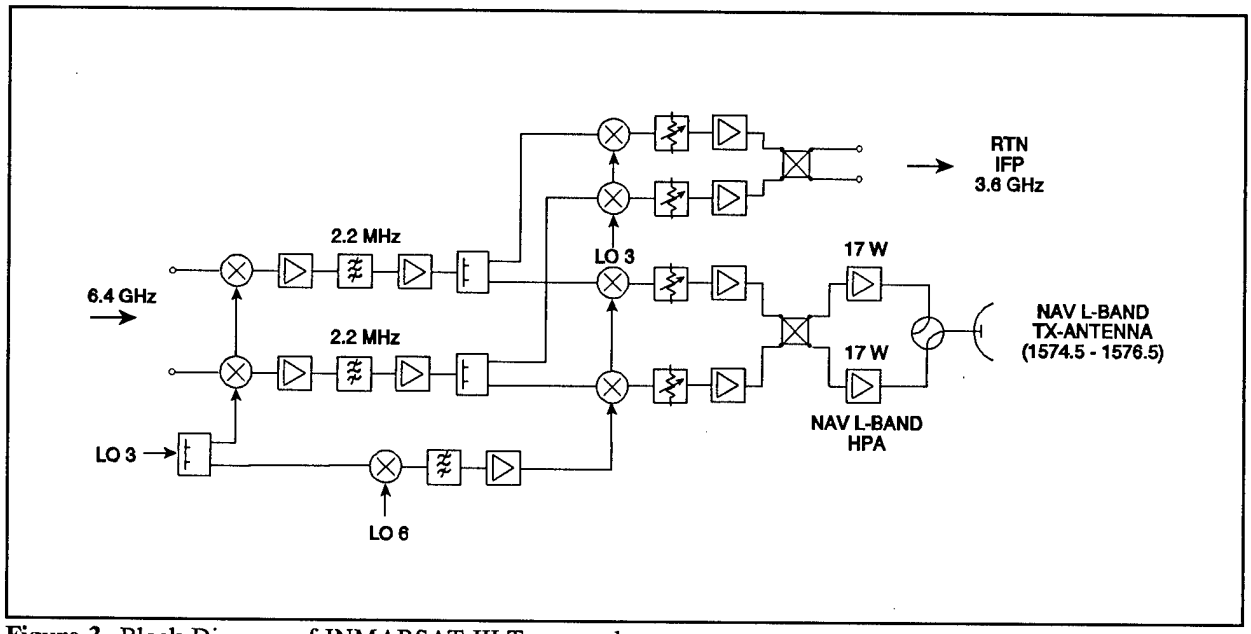


Figure 3 Block Diagram of INMARSAT-III Transponder

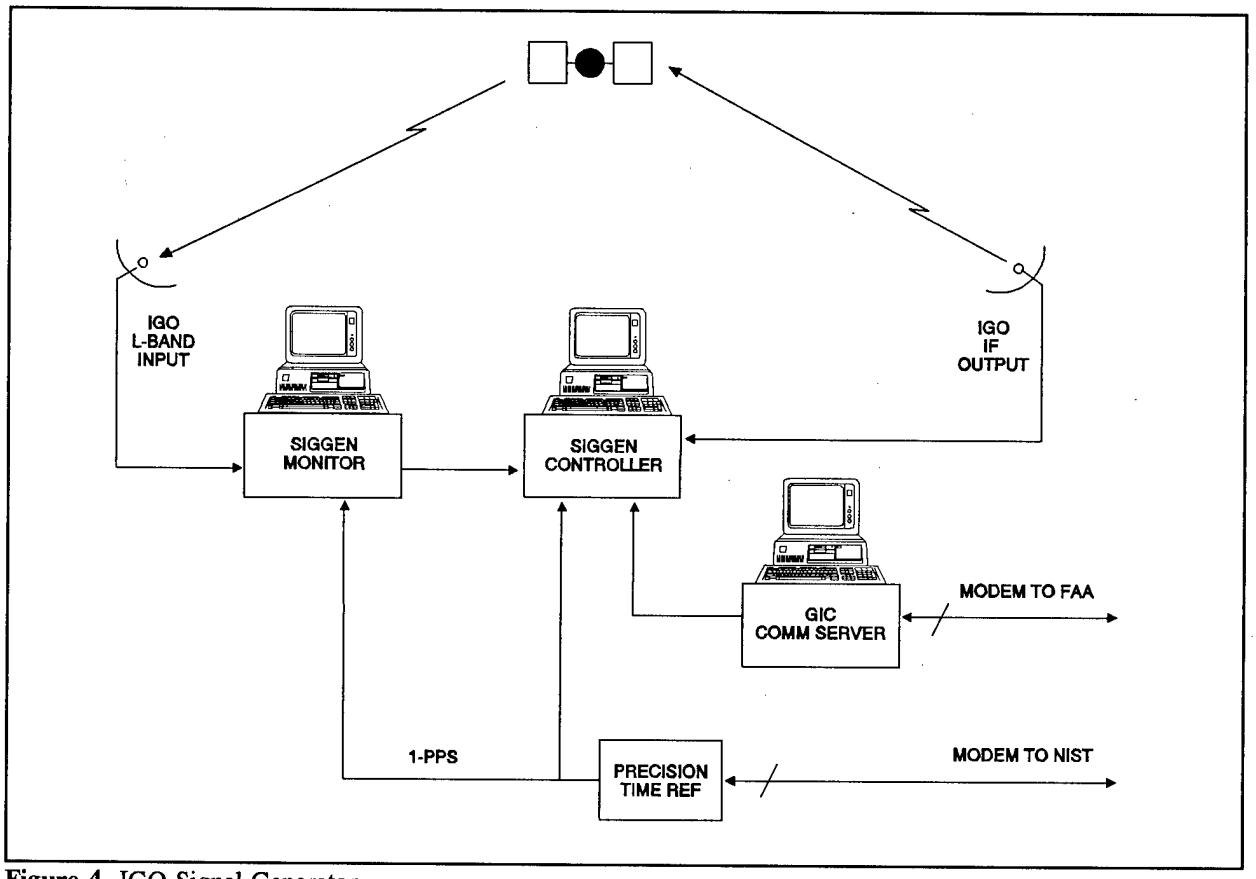


Figure 4 IGO Signal Generator

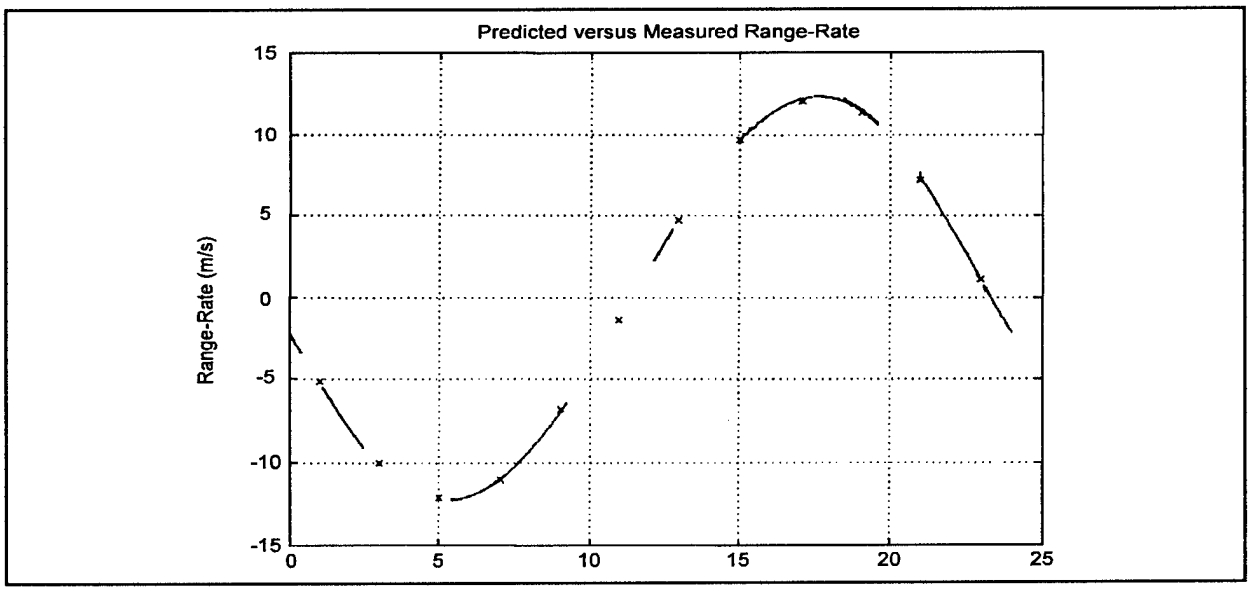


Figure 5 Range-Rate Estimation

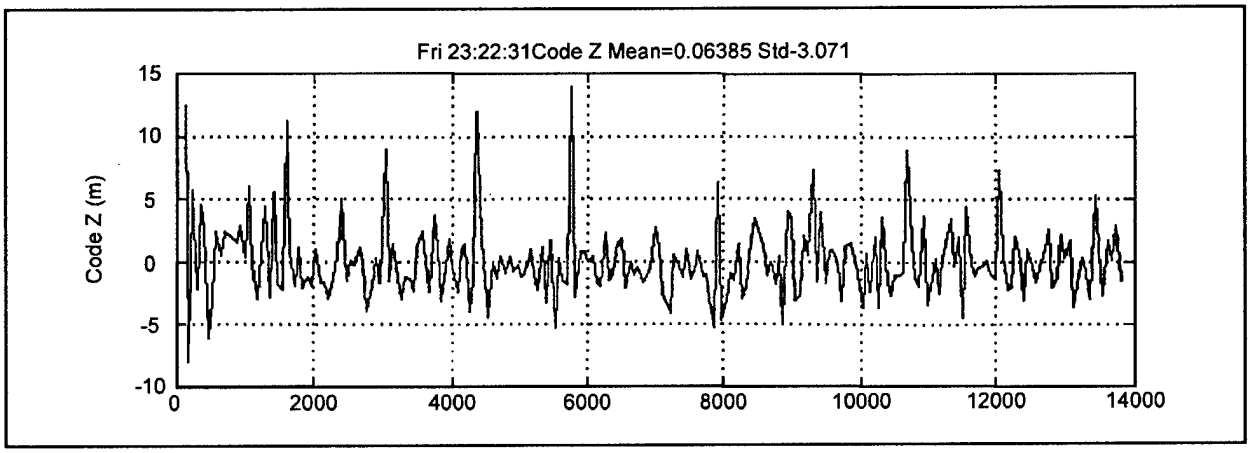


Figure 6 Code Control Loop Residual

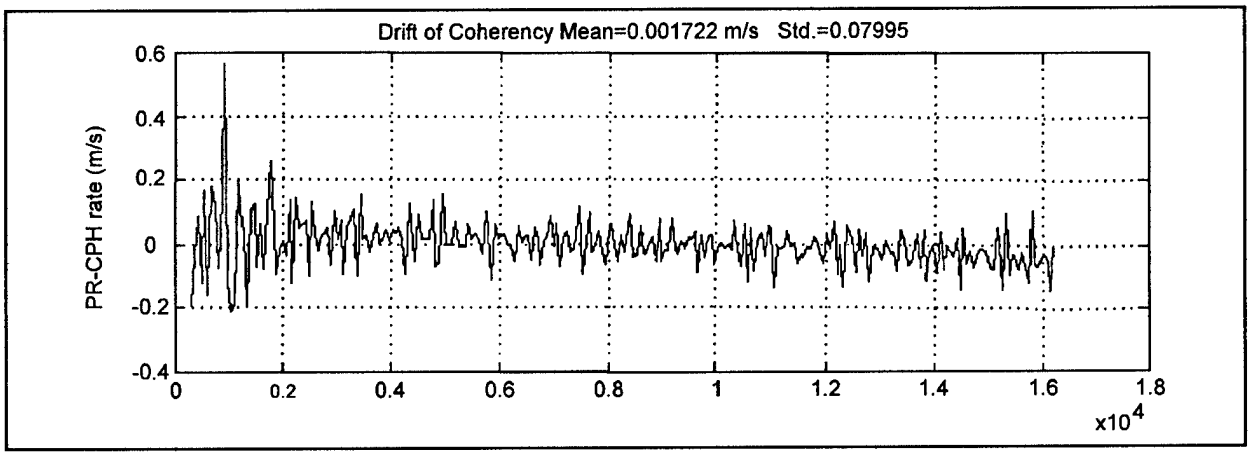


Figure 7 Control Loop Frequency Offset

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

COMPARATIVE ANALYSIS OF PARAMETERS OF GLONASS SPACEBORNE FREQUENCY STANDARDS WHEN USED ONBOARD AND ON SERVICE LIFE TESTS

Prof. Yuri G.Gouzhva, Prof. Arvid G.Gevorkyan, Dr. Arkady B.Bassevich
Dr. Pyotr P.Bogdanov, Arkady Y.Tyulyakov

Russian Institute of Radionavigation and Time (RIRT)
2, Rastrelli square, St.Petersburg, 193124, Russian Federation

Abstract

The condition for GLONASS proper functioning is the compliance of accuracy parameters of frequency standards when used onboard with prescribed specifications. It is shown that daily instability is no more than $3 \cdot 10^{-13}$ the most of oscillators. Presented are the measurement results of parameters of spaceborne oscillators in the course of its operation onboard as well as its comparison with the data of ground tests.

Introduction

While operating GLONASS, a correspondence of accuracy parameters of spaceborne references with its prescribed specifications is a condition of system proper functioning. In the course of system operation, the measurement aids of Control Centre perform a continuous monitoring of accuracy parameters of spaceborne oscillators by means of time comparisons with respect to the central synchronizer time. As was previously shown [1], according to the adopted procedure for frequency checking of spaceborne reference by means of measurement aids of synchronization system, the root-mean-square error is about $7 \cdot 10^{-14}$ for 1 day, which is quite enough to confirm that the daily frequency instability of spaceborne oscillator does not exceed $5 \cdot 10^{-13}$.

With regard to onboard measurements of accuracy parameters performed for satellites of constellation on the course of system operation over 1991, statistical data were processed and compared with certification results in ground tests.

Given are results of data processing involving oscillators' instability for measurement times of 10^5-10^6 s (Allan's variance) corresponding to the flicker floor. It is shown that this value does not exceed $3 \cdot 10^{-13}$ for the most of oscillators.

Comparative analysis of accuracy parameters of spaceborne oscillators in the course of its operation and in ground tests enables to conclude that the oscillators are resistant to destabilizing factors satellite onboard.

Estimate of oscillators' stability

Cesium beam tube frequency standards are used as spaceborne oscillators. The most important metrological parameters are oscillator frequency actual value and frequency stability that is considered as standard deviation, root-mean-square fractional frequency variation, or square root of Allan's variance. The main parameters of GEM spaceborne oscillator are given on Fig. 1.

In the course of system operation, one measures shifts of satellites' times with respect to central synchronizer time. Using results of time measurements the relative drifts of oscillator frequency actual value for 1 day are computed. The instrument error of determining these drifts for spaceborne oscillator is $7 \cdot 10^{-14}$ for 1 day [1]. As an example, Fig. 2 shows time dependences of oscillators' frequencies for a few satellites in 1991. The resultant estimates for relative drift of oscillators' frequency actual values enable also to evaluate Allan's variance for measurement times no less than 1 day. Time dependences of oscillators' frequency stability are given in Fig. 3. Stability values for various oscillators are within the area enclosed by curve lines A and B. Judging from given performances, the flicker floor for the worst oscillator is at the level about $2.5 \cdot 10^{-13}$, whereas for the best one it is below $1 \cdot 10^{-13}$. In ground processing

0-7803-0905-7/93 $3.00 © 1993 IEEE

data of frequency standards, in particular in life service tests, the similar performances were obtained, whose course and values were coincident with those in Fig. 3 (dotted area).

A spread in oscillators' frequency stabilities is due to the dispersion of figures of merit for cesium beam tubes (CBT), as to the value of the phase shift between oscillating fields, CBT frequency temperature coefficient, spread in CBT frequency shifts derived, by external magnetic fields, accuracy of temperature maintaining within the set.

An additional witness to the fact that the frequency stability of spaceborne oscillators used is good enough is the result of certification of SAPPHIRE ground CBT frequency standard constructed in the same manner with the same CBT. The certification took place in NIST in 1990, see Fig. 4.

While comparing given onboard performances of oscillators' frequency stability for 1991 with the similar data of P.Daly [2] for GLONASS spaceborne oscillators over 1986-1989, one can make a conclusion about improvement of performances expressed, first of all, in decreasing a flicker floor. The cause is a continuous parameter improvement of electronic assemblies of both frequency standard and CBT. It is seen in development results of GLONASS CBT frequency standards of the next generation. The work was carried out on further improvement of parameters of CBT as well as electronic assemblies in order to decrease an effect of destabilizing factors. As a result, the new oscillator MALAKHIT is developed which has gone through all the types of ground tests. Its parameters are given in Fig. 5. Shown in Fig. 6 is stability performance that took place on service life tests of MALAKHIT standards over 1 year at RIRT's reference base. As is shown in this figure, the flicker floor for these sets is at the level not exceeding $1 \cdot 10^{-13}$.

Conclusion

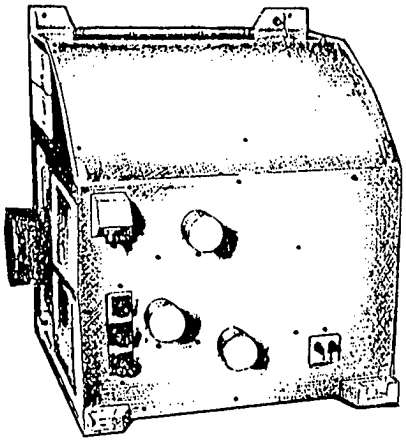
Onboard values of oscillators' frequency stability in GLONASS mid-orbital navigation system are at the prescribed level not exceeding $3 \cdot 10^{-13}$ and meet system requirements. A practical coincidence of onboard stabilities with those on ground tests allows to draw a deduction about good reproducibility of oscillators' performances as well as its fair immunity to destabilizing factors onboard.

References

1. Y.G.Gouzhva, A.G.Gevorkyan, P.P.Bogdanov. "Accuracy estimation of GLONASS satellite oscillators." - Proceeding of the 1992 IEEE Frequency Control Symposium 27-29 may 1992, pp. 306-309.
2. P.Daly, I.Kitching. "Characterization of NAVSTAR GPS and GLONASS onboard clocks." - IEEE AES Magazine July 1990.

GEM

Cesium Beam Frequency Standard for on-Board Space Equipment



Specifications

Output frequency	5 MHz
Relative frequency accuracy over life time	$\pm 1 \times 10^{-11}$
Stability, averaging time:	
1 s	5×10^{-11}
100 s	1×10^{-11}
1 hour	2.5×10^{-12}
1 day	5×10^{-13}
Temperature change of frequency, 1/° C	5×10^{-13}
Temperature range, °C	0 to + 40
Power Supply, DC	27 V
Power Consumption	80 W
Dimensions, mm	370x450x500
Weight, kg	39.6
Operational life, hours	17,500

Fig. 1

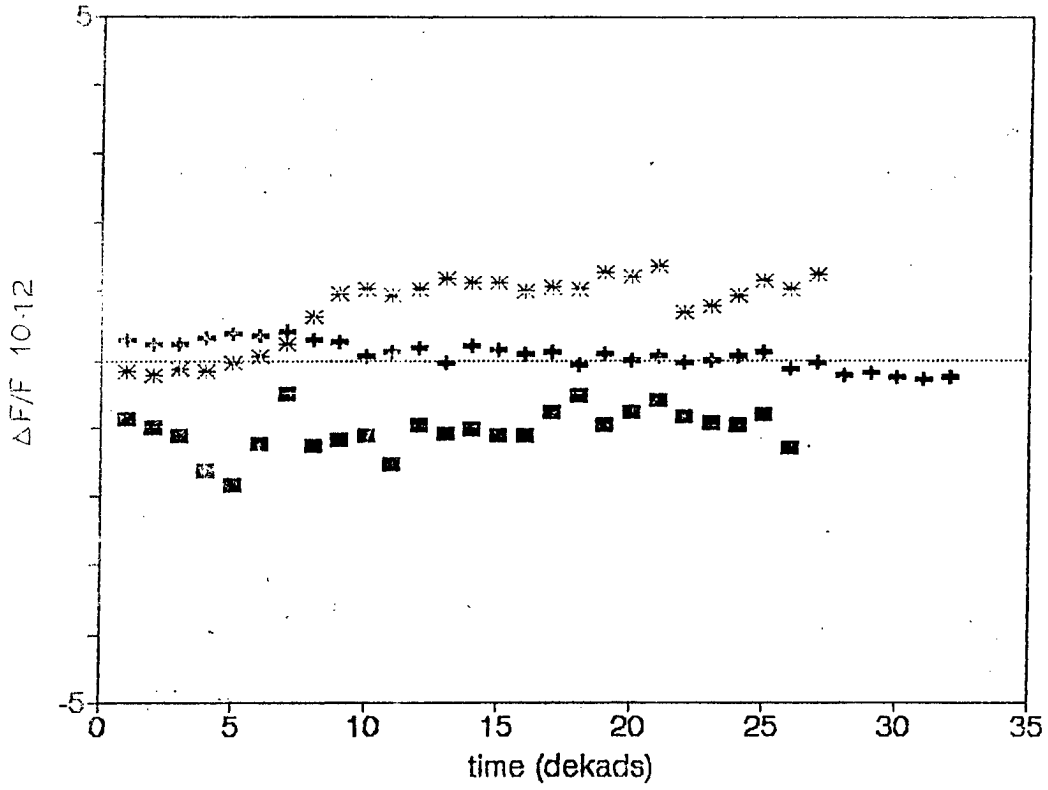


Fig.2 Time dependences of Oscillators frequencies

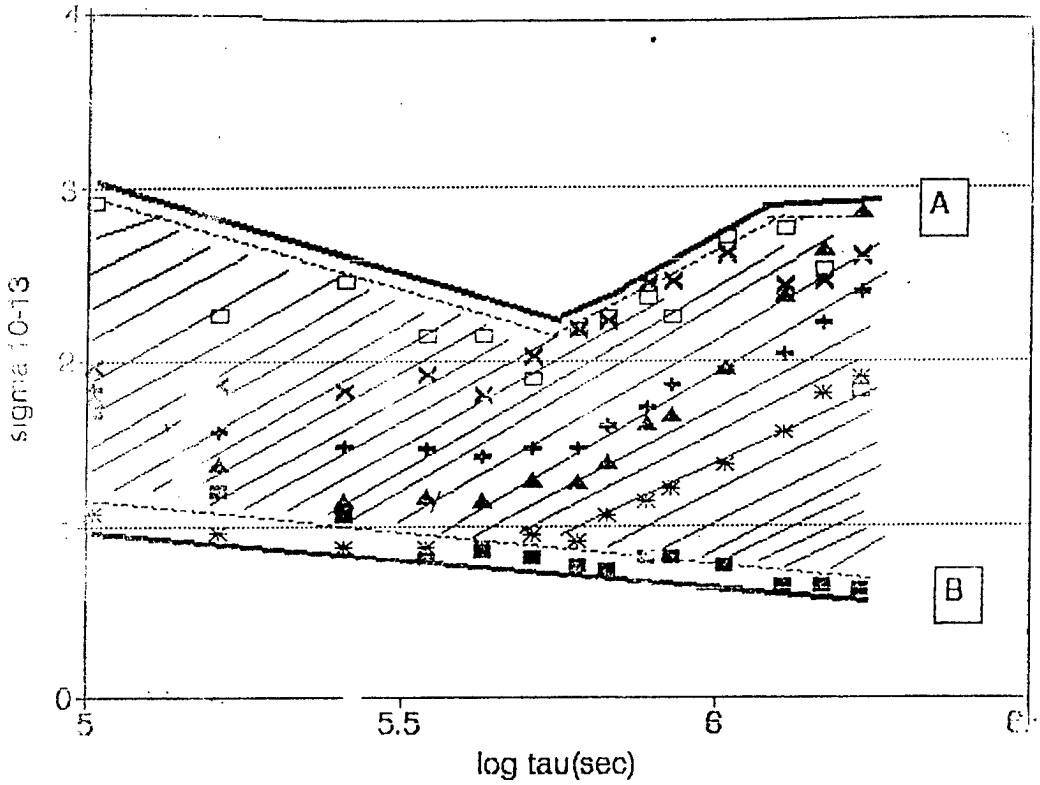


Fig.3 Allan variance for GEM standards.

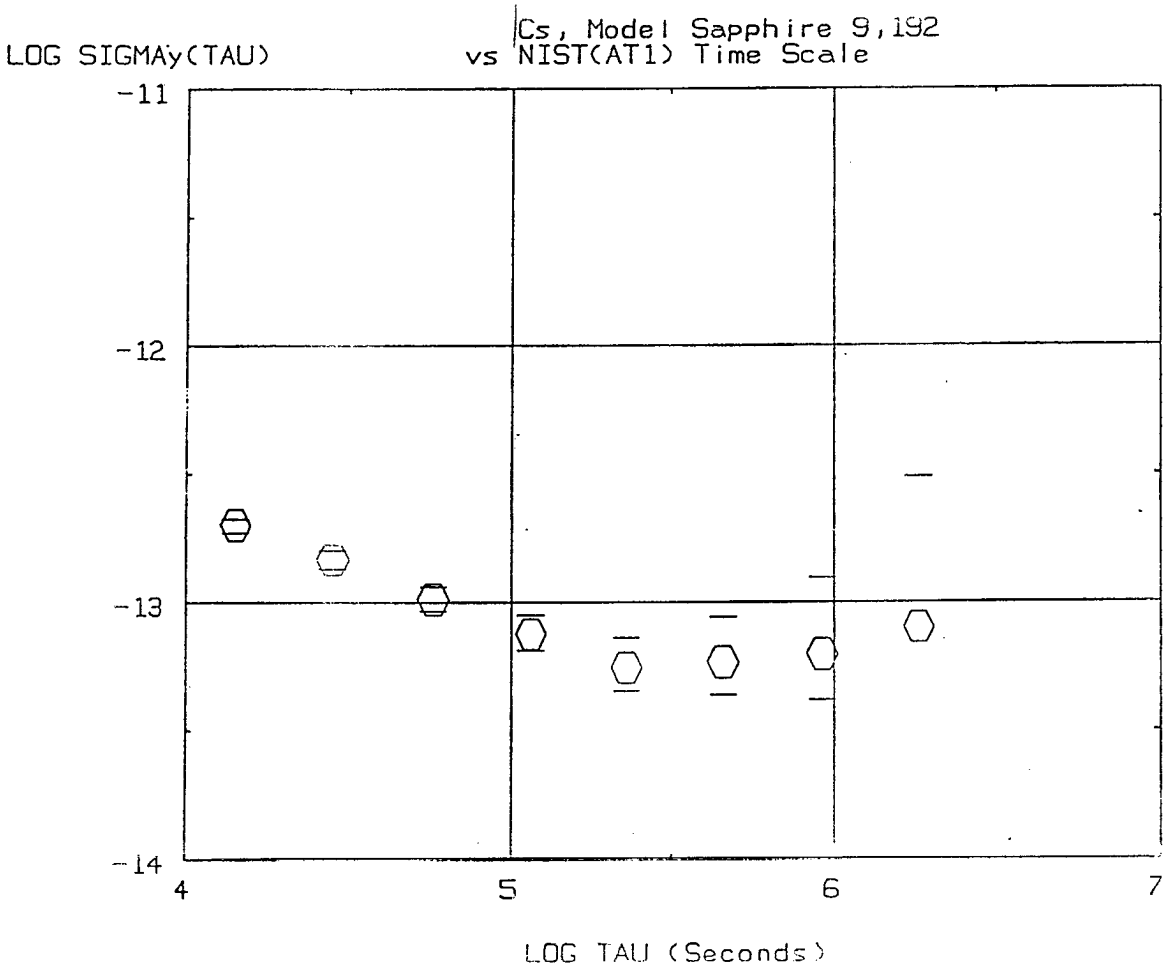
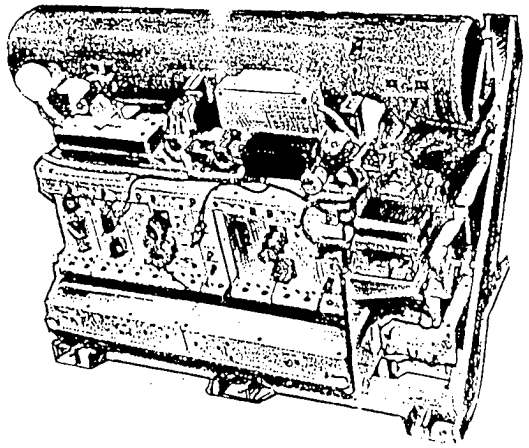


Fig. 4 Allan variance stability for SAPPHIRE standard

MALAKHIT

Cesium Beam Frequency Standard for On-board Space Equipment



Specifications

Output frequency	5 MHz
Relative frequency accuracy over life time	$\pm 1 \times 10^{-11}$
Stability, averaging time:	
1 s	2×10^{-10}
100 s	5×10^{-12}
1 hour	5×10^{-13}
1 day	1×10^{-13}
Temperature change of frequency, 1/°C	2×10^{-13}
Temperature range, °C	0 to + 40
Power Supply, DC	27 V
Power consumption	90 W
Dimensions	421x414x655
Weight, kg	52
Operational life, hours	28,500

Fig. 5

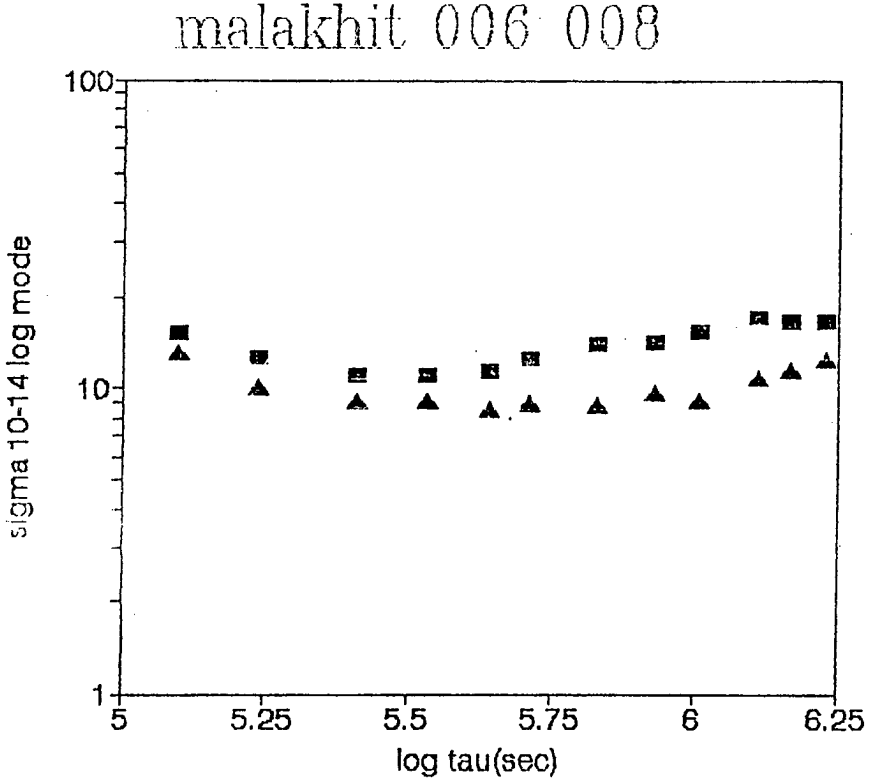


Fig. 6 Allan variance for malakhit standards.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

NIST-7, THE NEW US PRIMARY FREQUENCY STANDARD*

R. E. Drullinger, J. P. Lowe, D. J. Glaze and Jon Shirley
National Institute of Standards and Technology
Time and Frequency Division
Boulder, Colorado
USA

Abstract

NIST-7, an optically pumped, cesium-beam frequency standard has replaced NBS-6 as the official US primary frequency standard. The present short-term stability of the standard, measured with respect to an active hydrogen maser, is characterized by $\sigma_y(\tau) \approx 8 \times 10^{-13}\ \tau^{-1/2}$. Our first evaluation has resulted in an uncertainty of 4×10^{-14}. An improved servo-electronic system is being developed and this should improve stability and allow for more precise evaluation of the various systematic errors.

INTRODUCTION

An optically pumped, thermal atomic-beam frequency standard developed at NIST and known as NIST-7 has officially replaced NBS-6 as the US primary frequency standard. In the preliminary operation reported here, the standard has been shown to have both accuracy and stability that are several times better than NBS-6. Furthermore, these parameters are expected to improve in the near future with improved servo electronics.

The standard has been described elsewhere [1]. Briefly, the Ramsey cavity is 1.55 m long, and the atomic beam is 3 mm in diameter. An axial C-field geometry is employed and the atomic beam goes through the X-band waveguide cavity parallel to the long dimension of the waveguide. The microwave field, therefore, varies as a half-sine wave in the direction of the atomic beam; this results in Rabi lineshapes with very smooth and rapidly damped tails (see Fig. 1). The cavity ends are designed so the Poynting vector vanishes in the center of the atomic beam window [2], thus minimizing distributed-cavity phase shift.

Contribution of the U. S. Government, not subject to copyright.

For the evaluation reported here, a single, grating-feedback diode laser (linewidth $\leq$ 100 kHz) was frequency-locked to the $F = 4 \rightarrow F' = 3$ saturated-absorption feature in an external cesium cell. Approximately 0.1 mW in a 3 mm diameter ($2w_0$) light beam with linear polarization was used for state preparation. The optical pumping beam was orthogonal to the atomic beam and was retroreflected with its polarization rotated by 90°. This creates a zone of "randomized polarization" which leads to complete optical pumping ($\geq$ 99.9%) and avoids the problem of coherence trapping [3]. An acousto-optic device was used to synthesize a second light beam with the $\approx$ 450 MHz offset necessary to drive the $F = 4 \rightarrow F' = 5$ cycling transition in the detection region. The power density in this beam was adjusted to scatter approximately 10 photons per atom. The overall fluorescence collection and detection efficiency is $\geq$ 40%.

Measurements of frequency shifts reported here were made relative to an active hydrogen maser. At present, the stability of the standard is characterized by $\sigma_y(\tau) \approx 8 \times 10^{-13}\ \tau^{-1/2}$ (Fig. 2) and is limited by phase noise in the microwave radiation. The standard is designed to operate with an oven temperature of 110°C at which point the atomic shot-noise limited stability should be $\approx 3 \times 10^{-13}\ \tau^{-1/2}$. The frequency biases and their associated uncertainties are quoted throughout this paper in terms of fractional frequency change in the standard.

EVALUATION

The first evaluation of the standard has been reported in detail elsewhere [4] and will only briefly be presented here with the results summarized in Table 1. The correction for second-order Zeeman effect is made by measuring the first-order Zeeman splitting. Errors caused by field inhomogeneity are less than 10^{-15}. At present, the correction for the second-order Doppler shift has been made by a theoretical calculation which takes into account the

U.S. Government work not protected by U.S. copyright

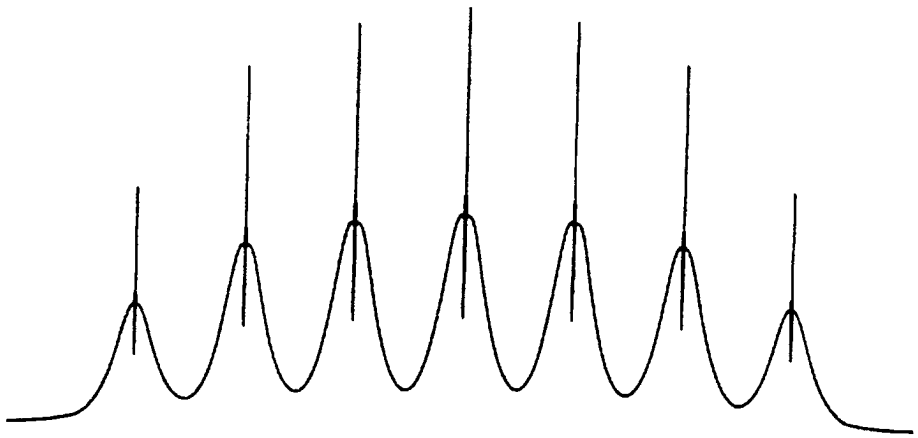


Fig. 1. Zeeman spectrum of the F = 3 to F = 4 transition. The microwave power level is about 7 dB below optimum and the C-field corresponds to a first order Zeeman splitting of 24.2 KHz ($\nu_{0,0}$ - $\nu_{1,1}$).

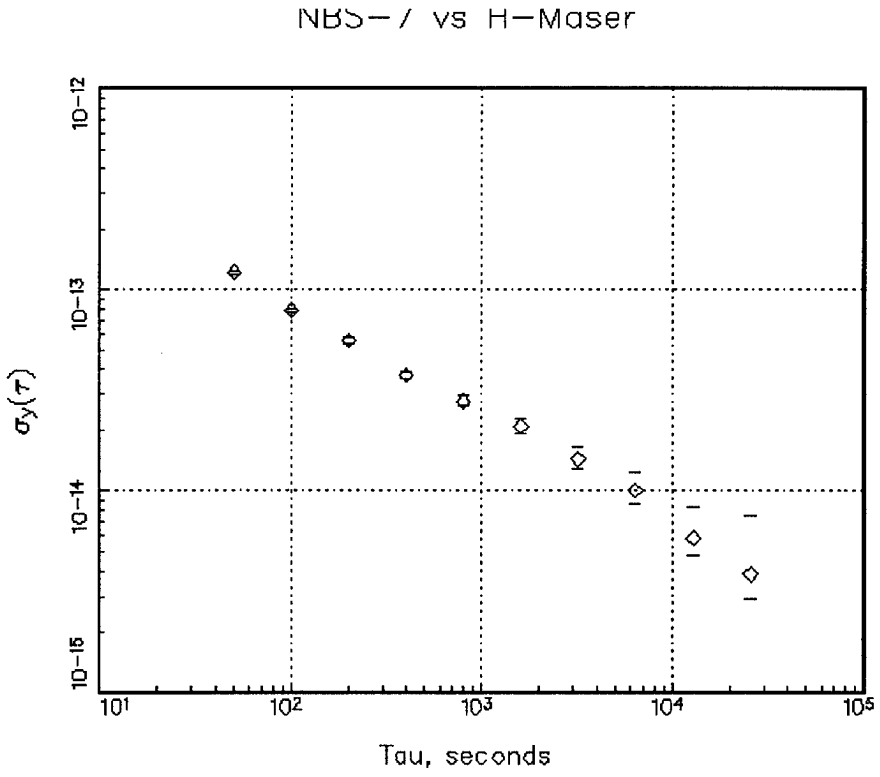


Fig. 2. Frequency stability of NIST-7 measured against an active hydrogen maser. The error bars represent 1σ on the statistics of the measurement.

velocity distribution, the microwave power, and the modulation parameters. A thermal velocity distribution, weighted by 1/v for detection by a cycling transition, was assumed. It predicts a Ramsey lineshape differing by only 1 or 2 % from the measured lineshape. The quality of this fit together with uncertainties in microwave power, and modulation parameters, leads to an uncertainty in the second-order Doppler correction of $\sim 2 \times 10^{-14}$.

The end-to-end cavity phase shift has been measured by beam reversal. The frequency shift on beam reversal at optimum power was 1.52×10^{-12} with an uncertainty of 3×10^{-14}. A search for distributed-cavity phase shift by alternately blocking one-half of the atomic beam near one end of the Ramsey cavity or the other showed no effect at the 1×10^{-14} level. The cavity Q of 600 and measured mistuning lead to cavity-pulling errors of less than 10^{-14}.

Line overlap shifts have not been evaluated in depth. But in an optically pumped standard with the high spectral symmetry demonstrated in Fig. 1, these effects are expected to lead to errors of less than 10^{-15}. The shift due to blackbody radiation [5] can be calculated to very high accuracy since the temperature of the atomic beam tube of NIST-7 is regulated. The fluorescence light shift in NIST-7 is expected to be $\leq 10^{-16}$ [6]. Using different optical powers and atomic beam fluxes we have observed no frequency shifts at the 10^{-14} level.

Frequency errors coming in through the electronics and arising from RF spectral purity, modulation distortion or offsets in the integrators or DC gain stages have been shown to be less than 1×10^{-14}.

SUMMARY

NIST-7 has undergone a preliminary evaluation and been shown to be several times more accurate than NBS-6. As of January 1, 1993, it has become the official US primary standard for frequency. It will be used to steer the long-term behavior of the NIST time scale while it continues to be developed and improved. As the error budget is reduced, international comparisons will be performed with other high-accuracy standards.

ACKNOWLEDGEMENTS

The authors thank Fred Walls and Andrea DeMarchi for their numerous and invaluable contributions throughout this work.

REFERENCES

[1] Robert E. Drullinger, David J. Glaze, J. P. Lowe and Jon H. Shirley, "The NIST optically pumped cesium frequency standard," IEEE Trans. Instrum. Meas., vol. 40, pp. 162-164, April 1991.

[2] A. DeMarchi, J. Shirley, D. J. Glaze, and R. Drullinger, "A new cavity configuration for cesium beam primary frequency standards," IEEE Trans. Instrum. Meas., vol. 37, pp. 185-190, June 1988.

[3] G. Théobald, N. Dimarcq, V. Giordano, and P. Cérez, "Ground state Zeeman coherence effects in an optically pumped cesium beam," Optics Comm., vol. 71, no. 5, pp. 256-262, June 1989.

[4] R. E. Drullinger, Jon H. Shirley, J. P. Lowe and D. J. Glaze, "Error Analysis of the NIST Optically Pumped Primary Frequency Standard," to appear in IEEE Trans. Instrum. Meas., April 1993.

[5] Wayne M. Itano, L. L. Lewis, and D. J. Wineland, "Shift of $^2S_{1/2}$ hyperfine splittings due to blackbody radiation," Phys. Rev. A, vol. 25, pp. 1233-1235, February 1982.

[6] Jon Shirley, "Fluorescent light shift in optically pumped cesium standards," in Proc. 39th Freq. Cont. Symp., 1985, pp. 22-23.

*This paper will also be published in the Proc. of 7th European Frequency and Time Forum, Neuchâtel, Switzerland, March 16-18, 1993.

Effect	Bias	Measurement Uncertainty (1σ)
2nd-order Zeeman	$\approx 10\ 000$	1
2nd-order Doppler	≈ 30	2
Line overlap		---
AC Stark	1.9	1
Cavity phase	76	3
Electronics		1

Table 1. Summary of systematic errors in NIST-7. The size of the resulting bias is given together with the associated uncertainty (all in units of 10^{-14}).

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN ANALYSIS OF MAJOR FREQUENCY SHIFTS IN THE LPTF OPTICALLY PUMPED PRIMARY FREQUENCY STANDARD

G.D. Rovera, E. de Clercq, A. Clairon
Laboratoire Primaire du Temps et des Fréquences
BNM-Observatoire de Paris
61, avenue de l'Observatoire, 75014 PARIS, France

Abstract

An analysis of the major frequency shifts of our optically pumped Cesium beam primary frequency standard is presented, toghether with a description of the most relevant characteristics of the device. The short term frequency stability was measured against an Hydrogen Maser. The square root of the Allan variance $\sigma_y(\tau)$ is at present $5.8{\cdot}10^{-13}\tau^{-1/2}$, with a flicker level at $3.3{\cdot}10^{-15}$. This extremely good frequency stability enables a highly precise analysis of the most important systematic frequency shifts to be performed in a short time. Our first accuracy evaluation results in an uncertainty (1σ) of $1.1{\cdot}10^{-13}$.

1 Introduction

Optically pumped Cs beam frequency standards [1] are expected to have interesting features compared to the classical Rabi type magnetic selection standards. The most relevant advantages are a high signal-to-noise ratio, which leads to a good short term stability, a symmetric microwave spectrum, a more confident retracing of atomic trajectories when the beam is reversed, and some additional degree of freedom given by the choice of the light beam parameters (frequency, polarization,...). Another interest is the realization of the second by a way different from the classical one. This could allow a check of possible systematic effects in its realization. On the other hand optically pumped standards have to take into account a broader velocity distribution, and therefore frequency shifts more dependent on microwave power and light induced shifts.

At present time the first optically pumped Cs primary standards are being evaluated [2, 3]. In this paper we present the main features of our standard. The frequency stability measured against an Hydrogen maser is reported. We study then the main frequency biases and report our first evaluation of accuracy.

2 Description of the device

2.1 Apparatus

A more detailed description of the Cs tube is given in [4, 5] and only a summary is presented here. The atomic beam is ribbon shaped ($3 \times 8\,\mathrm{mm}^2$), collimated by a multichannel array made of a stack of crinkled foils and the overall oven to oven length is 2,3 m. The static magnetic field, traditionally called the C field, is directed vertically and is produced by four current carrying rods surrounded by three magnetic shields, covering the cavity length and the optical interaction areas. Between each oven and the microwave cavity there are two optical interaction areas where laser beams can cross the atomic beam. They are 340 mm and 240 mm away from each extremity of the Ramsey cavity. The fluorescence light is collected and focused on a large area ($100\,\mathrm{mm}^2$) Si photodiode by means of a spherical mirror in front of two fast aspheric lenses followed by a light pipe.

The Ramsey cavity is of the conventional E-bend type with a drift length between the two arms of 1.01 m. The atomic beam passes through $3 \times 8\mathrm{mm}^2$ slots in the cavity arms at a distance $\lambda g/2$ from the terminal short circuits. The balance of the electrical length of the two cavity arms has been adjusted with an uncertainty of a few tens of microns using the detection of modes with an odd number of half wavelengths [6]. The RF power is fed in the third symmetric arm of the cavity by means of a coaxial cable.

2.2 Optical setup

The $4-4\,\sigma$ polarized transition of the Cs D_2 line, which gives the best population inversion [7, 8], is used for optical pumping. Detection is performed with the circularly polarized 4-5 cycling transition. The use of this transition has two advantages: a velocity distribution weighted by $1/v$ [9],and a better signal-to-noise ratio if

0-7803-0905-7/93 $3.00 © 1993 IEEE

the laser linewidth is narrower than the natural width of the Cs D2 line [10, 11].

We use a single extended-cavity semiconductor laser locked to the 4-5 transition using the saturated absorption spectrum of a Cs cell. A part of the laser beam is frequency shifted on the 4-4 transition by means of an acousto-optic modulator. The laser diode is a 850 nm, 150 mW single mode GaAlAs diode from the SDL company and has no additional anti-reflection coating. The laser temperature is regulated to 1 mK [12] and the whole mounting box to 100 mK. The output light is collimated and directed to a Littrow-mounted diffraction grating which terminates the extended cavity. The cavity length is about 10 cm. The wavelength is first selected by rotating the grating and tuning the temperature and injection current. For fine frequency tuning and locking, purpose the cavity length is then adjusted moving the grating by means of a Piezoelectric Ceramic. The measured FWHM linewidth of a beat-note between two identical devices is 70 KHz. The device reliability is good, the laser frequency can be kept locked for a month without trouble.

The perpendicularity of the laser beams and the atomic beam is adjusted to better than 1 mrad and we use standing waves in order to avoid deflection of the atomic beam and possible velocity selection. In these experiments the laser beam intensities are near 1.5 mW/cm^2 for the pumping beam and 0.1 mW/cm^2 for the detection beam.

The pumping efficiency is very good, the rate of unpumped atoms is only $2{\cdot}10^{-4}$ of the initial F = 4 level population. At detection, the stray background light is only 2 % of the clock transition amplitude.

Both east and west ovens are continuously heated (85 °C). Each atomic beam can be cut by a shutter. The beam reversal is simply performed by rotating the shutters and acting on the variable laser beam splitters. In this way, the atomic beam can be reversed in only a few minutes. The two atomic beams can also be used simultaneously.

2.3 Frequency control system

We use two different frequency control units in order to check for errors introduced by the electronic servo system. The first one has been borrowed from PTB and is similar to the control unit of PTB's Primary Frequency Standard [13]. The second one is a digital control loop that directly compares the frequency of the standard against a commercial Hydrogen Maser (Sigma Tau Company).

A block diagram of the servo system is shown in

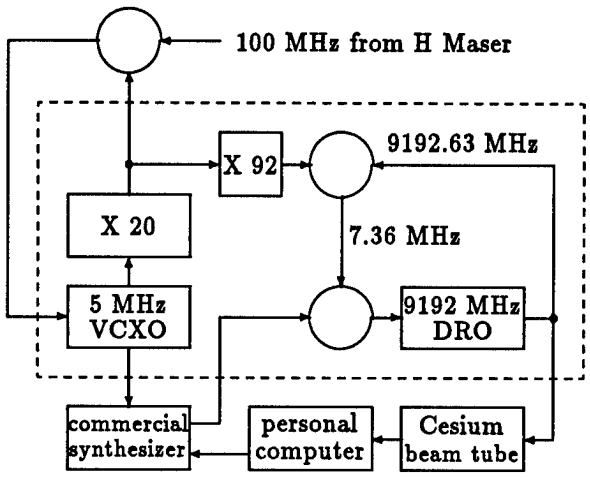


Figure 1: Block diagram of servo loop system. The dashed line encloses the PTB frequency multiplication chain.

Fig. 1. The 5 MHz Quartz oscillator of the PTB multiplier chain is locked to the 100 MHz coming from the maser with a second order loose loop (1 Hz) in order to reject the 29 Hz sidebands that are present in the maser signal. The 7.36 $\cdots$ Mhz is generated by a commercial HP synthesizer with 1 mHz resolution connected to a personal computer via a IEEE bus. The computer also drives a microwave attenuator in order to set the microwave power at the desired level or to servo the power at the optimum level. The fluorescence signal is then measured via an analog I/O board.

The modulation used is a slow square wave (1-2 Hz) with blanking time and it is performed in a totally asynchronous way by software. The error signal, given by the difference of the fluorescence signal in the two sides of the Ramsey central fringe, is digitally integrated by simple addition. In this way it is possible to realize a true integration that avoids all the problems of an analog integrator with finite gain. The value of the integrator register is then used to steer the central value of the synthesizer which is stored on the hard disk of the PC. Analyzing this data a posteriori, we can calculate the average frequency and variance of the measuring session. With this system we have to take care of the quantization noise of the I/O board and of the resolution of the synthesizer, but in both cases quantization error is lower than other noise.

3 Results

3.1 Stability

The square root of the overlapping Allan variance

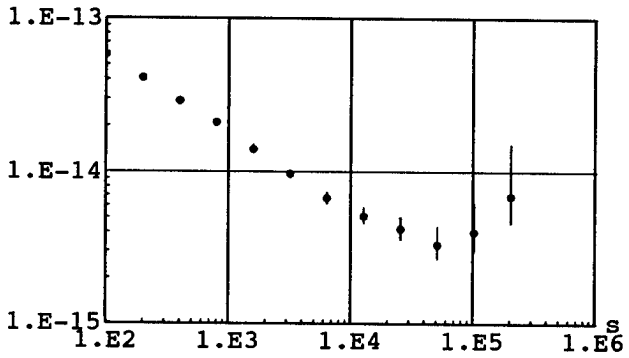


Figure 2: Square root of overlapping Allan variance.

of the relative frequency difference between our standard and the Hydrogen maser is reported in Fig. 2. For sampling times of up to 10^4 s the frequency instability shows the expected behaviour of a white frequency noise process. The measured data fit very well the line $\sigma_y(\tau) = 5.8{\cdot}10^{-13}\tau^{-1/2}$. The measured Signal-to-Noise ratio in a 1 Hz bandwidth at the modulation frequency is 20000. The shot noise limited S/N of 26000 is not reached owing to the AM noise of the laser intensity. The experimental S/N ratio and the transition linewidth of 100 Hz lead to a theoretical short-term stability $\sigma_y(\tau) = 2{\cdot}10^{-13}\tau^{-1/2}$. The discrepancy with the experimental value might possibly be explained by Audoin's effect of intermodulation [14],but needs further investigations. Anyway, to our knowledge, this short term stability is the best reported for a Cs standard and confirms the possibilities of optical pumping. The stability reaches a floor of $3{\cdot}10^{-15}$ for an averaging time of $5{\cdot}10^4$ s. The degradation for long times has not yet been investigated. It may come either from maser frequency drift or from slow fluctuation of the C field current. Indeed this outstanding stability allows us to perform accurate frequency measurements in a very short time.

3.2 Static magnetic field shift

The C field homogeneity in the transverse or longitudinal direction measured in previous experiments is better than $5{\cdot}10^{-4}$. In our case the evaluation of the quadratic Zeeman shift is limited at the 10^{-15} level by the difference between the mean square value of the magnetic field averaged over each arm of the cavity and averaged over the drift space between them [15]. Then the quadratic Zeeman shift is calculated from the linear Zeeman splitting between microwave lines (50.7 KHz) with an uncertainty of $3{\cdot}10^{-15}$.

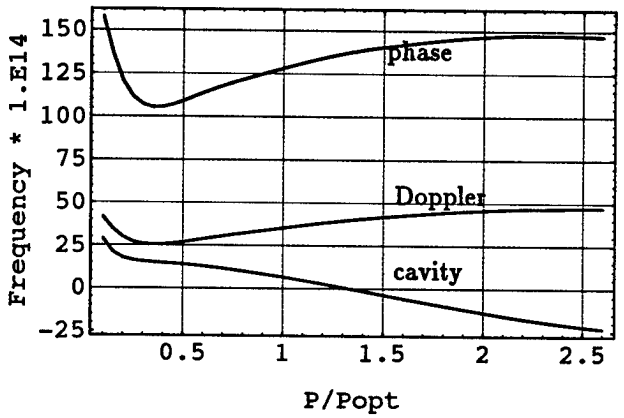


Figure 3: Microwave Power dependent shifts. Second order Doppler and cavity shift are represented with inverted sign.

3.3 Second order Doppler shift

In an optimized optically pumped thermal Cs beam the second order Doppler shift will probably be the limiting effect. In order to estimate the error in this case it is necessary to know the absolute field experienced by atoms inside the cavity, or at least relative to some definition of optimum power which can be experimentally identified. At present we use a method that enables us to search for the optimum power, defined as the power which maximizes the CW detected signal at resonance, and we can give better than 1 % accuracy [16]. However our present implementation is limited to 5 % accuracy. Taking into account a Maxwellian weighted by $1/v$ velocity distribution, confirmed by fitting of experimental and calculated Ramsey patterns, the second order Doppler variation resulting from a 5 % variation of microwave power is $0.8{\cdot}10^{-14}$. As the real velocity distribution has not been measured, the 2^{nd} order Doppler uncertainty is conservatively estimated to be $3{\cdot}10^{-14}$, see Fig. 3. The second order Doppler shift is $-35{\cdot}10^{-14}$ with $3{\cdot}10^{-14}$ uncertainty.

3.4 Cavity-related errors

The cavity loaded quality factor Qc is about 4500. The cavity detuning is -1.2 MHz.

Considering the high Q and the large detuning of our cavity the calculation of the cavity pulling effect using the formula given in [17], making the approximation of small detuning, is no longer valid. The formula we actually use is:

$$\left(\frac{\delta_\nu}{\nu}\right)_c = \Lambda_c \frac{\nu}{2Q^2} \frac{1}{p}\frac{\mathrm{d}p}{\mathrm{d}f} \qquad (1)$$

which is one of the intermediate stages leading at

the formula (9) given in [17]. In (1) p is the microwave power, Q the quality factor of the line and Λ_c a correction coefficient that takes into account modulation parameters. The relative variation of the microwave power versus frequency, around the operating point, is evaluated to be $-1{\cdot}10^{-6}$ per Hz, with an uncertainty of 5 %. Cavity pulling, as function of power is given in Fig. 3. Taking into account the present uncertainty in power measurement the cavity pulling effect is $-6.6{\cdot}10^{-14} \pm 1.5{\cdot}10^{-14}$.

The end-to-end phase shift is at present the biggest cause of uncertainty in our standard for the following reasons: firstly from our preliminary experiments on distributed phase shift we deduce a frequency shift equal to $71{\cdot}10^{-14}$ per mm of displacement of the center of gravity of the atomic beam. As the ovens are aligned with an accuracy better than 0.1 mm the uncertainty we have to account for this effect is $7{\cdot}10^{-14}$. The experiment is performed blocking a part of the atomic beam by a movable shutter in the vicinity of the oven exit, and then measuring frequency versus the shutter position. Secondly, the shift itself is as big as $128{\cdot}10^{-14}$ and therefore a 5 % uncertainty in determining microwave power can give a $2{\cdot}10^{-14}$ error. Finally, as the frequency measurements where performed with a $3{\cdot}10^{-14}$ uncertainty, the claimed total uncertainty for the phase shift is $8{\cdot}10^{-14}$. The end to end phase shift measured by beam reversal is then $128{\cdot}10^{-14} \pm 8{\cdot}10^{-14}$ at optimum microwave power.

3.5 Other shifts

Use of laser beams in a Cs clock can be the origin of new frequency biases via the light shift, the modification of the atomic trajectories, or via unexpected effects. The light shift due to the fluorescence light of the beam is calculated in our case to be less than $1{\cdot}10^{-15}$ using the results of Oshima et al [18]. However the light shift due to stray light is not a priori known, because the amount and direction of stray light propagating in the Ramsey cavity region is not known. In order to check for a possible stray light effect and other light related effects we have undertaken a set of frequency measurements varying the power of the pumping or detection beam, laser frequency, polarization of the detection beam, angle between laser beams and atomic beam (of a few milliradians), and using travelling waves instead of standing for the pumping or the detection beams. At present we have not found a simple correlation between optical parameters and the measured frequency, so we take as the uncertainty the standard deviation of the ensemble of all measurements in both beam directions, that is: $6{\cdot}10^{-14}$.

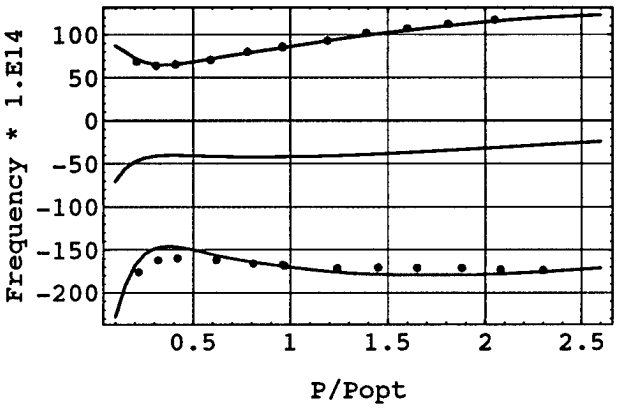


Figure 4: Microwave Power shift

The relative asymmetry of the neighbouring lines is always smaller than 1 % in our case. The Rabi pulling calculated [19] with the operating conditions for a such asymmetry, with the realistic assumption that the velocity distribution of the atoms in the m ± 1 sublevels is also Maxwellian multiplied by $1/v$, is $0.3{\cdot}10^{-14}$. We take this value as the uncertainty in order to be cautious.

Shifts due to the electronic system have been tested using two different servo systems, one totally analogic loaned from PTB, and the second one is the digital presented in this paper. The difference between the two systems is found to be $3{\cdot}10^{-14}$ and this value is taken as the uncertainty.

The shift due to the black-body radiation is not considered here. All other shifts are expected to be smaller than 10^{-14} and are not taken into account for this evaluation.

4 Total accuracy evaluation

Nearly all the frequency shifts depends on the microwave power actually experienced by the atoms. Studying the frequency shift versus the microwave power is a way of checking if the different frequency biases are correctly evaluated. Fig. 4 shows the measured relative frequency versus the normalized microwave power. The solid lines are calculated taking into account the second order Doppler effect, cavity pulling, and a 0.30 mrad end to end phase difference. The upper (lower) curve is for W-E (E-W) beam direction, middle curve is obtained excluding the phase shift. Fig. 5 and Fig. 6 are enlarged views of Fig. 4. The theoretical curve fits very well the experimental data for the W-E direction. the agreement in not so good for the E-W direction, but the maximum discrepancy is only

Type of Frequency Bias	Fractional Frequency Bias $\times 10^{-14}$	Fractional Uncertainty $\times 10^{-14}$
C Field	24020	0.3
Second order Doppler	−35	3
Cavity phase shift	128	8
Cavity pulling	−6.6	1.5
Rabi pulling	±0.3	.3
Unexplained Shifts	–	6
Electronics	–	3
$\sqrt{\Sigma x_i^2}$	–	11

Table 1: Systematic major frequency shifts in the LPTF OP primary frequency standard and uncertainties in parts in 10^{-14} at optimum power.

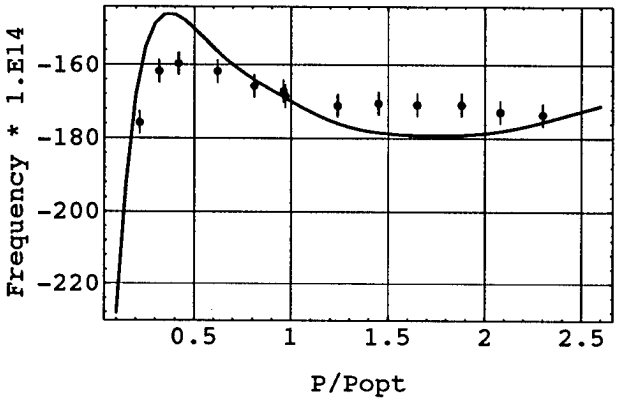


Figure 5: Microwave Power shift for E-W beam

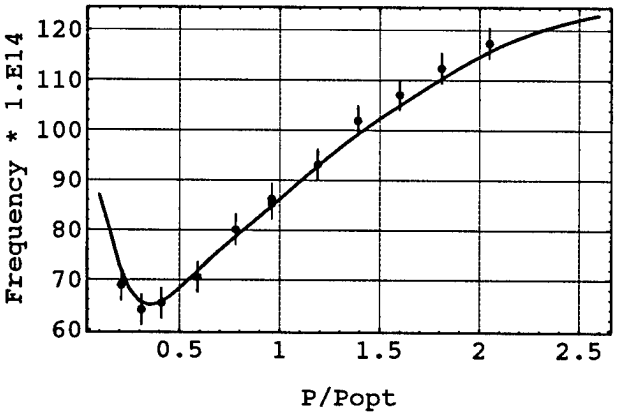


Figure 6: Microwave Power shift for W-E beam

$1{\cdot}10^{-13}$. This discrepancy is not explained at present time. Finally the different considered biases are summarized in Table 1 with their uncertainty. The final accuracy (1σ) is given by the quadratic sum of the uncertainties: $\sigma = 1.1{\cdot}10^{-13}$. Our standard has been compared indirectly with UTC PTB. The comparison is realized through the OP Hydrogen maser and UTC OP. The frequency of the maser measured against the LPTF standard is:

$$\nu_{cs} - \nu_{maser} = 25{\cdot}10^{-14} \pm 11{\cdot}10^{-14} \qquad (2)$$

The comparison between UTC PTB and the maser via UTC OP and GPS gives:

$$\nu_{PTB} - \nu_{maser} = 17{\cdot}10^{-14} \pm 5{\cdot}10^{-14} \qquad (3)$$

5 Conclusion

This paper describes the present characteristics of our standard. The biases that limit the actual accuracy at $1.1{\cdot}10^{-13}$ can be easily overcome after modification of some hardware parts. For example the distributed phase shift can be reduced using two slits in front of cavity slots. The accuracy can then be pushed to few parts in 10^{-14}. The evaluation presented here, joined with the evaluation of other optically pumped standards [2, 3], show that such technique comes to maturity. A international comparison between the different primary frequency standards is coming of great interest.

6 Acknowledgment

Many discussions with A. De Marchi about many aspect of this evaluation are gratefully acknowledged.

References

[1] M. Arditi, J.L. Piqué, "A cesium beam atomic clock using laser optical pumping. Preliminary test," *J. Phys. Lett.*, vol. 41, pp. L379-381, August 1980.

[2] Y. Koga, S. I. Oshima, Y. Nakadau and T. Ikegami, "Performance of an optically pumped cesium frequency standard," in *Electromagnetic Metrology, Proc. of International Symposium on Electromagnetic Metrology (ISEM '89)* Beijing, China, pp. 267-273, Pergamon Press, 1989.

[3] R. E. Drullinger, J. P. Lowe, D. J. Glaze and Jon Shirley, "NIST-7 the new US primary frequency standard," to appear in *Proc. of 7th European Frequency and Time Forum,* Neuchatel, March 1993.

[4] E. de Clercq, A. Clairon, B. Dahmani, A. Gérard and P. Aynié, "Design of an optically pumped Cs laboratory frequency standard," in Frequency Standards and Metrology, Proc. of the fourth symposium, Ancona, Italy, pp. 120-125, Springer-Verlag, 1989.

[5] E. de Clercq, G. D. Rovera, S. Bouzid, and A. Clairon, "The LPTF optically pumped primary frequency standard," to appear in *IEEE Trans. Instrum. Meas.*, April 1993.

[6] A. de Marchi, O. Francescangeli and G. P. Bava, "Feasibility of end to end phase shift correction from the outside of sealed cesium beam tubes," *IEEE Trans. Instrum. Meas.*, vol. 40, pp. 165-169, April 1991.

[7] G. Avila, V. Giordano, V. Candelier, E. De Clercq, G. Theobald and P. Cerez, "State selection in a cesium beam by laser diode optical pumping," *Phys. Rev.*, vol. A36, pp. 3719-3728, October 1987.

[8] E. de Clercq and P. Mangin, "Optical pumping of a cesium beam and population trapping," *Proc. 3rd European Frequency and Time Forum,* pp. 281-292, 1989.

[9] G. Avila, E. de Clercq, M. de Labachellerie and P. Cerez, "Microwave Ramsey resonances from a laser diode optically pumped cesium beam resonator," *IEEE Trans. Instrum. Meas.*, vol. IM-34, pp. 139-143, June 1985.

[10] S. Oshima, Y. Koga, Y. Nakadan, L. Hollberg and R. Drullinger, "The effect of laser line narrowing on the performance of optically pumped cesium atomic beam frequency standards," in *Proc. 2nd European Frequency and Time Forum,* pp. 531-536, 1988.

[11] N. Dimarcq, V. Giordano, P. Cerez and G. Theobald, "Analysis of the noise sources in an optically pumped cesium beam resonator," to appear in *IEEE Trans. Instrum. Meas.*, April 1993.

[12] G.D. Rovera, A. De Marchi, G. Santarelli, "A simple temperature control for semiconductor lasers," *Proc. 4th European Frequency and Time Forum,* pp. 456-459, Besançon, 12-14 March 1991.

[13] R. Schröder, "Frequency synthesis in primary cesium clocks," *Proc. 5th European Frequency and Time Forum,* 1991, pp. 194-200.

[14] C. Audoin, V. Candelier, N. Dimarcq, "A limit to frequency stability of passive frequency standards due to intermodulation effect," *IEEE Trans. Instrum. Meas.*, vol. 40, pp. 121-125, 1991.

[15] J. Vanier and C. Audoin, "The Quantum Physics of atomic frequency standards," Bristol, U.K., Adam Hilger, 1989.

[16] G. D. Rovera et al. To be published.

[17] A. de Marchi, G.D. Rovera and A. Premoli, "Effects of servo loop modulation in atomic beam frequency standards employing a Ramsey cavity," *IEEE Trans. Ultrason. Ferroelec. Freq. Contr.*, vol. 34, pp. 582-591, November 1987.

[18] S. Oshima, Y. Nakadan, T. Ikegami and Y. Koga, "Light Shift in an optically pumped Cs beam frequency standard," *IEEE Trans. Instrum. Meas.*, vol. 40, pp. 1003-1007, Dec 1991.

[19] A. De Marchi, G.D. Rovera, A. Premoli: "Pulling by neighbouring transitions and its effects on the performance of Caesium beam frequency standards" *Metrologia,* vol.20, pp.37-47, 1984.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

REDUCING THE EFFECT OF LOCAL OSCILLATOR PHASE NOISE ON THE FREQUENCY STABILITY OF PASSIVE FREQUENCY STANDARDS

C. SZEKELY, F. L. WALLS, JOHN P. LOWE, R. E. DRULLINGER, A. NOVICK
Time and Frequency Div., National Institute of Standards and Technology
325 Broadway, Boulder, CO 80303

Abstract

We report on the experimental test of a new concept for reducing limitation on short-term frequency stability of passive frequency standards due to local oscillator phase noise. Systems that use sinewave modulation to interrogate a stable resonance are limited in short-term frequency stability by phase noise at the second harmonic of the modulation, f_m. This effect limits the fractional frequency stability to approximately $\sigma_y(\tau) = 0.9\ f_m/\nu_0\ (S_\phi(2f_m))^2$, where ν_0 is the carrier frequency and $S_\phi(2f_m)$ is the phase noise at twice the modulation frequency. This new concept uses notch filters at $\pm\ 2f_m$ from the carrier to reduce this effect. Tests on a modified passive rubidium standard demonstrate an improvement of approximately 18 in $\sigma_y(\tau)$. The dual notch filters proved to be feasible and were obtained commercially. Measurements suggest that ultimate performances of approximately $2 \times 10^{-14}\tau^{-1/2}$ are possible if the atomic resonance has sufficient quality. Additional refinements may reduce this limitation even further.

INTRODUCTION

The short-term frequency stability of passive standards using sine wave modulation is often limited by phase noise at the second harmonic of the modulation frequency f_m [1-4]. When examined in detail the limit on short-term frequency capability depends on the modulation index, linewidth, and all even harmonics of f_m [4].

$$\sigma_y(\tau) = \left[\sum_{n=1}^{\infty} \left(\frac{P_{2n-1} - P_{2n+1}}{P_1} \right)^2 S_y(2nf_m) \right]^{1/2} / 2\tau^{1/2}, \qquad (1)$$

Contribution of the U.S. Government, not subject to copyright.

where $\sigma_y(\tau)$ is the Allan deviation, $S_y(f)$ is the spectral density of fractional frequency fluctuations, and P_{2n+1} is the Fourier coefficient of rank $2n+1$ in the response of the resonance. For most systems the largest contribution is due to the noise at $\pm\ 2f_m$ from the carrier ($n = 1$). This first term can be rewritten as

$$\sigma_y(\tau) = 0.9\ \frac{f_m}{\nu_0} \sqrt{S_\phi(2f_m)}\ \tau^{-1/2} \qquad (2)$$

where $S_\phi(f)$ is the spectral density of phase noise. For example, consider a system with $f_m = 137$ Hz and $S_\phi(274\ \text{Hz}) = 10^{-15.2}$ at a carrier frequency of 5 MHz. The limit to $\sigma_y(\tau)$ due to $S_\phi(274)$ is approximately $\sigma_y(\tau) = 6.2 \times 10^{-13}\ \tau^{-1/2}$. This effect is a very serious problem for proposed diode-laser-pumped, passive rubidium standards since a frequency stability of approximately $\sigma_y(\tau) = 1 \times 10^{-14}\ \tau^{-1/2}$ has been projected, based on signal-to-noise and linewidth [5,6]. A local oscillator with phase noise at $2f_m$ low enough not to significantly compromise this performance can presently be achieved only with cryogenic techniques [7-9].

In this paper we present details on a new approach that significantly reduces the effect of phase noise in the local oscillator on the frequency stability of passive frequency standards such as diode-laser-pumped rubidium [4,6]. This new room temperature approach makes use of a special filter with a notch at both the upper and lower $2f_m$ sidebands. To minimize the limitation on $\sigma_y(\tau)$, f_m should be small, the carrier frequency at which the filters are applied should be high, and the contribution of the phase noise of the filter and following synthesis chain should be low. As a first choice we have chosen $f_m = 37.5$ Hz and $\nu_0 = 10$ MHz, to avoid power line frequencies, to be consistent with available Q-factors in SC-cut quartz resonators, and to be approximately compatible with the expected linewidths in the new passive rubidium standards. This choice also reduces

U.S. Government work not protected by U.S. copyright

the contribution from the multiplier chain by 6 dB as compared to filtering at 5 MHz. We show that the available attenuation of the noise around $\pm 2f_m$ approaches 30 dB. From the theory developed by the Laboratoire de l'Horloge Atomique group at Orsay [4], we expect that the limits to the short-term frequency stability due to local oscillator phase noise can be improved by approximately a factor of 15-30 over traditional approaches. Experimental measurements confirm this view and indicate that short-term frequency stabilities of better than $2 \times 10^{-14}\ \tau^{-1/2}$ can be expected if the atomic signal have sufficient quality. Further improvements seem likely.

DUAL NOTCH FILTERS

The special dual notch filters proved to be feasible and were obtained commercially. They were assembled using two pairs of third overtone 10 MHz SC-cut resonators with nominal Q factors of 10^6 [10]. The resonators were selected to have matched temperature turnover points at approximately 60°C. Figure 1 shows the insertion loss and the phase shift across the resonator as measured on a 50 Ω network analyzer.

The insertion loss through the filter at 10 MHz is about 1 dB while the insertion loss at the bottom of the notch approaches 30 dB. Since the 10 MHz carrier signal is far away from the frequency of the notches, the crystal resonators carry very little power, basically only the noise power. This means that approximately 30 dB more power can be transmitted through the dual notch filter than could be transmitted through a band-pass filter made from 10 MHz resonators of similar quality. High output power is critical in minimizing the noise in the multiplier chain following the filter. The phase variation with frequency at 10 MHz is at least a factor of 30 lower than the phase slope at the center of the notches. This means that the phase of the 10 MHz carrier is much less affected by the frequency variations of the resonators than if a band-pass filter had been used. Some temperature control is needed but not the same thermal regulation that would have been required for a traditional band-pass filter.

Figure 2 shows the block diagram of the system used to measure the phase noise of the resonator. The sensitivity of the mixer and gain of the amplifier were calibrated using the PMCAL approach [11]. The

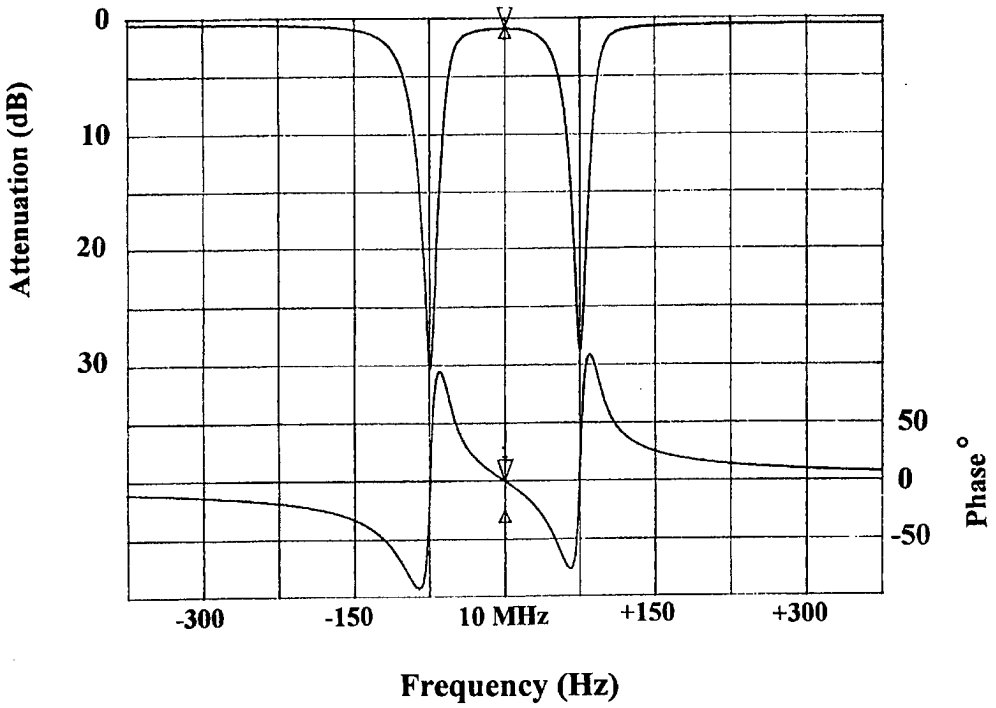


Figure 1. Transmission loss and phase shift across dual notch filter at 10 MHz.

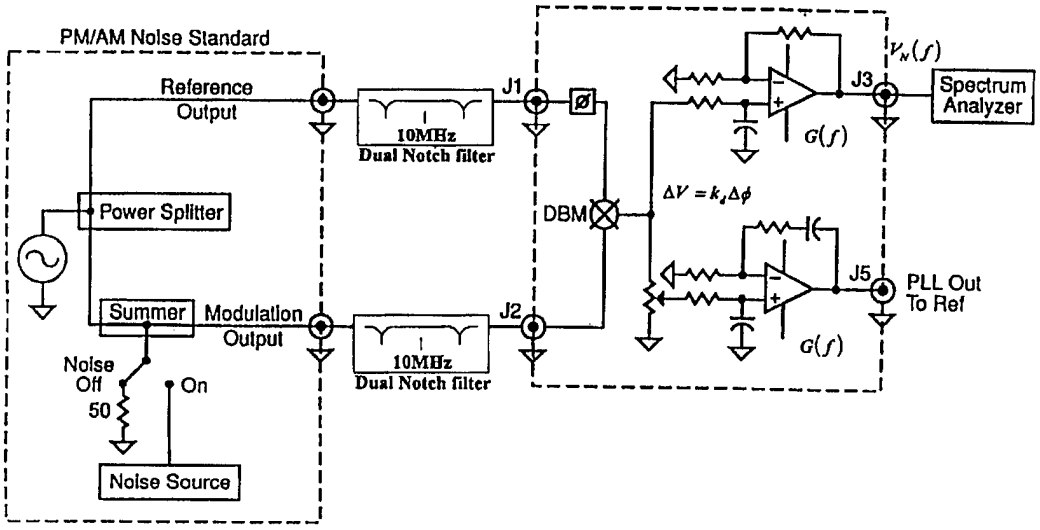


Figure 2. Block diagram of system used to measure phase noise of reference source and dual notch filter.

accuracy of the calibration is typically ±0.5 dB. Two notch filters were used in the measurement to equalize the delays on both sides of the phase bridge. This significantly improves the rejection of the phase noise of the source oscillator resulting in a lower noise floor for the measurements. The phase noise of a single filter (assuming equal contributions) was measured to be flicker phase at a level of $S_\phi(f) = 10^{-14}/f$ (-143 dBc/Hz at 1 Hz) and $S_\phi(75\ Hz) = 10^{-16.4}$ (-167.4 dBc/Hz) at the center of the notch. These results for the phase noise near the carrier are extraordinarily low and further proof of the advantages of notch filters over bandpass filters.

EXPERIMENTAL TEST OF SHORT-TERM FREQUENCY STABILITY

To test the concept of using a dual notch filter for improving the short-term frequency stability of passive standards, we modified the electronics of a small commercial rubidium standard to use an external 10 MHz local oscillator and a modulation frequency of 37.5 Hz. The discharge lamp was removed and the cell was optically pumped using a diode laser that was both frequency and amplitude stabilized [12]. The system was adjusted to accommodate the different noise levels in the electronics.

Figure 3 shows the block diagram 10 MHz source used for testing the effectiveness of the dual notch filter to improve the short-term frequency stability of passive standards. An external noise source with white noise from approximately 1 Hz to 100 Hz can be used to drive a dc-coupled phase modulator. This is used to artificially increase the noise of the 10 MHz local oscillator. The output from the oscillator can be taken before or after the dual notch filter. Figure 4 shows the phase noise after the filter with noise on.

The frequency stability of the diode-laser-pumped passive rubidium standard without noise and without notch filter is shown as curve A in Figure 5. The frequency stability with noise added to the oscillator and no notch filter is shown in curve B, while the frequency stability with noise and with filter is shown in curve C. The frequency stability achieved with the filter and noise on is virtually identical with that achieved with no noise, demonstrating that the dual notch filter is very effective in reducing the effect of the noise at $\pm\ 2f_m$ on the short-term frequency stability. The frequency stability achieved with the dual notch filter and noise on is not limited by the notch performance but by other details in the small commercial physics package.

The phase noise at $2f_m$ was measured to be $S_\phi(2f_m) = 10^{-9.7}$ rad^2/Hz with noise modulation and no filter. Using this value in Eq. (2) yields a frequency stability of $\sigma_y(\tau) = 5 \times 10^{-11}\tau^{-1/2}$, which agrees well with the measured value of $\sigma_y(\tau) = 7.5 \times 10^{-11}\tau^{-1/2}$.

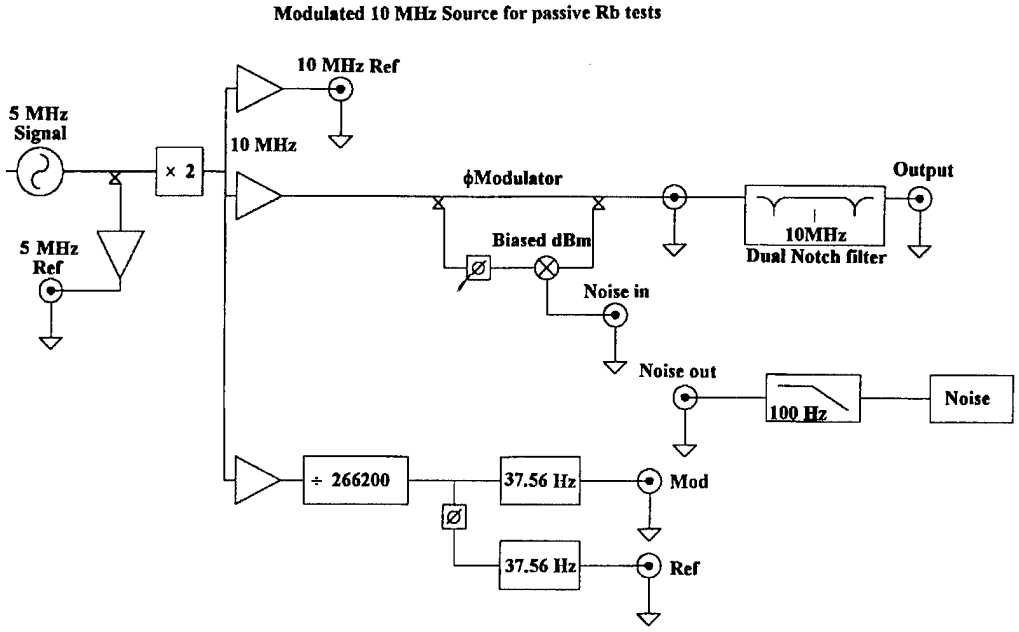


Figure 3. Block diagram of the external 10 MHz source used to drive the modified commercial passive rubidium frequency standard.

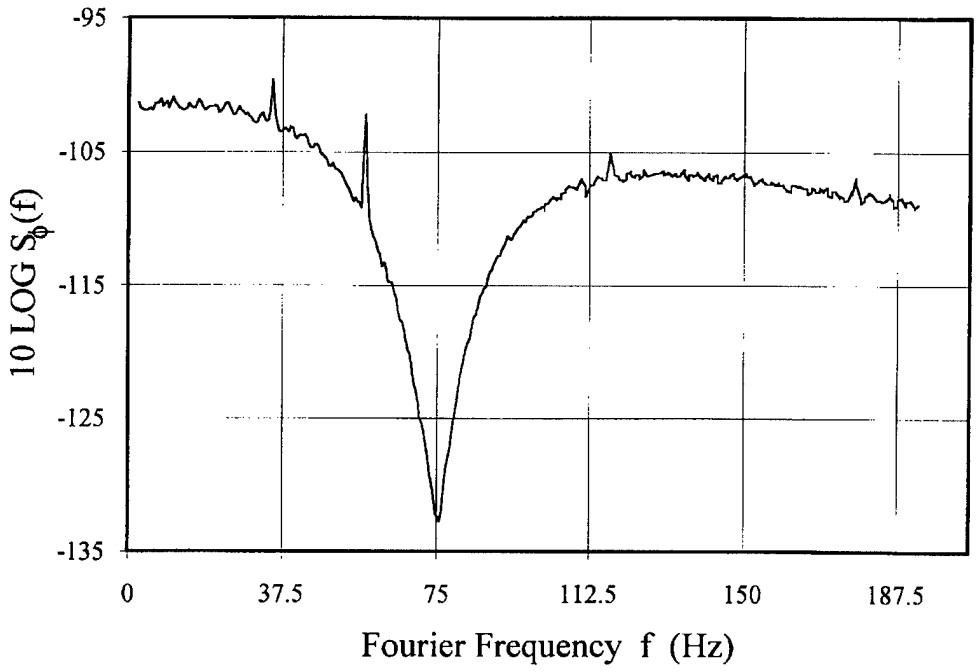


Figure 4. Phase noise of the local oscillator with noise on measured after the dual notch filter.

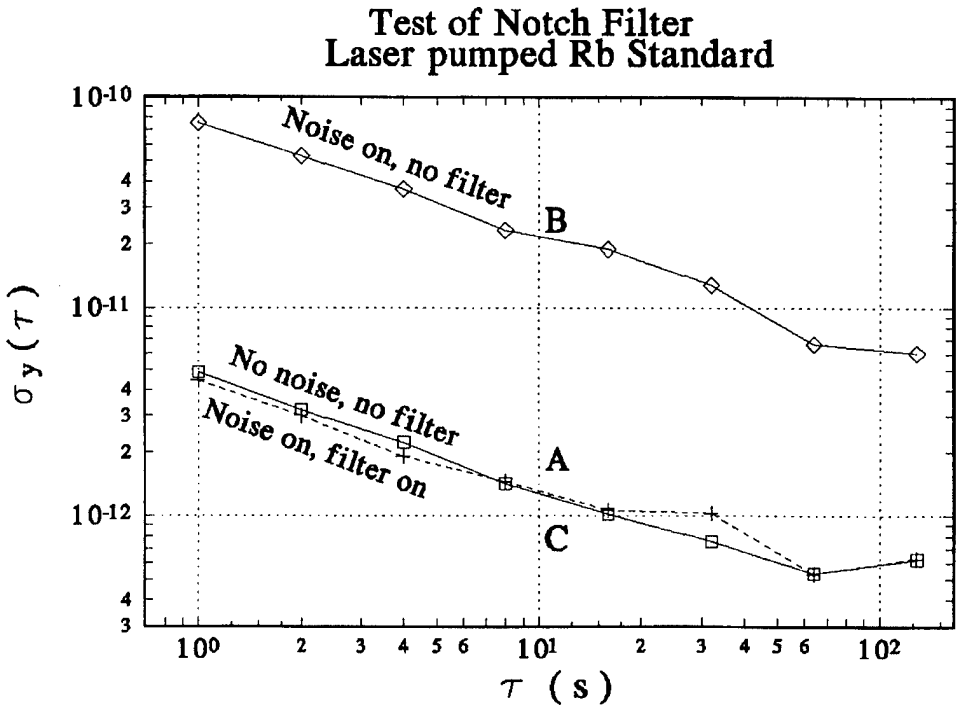


Figure 5. Short-term frequency stability of modified commercial passive rubidium frequency standard. Curve A shows the performance with no noise and no filter. Curve B shows the performance with noise on and no filter. Curve C shows the performance with noise on and filter on.

Demonstrating a full electronics system capable of $\sigma_y(\tau) = 1 \times 10^{-14}\tau^{-1/2}$ will have to wait further work on a rubidium physics package designed to take full advantage of the diode-laser pumping. We have made several 10 MHz to 100 MHz multiplier chains that do not significantly degrade the measured noise floor at the bottom of the notch filter. Using the presently measured performance of the dual notch filter and Eq. (2), we project a lower limit to the short-term frequency stability due solely to the electronics of $2 \times 10^{-14}\tau^{-1/2}$. The actual limitation is probably considerably less. Additional measurements should confirm this. If necessary, additional reductions in the phase noise at the bottom of the notches may also be obtained by selection of lower noise resonators and/or passive components.

CONCLUSION

This work confirms that the concept of using a dual notch filter at $\pm 2f_m$ from the carrier to substantially improve the short-term frequency stability of some passive standards and that the theory advanced in [4] is substantially correct. Further, we have shown that appropriate notch filters for passive rubidium standards can be constructed from available quartz resonators. Such filters have many practical advantages such as less effect on the phase noise of the carrier and higher power handling capability than traditional band-pass style filters. Measurements on prototype filters and multiplier chains indicate that this approach could support diode-laser-pumped passive rubidium standards operating at approximately $\sigma_y(\tau) = 2 \times 10^{-14}\tau^{-1/2}$.

ACKNOWLEDGEMENTS

The Authors thank Franklin Ascarrunz, Huascar Ascarrunz, and Lisa M. Nelson for assistance in constructing and testing the 10 MHz source and Roland Barillet for helpful discussions.

REFERENCES

1. G. Kramer, "Noise in Passive Frequency Standards," CPEM 1974 Digest, IEEE

Conference Publication No. 113, IEEE Cat. No. CH0770-817.

2. F. L. Walls and S. R. Stein, "Servo Technique in Oscillators and Measurement Systems," NBS Tech Note 692, U. S. Government Printing Office, Washington, D. C., SD Cat. No. C1346:692, 1976.

3. L. S. Cutler, Hewlett Packard, P O Box 10350, Palo Alto, CA 94303-0867, private communication, Sept. 1988.

4. C. Audoin, V. Chandelier and N. Dimarcq, "A Limit to the Frequency Stability of Passive Frequency Standards," IEEE Trans. Instr. Meas., vol. 40, pp 121-125, 1991.

R. Barillet, V. Giordano, J. Viennet, C Audoin, "Microwave Interrogation Frequency Noise and Clock Frequency Stability: Experimental Results", Proc. 6th EFTF, 1992.

5. J. C. Camparo and R. P. Frueholz, "Fundamental Stability Limits for the Diode-Laser-Pumped Rubidium Atomic Frequency Standard," J. Appl. Phys., vol. 59, pp 3313-3317, 1986.

6. J. P. Lowe, F. L. Walls, and R. E. Drullinger, "Ultra-High Stability Synthesizer for Diode Laser Pumped Rubidium," Proc. IEEE Freq. Cont. Symp., 1992.

7. S. R. Stein and J. P. Turneaure, "The Development of the Superconducting Cavity Stabilized Oscillator," Proc. 27th Annual Frequency Control symposium, 1973, p. 414.

8. G. J. Dick and J. Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Dielectric Ring Resonator for Ultra-Low Noise," IEEE Trans. Ultrason. Ferroelec. Freq. Contr., vol. 37, pp. 339-346, Sept. 1990.

9. A. G. Mann, A. N. Luiten, D. G. Blair and M. J. Buckingham, "Ultra-Stable Cryogenic Sapphire Dielectric Microwave Resonators," Proc. IEEE 46th Annual Frequency Control Symposium, 1992, pp 167-171.

10. Bob Smythe, Piezo Technology Inc. P.O. Box 547859, Orlando, Florida 32854-7859.

11. F. L. Walls, "Reducing Errors, Complexity, and Measurement Times for PM Noise Measurements," these proceedings.

12. C. Szekely, R.E. Drullinger, F.L. Walls, J.P. Lowe, and A. Novick, "Diode-Laser Pumped, Rubidium Cell Frequency Standards," submitted to Proc. 7th EFTF, Neuchatel, Switzerland, March 16-18, 1993.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

POTENTIAL FOR IMPROVING THE RUBIDIUM FREQUENCY STANDARD WITH A NOVEL OPTICAL PUMPING SCHEME USING DIODE LASERS

N. D. Bhaskar, The Aerospace Corporation
P. O. Box 92957, Los Angeles, CA 90009

Abstract

Frequency stability of a Rubidium Frequency Standard (RFS) is directly related to the signal-to-noise ratio (S/N) of the microwave-induced (6834 MHz) optical signal which is proportional to the fractional atomic population difference between the two $M_F=0$ ground state sublevels of ^{87}Rb. In the present RFSs this fractional population difference is small (<1%). S/N can be substantially improved by concentrating all of the atoms in one of the two $M_F=0$ sublevels. Potentially, this could lead to a significant improvement in the short-term performance of RFSs. We have developed a novel scheme for concentrating a large fraction of the Rb atoms in one of the two $M_F=0$ ground state sublevels. We optically pump the Rb vapor with circularly polarized light from a AlGaAs diode laser tuned to the D_1 transition (794.7 nm). Nearly all of the atoms are concentrated in one of the two high angular momentum states ($M_F=2$ or -2 sublevels depending on the handedness of the circular polarization). The pumping laser is switched off and two radio-frequency (RF) π-pulses are applied sequentially. The first π-pulse transfers the atoms from the 2, 2 (F, M_F) sublevel to the 2, 1 sublevel and the second π-pulse transfers the atoms from the 2, 1 sublevel to the 2, 0 sublevel. The resulting population distribution is diagnosed using a second AlGaAs diode laser (weak probe) in conjunction with a microwave field tuned to the 0-0 transition (6834 MHz). We obtain a fractional population difference of 0.7-0.9 between the two $M_F=0$ sublevels. This should result in an improvement in the S/N by a factor of 70-90 over the lamp pumped RFSs. This could potentially be of considerable importance towards the development of future RFSs. Various relaxations and field inhomogenieties limit the transfer efficiency from being 100%. The details of the experimental technique and possible applications are discussed.

I Introduction

Atomic Frequency Standards (AFSs) find extensive applications in satellite navigation and communication. The remarkable success of the GPS NAVSTAR satellites is primarily due to the on-board AFSs. They also find applications for testing fundamental physical laws. In general the reliability of AFSs has improved significantly over the last decade. In light of this, the GPS Block IIR satellites will carry RFSs in addition to Cesium Beam Frequency Standards. The basic design and architecture of the physics package in the AFSs have remained essentially the same since the atomic standards were first introduced in the early fifties. In the RFS the physics package consists of a ^{87}Rb resonance cell optically pumped by hyperfine filtered light from a RF excited ^{87}Rb discharge resonance lamp. The hyperfine filter consists of a ^{85}Rb cell. One of the two hyperfine components of the resonance light from the lamp is preferentially absorbed by the atoms in the filter cell. Thus, the transmitted light from the lamp-filter assembly mostly contains light which excites the ^{87}Rb ground state atoms in the F=1 hyperfine level, thereby optically pumping a small fraction of the atoms into the F=2 hyperfine level. The applied microwave radiation is tuned to induce atomic transitions between the between the two $M_F=0$ magnetic (Zeeman) sublevels of the two hyperfine levels. This is shown schematically in Fig. 1. The fractional population difference between the

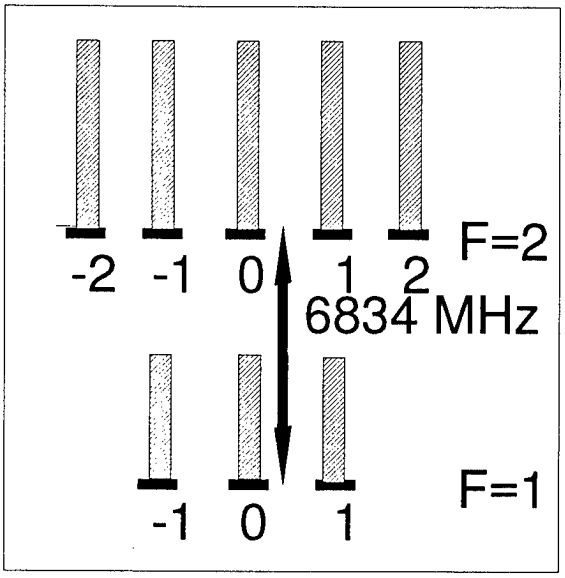


Fig. 1: Qualitative representation of atomic population distribution when optically pumped with a resonance lamp.

0-7803-0905-7/93 $3.00 © 1993 IEEE

two $M_F=0$ sublevels is defined by

$$\Delta = (N(2, 0)-N(1, 0))/N \quad \text{.....(1)}$$

where N is the total population of ^{87}Rb atoms. Typically, in a lamp pumped clock $\Delta<0.01$. This population difference is directly responsible for the error signal in the feedback loop. The short-term frequency stability of a RFS is expressed through the Allan Standard Deviation as

$$\sigma_y(\tau) = A/((S/N) \times Q \times (\tau)^{1/2}) \quad \text{.....(2)}$$

where A is a numerical constant, $Q=\Delta\nu/\nu_0$; $\Delta\nu$ is the FWHM of the microwave resonance signal, ν_0 is the center frequency and τ is the averaging time. The short-term stability is directly proportional to S/N and the signal S is proportional to Δ, the fractional population difference between the two $M_F=0$ sublevels. Increasing the population difference will directly improve the short-term frequency stability of the RFS. In spite of the small population difference the superior performance of the RFS is largely due to the state-of-the-art signal extraction electronics. Over the years many incremental improvements have occurred in a wide range of areas such as thermal design, mechanical stability, integrated electronics, and cavity design. This has contributed to the steady improvement in the short-term stability of the RFS. For some RFSs $\sigma_y(\tau)$ of $2\times10^{-12}/(\tau)^{1/2}$ has been reported [1]. Generally the RFSs are characterized by $\sigma_y(\tau) = (1\text{-}4)\times10^{-11}/(\tau)^{1/2}$.

The advent of tunable diode lasers has stirred considerable excitement for AFS applications [2]. A number of authors have discussed the potential improvement in the short-term stability that can be achieved in a RFS by using diode lasers for optical pumping [2-5]. Considerable advances have been reported in laser pumped cesium frequency standards [6-8]. However, such a dramatic progress has not yet been reported for diode laser pumped RFSs. Theoretical estimates predict at least two orders of magnitude improvement in the short-term performance of a laser diode pumped RFS over the lamp pumped RFSs [5]. The experimentally demonstrated improvement in the S/N achievable in this way, has thus far been, relatively modest [9]. It is generally recognized that achieving the full stability potential using laser diode pumping would be considerably harder than was originally anticipated, primarily due to the optical frequency stability requirements imposed on the diode lasers used for optical pumping and detection. The frequency stability of a laser pumped RFS is impacted by the stability of the laser through the ubiquitous light shifts [10]. More recently development of a diode laser pumped cesium cell frequency standard has been reported [11].

Laser optical pumping can easily redistribute the atomic population in a Rb atomic vapor so as to produce a large population imbalance between the F=2 and F=1 hyperfine levels. This is commonly referred to as hyperfine optical pumping. Using lasers nearly all of the atoms can be pumped into one of the two ground state hyperfine levels, with a uniform distribution among the 2F+1 magnetic sublevels. However, only the atoms in the two $M_F=0$ sublevels participate in the microwave transitions to produce a resonance microwave clock signal. This implies that, in this new population distribution, only a fraction 1/(2F+1) of the atoms present in the resonance cell contribute to the microwave transitions whereas the atoms in the other 2F $M_F\neq0$ sublevels contribute to the noise. The ideal case is where all of the atoms are optically pumped into one of the two $M_F=0$ sublevels, leaving all other sublevels unoccupied. We have developed a novel scheme for achieving such an atomic population distribution. In Section II we describe our experimental procedure and in Section III discuss the results. In Section IV we discuss how this scheme can be incorporated for potential development of a pulsed RFS. In Section V we conclude with a short summary of our results.

II Experimental Procedure

a) Apparatus

A schematic of the apparatus is shown in Fig. 2. A sealed cylindrical Pyrex glass cell containing a small amount of ^{87}Rb, 200 torr helium and 10 torr N_2 is placed at the center of a three axis Helmholtz coil assembly. A DC magnetic field (C-field) is established in one direction (designated as the Z direction), using a pair of Helmholtz coils. The component of the magnetic field in the X-Y plane is cancelled using the other orthogonal pairs of Helmholtz coils. The glass cell is heated to about 60° C using hot air. A AlGaAs diode laser tuned to the Rb D_1 transition ($5^2S_{1/2}$-$5^2P_{1/2}$; 794 nm) is used for optical pumping of the atomic vapor mixture. The pump laser beam is directed along the Z-direction. The collimated pump beam is circularly polarized using a quarter wave plate. The intensity of the transmitted pump beam is monitored using a photodiode.

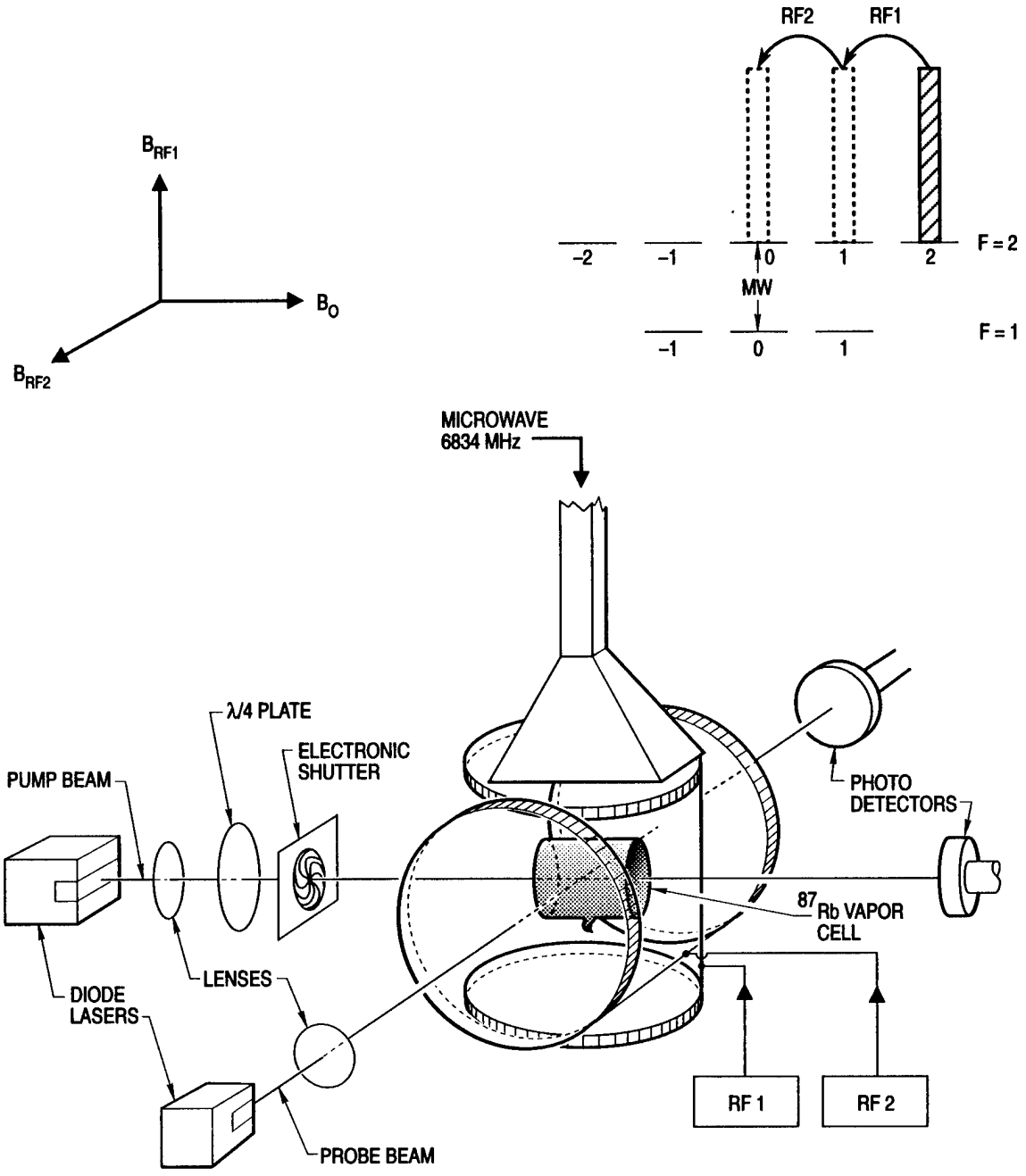


Fig. 2: A schematic of the apparatus.

b) Population Concentration in the 2, 2 Sublevel: Spin Polarization

The relative transition probabilities for exciting the various magnetic sublevels of the ^{87}Rb ground state atoms using the circularly polarized D_1 light are shown in Fig. 3.

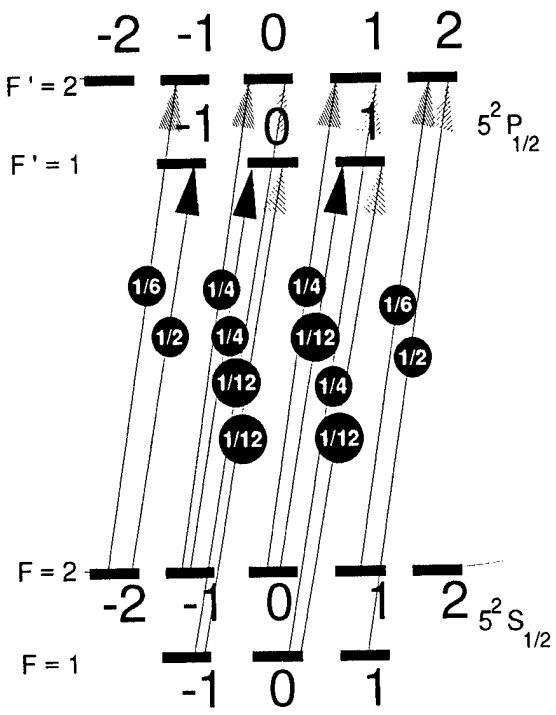


Fig. 3: Relative transition probabilities for exciting the magnetic sublevels of the ground state using D_1 right circularly polarized light.

The atoms in the 2, 2 sublevel do not absorb the σ_+ photons from the pump beam. In the excitation process all of the atoms other than those in the 2, 2 sublevel are excited to the $5^2P_{1/2}$ level. The excited atoms are relatively short lived (radiative lifetime of 28 ns). In the presence of 200 torr of He rapid collisional mixing of the excited atoms results in a uniform distribution of excited atoms among the sublevels of the excited hyperfine multiplet. In addition, the 10 torr of N_2 effectively quenches the resonance fluorescence. Consequently, the excited atoms return to the ground state (radiatively and non-radiatively) with equal probability for populating the ground state sublevels. The population imbalance among the ground sublevels occurs through the differences in the relative absorption probabilities and this is referred to as depopulation pumping [12]. After several pumping cycles, for pumping rate significantly larger than the total relaxation rate the atoms in the ground state are optically pumped into one of the two highest angular momentum states (M_F=2 or M_F=-2) depending on the handedness of the circular polarization. Since the atoms are pumped into an optically non-absorbing sublevel the transmission of the pump beam increases till it reaches a steady state. Using a diode laser, for conditions existing in our vapor mixture, we can easily concentrate nearly all of the atoms into the M_F=2 or -2 sublevel. The circularly polarized beam produces a high degree of spin polarization in the atomic vapor [12]. Rubidium atomic vapor, with a high degree of spin orientation is used in sensitive magnetometers [13].

c) Magnetic Resonance Spectroscopy

The resulting population distribution is diagnosed using the well known Zeeman magnetic resonance spectroscopy [14]. A strong RF magnetic field, orthogonal to the DC field, is applied to the atomic vapor using a pair of coils shown in Fig. 2. The RF fields are generated using frequency synthesizers and RF amplifiers. The RF field induces transitions between the adjacent Zeeman sublevels. The selection rule for the magnetic resonance transition is $\Delta F=0$, $\Delta M_F=\pm 1$. For ^{87}Rb, with nuclear spin 3/2, there are four magnetic resonance transitions for F=2, and 2 for F=1. For small magnetic fields (< a few gauss) the magnetic resonance frequencies for the six transitions are nearly equal. However, for an applied magnetic field of about 20 gauss the resonance frequencies for the six transitions become distinctly different. At resonance, the RF field tends to equalize the population of the two connecting levels and is detected by observing the reduced transmission of the pump beam. Typical magnetic resonance spectra for resonance lamp pumping and diode laser pumping are shown in Figs. 4a and 4b respectively. It is obtained by observing the intensity of the σ_+ pumping light transmitted through the vapor cell as the frequency of the RF field is scanned. When optically pumped with circularly polarized light from a resonance lamp, all of the six Zeeman transitions are observed with decreasing amplitude. Pairs of sublevels connected by the RF are also indicated in the spectra. The amplitude of the signal is directly proportional to the difference in population between the two connecting sublevels. From the spectra one can infer the resulting sublevel population distribution [12]. For the case of optical pumping with a resonance lamp the relative population distribution among the Zeeman sublevels is qualitatively shown in the inset in Fig. 4a. The magnetic resonance spectrum with laser optical pumping is dominated by a single peak, instead of the six seen for the case of pumping with the resonance lamp. With laser pumping over 90% of the atoms are concentrated in the 2, 2 sublevel (shown in the inset of Fig. 4b).

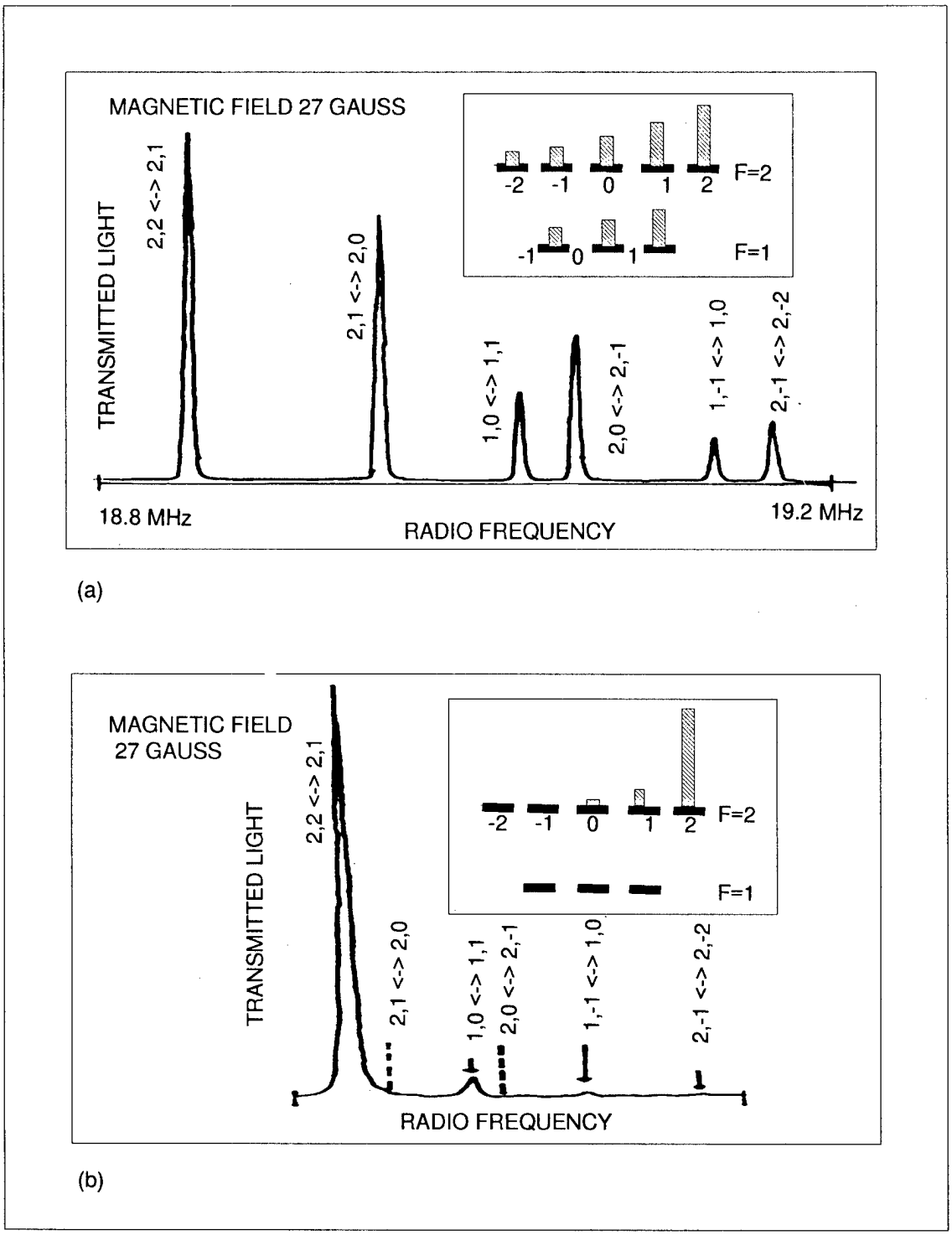


Fig. 4: Magnetic resonance spectra of optically pumped ^{87}Rb vapor mixture. (a) Optically pumped with a resonance lamp. (b) Optically pumped with a diode laser.

d) Population Transfer Using RF π-Pulses

This population distribution is not useful for the operation of a frequency standard, which requires a difference in the atomic population between the 2, 0 and 1, 0 sublevels. Therefore, a suitable scheme has to be devised to transfer, without much loss, atoms from the 2, 2 sublevel to the 2, 0 sublevel. This is a challenging problem. To accomplish this we use two pulsed RF magnetic fields. It is obvious that any

population transfer has to be carried out in the absence of the strong optical pumping radiation which, otherwise would, counteract the transfer. Therefore, the pump beam is turned off, and immediately after two strong RF π-pulses are applied in rapid succession. The first RF π-pulse is tuned to the 2, 2--2, 1 transition and the second RF π-pulse is tuned to the 2, 1--2, 0 transition. A π-pulse is characterized by $\omega \times \tau = \pi$ where ω is the Rabi frequency and τ is the pulse duration [15]. In terms of the RF magnetic field strength $\omega = \gamma \times H_1$ where H_1 is the average amplitude of the RF magnetic field at the center of the resonance cell and γ is the gyro magnetic ratio [15]. The purpose of the π-pulse is to invert the population of the two levels which it connects. For example, when the frequency of the RF field is tuned to the 2, 2--2, 1 transition, and the pulse duration set to satisfy the above condition, the π-pulse inverts the population between the 2, 2 and 2, 1 sublevels. Application of two π-pulses results in transferring a large fraction of the atoms, initially in the 2, 2 sublevel to the 2, 0 sublevel.

e) Population Diagnostics

To diagnose the final population distribution, a weak probe light from another AlGaAs laser is used. The probe laser beam is linearly polarized and is directed perpendicular to the pump beam. This laser is tuned to operate in the D_2 region ($5^2S_{1/2}$--$5^2P_{1/2}$; 780 nm) of the optical transition. The optical power of the probe beam is kept very low using neutral density filters. The transmission of the probe beam through the atomic vapor is monitored using a second photodiode as shown in Fig. 2. With 200 torr of He the absorption lines are substantially pressure broadened but the two ground state hyperfine levels remain fully resolved. However, the excited state hyperfine splittings are completely smeared. The probe laser is tuned to one of the two hyperfine components of the optical transition thereby enabling us to examine the time evolution of the atomic population of a single ground state hyperfine level. After the application of the second π-pulse a microwave pulse tuned to the 2, 0-1, 0 transition is applied. The microwave field is applied through a horn. The microwave field causes transitions of atoms between the 2, 0 and 1, 0 sublevels resulting in a change in the transmission of the probe beam.

A typical probe transmission signal is shown in Fig. 5. When the diagnostic microwave field is applied after the two RF π-pulses, the transmission of the probe through the vapor exhibits several damped oscillations. This is due to the coherent oscillation of the atomic population between the 2, 0 and 1, 0 sublevel induced by the applied microwave field. The frequency of oscillation is the characteristic Rabi frequency determined by the average strength of the microwave magnetic field in the cell region [16]. The amplitude of the oscillation decays rapidly with time due to collisional relaxation to thermal equilibrium and also due to the non-uniformity of the microwave field strength around the resonance cell. The initial amplitude of the oscillatory signal is directly proportional to the initial value of Δ just prior to the application of the microwave diagnostic pulse.

III Results

a) Population Transfer Efficiency-Calibration

To measure the transfer efficiency we employ the following procedure. We optically pump the atomic vapor using linearly polarized light tuned to excite the atoms in one of the two hyperfine levels. The resulting distribution is scanned using the weak probe with no RF pulses applied. This is shown in Fig. 6. When the pump laser is tuned to the F=1 level nearly all of the atoms are pumped into the F=2 level, with a uniform distribution among the five Zeeman sublevels. This corresponds to $\Delta = 0.2$. However, when the laser is tuned to the F=2 level nearly all of the atoms get pumped into the F=1 level, and uniformly distributed among the three Zeeman sublevels with $\Delta = -0.33$. This provides a scale for calibrating the initial amplitude of the microwave signal in terms of Δ and there by calculating the population transfer efficiency for the pulsed RF transfer scheme. With two RF π-pulses, theoretically one can transfer all of the atoms into the 2, 0 sublevel, assuming that initially all of the atoms were in the 2, 2 sublevel; this would result in $\Delta = 1$. Therefore, the initial amplitude of the microwave induced signal in the two RF π-pulse transfer case would be five or three times as large as the calibration signal (linearly polarized pump with no RF) depending on where the laser was tuned to obtain the calibration signal. The calibration signal is also shown in Fig. 5. Transfer efficiency of over 90% can be easily obtained.

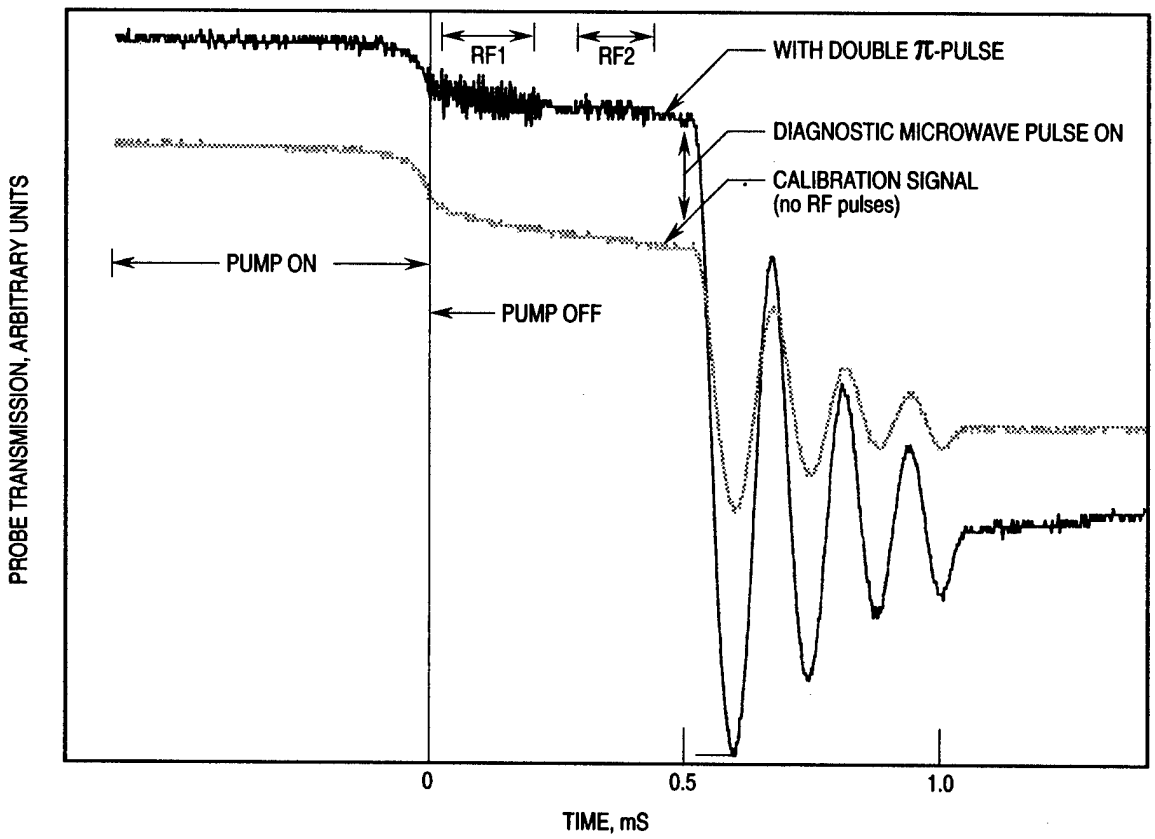


Fig. 5: Microwave induced signal: Comparison of the double π-pulse transfer scheme with the conventional laser optical pumping method. Microwave field is only for diagnostics.

The limitation in the transfer efficiency is primarily due to the inhomogeneities in the RF magnetic fields in the resonance cell. The RF magnetic fields are produced by two sets of coils whose dimensions are comparable to the cell dimension. At any instant of time the π-pulse condition is satisfied only over a fraction of the region the resonance cell due to the non-uniformity of the RF magnetic field strengths.

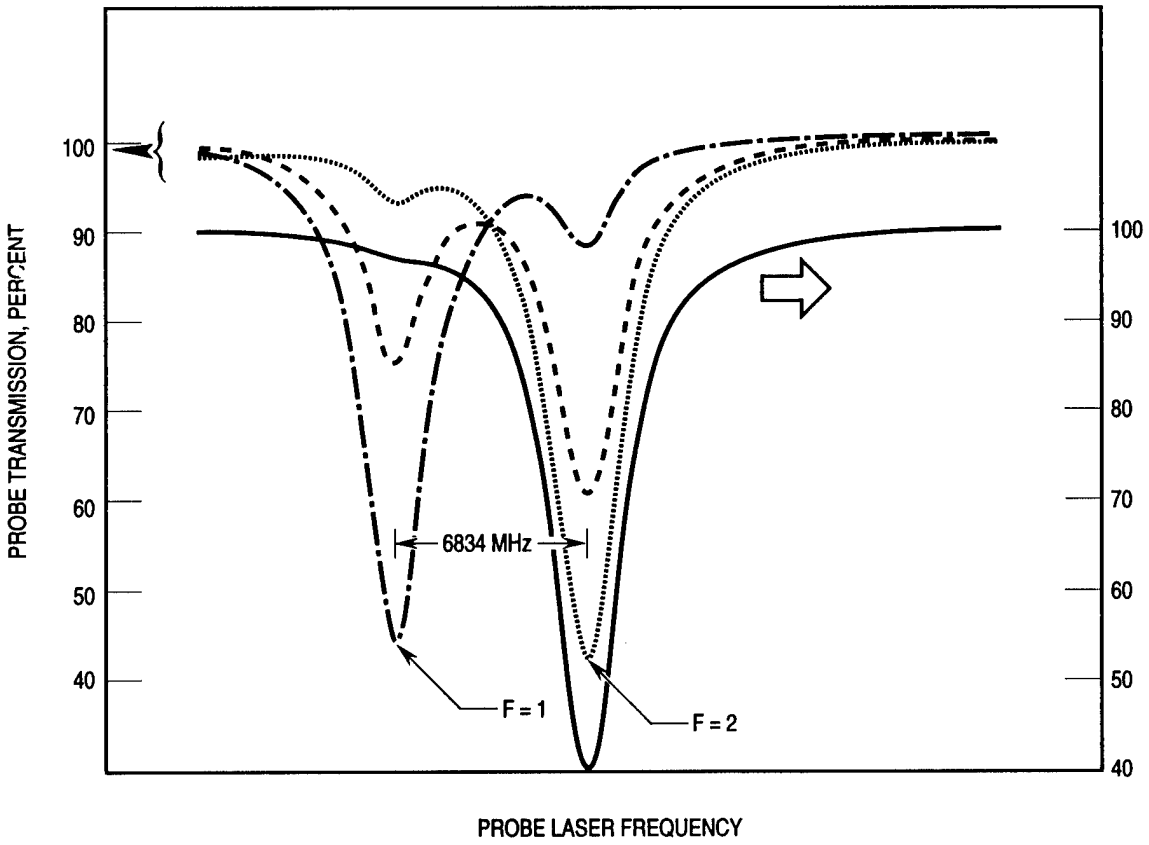


Fig. 6: Absorption profiles of the ^{87}Rb vapor mixture for various pumping conditions. Dotted line and broken-dotted line: Pump linearly polarized and tuned to excite F=1 and F=2 respectively. Dashed line: Pump off-thermal population distribution. Solid line: Pump circularly polarized and tuned midway between F=2 and F=1 transitions.

IV Pulsed Optical Pumping

a) Pulsed Frequency Standard

The fractional population difference Δ that can be obtained using a resonance lamp is typically less than 0.01, whereas in our present scheme we estimate Δ to be about 0.9. This could potentially result in a 90 fold improvement in the S/N, for all other factors remaining the same. The population transfer scheme discussed here is a pulsed one, unlike in the conventional lamp pumped clocks. Therefore, the microwave interrogation and the resulting error signal will also be pulsed. Consequently, this scheme would naturally lend itself to the development of a pulsed frequency standard. This should not be viewed as a draw back. Let us briefly review the various pulse parameters. The pump laser output power is about 10 mW. Typically it takes less than a millisecond to optically pump the Rb vapor, with nearly all of the atoms in the 2, 2 sublevel. The duration of each RF π-pulse is about 200 microseconds. The relaxation times of the various population distributions are found to be in the range of 20-50 milliseconds. In principle, this allows considerable time for the microwave interrogation in some form of a pulsed mode. [17]. A duty cycle of over 95% is achievable, the pump-transfer phase taking up only 5% of the total available interrogation time. One of the advantages of this method is that the interrogation is done with the pump beam off, which eliminates all the noises associated with the pump beam. Another major advantage is that the light shift effects can be potentially eliminated thereby improving the long-term stability. This also relaxes the requirement on the optical frequency stability of the pump laser. A pulsed pumping-interrogation scheme was investigated by English et al. [18] using a lamp for optical pumping. Their motivation was to eliminate the light shift effect. In their method, the pumping light and the microwave radiation were alternately pulsed on and off at a rapid rate in such a way that light and microwave were never simultaneously present in the resonance cell. They observed some residual shifts due to changes in the

pulsing parameters. Since their investigations in the late seventies considerable progress in digital electronics and data processing has occurred. In the context of such a large potential improvement that may be possible in the frequency stability in a RFS we believe that it is worth reinvestigating some of the interesting ideas using lasers for a pulsed RFS. Furthermore, current research and proposals for future advanced standards based on cooled neutral atoms, or cooled stored ion standards visualize utilizing pulsed schemes [19, 20].

b) Low C-Field Operation

The experiments reported here were performed in a C-field of about 20 gauss. This is required in order to resolve the magnetic resonance frequencies of the two RF π-pulses. In a low magnetic field of under a gauss (typical C-field used in a RFS) the six RF transition frequencies are very nearly equal. Application of a RF pulse at such a low magnetic field will only reverse the spin orientation produced by the circularly polarized light whereas our present application requires that the population be transferred to the 2, 0 sublevel from the 2, 2 sublevel. Instead of RF one could use microwave pulses to transfer the atoms from the 2, 2 to the 2, 0 level. Such a scheme would involve two microwave π-pulses, the first pulse tuned to transfer atoms from the 2, 2 to the 1,1 sublevel followed by a second pulse to transfer atoms from the 1,1 to the 2, 0 sublevel. Our preliminary experiments show that the two microwave transition frequencies are well resolved for C-fields as low as a few hundred milligauss. We believe that the pulsed population transfer scheme can be effectively implemented for low C-fields with microwave π-pulses. When optically pumped with circularly polarized light the atomic vapor acquires a macroscopic magnetization and readily becomes a magnetometer. The frequencies of the π-pulses depend crtically on the magnetic field. In principle, this arrangement with a built in magnetometer can be used for accurate control of the C-field thereby providing an accurate correction for the magnetic shift.

V Summary and Conclusions

We have experimentally demonstrated that using diode lasers for optical pumping and RF π-pulses for population transfer nearly all of the ground state Rb atoms in a vapor cell can be concentrated into one of the two $M_F=0$ sublevels. The fractional population difference Δ between the two $M_F=0$ levels can be made as high as 1. This should result in a very significant improvement (about a factor of 100) in the S/N over the conventional discharge lamp pumped RFSs. This would lead to a substantial improvement in the short-term stability of the RFS. This is a pulsed population transfer scheme. The microwave interrogation takes place in the absence of the strong pumping light. This has the potential for eliminating the light shift and thereby improve the long-term frequency stability of a RFS. Though these experiments were carried out at a C-field of about 20 gauss our preliminary investigations show that this population transfer scheme can be implemented just as efficiently in a low C-field using microwave π-pulses. The parameters of the scheme point to the viability of this approach for developing a pulsed RFS. However, considerable work remains to be done. We are currently engaged in the study of relaxations of various types of population distributions obtained by this technique. This is of considerable fundamental interest.

Acknowledgement

We thank James Sutter and Eric Abendroth for all the technical assistance. This work was supported by the U.S. Air Force Space and Missile Systems Center under contract No. F04701-88-0089.

References

1. F. Danzy and W. Riley, "Stability Test Results for GPS Rubidium Clocks," Proceedings of the 19th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp 267-274, December 1-3, 1987.

2. L.L. Lewis and M. Feldman, "Optical Pumping by Lasers in Atomic Frequency Standards," Proc. 35th Ann. Freq. Control Symp., 1981, pp 612-624.

3. M. Hashimoto and M. Ohtsu, "Experiments on a Semiconductor Laser Pumped Rubidium Atomic Clock," IEEE J. Q. Electr., vol. QE-23, 1987, pp 446-451.

4. M. Hashimoto and M. Ohtsu, "Modulation Transfer and Optical Stark Effect in a Rubidium Atomic Clock Pumped by a Semiconductor Laser," J. Opt. Soc. Am. vol. B6, 1989, pp 1777-1789.

5. J.C. Camparo and R.P. Frueholz, "Fundamental Stability Limits for the Diode-Laser Pumped Rubidium Atomic Frequency Standard", J. Appl. Phys., vol. 59, 1986, pp 3313-3317.

6. R.E. Drullinger, J.P. Lowe, D.J. Glaze and Jon Shirley, NIST-7, "The New US Primary Frequency Standard," presented at the 1993 IEEE International Frequency Control Symposium, Salt Lake City, Utah, 2-4 June, 1993.

7. G.D. Rovera, E. de Clercq and A. Clairon, " An Analysis of Major Frequency Shifts in the LPTF Optically Pumped Primary Frequency Standard," presented at the 1993 IEEE International Frequency Control Symposium, Salt Lake City, Utah, 2-4 June, 1993.

8. K. Hisadome and M. Kihara, "Prototype of an Optically Pumped Cesium Beam Frequency Standard," Proc. 45th Ann. Freq. Control Symp., 1991, pp 513-520.

9. R. E. Drullinger, Csaba Szekely and J.C. Camparo, "Diode-Laser Pumped Gas Cell Atomic Clocks," Proc. 46th IEEE. Freq. Control. Symp., 1992, pp 104-107; J.C. Camparo, private communication.

10. J. Vanier and C. Audoin, The Quantum Physics of Atomic Frequency Standards, Adam Hilger, Bristol and Philadelphia, 1989, pp 544-551.

11. P.J. Chantry, B.R. McAvoy, J.M. Zomp, and I. Liberman, "Towards a Miniature Laser-Pumped Cesium Cell Frequency Standard," Proc. 46th IEEE Freq. Control Symp., 1992, pp 114-122.

12. W. Happer, "Optical Pumping" Reviews of Modern Physics, 1972, pp 169-249.

13. A.L.Bloom," Optically Pumped Rubidium Magnetometer," Appl. Opt., vol. 1, 1962, p61.

14. L.W. Anderson and A.T. Ramsey, "Study of the Spin-Relaxation Times and the Effects of Spin-Exchange Collisions in a Optically Oriented Sodium Vapor," Physical Review, vol. 132, 1963, pp 712-723.

15. C.P. Slichter, "Principles of Magnetic Resonance," Springer-Verlag, New York, 1978, ch. 1, pp 23-24.

16. K. Shimoda, "Introduction to Laser Physics," Springer-Verlag, New York, 1986, ch. 7, pp 128-147.

17. M. Arditi and T. Carver, IEEE Trans. Instr. Meas., June-Sept. 1964, p 146.

18. T. English, E. Jechart and T. M. Kwon, "Elimination of the Light Shift in Rubidium Gas Cell Frequency Standards Using Pulsed Optical Pumping," Proceedings of the 10th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Nov. 1978, pp147-168. Published as NASA Technical Memorandum 80250.

19. W.D. Philips and S.L. Rolston, "Laser-Cooled Neutral Atom Frequency Standards," Proc. 46th IEEE Freq. Control Symp., 1992, pp 56-57

20. J.D. Prestage, R.L. Tjoelker, G.J. Dick and L. Maleki, "Ultra-Stable Hg^{+} Tapped Ion Frequency Standard," Proc. 44th Ann. Freq. Control. Symp., 1991, 572-581.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PROSPECTS FOR DEVELOPING GLONASS SPACEBOARD ATOMIC BEAM FREQUENCY STANDARD WITH LASER PUMPING

Prof. Arvid G.Gevorkyan, Dr.Alla N.Bessedina
Dr.Arkady B.Bassevich

Russian Institute of Radionavigation and Time (RIRT)
2, Rastrelli square, St.Petersburg, 193124, Russian Federation

Abstract

An improvement of accuracy parameters of passive atomic frequency standards may be realized using effective techniques of laser pumping and detection. Under these conditions, such parameters are drastically increased as a factor of atomic beam utilization, a difference in population used in sublevels' transition and an efficiency of detection. The paper presents the results of investigation and development of a rubidium beam frequency standard with semiconductor lasers used for pumping and detection. A schematic representation of a rubidium beam frequency standard is given. The discriminator is realized using modified production-type compact spaceborne rubidium beam tube and customized single-frequency injection lasers having frequency control/checking systems. For Ramsey's line width being approximately equal to 700 Hz, a figure of merit of the atomic beem tube exeeds 20. The ways are determined that enable to increase this parameter by at least one order of magnitude. This makes possible to develop the GLONASS spaceboard frequency standard with the daily frequency instability less than $1 \cdot 10^{-14}$.

Introduction

Accuracy parameters of GLONASS, as well as its reliability and duration of self-contained operation, are directly dependent on stability of frequency standards installed space vehicles onboard. The topical problem is to realize the daily frequency instability of 10^{-14}-10^{-15} onboard. Along with developments of spaceborne hydrogen masers, it is of interest to estimate a possible use of frequency standards based on the principle of atomic beam laser pumping and detection. Theoretical and experimental results show a possibility to achieve for frequency standards based on this principle a value of daily frequency instability better than $1 \cdot 10^{-14}$ [1,2]. The paper presents the results of investigation of reference physical packages and electronic circuits of frequency standard breadboards using the principle of optical pumping of Rb^{87} atomic beam by semiconductor lasers. The realized values of discriminator figure of merit (>20) make it possible to achieve an instability $\sigma = 5 \cdot 10^{-12} \cdot \tau^{-1/2}$. At present a compact frequency standard is being developed using a modified onboard Rb^{87} beam tube and semiconductor injection lasers. A real possibility is shown to achieve a figure of merit about 300 and to realize an instability $\sigma \approx 3 \cdot 10^{-13} \cdot \tau^{-1/2}$.

Investigation of discriminator breadboard using Rb^{87} atomic beam with laser pumping and detection

Parameters of atomic beam discriminators with laser pumping are being determined by proper choosing pumping/detection schemes, laser radiation performances, atomic beam tube design and quality of electronic circuits for laser stabilization and control, as well as by choosed working material and atomic beam parameters.

An examination of Rb^{87} atom hyperfine structure shows a simplier energy level structure in comparison with that of cesium (Fig. 1). An efficiency analysis of various pumping/detection systems in Rb^{87} atomic beam with one and two lasers, depending on choosed hyperfine transition, intensity, polarization and time of action of exciting light, made it possible to find optimal versions of Rb discriminator allowing to achieve the maximum figure of merit [3].

The greatest efficiency of

0-7803-0905-7/93 $3.00 © 1993 IEEE

one-frequency pumping of |2,0> and |1,0> sublevels is possible while using 2-2′σ(ĖıĤ) excitation scheme, for instance, D_2-line and equals 41% [3]. This is nearly three time above the maximum pumping efficiency which can be realized for |4,0> and |3,0> states in Cs beam (15,5% for 4-4′σ(ĖıĤ) excitation scheme of D_2-line). The most effective transition for detection is the cyclic |2>→|3′> transition.

Preliminary investigations required for the new frequency srandard realization were performed using a discriminator breadboard comprising an atomic beam tube with Ramsey cavity and external cavity diode injection lasers (ECDL) having systems for control and stabilization of radiation parameters (Fig. 2) [3,4]. An atomic flux in atomic beam tube was being generated by a multi-channel collimator; its value was $2 \cdot 10^{15}$ atom/s at $T_s = 140^0$C. The transit length of U-shape cavity was 13 cm, H_c field ~ 60 mOe.

Pumping and detection were performed using ECDL developed by Lebedev Institute of Physics of Russian Academy of Sciences [5]. These lasers were tuned to Rb^{87} D_2-line (780 nm) and had following parameters: continuous tuning range - more 10 nm, fundamental mode power - up to 3 mW at contrast > 10^2 and generation line width in single-frequency mode < 1 MHz. The temperature stabilization of diode laser was performed with a thermoelectric cooler within 0.1^0C. Microobjective and diffraction grating were mounted on piezoceramics. The full length of the composite cavity was 5 cm. The pumping laser was being tuned to the F=2→ℱ=2 transition and stabilized with atomic beam fluorescence. The detection laser was being tuned to the F=2→ℱ=3 cyclic transition and stabilized using an atomic beam fluorescence of the second (reference) atomic beam tube. The modulation frequency of external cavity length in laser frequency lock loop is 19 kHz. In order to analyze the laser generation spectrum, as well as for coarse frequency tuning to the atomic line, a control system is used comprising a monochromator with 0.1 nm resolution and spectral lamp with Rb^{87} isotope. For search and observation of the absorption line shape and fluorescence the laser is being scanned at 14 Hz. The check of laser frequency tuning to the choosed atomic transition is effected using absorption signals or fluorescence in cells with Rb^{87} vapor having neither coating no buffer gases. Fig. 3a gives an oscillogram of hyperfine components of beam fluorescence within the detection area which took place while scanning a frequency of the indication laser at F=2→ℱ=1,2,3 transitions. The high intensity of F=2→ℱ=3 component is due to multiple (no less than 20) photon reradiation in the detection area. The pumping efficiency for F=2 and F=1 levels was 98%. In order to decrease an amplitude noise influence of the indication laser, the discriminator signal registration at the modulation frequency of Ramsey transition 167 Hz was transferred into the high-frequency area. In such a case, the second harmonic of pumping laser modulation frequency in its stabilization mode (38 kHz) was used. This harmonic was also present in the fluorescence signal. The signal-to-noise ratio measured while using one-frequency pumping and fluorescent cyclic detection and adjusted to 0.25 Hz - band was 15.000, Ramsey line width was 700 Hz, figure of merit exceeded 20. A real possibility was demonstrated to achieve for the discriminator the value of short-term stability $5 \cdot 10^{-12} \tau^{-1/2}$ ($1 \text{ s} < \tau < 10^5$ s) [4].

Breadboard of atomic beam frequency standard with laser pumping and detection

The performed investigations have provided a basis for developing onboard atomic frequency standard with laser pumping of the beam whose circuit is given in Fig. 4. Atomic beam tube (ABT) is designed by means of modification of a compact ABT with Rb^{87} atomic beam. A coaxially-cylindrical quartz resonator with metallization and six-pole magnets are used in this ABT which was developed in 70s for aircrafts and was installed the first GLONASS space vehicles onboard. The main parameters of modified ABT with beam laser pumping are as follows (See Fig. 5): length - 44 cm, atomic transit length between two oscillating fields - 17 cm, beam diameter - 5 mm, beam source temperature - 140^0C. The ATB has three optical chambers with a possibility to generate a local magnetic field. A fluorescence within the detection area is registered by two photodiodes installed at 6.5 mm from the beam axis. The optical detector efficiency is about 30%.

The prototypes of single-frequency laser diode (See Fig. 6) are developed with following parameters: wave length - 780±3 nm, radiation power > 2 mW, generation line width - (25-30) MHz at the operating component/adjacent component ratio being no less than 20.

The frequency standard has a conventional frequency lock loop.

In order to stabilize a frequency of laser diodes the electronic circuits are developed providing an injection current stability about 10^{-6} for operating current range from 35 to 80 mA and an accuracy of maintaining an operating temperature of diode laser $(10^{-2}-5\cdot10^{-3})^{0}$C. Also developed is a system for search and automatic locking laser frequency (using injection current) to the atomic beam fluorescent line. Injection current modulation frequencies for pumping and detection lasers are choosed at 17 and 14 kHz correspondingly. The pumping laser is tuned to $F=2\rightarrow\mathcal{F}=2$ transition. The detection laser is tuned to the frequency of $F=2\rightarrow\mathcal{F}=3$ cyclic transition and stabilized using beam fluorescence in an additional optical chamber placed between the beam source and the pumping area. This enables to exclude a microwave resonance effect on control fluorescence signal within the frequency stabilization loop of the indication laser. Fig. 3b gives an oscillogram of hyperfine components of beam fluorescence with good enough resolution in case of using single-frequency laser diode. The pumping efficiency for F=2 and F=1 levels is up to 90%.

The discriminator design permits to predict an increasing of a figure of merit up to 300 due to rise of ABT output signal by means of:

- detection with two photodiodes (2 time);
- enlargement of the photon collection angle (10 time);
- increase of a number of atoms within the detection area (10 time).

Consequently, an expected rise of ABT output signal is 200 which can provide for the same noise level an instability $\sigma\simeq3\cdot10^{-13}\tau^{-1/2}$.

Conclusion

The promising results are obtained from the experiment with Rb^{87} beam laser pumping and detection. While using a production - type onboard ABT with Rb^{87}, an ABT set with optical pumping and detection is designed, as well as single-frequency laser modules comprising laser diode, microobjective, thermoresistor, feedback photodiode and thermoelectric cooler. The investigations of breadboard of the onboard ABT frequency standard have been carried out which enable to expect a realization of frequency instability not exceeding $1\cdot10^{-14}$ for 1 day.

References

1. P.Cerez, G.Theobald, V.Giordano, N.Dimarcq et al. "Laser diode optically pumped cesium beam frequency standard investigations at LHA." - IEEE Trans. Instrum. Meas., 1991, v.40, N 2, pp. 137-141.
2. R.Drullinger, D.Glaze, J.Lowe, J.Shirley. "The NIST optically pumped cesium frequency standard." - IEEE Trans. Instrum. Meas., 1991, v. 40, N 2, pp. 162-164.
3. A.N.Bessedina. Thesis. - RIRT, L., 1981; A.N.Bessedina, V.L.Velichanskii, A.S.Zibrov et al. "Rb^{87} beam frequency discriminator with optically pumping and detection by injection lasers". - Proc. of the 2nd seminar on the optical orientation of atoms and molecules (SOOAM-II), p. 15-16, Leningrad, 1989.
4. A.N.Bessedina, V.L.Velichanskii, A.S.Zibrov, V.S.Zholnerov. "Development and investigation of a quantum fequency discriminator utilizing an Rb^{87} atomic beam with pumping and detection by injection lasers." - Sov. J. Quant. Electr., 1991, v. 18, N 3, pp. 367-373.
5. A.M.Akulshin, V.Yu.Bazhenov, V.L.Velichanskii, A.S.Zibrov et al."An anomalously great continuous tuning of the frequency emitted by an injection laser with an external selective resonator". - Sov. J. Quant. Electr., 1986, v. 13, N 7, pp. 1391-1400.

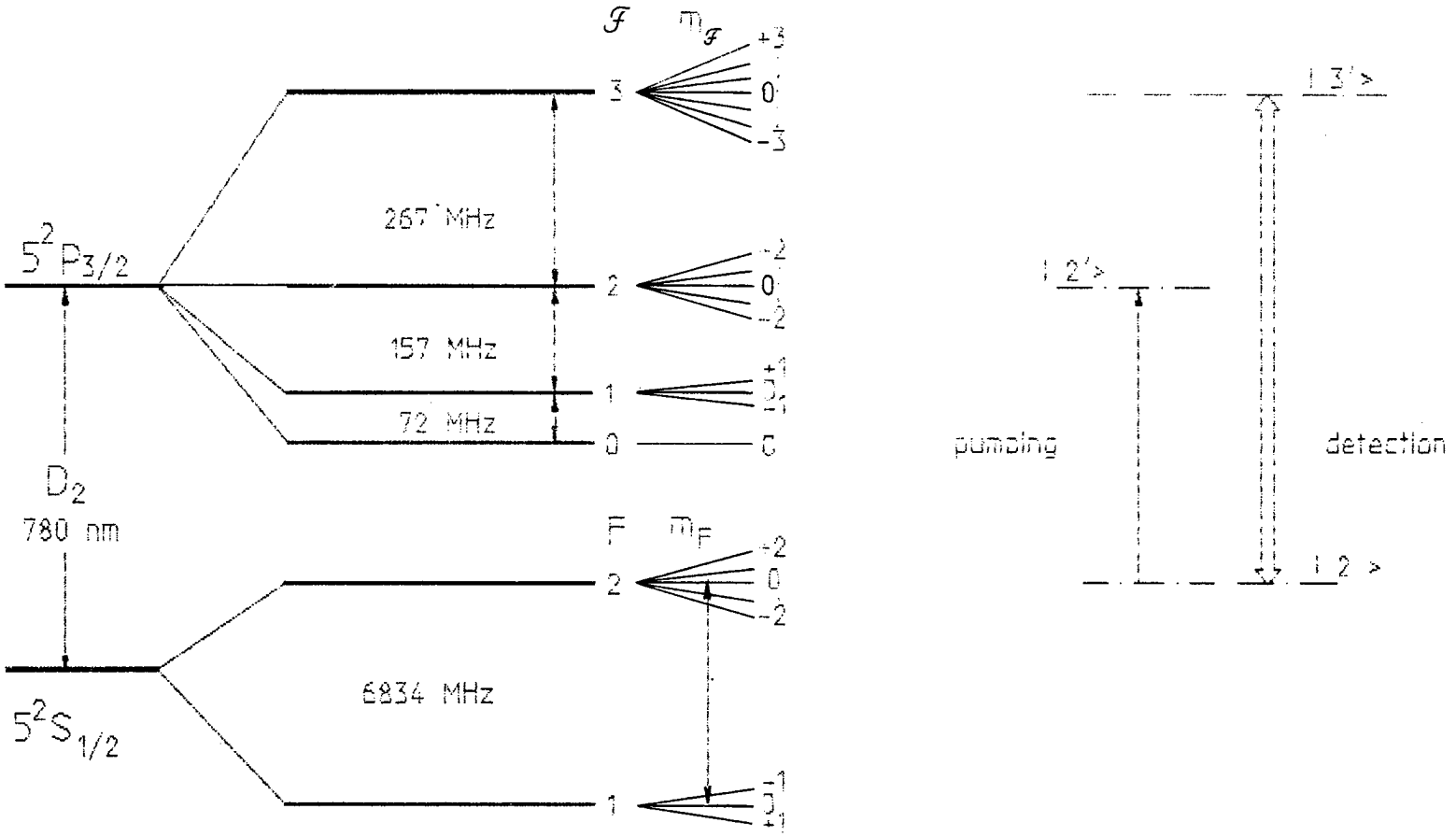


Fig.1. Used energy levels of Rb^{87} atom.

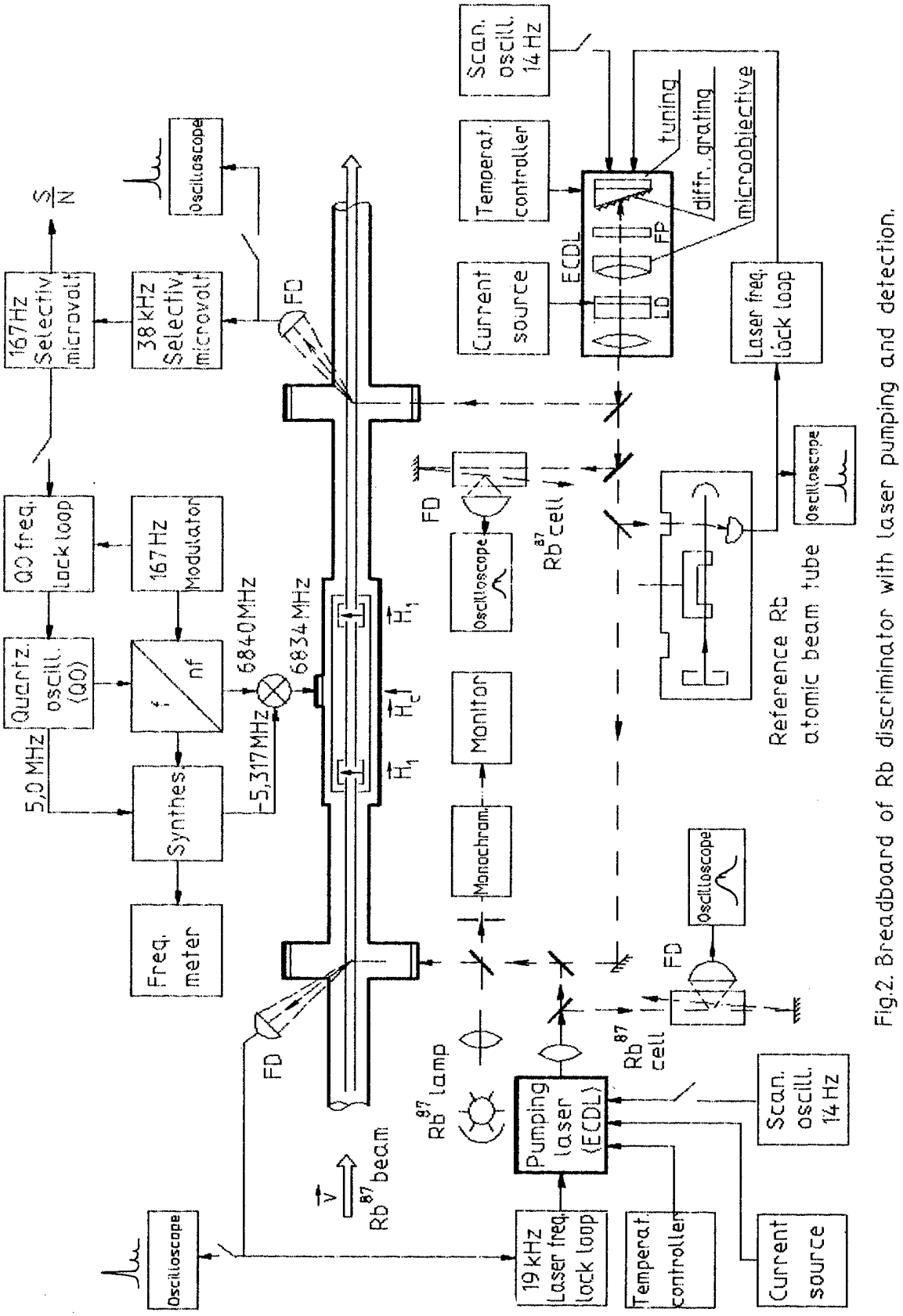
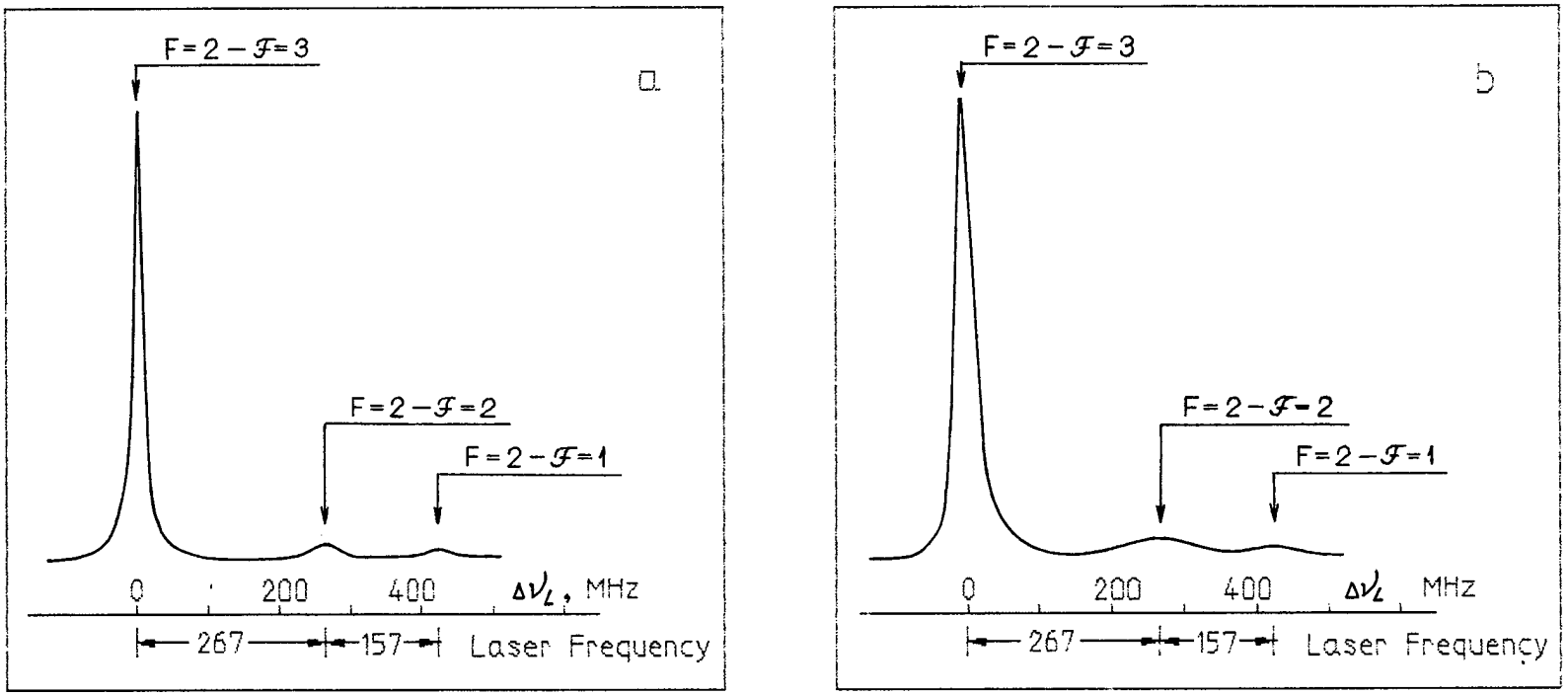


Fig.2. Breadboard of Rb discriminator with laser pumping and detection.



Oscillograms of Rb^{87} atomic beam fluorescence resonances at $F=2-\mathcal{F}$ optical transitions within the detection area:

a - diode laser with external cavity,

b - one-frequency laser diode.

Fig. 3

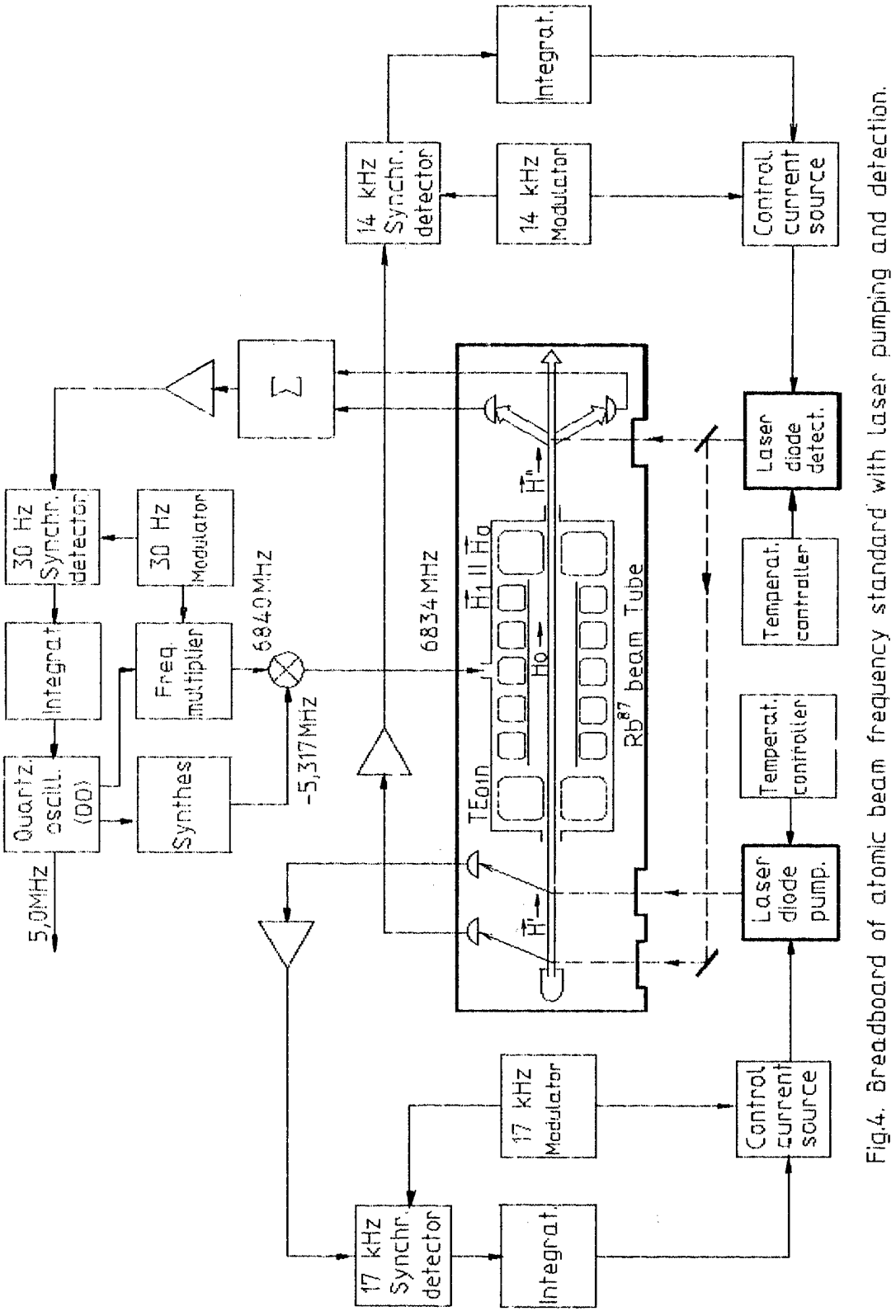


Fig.4. Breadboard of atomic beam frequency standard with laser pumping and detection.

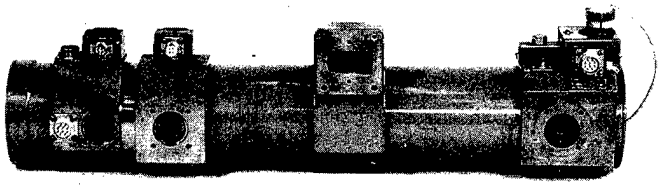


Fig.5. Rubidium Atomic beam Tube
L=44 cm

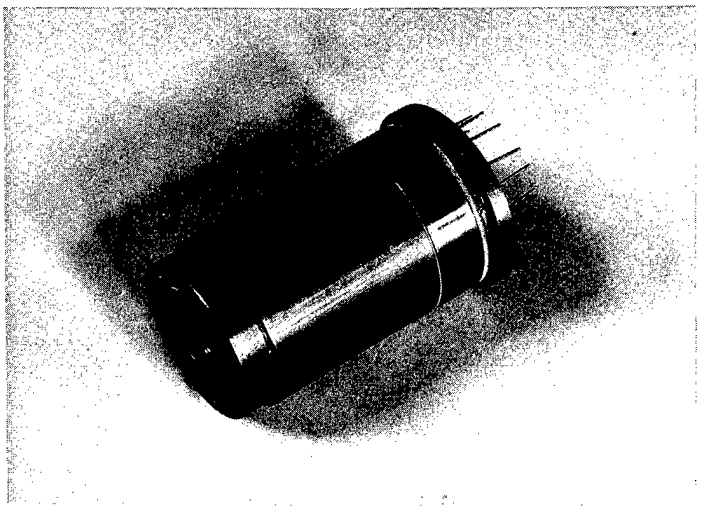


Fig.6. Laser diode module system for a diode laser-pumped Rb beam atomic clock
L=50 mm, ϕ30 mm.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A DUAL FREQUENCY SYNTHESIS SCHEME FOR A HIGH C-FIELD CESIUM RESONATOR

E. Rubiola, A. Del Casale, A. De Marchi
Politecnico di Torino, Dipartimento di Elettronica
c.so Duca degli Abruzzi n. 24, I-10129 Torino, Italy

Abstract

The high C-field concept for a Cesium resonator is based on the idea that a true two level system can be obtained, without having to observe excessive field stabilities, if the C-field is set at the turning point of the frequency of a Zeeman component. This is due to the fact that the reference transition has no linear Zeeman effect at that field.

Nevertheless, because of the second order Zeeman coefficient, a $5{\cdot}10^{-7}$ field accuracy must be realized for a Zeeman bias effect smaller than 10^{-14}. Therefore the C-field must be under closed loop control based on frequency measurements on a field dependent transition.

In this paper a synthesis scheme is proposed, which can provide two frequencies with the necessary resolution: two different frequency lock loops are used to measure the ν_{-1} and ν_0 transitions. The former controls a high stability quartz oscillator with a relatively long time constant. The latter is used to measure C-field and can be faster, if needed to enact a tight control of the magnet's current power supply.

1 High C-field physics and its implications on electronics

In the high C-field standard, as proposed in [1], the field level ($B_0 \simeq 82$ mT) is chosen in such way as to minimize the frequency ν_{-1} of the $m_F = -1$ transition, which is used as the reference. In this conditions the 1st order Zeeman effect vanishes and, from the Breit Rabi Formula, the clock frequency turns out to be

$$\begin{aligned} \nu_{-1\,\min} &= \nu_{00}\sqrt{15/16} \\ \nu_{-1}(B) &\approx 8.900,727,438,257{\cdot}10^{9}\,\mathrm{Hz} + \\ &\quad +(B-B_0)^2 \cdot 4.41{\cdot}10^{-10}\,\mathrm{Hz/T^2} \quad (1) \end{aligned}$$

where ν_{00} is the resonance frequency of the unperturbed atoms.

In the error budget, a bias of $3{\cdot}10^{-14}$ due to C-field can be tolerated for a residual uncertainty below 10^{-14}. For this target, the field must be set to within $5{\cdot}10^{-7}$. Therefore, the C-field must be under closed loop control based on frequency measurement of a field dependent transition. The selected one is $m_F = 0$, whose frequency at B_0 is given by

$$\begin{aligned} \nu_0 &= \nu_{00}\sqrt{15/16} \\ \nu_0(B) &\approx 9.475,548{\cdot}10^{9}\mathrm{Hz} + \\ &\quad +(B-B_0)\cdot 2.1{\cdot}10^{10}\mathrm{Hz/T} \quad (2) \end{aligned}$$

A cavity which resonates at both frequencies ν_{-1} and ν_0 was designed and realized [2]. Because of the wide frequency separation of the two modes, clock operation and C-field measurement can take place at the same time without appreciable accuracy loss due to AC Zeeman effect. In order to accomplish this, two separate sinthesizers, each one with its own modulation frequency, are needed.

The synthesizers' main specifications can be drawn from the results of a theoretical analysis

0-7803-0905-7/93 $3.00 © 1993 IEEE

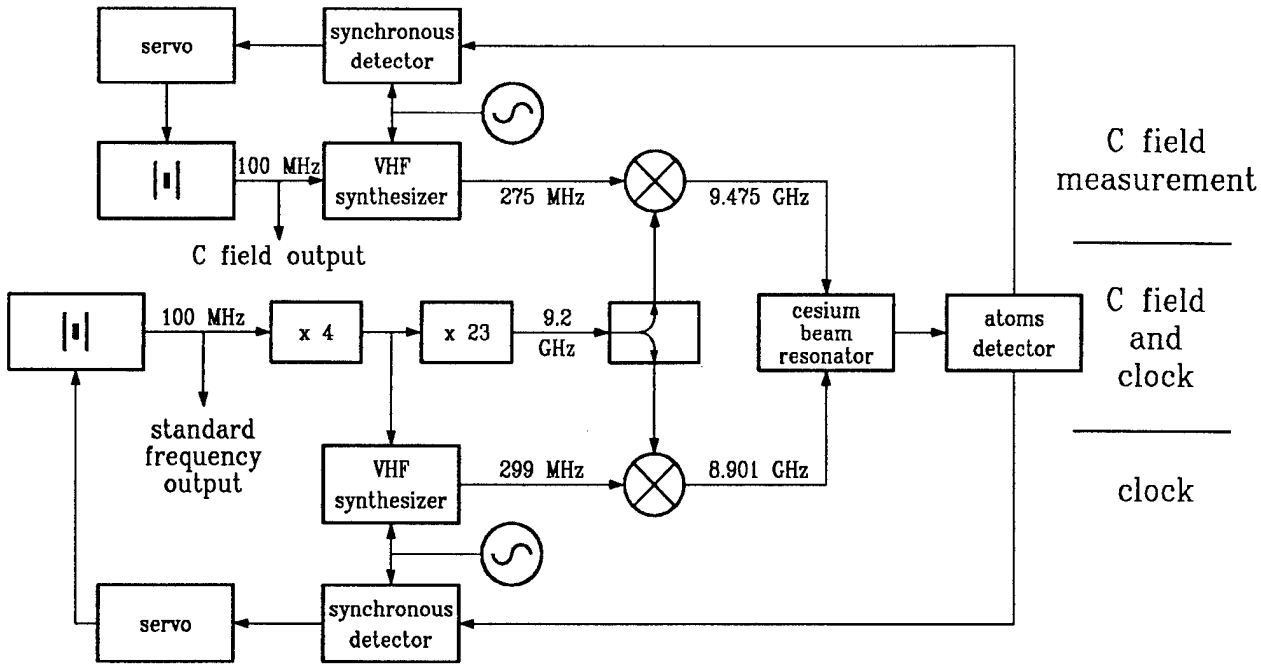


Figure 1: Block diagram of the dual frequency synthesis system under development.

of the beam tube which leads to the following expected parameters: accuracy in the 10^{-14} range, white frequency noise limitation for σ_y at several $10^{-13}/\sqrt{\tau}$, atomic linewidth of 160 Hz and cavity bandwidth of 300 kHz.

Resolution is not a critical parameter for ν_0. From equation (2), it can be stated that 100 Hz it all what's needed if a C-field accuracy of 10^{-7} must be achieved. The clock frequency ν_{-1} should be synthesized at least with the same resolution as for the accuracy limit of the standard, i.e., 10^{-4} Hz, although in principle this is not mandatory.

2 Synthesis scheme

The synthesis scheme is based on the idea that the cesium beam tube can be seen as two separate resonators. Then, two quartz oscillators are frequency locked for clock operation and C-field measurement.

The practical scheme, shown in fig. 1, uses a single multiplication chain for rising the quartz output frequency to 9.2 GHz, and two separate synthesizers for obtaining the final frequencies. The value was chosen to be 9.2 GHz because it is close to $(\nu_0 + \nu_{-1})/2$ and a suitable multiple of the main quartz (100 MHz). Moreover, most of this scheme can be reused for low field operation at 9.193 GHz, if this option is needed.

The multiplication from 100 MHz to 9.2 GHz needs two steps. The first stage ($\times 4$), which can be made with low noise double balanced mixers or with a classic transistor multiplier, is filtered for the minimum 100 MHz spurious output, and the second one ($\times 23$) is based on a SRD for the lowest noise. A single stage 100 MHz $\rightarrow$ 9.2 GHz must be avoided because the SRD multiplier would produce an appreciable power at any frequency multiple of 100 MHz, including 8.9 GHz; this last, close to ν_{-1}, is almost in the cavity bandwidth.

A selection on high performance commercially available oscillators shows that one can choose a low frequency device (5–10 MHz) for the best frequency flicker, or a high frequency one ($\geq$100 MHz) for the lowest phase noise. The phase noise S_ϕ expected at 8.9 GHz from two representative oscillators is shown in Fig. 2, together with the expected white frequency noise of the tube. Since at all frequencies below the atomic linewidth the

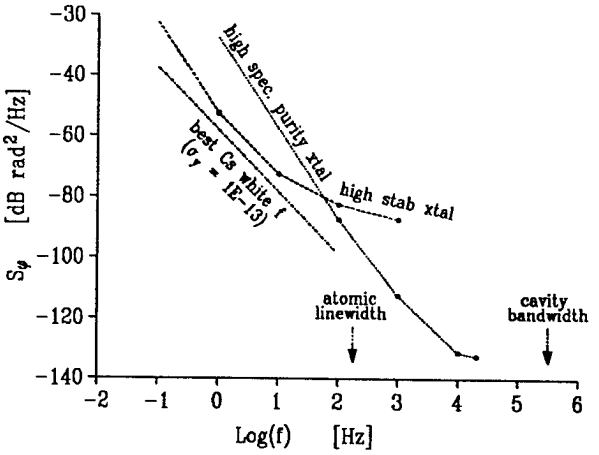


Figure 2: Phase noise of the oscillators at 8.9 GHz, compared to the expected tube noise.

considered oscillators' noise is higher than the white noise expected in closed loop from the beam intensity limitation, the overall noise will be limited by the oscillator. Then it was decided to take full benefits of the technology by using a 100 MHz low noise oscillator phase locked to a 5 MHz low flicker one. This PLL is seen as the main 100 MHz oscillator. A buffer, with a forward/backward gain ratio of 140 dB [3] provides the clock output.

The 299 MHz VHF synthesizer provides a modulated signal and it is responsible of the standard frequency tuning. In a first stage of the experimental work it can be replaced by commercial equipment, loosing in resolution.

The second oscillator is locked to the ν_0 line, thus providing continuous monitoring of the C-field. When the C-field loop is open, this oscillator allows numerically controlled precision scanning of the ν_{-1} Zeeman curve.

3 VHF synthesizer

The main concerns in design the 299 MHz VHF synthesizer are:

1. Spectral purity. The estimated microwave noise floor is −130 dB rad^2/Hz, due to the quartz (see Fig. 2), plus the noise of the multiplication chain. This synthesizer should not degrade the spectral purity of the microwave signal. A limit of −135 dB rad^2/Hz seems reasonable.

2. Spurious signals. They should be kept at a negligible level in the cavity bandwidth. Moreover, spurious signal around 100 MHz are to be carefully avoided because a multiplication by three can take place when mixing this signal with the 9.2 GHz microwave, thus yelding to spurious near ν_{-1}.

3. Resolution should be about 10^{-4} Hz.

The proposed scheme is shown in Fig. 3, where an high resolution 207 kHz synthesizer is up converted to 299 MHz in several steps.

The number of conversion stages and the intermediate frequencies have been chosen for the lowest spurious signals, expecially for those which are close to the carrier, avoiding high Q filters or resonators because of their phase instability. These unwanted signals have been evaluated using double balanced mixers intermodulation tables reported in [9, 10] and experimental data, taking into account all harmonics up to the 11th order.

In order to keep the output clean from 100 MHz spurious signals, the last stage is driven by a 266.$\overline{6}$ MHz signal made by a regenerative converter. This converter is in principle a regenerative divider in which the output is taken from the multiplier output instead of the input. The regenerative scheme is well known for its low noise performances [4, 5].

The lower frequency conversion stages are less critical and can be based on digital dividers. With these dividers, a white noise floor less than −140 to −150 dB can be achieved [5, 6, 7, 8]. Flicker and higher slope phase noise types are not a problem, as compared to the quartz multiplied to 9.2 GHz.

The low frequency synthesizer design suffers from the difficulty of obtaining sub-millihertz resolution and a sufficient spectral purity. NCO based schemes, that have excellent resolution, show a pseudo random distribution of spectral lines spaced from the carrier by multiples of the

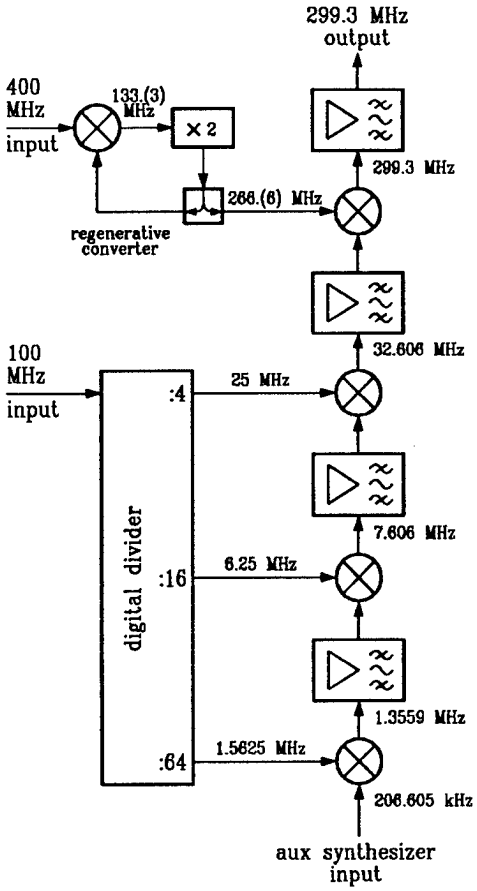


Figure 3: Main VHF synthesizer.

resolution; these signals are within the main FLL bandwidth. In order to overcome these problems, two schemes are under study.

Following the first approach, the 207 kHz signal is obtained dividing the output of a commercial synthesizer by a suitable factor, say 1000, and shaping the output with a digital to analog converter. In the second one, the 207 kHz is implemented as a infinite Q IIR filter in a DSP driven by the standard frequency; a slow software gain control avoids saturation problems.

References

[1] A. De Marchi, "The high C-field concept for an accurate cesium beam resonator," *Proc. 7th European Frequency and Time Forum*, Neucâtel, Switzerland, March 1993 (to be printed).

[2] A. De Marchi, E. Bava, P. Tavella, "Accuracy analysis of the atom-microwave interaction in a TE_{017} cylindrical Rabi cavity for a high C-field cesium beam resonator," *Proc. 7th European Frequency and Time Forum*, Neucâtel, Switzerland, March 1993 (to be printed).

[3] A. De Marchi, F. Mussino, M. Siccardi, "A high isolation low noise amplifier with near unity gain up to 100 MHz," *Proc. 47th Frequency Control Symposium*, 1993.

[4] E. Rubiola, M. Olivier, J. Groslambert, "Phase noise in the regenerative frequency dividers," *IEEE Trans. Instr. & Meas.*, vol. 41 no. 3, pp. 353–360, Jun 1992.

[5] M. M. Driscoll, T. D. Merrell, "Spectral performance of frequency multipliers and dividers" *Proc. 46th Frequency Control Symposium*, 1992, pp. 193–200.

[6] W. F. Egan, "Modeling phase noise in frequency dividers" *IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control*, vol. UFFC-37 no. 4, July 1990, pp. 307–315.

[7] W. F. Egan, "Phase noise modeling in frequency dividers" *Proc. 45th Annual Frequency Control Symposium*, May 1991, pp. 629–635.

[8] F. W. Walls, C. M. Felton "Low noise frequency synthesis," *Proc. 41st Annual Frequency Control Symposium*, May 1987, pp. 512–518.

[9] B. C. Henderson, "Mixers: part 1 — characteristic and performances," "Mixers: part 2 — theory and technology," *RF microwave components designers' handbook*, Watkins-Johnson Company, 1990-91.

[10] D. Cheandle "Selecting mixers for best intermod performance" (part 1 and 2), *RF microwave components designers' handbook*, Watkins-Johnson Company, 1990-91.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PROSPECTS FOR THE USE OF MULTILAMBDA RABI CAVITIES IN ATOMIC BEAM FREQUENCY STANDARDS

E. Bava*, A. De Marchi#, P.Tavella!

* Dip. Elettronica e Informazione, Politecnico di Milano, Italy
# Dip. Elettronica, Politecnico di Torino, Italy
! Istituto Elettrotecnico Nazionale G. Ferraris, Torino, Italy

Abstract: Multilambda Rabi cavities have been pointed out to be a viable solution for the reduction of cavity phase shift and other Doppler related effects in atomic beam frequency standards. For this reason they may be advisable in devices in which operations apt to measure or correct cavity phase shift are impossible or undesirable.
In this paper a discussion is presented of configurations that may gain by the use of such microwave cavities, and of the cavity modes which are most suitable for this application. Analytical and numerical results are reported on particular examples of cavity. They support the idea that a Ramsey cavity is not always the only solution for guiding microwaves in a beam standard.

Introduction

In an era in which accuracies as good as 10^{-14} are becoming a common goal for laboratories involved in the development of primary atomic frequency standards, it is important to reexamine accuracy problems and possible solutions with regards to the design of the microwave structure.
In particular it is not so obvious that the after-the-fact correction to the output frequency of the standard, needed to take into account end-to-end cavity phase shift in Ramsey cavities, may be reliable to that level of accuracy. A bias effect, due to this reason, smaller than 10^{-13} is very difficult to obtain in a beam standard using such a cavity, and a correction accurate to better than 10% is therefore needed for the cavity phase shift in order to claim 10^{-14} overall accuracy for the standard. This is particularly difficult in traditional tubes, employing magnetic selection, because of the coupling between distributed phase shift and atomic angular velocity dispersion, which intrinsically prevent correct beam retrace in beam reversal experiments, but is not obviously feasible in other configurations either.
Another problem which is difficult to curb in Ramsey cavities is the analysis and construction of the two waveguide bends needed to form the U-shaped structure. These in fact are difficult to characterize in usable transmission line models, and are difficult to measure to the necessary level of accuracy.
For these reasons the need has been felt for some time to find a better solution, and analysis of multiwavelength Rabi cavities has been initiated with the purpose of understanding if they can represent a viable solution [1 - 3]. Promising preliminary results have been obtained from the theoretical analysis.
Another situation in which this solution might be interesting both for its relative immunity to asymmetries, and for the much better ease of construction over a Ramsey cavity, is the field of short commercial Cesium beam tubes, where it could also introduce interesting volume savings.
Cavity modes of different kind can be employed in a usable Rabi cavity. Only modes with a longitudinal microwave magnetic field have been so far analyzed, although it is possible that modes with a transversal magnetic field may not be much worse. The number of half wavelengths considered has been so far only odd (up to seven halves), which results in normal interference patterns with a maximum at center frequency, but it could easily be even as well, which would produce inverted patterns with a minimum at center frequency. This may have some advantages in this type of cavity because of the intrinsic capacity of inverted patterns to reject cavity pulling effects.
The most suitable cavity modes appear to be a TE_{01n} mode in a cylindrical cavity, which is the one first studied for the application to primary standards, and the TE_{02n} mode in a rectangular cross section cavity, which might be particularly interesting for use in short commercial tubes.
In this paper the theoretical analysis of Doppler and power effects in such a cavity is shortly illustrated and compared with the results of calculations of the atom-microwave interaction obtained by direct step-by-step numerical integration of the Schrödinger equation. A monokinetic beam has been used so far for these calculations. Integration over the velocity distribution is still in the works, but does not appear to be able to produce dramatic variations in the results. A comprehensive analysis will be published elsewhere.

Description of the cavity

The structure of the TE_{01n} cavity with its feeder and the magnetic field map is sketched in Fig. 1.

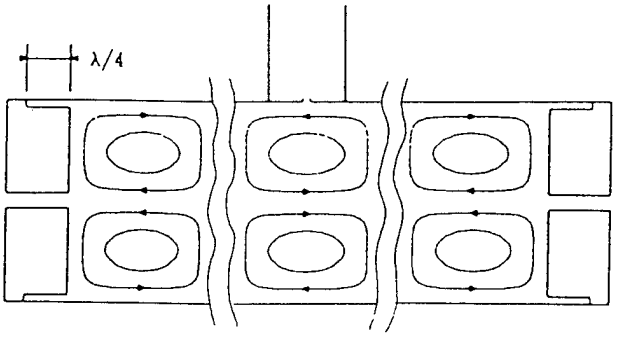


Fig. 1: Cross section of the cavity with magnetic field lines.

0-7803-0905-7/93 $3.00 © 1993 IEEE

In order to excite the desired π transition in the atomic beam, the structure is placed in a quantization field parallel to the cavity axis (x) so that the relevant electromagnetic field component is the longitudinal magnetic field induction $B_x(x)$. The analysis of such a field distribution along x is performed by means of the scattering matrix formalism with reference to the transmission line model given in Fig. 2. Asymmetries are introduced in the losses of the two end shorts ($\Gamma_{1,2} = \Gamma e^{\pm\chi}$, where + (-) refers to #1 (#2) section, and $\Gamma = -e^{-\xi}$ with $\xi = \xi' + i\xi''$ and $\chi = \chi' + i\chi''$) and in the different cavity-arm lengths ($l_{1,2}$ with $l_1 - l_2 = 2\,d$ and $l_1 + l_2 = 2\,l = L$), whereas the propagation constant $\gamma = \alpha + i\beta$ is supposed independent of the x coordinate. Forward and backward transverse-electric-field wave amplitudes a_j and b_j, with respect to the three-port junction, are linked by a scattering matrix S which, due to the symmetry of the structure, may be written as:

(1)
$$\mathbf{S} = \begin{vmatrix} R & M & T \\ M & R & T \\ T & T & K \end{vmatrix}$$

where R and K are the reflection coefficients from the two guide sections of the cavity and from the center feed, respectively, and M and T are the transmission coefficients between the two sides of the cavity and between one side and the feeder.

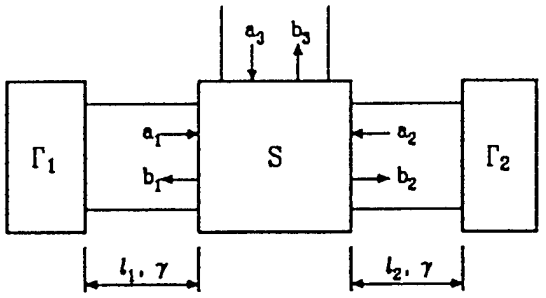


Fig. 2: Model used for field calculation.

Without any loss of generality T can be assumed real and the reference planes 1 and 2 coincident (position of plane 3 is irrelevant). For a lossless junction S is unitary and, among others, the condition $|M - R| = 1$ is obtained; as a consequence we can set $M - R = e^{i\delta}$. In the limit of weak coupling between feeder and cavity, $M \to 1$ and $R \to 0$ and δ is vanishingly small. However even with a 4 mm diameter circular hole, it can be shown [4] that $\delta = 2*10^{-3}$.

Analytical development

From the analysis of the circuit in Fig. 2 it turns out that the two travelling waves b_1 and b_2 feeding the two halves of the cavity, can be expressed as:

(2)
$$b_1 = b_s + b_a \qquad b_2 = b_s - b_a$$

where, by omitting terms in $(\gamma d)^2$ and χ^2, the symmetric and antisymmetric components are:

(3)
$$b_s \approx \frac{a_3 T}{\Delta_s} \qquad b_a \approx b_s(e^{2\gamma d - \chi} - 1)(1 - 1/\Delta_a)$$

with

(4)
$$\Delta_s = 1 - (M + R)\Gamma e^{-2\gamma l}$$

(5)
$$\Delta_a = 1 + (M - R)\Gamma e^{-2\gamma l}$$

The longitudinal magnetic field induction B_x, which is relevant to the problem at hand, is proportional to the transverse electric field and, therefore, its expression in the two cavity halves ($x < 0$ in the half #1 and $x > 0$ in the half #2) becomes:

(6)
$$B_{x1}(x) \propto b_1 e^{-\left(\gamma l_1 + \frac{\xi - \chi}{2}\right)} \sinh\left[\gamma(l_1 + x) + \frac{\xi - \chi}{2}\right]$$

(7)
$$B_{x2}(x) \propto b_2 e^{-\left(\gamma l_2 + \frac{\xi + \chi}{2}\right)} \sinh\left[\gamma(l_2 - x) + \frac{\xi + \chi}{2}\right]$$

These expressions are to be used introducing the actual cavity dimensions and operation frequency with the resonance condition for the considered mode, that is $2\beta l + \xi'' = n\pi$. As a consequence $\Delta_a \approx 1 + e^{i\delta}$.

After substitution of expressions (2) into (6) and (7), taking into account (3), (4), and (5), B_{xl} becomes:

(8)
$$B_{xl}(x) \propto b_s N_1 \cos\left[\left(1 - i\frac{\alpha}{\beta}\right)\beta x + r_1\right]$$

with

(9)
$$N_1 = \left(1 + \frac{b_a}{b_s}\right) e^{-i(\beta d - \chi''/2)} e^{-[\alpha(l+d) + (\xi' - \chi')/2]} \approx e^{-(\alpha l + \xi'/2)}$$

(10)
$$r_1 = \beta d - \frac{\chi''}{2} - i\left[\alpha(l + d) + \frac{(\xi' - \chi')}{2}\right]$$

A similar expression holds for B_{x2} with:

(11)
$$N_2 = \left(1 - \frac{b_a}{b_s}\right) e^{i(\beta d - \chi''/2)} e^{-[\alpha(l-d) + (\xi' - \chi')/2]} \approx e^{-(\alpha l + \xi'/2)}$$

(12)
$$r_2 = \beta d - \frac{\chi''}{2} + i\left[\alpha(l - d) + \frac{(\xi' + \chi')}{2}\right]$$

The last part of expressions (9) and (11) (where $N_1 \approx N_2$) are approximated to first order, by considering very small compared to unity all terms containing a product of any two of the quantities $\gamma d, \chi, \delta, \alpha l$ and ξ'. The term αd is also dropped, as calculations show that its effect is small in the cylindrical cavity, for reasonable values of d. From here on it will be assumed $N_1 = N_2 = N$.

The field time dependence is introduced by taking the real part of $B_{xj}e^{i\omega t}$, for $j = 1,2$, being ω the field angular frequency.

In the study of the atom-field interaction, since $\omega \approx \omega_0$, the rotating wave approximation is used. The relationship between time origin and x coordinate is chosen as $x = vt$, with v the atom velocity along the cavity axis.

The magnitude of the frequency shift due to the 1st order Doppler effect is obtained by evaluating the atom transition probability as a function of detuning from the natural resonance ω_0 of the atom at rest. To this aim we consider the two-level atom Hamiltonian with coupling due to a magnetic-dipole π transition.

The Magnus expansion method [5] appears to be a useful tool to evaluate the atom evolution in a multiwavelength cavity when field expressions are complicated by the presence of losses in the walls. This method has been used to study many problems and in particular the interaction of atoms moving along a pure standing-wave [6] and it has been

shown that convergence problems in the useful range of field amplitudes are avoided if the interaction representation is used and if the cavity is divided in zones, each a half wavelength long. Therefore each zone is labelled with the integer k which ranges from -(n-1)/2 to (n-1)/2, n being the number of halfwavelengths in the cavity. For each k value a new origin is established at $x = x_k$, the antinode coordinate of the magnetic field standing-wave in the lossless case (Im $r_j = 0$).

The x_k values are given by the condition $\beta x_k = -\beta d + \chi''/2 + k\pi$ and the boundaries of each zone are $\beta x_b = -\beta d + \chi''/2 + (2k \mp 1)\pi/2$. By means of a k dependent unitary transformation, the zone boundaries become $-\pi/2$ and $\pi/2$, therefore expressions easier to handle are produced and, most important, the time of flight between adjacent zones is taken into account exactly, independently of the iteration level at which the Magnus expansion is truncated. This goal is accomplished in the following steps. After introducing the equivalent time

$$\tau = \int_{-\pi/2\beta v}^{\pi/2\beta v} \cos\beta v t \, dt = \lambda_g/\pi v$$

and the normalized detuning $\theta = (\omega - \omega_0)\tau/2 = (\omega - \omega_0)/\beta v$, the expression of the interaction Hamiltonian $H_I(x)$ becomes:

(13) $$H_I(x) = -\hbar\frac{b_c\tau}{4}N\begin{vmatrix} 0 & \cos\left[\left(1+i\frac{\alpha}{\beta}\right)\beta x + r_j^*\right]e^{-i\theta\beta x} \\ \cos\left[\left(1-i\frac{\alpha}{\beta}\right)\beta x + r_j\right]e^{i\theta\beta x} & 0 \end{vmatrix}$$

where b_c is the Rabi angular frequency corresponding to the amplitude B_p of an ideal standing wave (the notation here for b_c is the same used in [7]). With the substitution $y_k = \beta(x - x_k)$, in the k-th zone the transformed Hamiltonian $H'_I(y_k)$ and the evolution operator $U'_I(y_k)$ satisfy the differential equation:

(14) $$i\hbar\frac{dU'_I(y_k)}{dy_k} = H'_I(y_k)U'_I(y_k)$$

Because of the unitarity condition, the operator $U'_I(y_k)$, describing the evolution between the beginning and the end of each zone, will be expressed as:

(15) $$U'_I(y_k) = \begin{vmatrix} F & iG \\ iG^* & F^* \end{vmatrix}$$

First-order solutions

The amount of the 1st order Doppler effect is expected to be rather small and, as a consequence, the Magnus method should provide some interesting information just at the 1st step, although at this approximation level Im(F) = 0. Moreover the 1st iteration should provide a satisfactory approximation for low field levels. For a satisfactory evaluation of the power dependence of the shift the first iteration seems insufficient.

By retaining only first order terms in perturbation parameters and the first power of θ in the products containing θ and the same perturbation parameters, the evaluation of the matrix elements in (15) leads to the following expressions:

(16) $$F_k^{(1)} \approx \cos L_0 - \mathrm{Re}[\eta_k]L_0\sin L_0$$

(17) $$G_k^{(1)} \approx (-1)^k(\sin L_0 + \mathrm{Re}[\eta_k]L_0\cos L_0 + i\,\mathrm{Im}[\eta_k]\sin L_0)$$

where subscripts k refer to the zone number and superscripts (1) to the iteration order.

In (16) and (17) $|L_0|$ has been substituted with L_0 because we are interested only in the range of small θ values and:

(18) $$L_0 = 2N\frac{b_c\tau}{4}\frac{\cos\theta\pi/2}{1-\theta^2}$$

(19) $$\eta_0 \approx \theta\frac{\chi'}{2} + i\left(-\frac{\alpha}{\beta} + \alpha l + \frac{\xi'}{2}\right) \qquad \text{for} \quad k = 0$$

(20) $$\eta_k \approx \theta\left(\frac{\chi'}{2} - \frac{\alpha}{\beta}k\pi + \alpha l + \frac{\xi'}{2}\right) - i\frac{\alpha}{\beta} \qquad \text{for} \quad k > 0$$

(21) $$\eta_k \approx \theta\left(\frac{\chi'}{2} - \frac{\alpha}{\beta}k\pi - \alpha l - \frac{\xi'}{2}\right) - i\frac{\alpha}{\beta} \qquad \text{for} \quad k < 0$$

As expected, the expression for L_0 is coincident with that of L in [6] where neither asymmetry nor loss parameters where introduced.

In a halfwavelength interaction structure (k = 0, n = 1) the transition probability $P^{(1)}(1)$ is estimated as:

(22) $$P^{(1)}(1) = G_0^{(1)}(1)G_0^{(1)*}(1) \approx \sin^2 L_0 + 2\mathrm{Re}[\eta_0]L_0\sin L_0\cos L_0$$

At this approximation level, it depends only on χ', in fact the other perturbation terms, retained to the first order, are cancelled.

By keeping in (18) terms up to θ^2, the normalized frequency shift θ_M due to the 1st order Doppler effect turns out to be:

(23) $$\theta_M^{(1)}(1) \approx \frac{\chi'}{4\left(\frac{\pi^2}{8} - 1\right)}$$

It is independent of the power level: this seems to be a feature of the 1st iteration. Note that the denominator in (23) is almost 1 and therefore $\theta_M(1) \approx \chi'$. Such expression is in agreement with the semi-empirical evaluation reported in [3].

In a three halfwavelength cavity (n = 3) the transition probability is evaluated by using a similar procedure and the normalized frequency shift results:

(24) $$\theta_M^{(1)}(3) \approx \frac{\chi'}{4\left(\frac{\pi^2}{8} - 1\right) + \frac{2\pi^2}{L_0}\sin 2L_0}$$

which turns out to be dependent on the power level only because of the accumulated atom polarization in the different zones. The expression (24) shows that also in this case χ' is the only perturbation parameter affecting the shift to 1st order. In fact in each zone labelled with $k < 0$ the perturbations due to α/β, αl, and ξ' give contributions to the Doppler shift, but their effects are cancelled (to the first order) by the correspondent zones with $k > 0$.

Moreover by comparing the expressions of $G(1)$ and $F(1)$ with those of $G(3)$ and $F(3)$ and $G(5)$ and $F(5)$ it turns out that this property holds also when the computations are extended to higher k values, therefore the perturbation parameter mostly affecting the Doppler shift remains χ' also for a higher (odd) number of halfwavelength in the cavity.

Numerical evaluations and discussions

The analytical development, though approximated, has shown that the asymmetry parameter mostly influencing the Doppler shift, i.e. shifting the maximum of the transition probability, is χ', the semi-difference in the reflection

coefficient amplitudes at the reference planes 1 and 2. Numerical evaluations confirm this behavior, yielding shift linearly dependent on χ'.

By setting C_2 and C_1 the probability amplitudes of the two atomic levels, numerical solutions have been obtained by the integration of the Schrödinger equation with the Hamiltonian (13). Such integration has been evaluated with a 4th order Runge-Kutta method, checking for accuracy at each integration step up to 10^{-12} level, for different values of the detuning parameter θ. The evaluation has been repeated for different cavity lengths, setting the initial and final values of the integration variable $y = \beta x$ to $-\beta d + \chi''/2 - n\pi/2$ and $-\beta d + \chi''/2 + n\pi/2$ respectively and appropriately setting the other parameters dependent on n. The initial values are $C_1 = 1$, $C_2 = 0$ in all cases.

Numerical evaluations confirm the major dependence on χ' and show a clear dependence on power for the Doppler shift.

Recording the value of $C_2C_2^*$ at the end of the cavity for different detuning values, the line-shape as well as the line-width can be inferred. Such evaluation repeated for different n and power values illustrates the advantages of using a larger number of halfwavelength zones and a power level below the optimal to reduce the line-width.

In Fig. 3 the line-shapes versus the detuning θ are represented for two different values of power and for different n values. It can be seen that with the optimal power level the fringe envelope of the cases $n = 7$ is almost coincident with the line-shape of the case $n = 1$.

From the line-shapes obtained in this way, for various n values, cavity asymmetries and microwave power conditions, the Doppler shift is evaluated as the displacement of the line center. Such a displacement can be inferred by two different procedures. Firstly it can be estimated from the central value of the line-width at half maximum height, as a second method it can directly be estimated by exploring the output values of $C_2C_2^*$ in the neighbourhood of the maximum, by varying θ. The two methods lead to shift evaluations very close to each other and differ only for power level over the optimal value.

The shift turns out to be dependent on power and with different sensitivity to the asymmetry parameters of the cavity as reported in [2], but the most influencing asymmetry factor is χ', on whose value the shift is linearly dependent in complete agreement with the analytical survey.

The comparison with the analytical results points out that there is a power dependence as expected but that the first analytical iteration is not sufficient to satisfactory report such a dependence. In Fig. 4 the different results of the Doppler shift power dependence in the case of $n = 1$, $n = 3$, and $n = 7$ are reported showing that, as expected, the analytical expression are valid mostly for low power level.

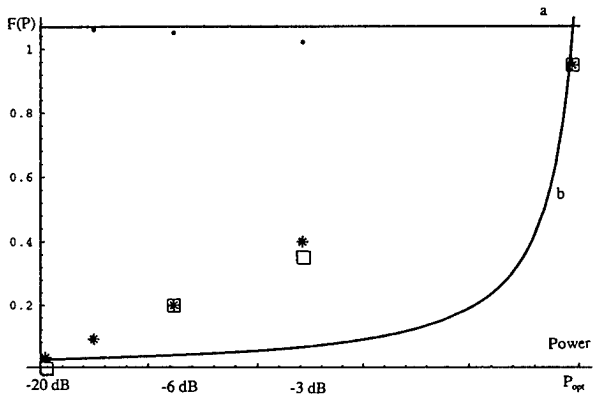


Fig. 4: Doppler shift power dependence as obtained from numerical integration and from the first analytical iteration.
a) analytical (• numerical) estimation for $n = 1$
b) analytical (* numerical) estimation for $n = 3$
□ numerical estimation for $n = 7$.

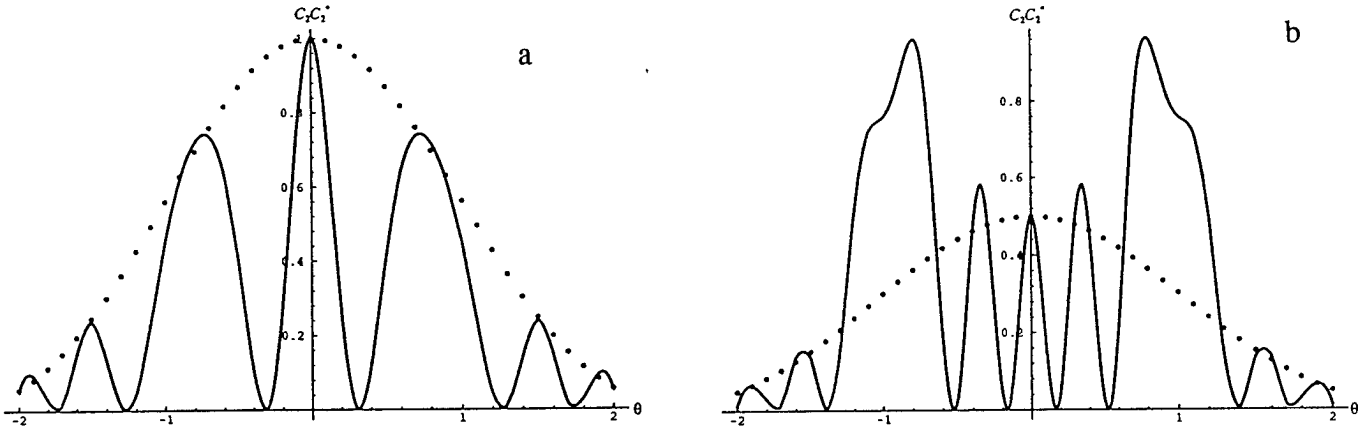


Fig. 3 a): Central fringes of the transition probability for $n = 1$ cavity (dotted line) and $n = 7$ (solid line) at optimal power level.
b): The same as a) at power level 6 dB below optimal.

Conclusions

In this paper an analysis is reported of the transition probability for an atomic beam excited by resonant microwave radiation in a TE_{01n} cavity fed in the center. Both the results of numerical integration of the Schrödinger equation, and an analytical approximation obtained by means of a Magnus expansion are given. The agreement is satisfactory up to the $n = 7$ case, which is the highest order examined. No asymmetry was assumed in the center feed, mainly because its effect can be shown to be very small for a cylindrical cavity, since the TE_{01n} mode in that kind of structure has very low losses on the side walls. A side effect of this consideration is that there is no need for highly precise centering of the feed in a cavity based on that mode. The analysis would therefore need to be completed with consideration of the length asymmetry, in order to apply the theory to a different mode, like a TE_{02n} mode in a cavity with rectangular cross section, which might be interesting for a short commercial tube.

References

1 A. De Marchi, R. E. Drullinger, J. H. Shirley, "Interference fringes from single cavity excitation of an atomic beam", in Proc. 44th Frequency Control Symp., pp. 34-38, 1990.

2 E. Bava, P. Tavella, A. De Marchi, "Doppler effects in a multi-wavelength Rabi cavity for the excitation of atomic beams in microwave frequency standards", in Proc. 46th Frequency Control Symp., pp. 81-85, 1992.

3 A. De Marchi, E. Bava, P. Tavella, "Accuracy analysis of the atom-microwave interaction in a TE_{017} cylindrical Rabi cavity for a high C-field Cesium beam resonator", to appear in Proc. 7th European Time and Frequency Forum, 1993.

4 R. E. Collin, *Field theory of guided waves*, IEEE Press, New York, 1991.

5 W. Magnus, "On the exponential solution of differential equations for a linear operator", Commun. Pure Appl. Math. 7, p. 649, 1954.

6 E.Bava, A.G.Shahian, "Interaction between atom and radiation with sinusoidally shaped amplitude: application to frequency standards" Appl. Phys. B 54, pp. 500-505, 1992.

7 J. Vanier, C. Audoin, *The quantum physics of atomic frequency standards*, Adam Hilger, Bristol, Philadelphia; pp. 636-647, 1989.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Atomic Phase Delay in a Rubidium Atomic Clock

J. C. Camparo and R. P. Frueholz
M2-253

Electronics Technology Center
The Aerospace Corporation
P. O. Box 92957
Los Angeles, CA 90009

Abstract

In a rubidium atomic clock, phase or frequency modulation of the probing microwave signal induces atomic population oscillations at the same frequency. These population oscillations result in transmitted lamp intensity oscillations, which are in turn used to generate the error signal that stabilizes the frequency of the probing microwave field. In this study we examine the phase-delay that exists between the oscillating atomic system and the probing microwave field. This phase-delay must be accounted for in the feedback control circuitry of the clock in order to maximize the device's short-term stability. Our motivation is a better understanding of the effects that can slowly degrade atomic clock performance over long, multiyear operational lifetimes, in particular slow microwave power variations. We find both experimentally and theoretically that the atomic phase-delay can vary by more than 100°, and that in the regime of normal clock operation it is a rapidly changing function of microwave power. Defining P_o as the normal operating power entering the clock cavity, our results indicate that at P_o the phase-delay sensitivity to microwave power variations is maximized and on the order of ten degrees per dB. Moreover, our results show that it is unadvisable to try to operate the clock at some different microwave power where this sensitivity is smaller, since the amplitude of the first harmonic signal is also maximized at P_o. However, with regards to multiyear Rb clock operation, so long as microwave power level changes do not exceed roughly 3 dB there should be little effect on the clock's short-term stability.

Introduction

When an atomic system interacts with a resonant field whose amplitude or phase is modulated, it is well known that the atomic population among various eigenstates will oscillate at the fundamental modulation frequency and its harmonics. It is also true, however, that the atomic oscillations occurring at the fundamental frequency will have a phase lag with respect to the driving field, and that this atomic phase-delay must be accounted for in the feedback control loop of an atomic clock in order to optimize the clock's short-term stability. Our primary motivation in the present study is a better understanding of the effects that can slowly degrade atomic clock performance over long, multiyear operational lifetimes. Specifically, if conditions were to permit the atomic phase-delay to change slowly with time, short-term performance could be degraded. In the following sections of this paper, we will describe both our theoretical and experimental investigations of this issue.

Overview of Rb Atomic Clock Operating Principles

Figure 1 illustrates some of the generic Rb atomic clock elements. The output of a voltage-controlled-crystal-oscillator (VCXO) is multiplied to the 10^2 MHz range, and is then subjected to phase modulation at a 10^2 Hz rate. This phase modulated signal is applied to a Step Recovery Diode, creating an X_B band microwave signal, which excites the resonant mode of a microwave cavity containing a vapor of Rb atoms in a resonance cell. When the microwave frequency is reasonably close to the 0-0 atomic hyperfine resonance (within ~ 1 kHz), the phase modulation of the microwaves causes the atomic population imbalance between the 0-0 hyperfine levels (created by optical pumping) to oscillate, and this oscillation in turn causes a modulation of the light intensity monitored by the photodetector. Since the optical modulation amplitude is proportional to the detuning of the microwaves from the atomic resonance frequency,[1] the IF output of the mixer shown in Fig. 1 can provide an error signal to stabilize the VCXO's rf frequency, so long as the phase difference between the LO and RF inputs is

0-7803-0905-7/93 $3.00 © 1993 IEEE

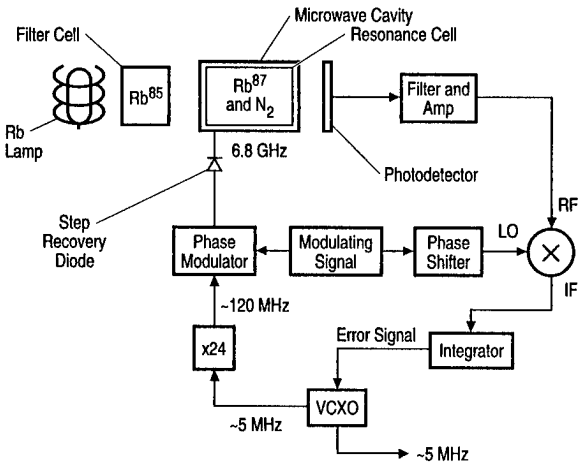


Fig . 1: Schematic diagram of typical rubidium atomic clock.

not close to 90^{o}.

Theoretical Analysis

In the present work we are concerned with the phase offset between the modulated signals at the LO and RF inputs to the mixer of Fig. 1 caused by atomic processes occurring within the resonance cell. To investigate this question we consider sinewave modulation of the phase:

$$\Theta(t) = a \sin(\omega_m t)$$

Here, $\Theta(t)$ is the microwave signal's phase, a is the modulation index, and ω_m is the modulation frequency; note that sinewave phase modulation implies cosinewave frequency modulation. The LO signal that is input to the mixer may then be written as:

$$S_{LO} = a \sin(\omega_m t + \phi + \pi/2) = a \cos(\omega_m t + \phi)$$

where $(\phi + \pi/2)$ is the phase added by the phase shifter of Fig. 1. The RF input comes from the photodetector, and we write this as:

$$S_{RF}(t) = A \cos(\omega_m t + \psi_1) + B \cos(2\omega_m t + \psi_2),$$

where A and B are the signal amplitudes of the 1st and 2nd harmonic components of the RF signal, respectively, and ψ_1 and ψ_2 are the phase-delays induced by the atomic system for the 1st and 2nd harmonic components, respectively. We again note that for X_B band radiation within about 1 kHz of the atomic resonance, A is proportional to the frequency difference between the microwaves and the atomic resonance. The choice of a cosine function for the RF signal, rather than a sine function, is motivated by the fact that in the limit of very high microwave power the atomic system should be nearly in phase with the microwave field's frequency modulation, which is the derivative of the phase modulated signal.

Considering the case of no intermodulation products at the mixer output, the IF signal is the simple product of the RF and LO signals:

$$S_{IF}(t) = (Aa/2)[\cos(\phi - \psi_1) + \cos(2\omega_m t + \phi + \psi_1)] +$$

$$(Ba/2)[\cos(\omega_m t + \psi_2 - \phi) + \cos(3\omega_m t + \psi_2 + \phi)].$$

Averaging (via the Integrator of Fig. 1) $S_{IF}(t)$ over some time interval that is long compared to the modulation period then yields an expression for the error signal that is applied to the VCXO:

$$\text{Error Signal} = \langle S_{IF} \rangle_\phi = (Aa/2) \cos(\phi - \psi_1).$$

Clearly, the error signal is maximized, and the greatest clock stability is attained, when the phase offset provided by the Phase Shifter ϕ cancels the phase-delay associated with the atomic system, ψ_1.

To evaluate ψ_1, and hence the required value of ϕ, it is necessary to understand the dynamical behavior of an atomic quantum system interacting with a phase modulated field. Solutions to this problem have been obtained by various workers in the limit of weak microwave fields.[2] However, since rubidium atomic clocks typically operate in the regime of strong fields, where they achieve their greatest stability,[3] these previous analyses are of uncertain value for the determination of ψ_1. In strong fields the determination of ψ_1 necessitates the solution of the density matrix equations. Assuming a 2-level, homogeneously-broadened atomic system, these equations may be written as:

$$\dot{X} = -\gamma_2 X - \Delta Y + \Omega Z \sin[\Theta(t)]$$

$$\dot{Y} = \Delta X - \gamma_2 Y + \Omega Z \cos[\Theta(t)]$$

$$\dot{Z} = -\Omega X \sin[\Theta(t)] - \Omega Y \cos[\Theta(t)] + \gamma_1 (Z_0 - Z)$$

Here, X and Y are related to atomic coherence, while Z is the population difference between the upper and lower hyperfine states; γ_1 and γ_2 are the longitudinal

and transverse relaxation rates, respectively; Ω is the Rabi frequency (proportional to the square root of the microwave power level) and Δ is the field detuning. Nominal values of these parameters for a Rb clock are collected in Table I.

Table I: Values of the parameters required by the density-matrix equations.

Parameter	Value
γ_1	311 Hz
γ_1	311 Hz
ω_m	451.7 Hz
a	0.9

Once a means of calculating the atomic dynamics is in hand, and the appropriate input parameters are known, it is a relatively straightforward task to determine ψ_1. We numerically solve the density matrix equations using a fourth-order Runge-Kutta technique with adaptive step-size,[4] and calculate the atomic population in the F=2 ground state hyperfine level, which is proportional to the light intensity reaching the photodetector, as a function of time. To ensure a non-zero first harmonic component to the atomic signal, this calculation is carried out at a microwave detuning of -10 Hz. The atomic population at time t is then multiplied by the corresponding microwave phase $\Theta(t)$, and the product is averaged over some very long time (typically a hundred modulation periods) to yield $\langle S_{IF}\rangle_\phi$. Two calculations are performed, one with ϕ=0, and the other with an arbitrary value of ϕ (we take $\phi=10^o$). Forming the ratio r,

$$r = \langle S_{IF}\rangle_0/\langle S_{IF}\rangle_\phi = \cos(\psi_1)/\cos(\phi-\psi_1),$$

then allows a simple evaluation of ψ_1:

$$\psi_1 = \arctan[r^{-1}\csc(\phi) - \cot(\phi)].$$

Once ψ_1 is known it is easy to calculate the amplitude of the first harmonic signal A, and a similar analysis can be performed independently to determine ψ_2 and B.

Figures 2a and 2b show the results from these calculations for the first and second harmonic components, respectively. Here, circles correspond to the computed values of ψ_1 and ψ_2, while the solid lines represent the relative amplitudes of the IF signal; both of these quantities are plotted as a function of Rabi frequency. Considering Fig. 2a, note that the normal operating power of this clock, P_o, would be set so that $\Omega \cong 500$ Hz, since this is where the first harmonic signal is maximized. An

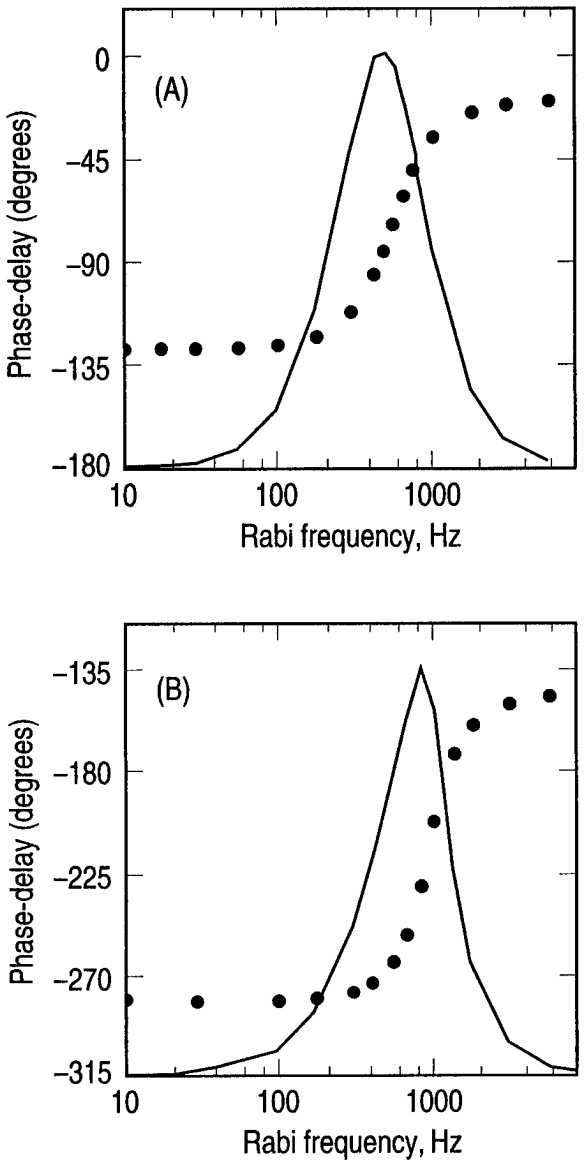


Fig. 2: Theoretical results. (a) Phase-delay of atomic signal's first harmonic versus Rabi frequency. Circles correspond to the calculated phase-delay data. The solid line is the relative amplitude of the first harmonic signal at a detuning of - 10 Hz. (b) Phase-delay of atomic signal's second harmonic versus Rabi frequency. Circles correspond to the calculated phase-delay. The solid line is the relative second harmonic amplitude, which was nearly constant for small detuning.

important point worth recognizing from the results is that the phase-delay changes by about 10^2 degrees as a function of Ω. Moreover, even at relatively high Rabi frequencies, where one might expect $\psi_1 \cong 0$ due to the adiabatic nature of the field-atom interaction,[5] the calculations predict a non-negligible atomic phase-delay of -20°. Thus, for optimum performance ϕ can never be set to zero. Finally, the calculations show that the atomic phase-delay ψ_1 at P_o, and hence the required phase shift ϕ to maximize clock stability, is a rapidly changing function of microwave power. In particular, the calculations indicate a microwave power coefficient for the phase-delay ($d\psi_1/d\log[P]$) of 8.9°/dB at P_o. In order to reduce any potential deleterious effects of microwave power variation on clock performance, one would want this coefficient to be as small as possible. Regarding Fig. 2a, one might therefore be tempted to operate the clock at either higher or lower values of the microwave power. However, any attempt to reduce or increase the microwave power in the cavity would lead to a decrease in the amplitude of the first harmonic signal and hence the clock performance. It thus appears from these calculations that a relatively high sensitivity of atomic phase-delay to microwave power is a necessary condition of atomic clock design.

Considering Fig. 2b, note that the phase-delay of the second harmonic signal ψ_2 also shows a rapid variation in the microwave power regime of normal clock operation P_o. Further, the second harmonic signal amplitude is a maximum in this regime. Note, however, that the microwave power level that maximizes the second harmonic signal amplitude B, is not exactly equal to the microwave power level that maximizes the first harmonic signal amplitude A; there is an approximately 5 dB difference in the optimum values of the microwave power. Consequently, if one were to use the second harmonic signal to optimize microwave power, one would need to choose a microwave power somewhat less than that which yielded the sharp maximum in the second harmonic signal's amplitude.

Experiment

While suggestive, the calculations presented above are an idealization of the actual physics that occurs in the gas-cell atomic clock. In particular, the calculations take no account of what is sometimes referred to as the *macroscopic scale* of physics.[6] The Rb atoms confined in the resonance cell of Fig. 1 are also in the presence of a nitrogen/noble-gas mixture. The buffer gas essentially "freezes" the rubidium atoms in place so that they only interact with microwaves and light in some spatially localized region.[7] Different Rb atoms in different regions of the resonance cell experience different microwave power levels, and would therefore be at different places on the curve of Fig. 2. As a result of this macroscopic scale of physics, an experimental examination of the atomic phase-delay is warranted.

Our experimental apparatus is illustrated in Fig. 3. Basically, a natural Rb vapor was contained with ~ 10 torr N_2 inside a cylindrical microwave cavity whose TE_{111} mode was resonant with the Rb ground state hyperfine transition. The cell temperature was maintained at 72° C, corresponding to a rubidium density of 1.4×10^{12} cm^{-3},[8] and an isotopically enriched Rb discharge lamp was used for optical pumping. The cavity and resonance cell were located inside a static magnetic field of 310 mG oriented along the cavity's axis of symmetry; this field defined the atomic quantization direction. The output of a VCXO at approximately 107 MHz, was multiplied up into the gigahertz range before being amplified by a 30 dB solid state amplifier. The microwave power entering the cavity could be controlled with variable attenuators, and these were calibrated to microwave Rabi frequency by measuring the linewidth of the hyperfine transition;[9] additionally, the dephasing rate was found through these linewidth measurements to be 311 Hz. The control voltage to the VCXO was sinusoidally modulated at 451.7 Hz, with the crystal detuned from resonance by -102 Hz. A photodiode was located inside the cavity to detect the lamp light transmitted by the resonance cell, and its output was directed through a pre-amp to an SRS 510 lock-in amplifier. Phase measurements of ϕ were made by zeroing the quadrature component of the lock-in output, referenced to the frequency modulating signal.

Figure 4 shows the experimental results of the lock-in phase offset ϕ and corresponding maximum first harmonic signal amplitude as a function of Rabi frequency. Quantitatively, the agreement between theory and experiment is quite good. Experiment confirms that the atomic phase-delay changes by about 10^2 Hz as a function of Rabi frequency, and

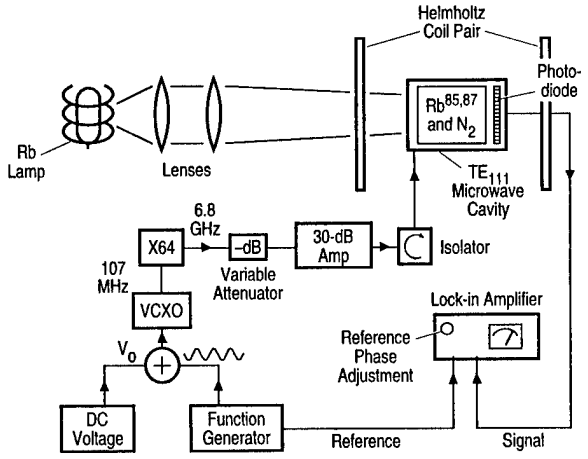


Fig. 3: Experimental apparatus described in the text that was used to measure the atomic phase-delay in an optically pumped Rb vapor.

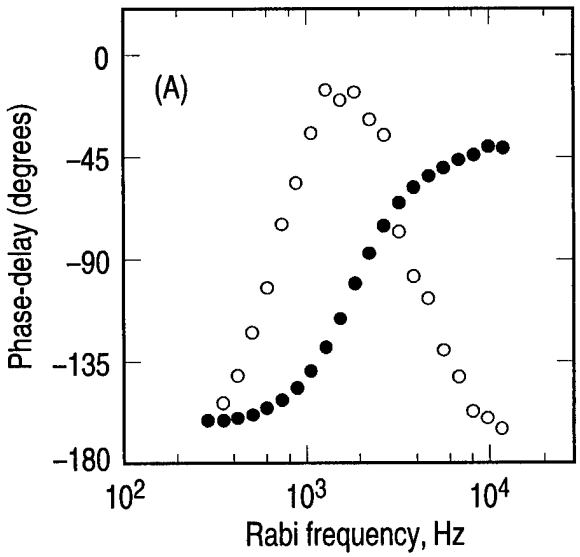


Fig. 4: Experimental results of atomic signal's first harmonic phase-delay versus Rabi frequency. Circles correspond to the measured phase-delay. The diamonds are relative amplitude of the first harmonic signal.

that in the limit of large Rabi frequency the phase-delay asymptotes to a nonzero value of ~ -40°. Moreover, at Rabi frequencies where the first harmonic component of the signal is maximized, $d\phi/d\Omega$ is also maximized. The experimental value of $d\phi/d\log[P]$ at P_o is found to be 7.1°/dB in good agreement with the theoretical result of 8.9°/dB. As a whole these results both confirm the theoretical predictions discussed in the previous section, and suggest that with regard to phase-delay calculations the approximation of a 2-level, homogeneously broadened atom is quite good.

Discussion

In terms of understanding atomic clock stability, and its sensitivity to microwave power changes, a significant result of the present study concerns the value of $d\psi_1/d\log[P]$ at P_o. Typically, operational atomic clock microwave power levels are chosen so that the atomic dynamics is just within the saturation regime. This power level yields both a large signal amplitude and a relatively narrow atomic resonance line, and previous results have shown that the shot-noise-limited Allan variance (σ_y^{shot}) is relatively insensitive to microwave power changes of less than about 10 dB (i.e., $d\sigma_y^{shot}/d\log[P] \cong 0$ at P_o).[3] Consequently, just considering shot-noise-limited performance, one would conclude that likely microwave power changes occurring over a Rb clock's multiyear life would have no effect on short-term performance.

However, σ_y should actually be seen as $\sigma_y^{shot}/\cos(\phi-\psi_1)$, which results in an additional sensitivity of short-term performance to microwave power changes. In particular, assuming an atomic phase-delay sensitivity of 9°/dB, a 10 dB change in microwave power would cause a 90° phase shift and an unlocking of the clock. Therefore, in order to ensure specified short-term performance over an atomic clock's multiyear mission life, it is more important to consider the effects of microwave power changes on ψ_1 than on σ_y^{shot}. Specifically, in order to ensure that σ_y does not change by more than about 10%, we would recommend specifying that microwave power changes be no larger than 3 dB over Rb clock operational lifetimes.

References

[1] J. Vanier and C. Audoin, The Quantum Physics of Atomic Frequency Standards, Vol. 2: Adam Hilger, 1989, Ch. 7.

[2] See for example: R. Karplus, "Frequency Modulation in Microwave Spectroscopy," Phys. Rev., vol. 73, pp.1027-1034, 1948; O. Haworth and R. E. Richards, "The Use of Modulation in Magnetic Resonance," Prog. Nucl. Mag. Res. Spectrosc., vol. 1, pp. 1-14, 1966.

[3] J. C. Camparo and R. P. Frueholz, "A Three-Dimensional Model of the Gas Cell Atomic Frequency Standard," IEEE Trans. Ultrason. Ferroelec. Freq. Contr. vol. UFFC-36, pp. 185-190, 1989.

[4] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, Numerical Recipes: Cambridge University Press, 1989.

[5] See for example: J. C. Camparo and R. P. Frueholz, "Parameters of adiabatic rapid passage in the 0-0 hyperfine transition of Rb^{87}," Phys. Rev. A, vol. 30, pp. 803-811, 1984; R. P. Frueholz and J. C. Camparo, Phys. Rev. A, vol. 47, pp. 4404-4412, 1993.

[6] J. C. Camparo, "Advances in the theory of gas-cell atomic frequency standards," in Frequency Standards and Metrology: Springer-Verlag, 1989, pp. 62-67.

[7] R. P. Frueholz and J. C. Camparo, "Microwave Field Strength Measurement in a Rubidium Clock Cavity Via Adiabatic Rapid Passage," J. Appl. Phys., vol. 57, pp. 704-708, 1985.

[8] T. J. Killian, "Thermionic Phenomena Caused by Vapors of Rubidium and Potassium," Phys. Rev., vol. 27, pp. 578-587, 1926.

[9] J. C. Camparo and R. P. Frueholz, "Linewidths of the 0-0 Hyperfine Transition in Optically Pumped Alkali-Metal Vapors," Phys. Rev. A, vol. 31, pp. 1440-1448, 1985; J. C. Camparo and R. P. Frueholz, "Saturation of the 0-0 Hyperfine-Transition Linewidth Enhancement Factor in Optically Pumped Alkali-Metal Vapors," Phys. Rev. A, pp. 1888-1889, 1985.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

LIFE, LINE Q, SPIN-EXCHANGE TUNING AND STABILITY OF ATOMIC HYDROGEN MASER FREQUENCY STANDARDS

H. B. Owings, P. A. Koppang, C. C. MacMillan, and Harry E. Peters

Sigma Tau Standards Corporation
1711 Holt Road, P.O. Box 1877
Tuscaloosa, Alabama 35403-USA

Abstract

A new and substantially different design of field operable atomic hydrogen maser was developed at Sigma Tau Standards Corporation (STSC) over nine years ago (1982-1984). This maser incorporated the first successful automatic cavity stabilization system for active oscillator hydrogen masers which did not require a second stable frequency source as a reference. The design also realized several other innovations, including very high hydrogen utilization efficiency, as well as relatively small size and low power consumption. The new design promised extended operating life with improved long term stability while maintaining the characteristic good short term stability of the classic hydrogen maser oscillator. Data has been accumulating over the years since the first deliveries began in 1985; this, as well as new data presented herein, shows that the STSC design of hydrogen maser may be expected to provide stable performance for periods of time measured in decades, without long term interruption for recoating of the storage bulb, changing of pumps or replenishment of hydrogen.

Introduction

A time and frequency standard should not only have exceptional stability, but it should also operate continuously and have a very long life. It is particularly noteworthy that the first developmental hydrogen masers to use the new cavity tuning servo and other features of the STSC hydrogen maser design are still being used as frequency standards, and the long term stability, over approximately eight years of continuous operation, has been favorably compared to the stability of the international time and frequency standards systems [1,2,3].

There are 36 STSC hydrogen masers presently operational. Thirty of these are in use at standards laboratories or radio astronomy stations and six are complete and in the final testing phase at STSC. Figure 1 gives the installation schedule and users of these masers. A great deal of stability data, much of it provided by independent laboratories, has previously been documented [1]. One of the most important questions remaining is how long will the masers continue to operate without major maintenance of the basic frequency determining systems, e.g. those which are contained within the vacuum system and which, if changed, would possibly cause a discontinuity in the standard output frequency of the maser.

The most important factor in fundamental hydrogen maser stability and reproducibility over very long periods of time is the condition of the Teflon coated quartz storage bulb within the cavity of the maser. If the bulb surface becomes contaminated with material that changes the wall shift or causes recombination of the hydrogen molecule, the basic output frequency may change, or the maser may even stop oscillating. With approximately 130 maser years of operation and no recalls for recoating of the storage bulb, it does not appear that bulb surface degradation is a serious life limiting factor. Line Q measurements on some older masers presented later in this paper confirm that the bulb surface is not degrading noticeably with time.

Another very important factor is the long term stability of the cavity automatic tuning system. Spin exchange tuning measurements on several of the older STSC masers, along with extensive independent measurements of long term stability, show that the

0-7803-0905-7/93 $3.00 © 1993 IEEE

tuning system is very reliable and stable. Other important factors, which could interrupt maser operation, are replacement of the vacuum pumps or refilling of the hydrogen supply bottle. As shown further on in this paper, the pump life and hydrogen utilization efficiency are extremely good in these masers. No pump elements have been replaced due to failure, or hydrogen bottles refilled to date, and in any event, they may be replaced or refilled without accessing the main maser vacuum or causing a later change in the output frequency of the maser.

Essentially all other problems that might occur and cause interruption of the maser, such as malfunction of electronic systems, are external to the vacuum system and relatively easy to repair. Upon repair of such systems the maser standard output frequency returns to the unperturbed value in a relatively short time with high reproducibility.

STSC Atomic Hydrogen Maser Description

The Sigma Tau Standards Corporation hydrogen maser is an active oscillator, with a natural output frequency of 1,420,405,751.xxx,x Hz, derived from quantum transitions between two of the four magnetic "hyperfine" levels of the ground electronic state of atomic hydrogen. This is the well known 21 cm line of atomic hydrogen. Figure 2 shows the general physical configuration of the maser and Figure 3 is a block diagram illustrating the physical and electronic systems.

A small storage bottle supplies molecular hydrogen under electronic servo control to the source discharge bulb where the molecules are dissociated into atoms. The atoms emerge from the source through a small elongated hole, the source collimator, and then pass through a magnetic "state selector" that directs a beam of atoms in the correct quantum state to a Teflon coated quartz storage bulb.

The storage bulb is located within a microwave cavity resonant at the hydrogen transition frequency and maser action results in CW emission. A portion of the signal is coupled from the cavity and passes to a high resolution receiver/synthesizer system. To provide the proper environment for maser action and minimize systematic perturbations, there are several levels of thermal control, four levels of magnetic shielding and two differentially pumped sections in the vacuum system maintained by sputter-ion pumps.

The maser has an automatic frequency control system of the cavity resonant frequency switching type to reduce cavity pulling of the maser output [4]. This cavity servo requires no other stable frequency source as a reference (a requirement in other types of maser automatic cavity tuning systems). Most importantly, the short term stability is not significantly degraded, while long term drift due to cavity pulling has been substantially corrected [5]. Organizations requiring the best long term stability and reproducibility, in addition to the ease of rapid, noise free comparisons with other standards, will find a maser with this auto tuning system a great improvement over other systems.

Line Q - Method

The primary basis for the exceptional stability and reproducibility of the atomic hydrogen maser is the large atomic line Q (Q_l), on the order of 2×10^9, which is derived from the large storage time of the atom within the storage bulb and the low interaction energy of the atom with the Teflon surface of the bulb [6]. There are several mechanisms affecting Q_l: the most basic limitation is the quality of the Teflon coating of the storage bulb which affects the recombination rate of atoms to molecules on the surface or decorrelation through phase shifting wall interactions [7]; then there is the escape rate of the atoms through the bulb entrance collimator, possible degradation by non-homogeneous magnetic fields and change in Q_l due to spin exchange collisions between atoms [8].

It is not the intent of this paper to analyze Q_l or atomic perturbation mechanisms, rather it is the stability of Q_l over long periods of time that is important here. With constant or slowly varying Q_l, in conjunction with long term stability data on the masers, it can be inferred that the bulb wall is not changing, or is changing slowly, and maser performance or operating lifetime is not being degraded by changes in the storage bulb.

The method used for measuring Q_l on the STSC masers is illustrated with the help of the following equation:

$$(F_m - F_o)/F_o = (Q_c/Q_l) \times (F_c - F_{co})/F_o. \qquad \{1\}$$

Here Q_c is the cavity Q, Q_l is the line Q, F_m is the maser oscillation frequency, F_o is the maser oscillation frequency when tuned, F_c is the cavity frequency and F_{co} is the cavity frequency when tuned. (Due to spin exchange interactions, which pull the maser frequency

in a direction opposite to cavity pulling, F_{co} is not equal to F_o.) To measure Q_l, the cavity frequency is varied and the resultant change in maser output frequency is measured. Since Q_c varies little and is well know from cavity measurements, Q_l is easily calculated.

Spin Exchange Tuning Method

The most important environmental perturbation to be overcome to achieve long term stability in hydrogen masers is the cavity pulling effect. The ratio, Q_c/Q_l, (Equation 1) has a typical value of about 2×10^{-5}; this low value explains why the hydrogen maser is so much more stable than oscillators based directly on a resonator, such as a cavity or quartz crystal, where the oscillation frequency is directly dependent on the resonance frequency. Still, after the quality of the bulb coating is assured, the primary long term perturbation to be addressed in the oscillating hydrogen maser is the cavity pulling effect.

In the STSC hydrogen masers, the automatic cavity tuning system stabilizes the cavity frequency. Spin exchange tuning techniques, as discussed in references [1] and [3], are used to set the cavity frequency initially so that $(F_c - F_{co})$ is close to zero, and then $F_m = F_o$, the unperturbed maser frequency. Later measurements of the "pressure shift" (a change in hydrogen beam flux which changes Q_l through spin exchange interactions) indicate whether the cavity auto-tuner has maintained the cavity at the spin exchange tuned position. The measurement is carried out consecutively with measurement of Q_l using Equation 1.

Line Q and Spin Exchange Tuning Measurements

It is not often that the opportunity arises to measure the hydrogen maser parameters in older STSC hydrogen masers, since they are in continuous operation at distant locations. Four masers have been checked recently. Two masers, "N-1" and "N-2" have been in continuous operation at the U. S. Naval Research Laboratory since they were delivered in September, 1985. At the invitation of NRL, one new STSC maser was brought from Tuscaloosa, Alabama to NRL on the occasion of the 1992 PTTI Meeting and used as a reference against which the NRL masers could be compared. The use of a third maser minimized the interruption of the NRL masers and provided the extra short term stability provided by the most recent STSC masers.

One other maser, NAV-1, was brought back to STSC in February, 1993 for testing after surviving hurricane Andrew at the USNO Florida substation. The maser was completely functional and stable soon after arrival and the Line Q and spin exchange offset were then checked. Another USNO hydrogen maser, NAV-6, was present at STSC for checking prior to being sent to the KOKEE Park Geophysical Observatory, Kauai, Hawaii in February, 1993. The overall results are given below.

Maser	Line Q	SE Offset Correction	Elapsed Time Years
N - 1	1.75×10^9	$+ 1.1 \times 10^{-13}$	7.5
N - 2	1.70×10^9	$+/-3.4 \times 10^{-14}$	7.5
NAV-1	1.78×10^9	$+/-1.0 \times 10^{-14}$	1.8
NAV-6	1.77×10^9	$+/-0.5 \times 10^{-14}$	1.0

The spin exchange offset (SE Offset) change from the value set at the beginning of the elapsed time period given above is the total maser output frequency correction calculated from the pressure shift and has an uncertainty of about $+/-1 \times 10^{-14}$ due to the combined instabilities of both the maser being tested and the comparison maser at both the beginning and end of the elapsed time period. Within the error of the measurements N-1 was the only maser requiring correction of the spin exchange offset.

The above results, as well as measurements on many masers during construction and initial test, indicate that the cavity tuning system is not responsible for the small amount of initial drift of the maser output frequency usually observed on new STSC hydrogen masers. When delivered, the drift of the maser is less than 2×10^{-15} $+/- 1 \times 10^{-15}$/day and is increasing in frequency. Later data from organizations using the masers indicate that the drift slows down with a time constant of about 6 months to a year[1].

This initial frequency drift is believed to be due to chemical clean up of the Teflon wall surface by the atomic hydrogen or to residual outgassing of adsorbed contaminants from the wall. The relative constancy of the line Q's implies that the bulb is not being contaminated with vacuum background substances.

Vacuum Pumps

One of the concerns with many other hydrogen masers is that the hydrogen pumps become saturated

and the vacuum system needs to be vented to atmospheric pressure to replace the pumping elements every few years. This invariably entails a lengthy period for evacuating the maser after the pump change and waiting for the temperature controls to balance, and it also becomes necessary to spin exchange tune the maser to recover stability and reproduce the previous output frequency.

Experimentally and theoretically, the hydrogen flow rate is exceptionally low in STSC hydrogen masers. The typical total current in the ion pumps is between 80 and 120 microampers, which is between 10 and 20 times lower than typical in other designs [9,10]. This pump current corresponds to approximately 0.01 moles of molecular hydrogen per year. The 20 liter/sec ion pumps used in the current masers will absorb approximately 0.6 moles of molecular hydrogen by previous pump tests [9], and will probably absorb more than this due to the very slow rate of hydrogen diffusion into the pump cathodes in the present design. On this basis it is expected that the present hydrogen pumps should last on the order of 60 years or more.

The present experimental evidence on pump lifetime does not contradict the above estimate [10]. The hydrogen pump on one of the first masers of the present type to be delivered (Maser N-1, presently at the U. S. Naval Research Laboratory) has only a ten liter/second hydrogen pump and has been operating continuously since early 1985 with no sign of pump saturation. On that basis it is expected that the 20 liter/second pumps used on later masers should last at least 16 years. In view of the fact that the pumps are readily changed it is clear that pump lifetime is not a problem for long term frequency continuity.

Hydrogen Supply

Refilling of the hydrogen bottle has also been required every few years in past designs of hydrogen masers. The hydrogen supply in the present masers is contained in a 75 cc stainless steel bottle at a maximum pressure of 250 PSI. The hydrogen is absorbed in an alloy of nickel and cerium free mishmetal which initially contains two moles of molecular hydrogen. No hydrogen bottle has ever had to be refilled on the STSC hydrogen masers due to exhaustion of the supply. At 0.01 moles per year there is enough hydrogen to last approximately 200 years. The bottle may also be refilled without stopping the maser oscillation.

Stability

Constancy of line Q and accuracy of tuning out cavity and spin exchange related variations do not by themselves indicate lack of drift of the maser output frequency. There are several other significant perturbations to the maser oscillation frequency, namely the 2nd order Doppler shift (special relativistic time dilation), the magnetic field dependency and the wall shift, as well as relatively smaller higher order factors. The second order Doppler effect depends on the velocity of the atoms in the storage bulb, and so ultimately on the temperature, with a temperature coefficient of approximately $1.5 \times 10^{-13}/^{o}C$. Since the temperature is controlled to a stability of approximately $10^{-5}{}^{o}C$/day to obtain the typical daily frequency stability, the Doppler shift is not a significant contributor to instability.

The magnetic field stability is also not a serious limitation to stability, either long term or short term. The typical operational Z-axis field in the maser interaction region is 500 microgauss in the STSC masers, and the frequency sensitivity to external field variations is typically on the order of 1×10^{-14}/gauss or less. The variation of the "wall shift" with time appears to be the main contributor to long term drift in these masers. The total wall effect is only about 3×10^{-11} and does not contribute to short term instability. However, at the long term stability levels presently being achieved due to the success of the automatic cavity tuning system, the slow change in the wall shift over the first one to two years of maser operation appears to be the main limitation to long term stability.

However, after the first year or so of operation, the typical apparent drift is so small that it can only be estimated by comparison, via GPS, radio astronomy, or other precise methods, with the international time and frequency system over several months of observation. Each maser is typically more stable than the usual selected commercial passive standard clock ensemble presently used in the majority of standards laboratories. This performance is best illustrated by data obtained on several STSC hydrogen masers used at the United States Naval Observatory. (The original data used in the following figures was obtained courtesy of the USNO [11] but the final plots and interpretation are solely the responsibility of the present authors.)

USNO Masers NAV-1 -- NAV-8

Some of the data given below was included in a paper by the present authors given at the 1992 Frequency Control Symposium [1]. The performance illustrated is qualitatively continuing at this writing [11] and fully illustrates one of the main points of the present paper. Eight STSC hydrogen masers have been delivered to the USNO. Five of these are presently used in conjunction with the other standards involved with the USNO time and frequency standards operation, the other three are located more remotely at USNO radio astronomy installations involved with geodesy, polar motion and similar applications.

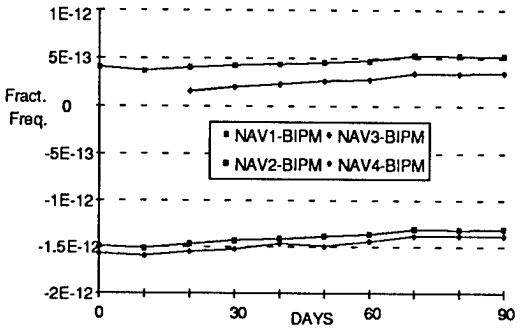


Figure 4. Frequency Stability of STSC Masers vs. BIPM (November 29, 1989 to February 27,1990)

Figure 4 shows the relative frequency of four masers during a 90 day period starting in November, 1989, versus the BIPM [1,2,3]. The frequency starting points on the vertical axis are synthesized and arbitrary. All four of the masers are increasing smoothly in frequency relative to the BIPM by approximately two parts in 10^{15} per day, while the masers are varying between themselves by one part in 10^{15} or less.

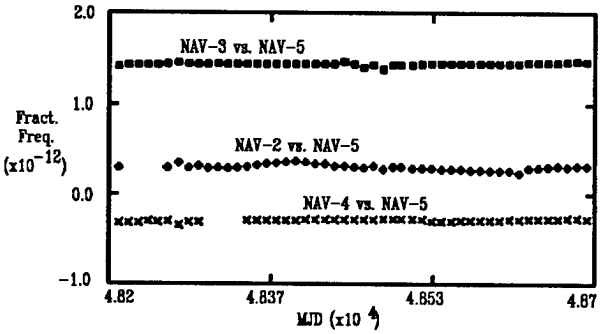


Figure 5. Frequency Plots, USNO - STSC Masers (Synthesizer frequency and phase changes removed)

Figure 5 shows the relative frequency of four STSC masers for 500 days beginning November, 1990. The drift between the masers is almost unnoticeable in this figure and calculations indicate that the variation is on the order of one part in 10^{16} per day or less.

Maser	Drift - Fractional Frequency
NAV - 2	+9.4 x 10-17/day
NAV - 3	+0.4 x 10-17/day
NAV - 4	-6.4 x 10-17/day
NAV - 5	+9.4 x 10-17/day

Indicated above is the drift of the four masers shown in Figure 5 relative to the BIPM for a period of 170 days ending in April, 1992, as reported by the USNO. The error in the drift between masers is very small due to the good short term stability of the masers, but the drift with respect the BIPM has an uncertainty of +/- 1 x 10^{-16} due to the errors associated with long distance international measurements for a time interval of only 170 days. This data illustrates the degree to which the maser frequency variations agree with the international atomic time scale (BIPM) a year after the data in Figure 4 was obtained.

Other Stability Data

A great deal of stability data referencing the STSC hydrogen masers against other international standards has been presented previously [1]. Communications with these laboratories indicate that the good performance described previously is continuing. In particular, the operation of masers N-1 and N-2 at the Naval Research Laboratory is continuing normally. (See ACKNOWLEDGMENTS.)

NRL has been comparing N-1 with the international time and frequency laboratories and within the error of measurement this maser has the same frequency now as when first installed. N-2, which had an unusually large drift of approximately 5 x 10^{-15}/day when first installed in 1985, is drifting less than 1 x 10^{-15}/day at present. Among the STSC hydrogen masers installed in standards laboratories, N-2 has had the largest drift reported. The drift of this maser is an exception to the generally much better performance of the other masers and is due to the experimental conditions existing in early 1985 when the bulb was processed.

Conclusion

One of the main goals of this paper was to show that the STSC atomic hydrogen masers have an inherently long life. In addition, instability or interruption due to

degradation of line Q or systematic errors associated with the cavity automatic tuner were shown to be very small. The masers do have some initial change in frequency with time, and this is believed to be due to changes in the interaction of the hydrogen atoms with the wall of the storage bulb. The initial drift is very small however, typically less than 2 x 10^{-15}/day, and disappears with a time constant of between six months and one year. Thus for most of the operational life of the maser, the long term drift is so small as to be inconsequential for all present applications.

ACKNOWLEDGMENTS

The authors are very grateful to the United States Naval Research Laboratory for providing access to STSC Hydrogen Masers N-1 and N-2 for the measurements included herein and particularly to Mr. Ed Powers and Dr. Joseph White for providing stability and instrumental data on the masers for many years as well as for many fruitful discussions on hydrogen maser technology.

The authors are also very grateful to Dr. G. M. R. Winkler, Mr. Paul Wheeler and many others at the United States Naval Observatory for providing access to STSC hydrogen maser performance data and for many very useful discussions on hydrogen maser applications.

REFERENCES

[1] H. B. Owings, P. A. Koppang, C. C. MacMillan and H. E. Peters, "Experimental Frequency and Phase Stability of the Hydrogen Maser Standard Output as Affected by Cavity Auto-Tuning," in Proceedings of the 1992 Frequency Control Symposium (48th Annual Symposium on Frequency Control).

[2] E. Powers, A. Gifford and P. Wheeler, "Hydrogen Maser Performance at the United States Naval Observatory and the Naval Research Laboratory," in Proceedings of the 47th Annual Symposium on Frequency Control, 1991.

[3] H. E. Peters, H. B. Owings and P. A. Koppang, "Hydrogen Masers with Cavity Frequency Switching Servos," in Proceedings of the 22nd Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 1990.

[4] H. E. Peters, "Design and Performance of New Hydrogen Masers Using Cavity Frequency Switching Servos," in Proceedings of the 38th Annual Symposium on Frequency control, 1984.

[5] Claude Audoin and William A. Diener, "Frequency, Phase and Amplitude Changes of the Hydrogen Maser Oscillation," in Proceedings of the 1992 Frequency Control Symposium (48th Annual Symposium on Frequency Control).

[6] H. E. Peters and P. J. Washburn, "Atomic Hydrogen Maser Cavity and Bulb Design Optimization," in Proceedings of the 16th Annual PTTI Meeting. 1984.

[7] Vanier, J. and Audoin, C. "The Quantum Physics of Atomic Frequency Standards," Bristol, Adam Hilger, IOP Publishing LTD, 1989.

[8] D. Kleppner, H. C. Berg, S. B. Crampton, N. F. Ramsey, R. F. C. Vessot, H. E. Peters, and J. Vanier, "Hydrogen Maser Principles and Techniques," Phys. Rev. 138, No. 4A, A972-A983, 17 May, 1965.

[9] H. E. Peters, T. E. McGunigal, and E. H. Johnson, "Hydrogen Standard Work at Goddard Space Flight Center." in Proceedings of the 22 Annual Frequency control Symposium, 1968.

[10] Harry E. Peters, "Characteristics of Advanced Hydrogen Maser Frequency Standards," in Proceedings of the 5th Annual PTTI Meeting, 1973.

[11] Data provided by the United States Naval Observatory.

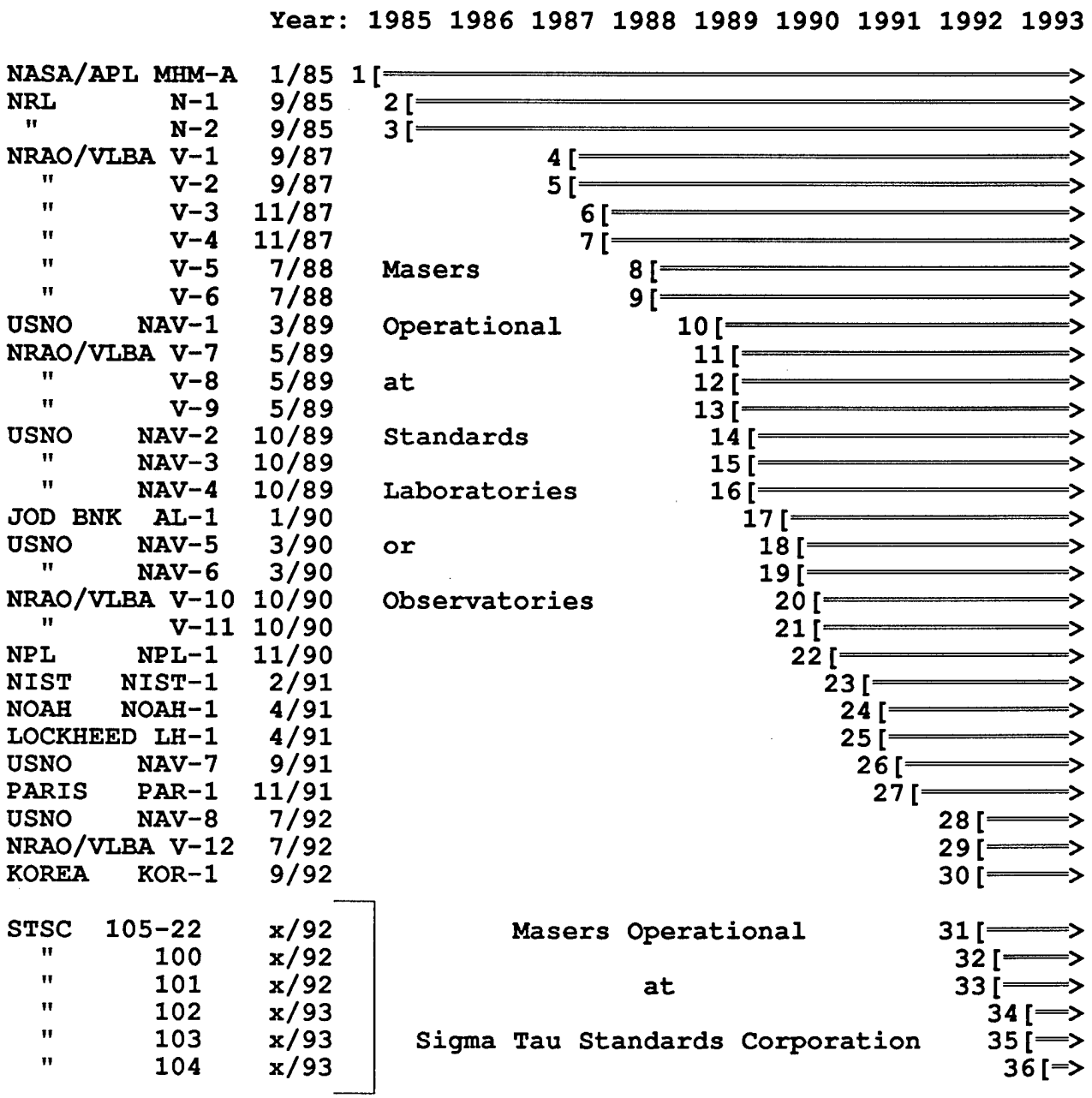


Figure 1. Installations with Sigma Tau Standards Corporation Atomic Hydrogen Masers

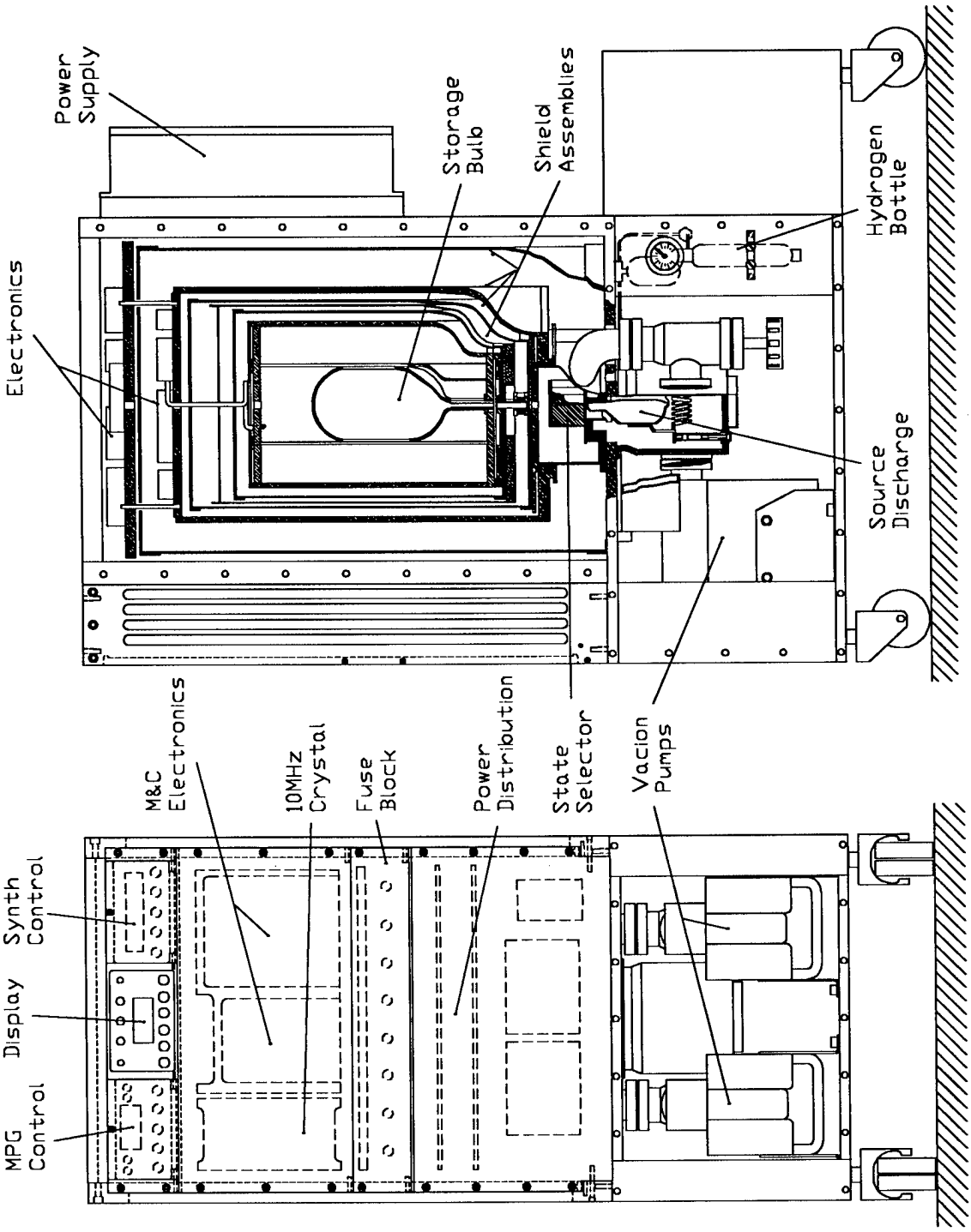


Figure 2. STSC Hydrogen Maser Physical Assembly.

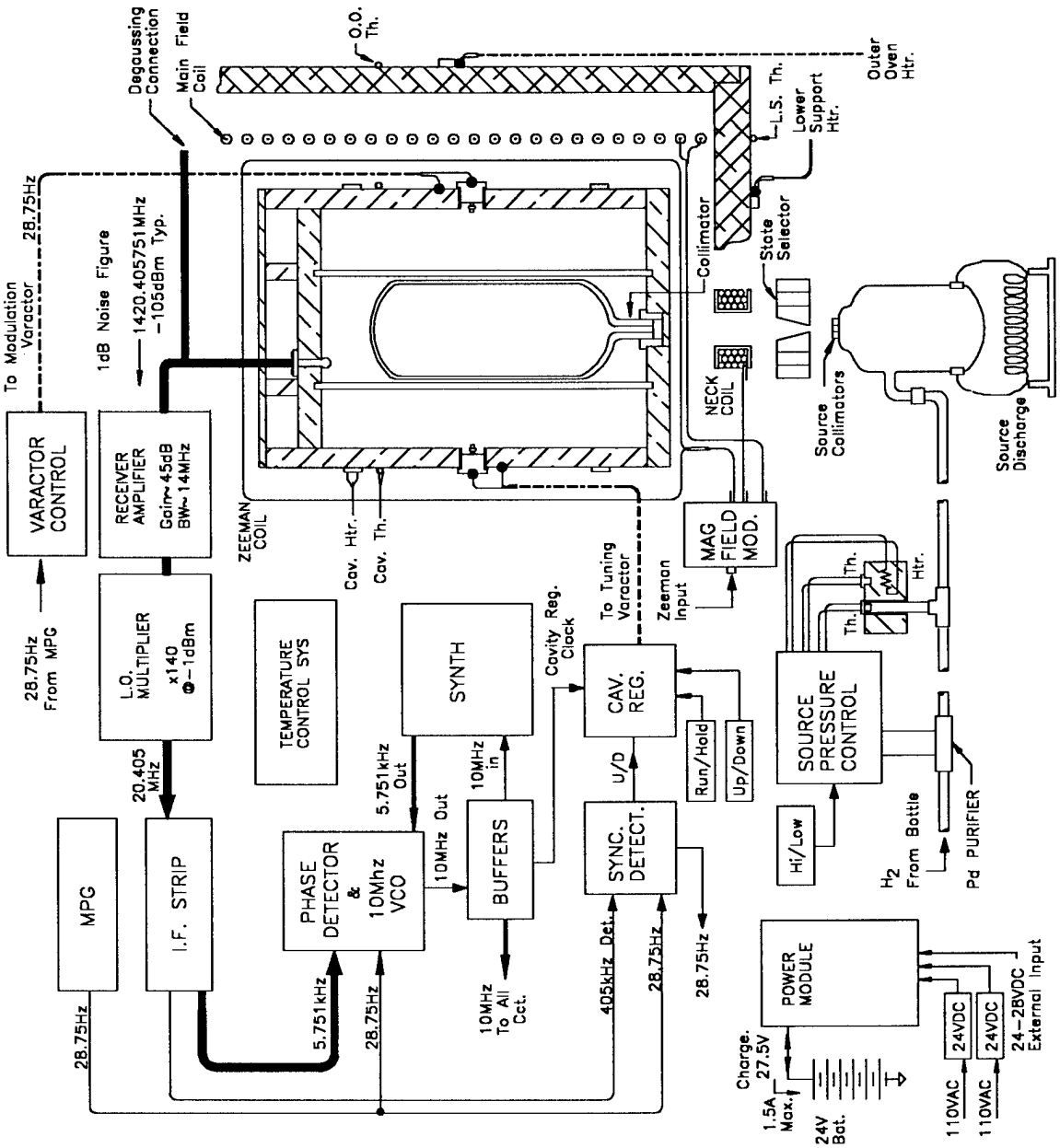


Figure 3. STSC Hydrogen Maser Block Diagram.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

RECENT INVESTIGATIONS WITH THE HARVARD-SMITHSONIAN CRYOGENIC HYDROGEN MASER*

R.L. Walsworth, E.M. Mattison, and R.F.C. Vessot

Harvard-Smithsonian Center for Astrophysics
60 Garden St., Cambridge, MA 02138

I.F. Silvera

Lyman Laboratory of Physics, Harvard University
Cambridge, MA 02138

Abstract

We have developed a hydrogen maser that operates near 0.5 K using a high-cooling-power ^{3}He refrigerator. We are using this cryogenic hydrogen maser to study low temperature atomic hydrogen spin-exchange collisions and hydrogen–helium interactions, and to investigate the performance of cryogenic hydrogen masers as possibly improved frequency standards. We report recent measurements on these topics.

Introduction

The atomic hydrogen (H) maser is the most stable time and frequency standard currently available for measurement intervals between 10 and 10^5 seconds. The frequency stability of state-of-the-art H masers is typically better than one part in 10^{15} for averaging intervals of 10^4 seconds. A cryogenic hydrogen maser (CHM) operating at temperatures below 1 K could have a frequency stability that is two to three orders of magnitude better than a room temperature H maser due to: (i) lower thermal noise; and (ii) a reduced H-H spin-exchange relaxation rate and the resulting higher operational signal power. In 1986 three groups [1,2,3], including ours, demonstrated oscillation of a CHM using superfluid helium-lined containers. It has subsequently been predicted by Verhaar, Koelman, Crampton, and co-workers [4,5] that new spin-exchange effects may make it more difficult to achieve a radical improvement in frequency stability using a CHM. In practice, the stability that can be realized by the CHM remains to be determined.

Harvard-Smithsonian CHM

The Harvard-Smithsonian CHM is currently operated in a new, high-cooling power recirculating ^{3}He refrigerator. This refrigerator provides 2 mW of cooling power at 0.5 K in the recirculating mode and with the CHM operating. A detailed discussion of the design and operation of the Harvard-Smithsonian CHM has been published previously [3]. To review, H atoms from a room temperature microwave discharge enter the ^{3}He refrigerator and are cooled to ~ 10 K. The atomic beam passes through a hexapole magnet that focusses the two upper hyperfine states into a microwave resonant cavity (at 1420 MHz) that is maintained ~ 0.5 K in a region of stable and homogeneous magnetic field. The cold H atoms are stored within the cavity for times $\gtrsim$ 2 seconds in a chamber coated with a superfluid helium film. With a modest input flux ($\gtrsim 10^8$ sec^{-1}) the Harvard-Smithsonian CHM performs a continuous maser oscillation. The CHM signal is detected inductively, isolated, amplified, and then piped out of the ^{3}He refrigerator. The frequency and power of this output signal are precisely measured. Room temperature H masers in our laboratory serve as frequency comparison standards for the CHM.

Recent results

The Harvard-Smithsonian CHM/^{3}He refrigerator system has been fully operational since early 1993. In this section we descibe our experimental results as of June, 1993.

We have operated the CHM with a maximum H input flux $\approx 10^{11}$ sec^{-1}, resulting in a maximum signal power $\approx 10^{-13}$ W and a maximum H density $\approx 10^9$ cm^{-3}. We have measured the CHM output frequency as a function of temperature (see Fig. 1). The CHM signal frequency is decreased by H collisions both with the He-film-coated walls of the storage chamber and with the accompanying He vapor. These two negative frequency shifts have opposite trends with temperature, since the H-wall residency time decreases with temperature and the He vapor pressure increases with temperature. Thus there is a

*Supported by the Air Force Office of Scientific Research, NASA, and the Smithsonian Institution.

0-7803-0905-7/93 $3.00 © 1993 IEEE

stable minimum in the magnitude of the total H-He frequency shift. For the current CHM geometry this minimum is near 0.53 K.

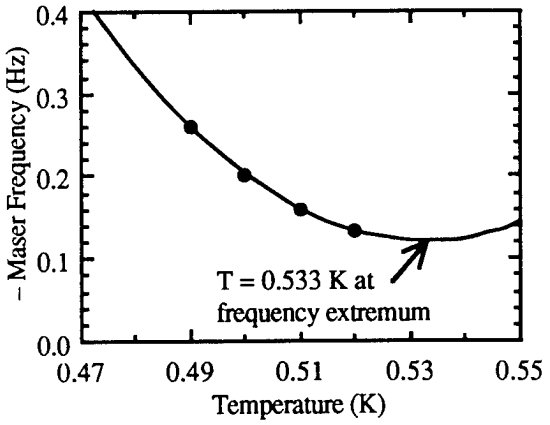


Figure 1. Measured dependence of CHM signal frequency on temperature, and a quadratic fit to data.

Our measurements show a stronger dependence of CHM frequency on temperature than that reported by the University of British Columbia (UBC) group [6]. We believe that this larger "wall shift" in our system is due to H interactions with paramagnetic impurities in the sapphire substrate that lies beneath the He film. (Note that there is also a 50 μm layer of Teflon between the sapphire and the He.) The hypothesis of strong H-wall interactions in our CHM is supported by two other measurements:

1) We have determined the absolute value of the H-wall frequency shift = – 0.297 (5) Hz at T = 0.490 (2) mK. The results of the UBC group, however, imply a much smaller frequency shift ≈ – 0.1 Hz.
2) We have measured our CHM's line-Q ≈ 2.3×10^9 with insignificant magnetic and collisional relaxation, implying $T_2 \approx 0.52$ seconds. The contribution to T_2 from H escape from the storage bulb is calculated ≈ 2.5 seconds from geometry. Thus we conclude that the observed T_2 is primarily due to wall relaxation.

We have also measured the CHM's frequency stability relative to a room temperature H maser. At T = 0.49 K, with 20 μK temperature control and a signal power ≈ 10^{-14} W, we found $\sigma_y(\tau) \approx 1.6 \times 10^{-13}/\sqrt{\tau}$, for 10 seconds < τ < 160 seconds. Here $\sigma_y(t)$ is the Allan, or two-sample, deviation [7], a measure of fractional frequency stability over an averaging interval τ. A typical set of data comparing CHM and room temperature H maser frequency stability is shown in Fig. 2. Also shown in Fig. 2 is the measured stability of the reference H maser (designated P13) to a second room temperature H maser (P12). It is clear from this data that measurement of the CHM's short-term frequency stability is already limited by the performance of the room temperature H masers.

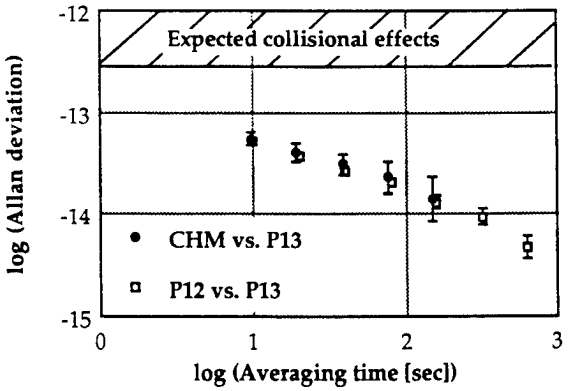


Figure 2. Typical measured short-term frequency stabilities of the CHM relative to a room temperature H maser (P13), and P13 relative to a second room temperature H maser (P12).

Finally, we have made preliminary measurements of H-density-dependent spin-exchange frequency shifts in the CHM. An example of such a measurement is shown in Fig. 3. In this case the CHM's microwave cavity was tuned ~ 10 kHz off-resonance, so that the observed density-dependent frequency shift gives information on the spin-exchange line-broadening cross section $\bar{\sigma}_{se}$. From several two-point runs at T = 0.49 K, such as that shown in Fig. 3, we find $\bar{\sigma}_{se} = 2\text{–}3 \times 10^{-16}$ cm^2, whereas recent calculations [5] predict $\bar{\sigma}_{se} \approx 8 \times 10^{-17}$ cm^2. (Note that the absolute H density calibration in this preliminary CHM work may be off by a factor of two to three, possibly explaining the discrepancy with theory.)

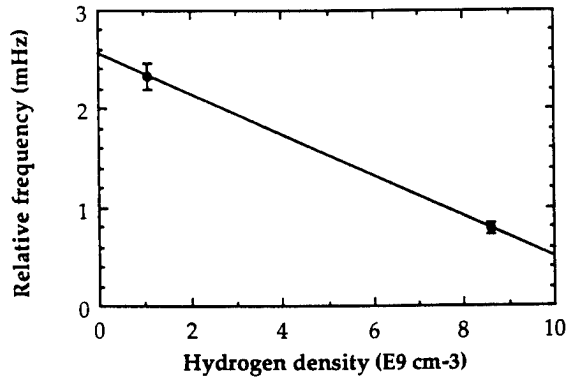


Figure 3. Example of preliminary measurements of H-density-dependent frequency shifts in the CHM.

Work in progress

We have recently made modifications to our CHM/^{3}He refrigerator system to improve the maximum H input flux into the maser, and to allow data-taking for extended periods of time (~ 1 day). These system improvements are currently allowing us to do the following:

1) Complete investigations of H-H spin-exchange collisions at temperatures ~ 0.5 K.
2) Determine the CHM's frequency stability, relative to room temperature H masers, out to 10^4 seconds.

We plan to complete these measurements within the next year.

References

[1] H.F. Hess, G.P. Kochanski, J.M. Doyle, T.J. Greytak, D. Kleppner, Phys. Rev. A **34**, 1602 (1986).

[2] M.D. Hürlimann, W.N. Hardy, A.J. Berlinsky, R.W. Cline, Phys. Rev. A **34**, 1605 (1986).

[3] R.L. Walsworth, I.F. Silvera, H.P. Godfried, C.C. Agosta, R.F.C. Vessot, E.M. Mattison, Phys. Rev. A **34**, 2550 (1986).

[4] B.J. Verhaar, J.M.V.A. Koelman, H.T.C. Stoof, O.J. Luiten, S.B. Crampton, Phys. Rev. A **35**, 3825 (1987).

[5] J.M.V.A. Koelman, S.B. Crampton, H.T.C. Stoof, O.J. Luiten, B.J. Verhaar, Phys. Rev. A **38**, 3535 (1988).

[6] R. Jochemsen, M. Morrow, A.J. Berlinsky, W.N. Hardy, Phys. Rev. Lett. **47**, 852 (1981).

[7] D. Allan, Proc. IEEE **54**, 105 (1966).

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

LONG TERM STABILITY OF Hg^+ TRAPPED ION FREQUENCY STANDARDS[1]

R. L. Tjoelker, J. D. Prestage, G. J. Dick, L. Maleki

California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

Frequency standards based on trapped ions have fundamental advantages over more established standards for long averaging times. We report the development of a second $^{199}Hg^+$ Linear Ion Trap (LIT) based frequency standard to provide a capability for measuring stability beyond all existing frequency standards. Increased signal using a second light collection system together with our previously demonstrated atomic line $Q \approx 2 \times 10^{12}$ yields a performance better than $\sigma_y(\tau) = 7 \times 10^{-14}/\tau^{1/2}$. Sensitivity to the leading perturbations are measured to identify regulation requirements to obtain a stability of 10^{-16}.

Introduction

High performance atomic frequency standards require good atomic line resolution ($Q \equiv f/\Delta f$) and signal to noise (S/N). The 40.5 GHz atomic ground state hyperfine transition of $^{199}Hg^+$ [1] has been measured in the JPL trapped ion frequency standard to have a corresponding $Q > 2 \times 10^{12}$[2,3]. Good S/N is achieved with as many as 3×10^7 ions in a linear ion trap(LIT)[4]. Previously we reported stability measurements using the first LIT based Hg+ frequency standard (LITS-1) while steering three different Local Oscillators(LO): a quartz crystal, H-maser, and the Superconducting Cavity Maser Oscillator (SCMO)[2,3,5]. In each case, measurement of the long term stability was apparently limited by the stability of available Hydrogen reference masers.

In this paper, we report the development of a second Hg+ trapped ion frequency standard (LITS-2). This standard incorporates improved light collection (increased S/N) and thermal regulation to maintain long term stability. Under typical operating conditions, preliminary measurements against H-masers show performance better than $\sigma_y(\tau) = 7 \times 10^{-14}/\tau^{1/2}$ with measurements spanning up to 10 days. Sensitivity to systematics have been characterized and preparations are currently underway for long term stability measurements between the two Hg+ linear ion trap based frequency standards LITS-1 and LITS-2.

Operation of the JPL Linear Ion Trap Standard

In the present operational configuration (Fig. 1), $^{199}Hg^+$ ions are formed inside the trap by an electron pulse which ionizes a neutral background vapor of ^{199}Hg. Neutral mercury is introduced into the vacuum by heating isotopically purified ($\approx$ 91%) HgO powder. Collisions with a helium buffer gas ($\approx 10^{-5}$ mbar) cool the ions to near thermal equilibrium with the vacuum enclosure at room temperature. Resonant ultraviolet (UV) radiation (194.2 nm) from a ^{202}Hg discharge lamp optically pumps the $^{199}Hg^+$ ions into the F=0 hyperfine level of the

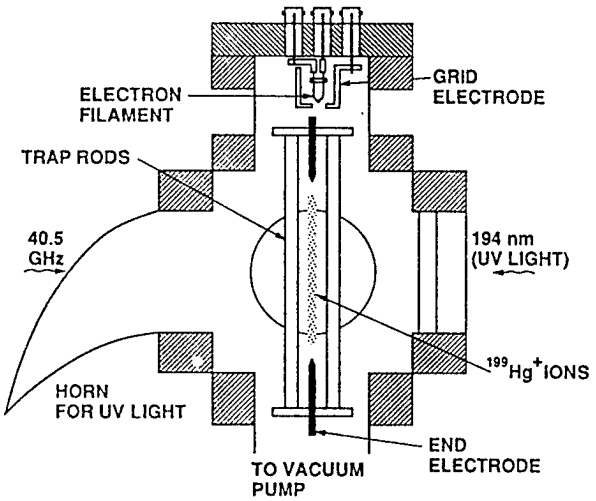


Figure 1: Linear ion trap frequency standard. Microwaves near 40.5 GHz enter the trap region from the left. State selection light from the ^{202}Hg discharge lamp enters from the right, focused onto the trapped ions, and collected in a pyrex horn. Fluorescence from the trapped ions is collected in a direction normal to the page.

[1]This work was performed by the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

0-7803-0905-7/93 $3.00 © 1993 IEEE

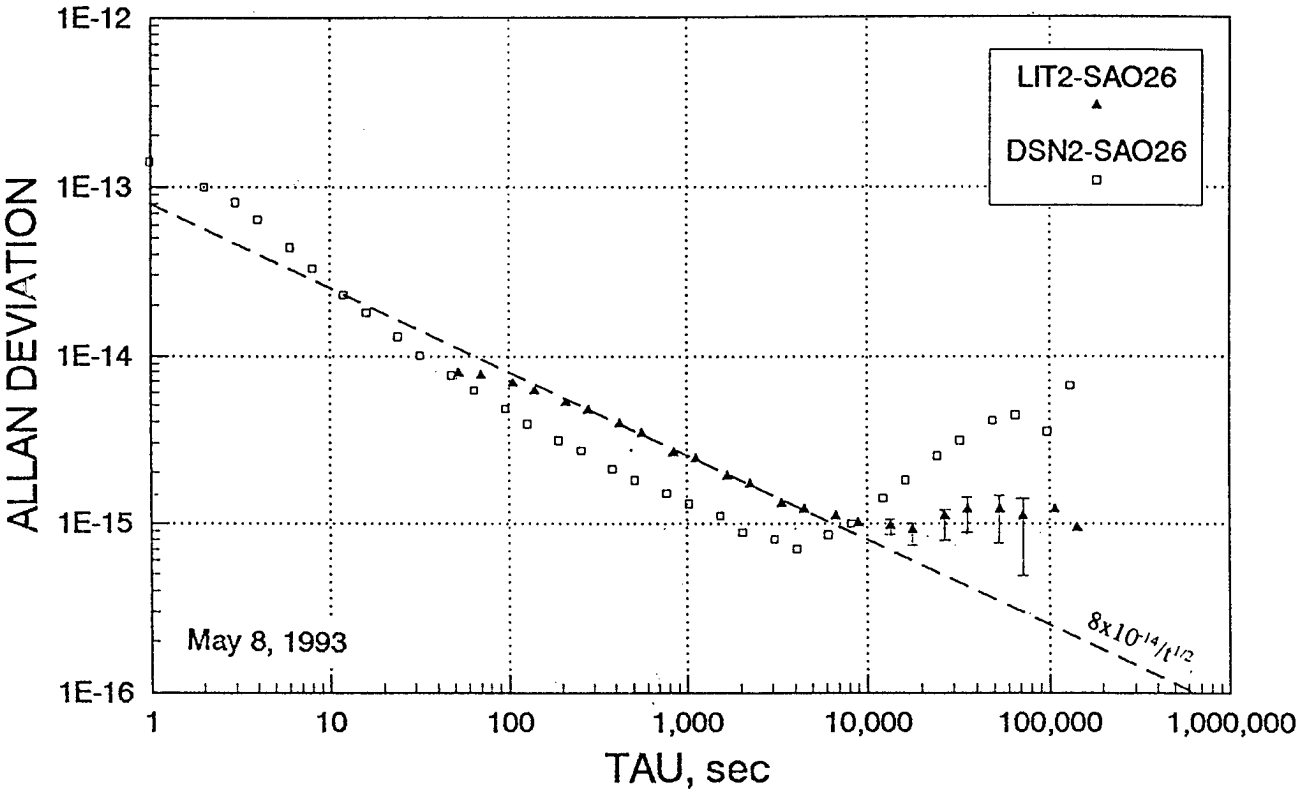


Figure 2: Four day stability comparison of LITS-2 vs. the H-maser SAO-26. The dashed line represents a performance of $8x10^{-14}/t^{1/2}$. Also shown for reference is the stability of two H-masers SAO-26 and DSN-2 over the same 4 day interval.

ground state. To minimize scattered light the state selection light is collected in a pyrex horn after passing through the trap. Microwave radiation (40.5 GHz) to transfer the population to the F=1 ground state enters the trap region through the same pyrex horn and in a direction opposite to the UV state selection/interrogation light. This allows collection of atomic fluorescence in both directions perpendicular to the incident pumping light. Previously, fluorescence was collected in only one of these two directions. Inclusion of a second light collection system has improved S/N by $2^{1/2}$.

After the $^{199}Hg^+$ ions are pumped into the $^2S_{1/2}(F=0,m=0)$ state, the technique of successive oscillatory fields[6] is used to probe the approximately 40.5 GHz hyperfine clock transition to the $^2S_{1/2}(F=1,m=0)$ state. The UV state selection/interrogation light level is reduced to near zero during the microwave interrogation period to prevent light shifts and broadening of the hyperfine transition. After application of the second microwave pulse, the UV light is again turned back on to determine if the microwaves were on resonance and successfully transferred the population to the F=1 state. The optical pumping/interrogation time is about 1.5 seconds. With two collection systems, a background photon count level between 60-100 kHz is subtracted to generate the resonance. With good collimation and filtering, signal to background of 17% is obtained, yielding S/N of the Ramsey interrogated clock transition as high as 50.

Long Term Stability of LITS-2

Stability measurements of the first JPL ion trap standard LITS-1 against H-masers have been previously published with $\sigma_y(\tau)=1x10^{-13}/\tau^{1/2}$ representing the best measured performance [2,3,5]. Initial stability measurements of LITS-2 show performance of $\sigma_y(\tau)=7x10^{-14}/\tau^{1/2}$. Figures 2 and 3 show the stability measured against the H-maser SAO-26[8a]. For times greater than ≈10,000 seconds the measurement is limited by drift in the H-maser. For reference, stability of SAO-26 against the H-maser DSN-2[8b] is also shown. Figure 2 shows data taken over a 4 day interval and Figure 3 over a 10 day interval.

Preparations to measure LITS-2 against LITS-1 involves upgrading LITS-1 to give it similar performance and stability as LITS-2. For investigations of long term stability and limiting systematics, the two ion trap

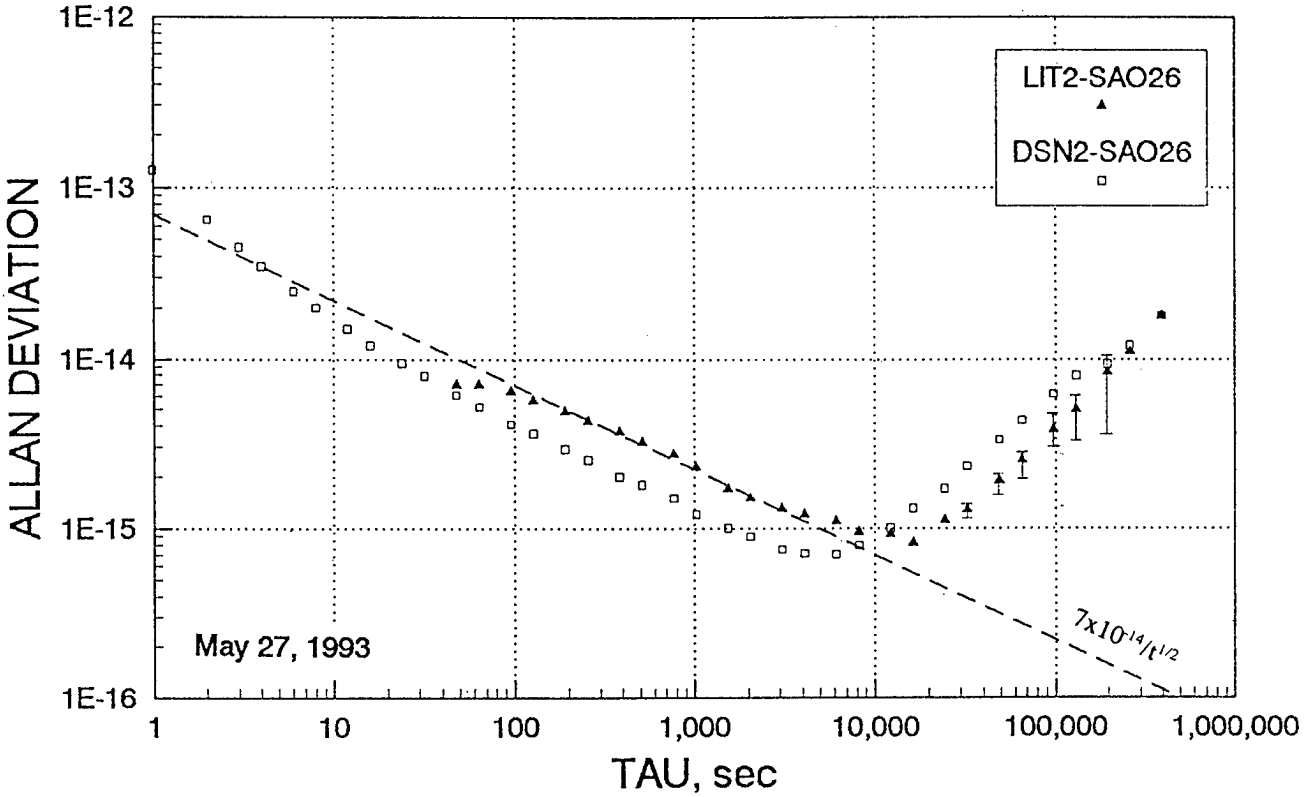


Figure 3: Ten day stability comparison of LITS-2 vs. the H-maser SAO-26. The dashed line represents a performance of $7x10^{-14}/t^{1/2}$. Also shown for reference is the stability of two H-masers SAO-26 and DSN-2 over the same 10 day interval.

standards will steer two VLG-11[8a] receivers locked to a common H-maser. Regulation to give a fractional frequency stability of 10^{-16} is currently being incorporated into both ion trap standards based upon systematic and sensitivity studies performed with LITS-1 [7].

Environmental Sensitivity and Systematics

The long term stability depends on sensitivities to environmental perturbations. Unless noted otherwise, the systematic data presented here was measured using LITS-1. Under typical operating conditions the largest frequency offsets and measured sensitivities are shown in Table 1. The limiting *stability* of the LITS units depends on how well these offsets are regulated. The obtainable *accuracy* of the system depends on how well these offsets can be modeled, understood, and corrected for. The accuracy of the Hg+ LIT standards is currently about 10^{-13} with the 2nd order Doppler shifts (ion number and temperature) the largest uncertainty. With better temperature and ion number measurements it may be possible to achieve 10^{-14} accuracies with the fieldable, room temperature, lamp based systems LITS-1 and LITS-2 (see Ref.[9]).

A brief discussion of the leading perturbations on the Hg^+ standards follow:

Magnetic Shifts

The ground state hyperfine energy levels of $^{199}Hg^+$ as a function of magnetic field is shown in Fig. 4(a). The clock transition $^2S_{1/2}(F=0,m=0)$ to $^2S_{1/2}(F=1,m=0)$ has a second-order field dependence of

$$f(B)= 40.507347997\ GHz + 97.2\ B^2\ [Hz/G^2]. \quad (3)$$

The fractional sensitivity of this transition to magnetic field variations is 800 times smaller than that of hydrogen and nearly 20 times smaller than cesium at the same operating field.

A Helmholtz coil centered on the trapping region is used to apply a fixed magnetic field along the trap axis. With B=35 mG the unshielded atomic sensitivity is $1.7x10^{-13}$/mG. To prevent ambient field disturbances from influencing the ion hyperfine frequency the trap region is surrounded by a triple layer magnetic shield. With a shielding factor of 10,000, ambient shifts as large as 5 mG would produce a

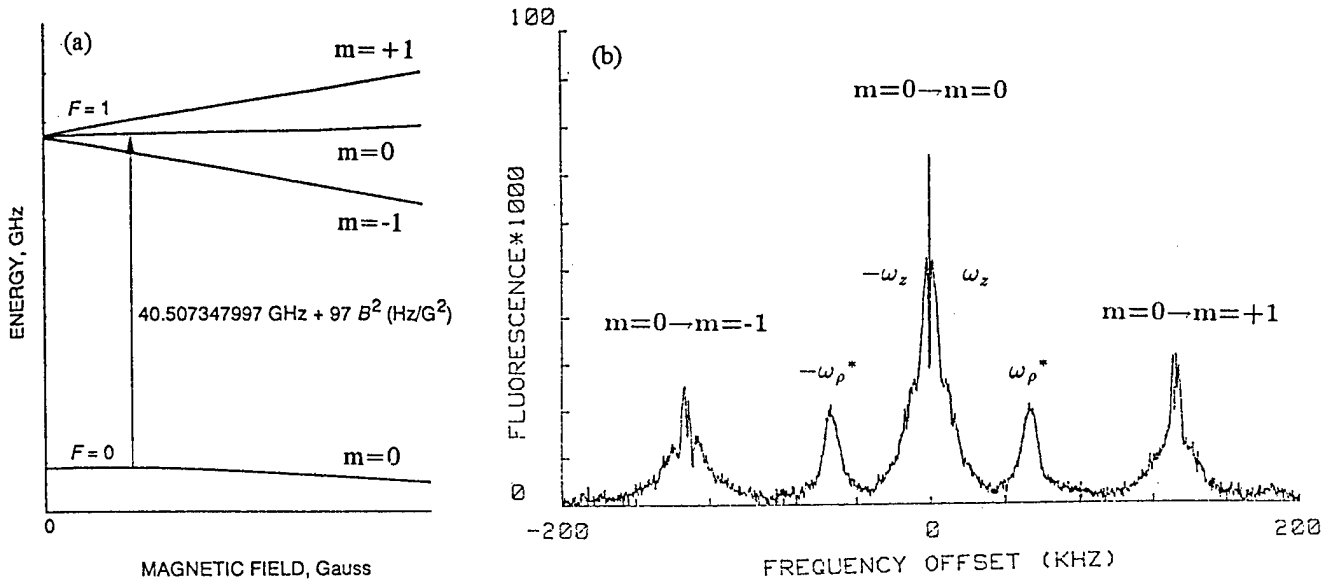


Figure 4: (a) Ground state energy levels of $^{199}Hg^+$ in an external magnetic field. (b) The Zeeman transitions and motional sidebands observed in LITS-2 with B=95 mG. The $^2S_{1/2}(F=0,m_F=0)$ to $^2S_{1/2}(F=1,m_F=0)$ clock transition is the sharp central peak and has been measured to be as narrow as 16 mHz [2].

fractional frequency shift less than $1x10^{-16}$.

Magnetic gradients across the ion cloud will degrade atomic coherence and limit line Q. The materials of the trap are carefully selected for their non-magnetic properties. Currently, residual field gradients do not limit atomic coherence (with interrogation times up to 30 seconds) though a related effect referred to as Zeeman coherence can result in loss of signal. This occurs when the Zeeman transitions (F=0,m=0) to (F=1,m=+-1) overlap the ion motional sidebands of the clock transition (F=0,m=0) to (F=1,m=0). Figure 4(b) shows the three ground state transitions between the F=0 and F=1 levels and motional sidebands at the ion secular frequency ($\omega_\rho \approx 60$ kHz) and the axial frequency ($\omega_z \approx 1$ kHz). This Zeeman coherence condition is avoided by operating with an applied magnetic field greater than ≈35 mG. An analysis of the Zeeman and motional sideband linewidth and location also yield information on ion number and temperature[9,10].

Second Order Doppler Shift

A property of RF ion traps is that a second order Doppler shift results from the driven ion motion in the trap (micromotion). While increasing the ion number improves S/N, the frequency shift also increases as the cloud radius grows. In order to improve S/N without increasing the second order Doppler shift the hybrid rf/dc linear ion trap (LIT) was developed[4]. This trap confines ions along a line of nodes of the rf field. The trapping force transverse to the line of nodes is generated by the ponderomotive force as in conventional Paul traps while the axial trapping force is provided by dc electric fields.

Since the distance from the node line determines the magnitude of the rf trapping field, the 2nd order Doppler shift is reduced from that of a hyperbolic trap for the same number of ions by a factor of $(5/3)R_{sph}/L$, where R_{sph} is the radius of a hyperbolic trap and L the length of the linear ion trap. For sufficiently dense ion clouds (Ref. [9], Appendix 1), the second order Doppler shift in a long linear trap is

$$(\Delta f/f)_{linear} = -(q^2/8\pi\varepsilon_o mc^2)(N/L), \quad (1)$$

where N/L is the linear number density of ions in the trap. For $^{199}Hg^+$, this fractional frequency shift is

$$(\Delta f/f)_{Hg} = -(3.8 \times 10^{-21} \text{ meters})(N/L). \quad (2)$$

For our trap dimensions, clouds consisting of 10^6 ions exhibit a second-order Doppler shift from trapping fields of $4x10^{-14}$. Larger clouds are often trapped giving increased S/N as a tradeoff for increased 2nd order Doppler shift.

For ultra-stable operation it is necessary to keep the number of ions in the trap fixed. In the room temperature apparatus the confinement lifetime, typically greater than

Table 1: The largest fractional frequency offsets and sensitivities of the $^{199}Hg^{+}$ clock transition under typical operating conditions.

SYSTEMATIC	OPERATION	OFFSET ($\times 10^{-14}$)	SENSITIVITY
MAGNETIC FIELDS:			
Applied Field	$B = 35mG$	+300	$\left(\frac{\Delta\nu}{\Delta B}\right)\frac{1}{\nu} = 1.7 \times 10^{-13}/[mG]$
Ambient Fluctuations	$\Delta B \approx 1mG/10^4$		$\left(\frac{\Delta\nu}{\Delta B}\right)\frac{1}{\nu} = 1.7 \times 10^{-17}/[mG]$
2nd ORDER DOPPLER:			
Ion Temperature	$T = 650K$	-45	$\left(\frac{\Delta\nu}{\Delta T}\right)\frac{1}{\nu} = -6.9 \times 10^{-16}/[°C]$
Ion Number:	2.5×10^6	-43	
(EC potential,$V \geq 10V$)	20V		$\left(\frac{\Delta\nu}{\Delta V}\right)\frac{1}{\nu} = -2.4 \times 10^{-15}/[V]$
(Fil. emission)	$300\mu A$		$\left(\frac{\Delta\nu}{\Delta I}\right)\frac{1}{\nu} = -1.2 \times 10^{-15}/[\mu A]$
(HgO heat,$V \geq 10V$)	$230°C$		$\left(\frac{\Delta\nu}{\Delta T}\right)\frac{1}{\nu} = -8 \times 10^{-15}/[°C]$
(Unit Temp.)	$23°C$		$\left(\frac{\Delta\nu}{\Delta T}\right)\frac{1}{\nu} = -9 \times 10^{-15}/[°C]$
HELIUM PRESSURE	$P = 10^{-5}$mbar	+11	$\left(\frac{\Delta\nu}{\Delta P}\right)\frac{1}{\nu} = 1.1 \times 10^{-13}/[10^{-5}mbar]$
LIGHT SHIFT	1000/s	+0.12	$\left(\frac{\Delta\nu}{\Delta\gamma}\right)\frac{1}{\nu} = 1.2 \times 10^{-18}/[photons/s]$

4 hours, is limited by the low but constant background pressure of neutral mercury (≈4x10^{-10} mbar). Currently, ion number is kept constant for days to better than 0.1% through repeated loading (topping) of the trap. Once each interrogation cycle (typical cycle time 20 to 30 seconds), a LaB^6 filament emits electrons to ionize neutral mercury in the trapping region. If the trap is not periodically topped off, the ion number decreases and the second-order Doppler shift is directly observed as shown in Fig. 5.

Stabilization of the ion number requires regulation of the neutral mercury pressure, filament emission, and trap potentials (Table 1). Some stabilization of ion number is also apparently achieved by filling the trap to a saturated limit imposed by the endcap electrodes. The 1.5 mm diameter "needle" electrodes, the endcap potential, and ion temperature together define the trap volume. Figure 6(a) shows the 2nd order Doppler frequency shift as a function of the endcap electrode potential V_{EC}. The trapped ion number seems limited and sensitivity to endcap potential reduced for endcap potentials greater than ≈10 V. The interpretation that the frequency offset in Fig. 6(a) is in part due to variations in Hg^{+} number is supported by Fig. 6(b) which shows the observed fluorescence of the hyperfine transition versus the same variation in endcap voltage. The apparent ability to passively regulate ion number is a unique property of the hybrid rf/dc linear ion trap. Further research is needed to determine the effectiveness of this passive regulation for very long averaging times.

Background Pressure Shifts

A buffer gas of neutral helium is used to cool and keep the Hg^{+} ions in thermal equilibrium with the room temperature vacuum enclosure (see temperature measurements in Ref.[9]). Cooling the ions reduces the radial extent of a given ion cloud, reducing the driven and thermal second order Doppler shift.

Collisions between the mercury and the helium buffer gas perturb the mercury hyperfine transition. The measured fractional frequency offset corresponding to helium pressure for a fixed set of loading conditions is shown in Fig. 8. The helium pressure shift on Hg+ was first measured and discussed by Cutler et.al. [10]. At low pressures the quantity of helium is not sufficient to cool the trapped ions. The ion temperature rises producing a thermal second order doppler shift of 3.2x10^{-7}/mbar. The ion number is not held constant for these measurements so the actual sensitivity at low helium pressure is larger (due to a reduction in ion number which shifts the frequency higher as shown in Figs. 5 and 6). For helium pressures larger than ≈6x10^{-6} mbar the sensitivity is reduced to 1.1x10^{-8}/mbar. In this region the mercury ions are in thermal equilibrium at ≈650 K and the predominant frequency dependence is a hyperfine pressure shift [10,11]. Currently the helium pressure is measured with an ionization gauge and the measured level is used to servo the heater of a quartz helium leak to hold the pressure constant.

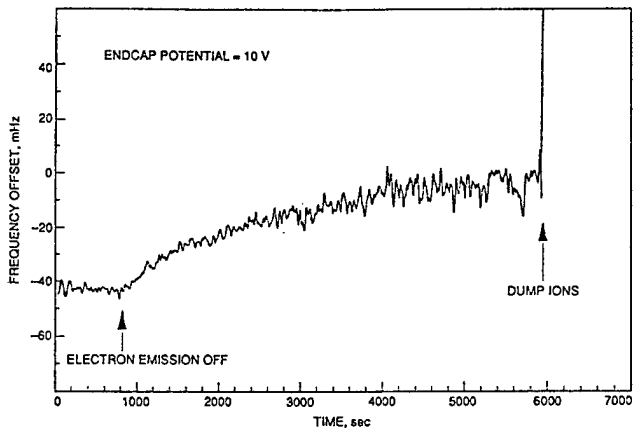


Figure 5: A direct measure of the 2nd order Doppler shift. As the trapped ion number diminishes the clock frequency increases.

Temperature Shifts

For the reported stability measurements, the standard was in a test chamber at the JPL frequency standards test facility with a regulated temperature to 0.05°C. At the present time the ion standard physics unit is not thermally regulated except for the helium leak and HgO heater. The fractional temperature sensitivity of the complete unregulated and uninsulated system is -9×10^{-15}/°C measured by varying the temperature in the environmental test chamber by 5°C.

The mercury lamp must also be temperature controlled as the 194 nm UV brightness depends the operating temperature. In the present version the temperature is controlled by blowing dry nitrogen gas over the tip of the bulb. The lamp has an average power consumption of ≈2 Watts (15 watts peak) and resides inside the magnetic shields. The equilibrium temperature depends on the on-off lamp cycle. The time constant for the lamp and surrounding apparatus to reach thermal equilibrium is about 3 hours after starting with new settings. The current lamp in LITS-1 is three years old with over two years of total operating time. The UV output does not appear to have degraded over this time.

The neutral mercury pressure in the trap region depends on the temperature of the HgO source. Temperature control on the heater is better than 0.1°C at 230°C. Higher regulation may be needed or perhaps active ion number control[10].

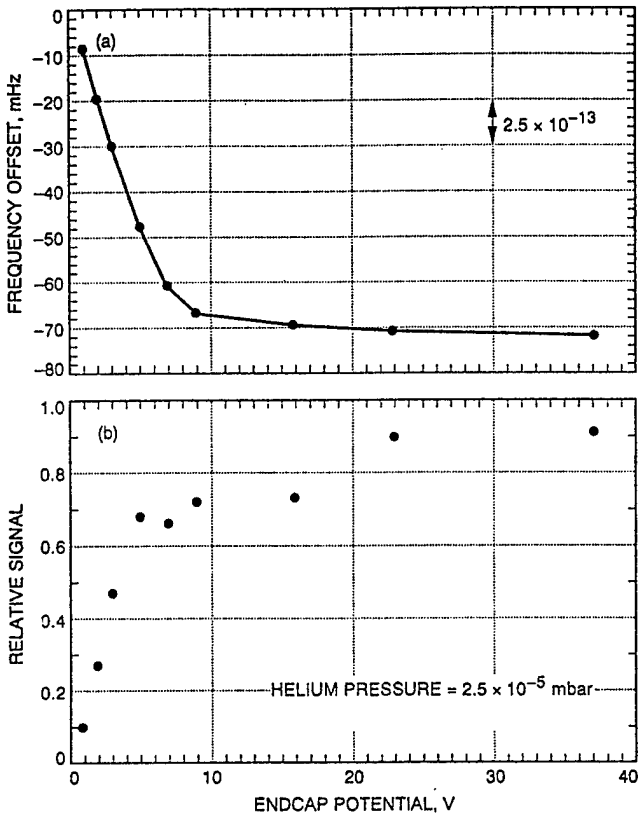


Figure 6: (a): Fractional frequency offset vs. dc potential on the LIT endcap electrodes for a large ion cloud. (b) Variations in the observed peak fluorescence as a function of endcap potential.

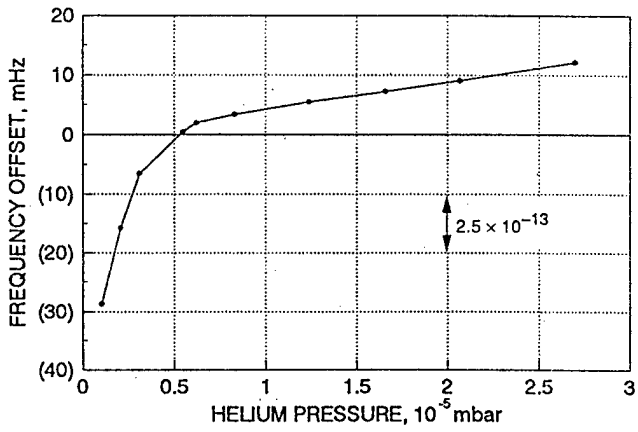


Figure 7: Fractional frequency offset vs. helium buffer gas pressure.

Light Shifts

For state selection and interrogation the lamp is switched in a bimodal operation between a bright and dim state. The lamp is kept dim during the microwave interrogation to avoid light shifts (the UV lamp spectrum and ion absorption spectrum do not perfectly overlap). In this dim state a light sensitivity of 1.2×10^{-18}/photon/sec is observed. Since the typical photon count rate during the dim state is 1 kHz, an offset of $\approx10^{-15}$ is present due to switching the lamp in this way. The use of a shutter eliminates this effect entirely.

Conclusion

A second Hg+ trapped ion frequency standard (LITS-2) has been developed to provide a capability for measuring stability beyond all existing frequency standards for long averaging times. Preliminary stability measurements between LITS-2 and H-masers show short term performance better than $\sigma_y(\tau)=7\times10^{-14}/\tau^{1/2}$ with measurements spanning up to 10 days. Current measurements were performed using a S/N=40 and interrogation time of 16 seconds. With anticipated S/N improvements (S/N≈60), and the already demonstrated interrogation time of 30 seconds ($Q>2\times10^{12}$), stability performance of less than $3\times10^{-14}/\tau^{1/2}$ is expected with the existing standards.

The limiting long term stability depends on the sensitivity of the atomic system to changes in environmental and operating parameters. The environmental sensitivity of the LIT frequency standards have been measured and current regulation on LITS-2 designed to allow stabilities down to 10^{-16}. Presently, comparisons have shown stabilities better than 8×10^{-16} before H-maser drift apparently limits the measurement. Long term comparisons between LITS-2 and LITS-1 and a determination of accuracy are currently underway.

References

[1] F. G. Major and G. Werth, "Magnetic Hyperfine Spectrum of Isolated $^{199}Hg^+$ Ions", *Appl. Phys.* **15**, 201-208, 1981.

[2] J.D. Prestage, R.L. Tjoelker, G.J. Dick, L.Maleki, "Ultra-Stable Hg^+ Trapped Ion Frequency Standard", IEEE Trans. Instr. and Meas. **42**, April (1993).

[3] J.D. Prestage, R.L. Tjoelker, G.J. Dick, and L. Maleki, "Hg+ Trapped Ion Standard Performance With the Superconducting Cavity Maser Oscillator as L.O.", Proc. 46th Ann. Symp. Freq. Control, 58-63, May 1992.

[4] J. D. Prestage, G. J. Dick, L. Maleki, "New Ion Trap for Frequency Standard Applications", *J. Appl. Phys.* **66**, No. 3, 1013-1017, August 1989.

[5] J.D. Prestage, R.L. Tjoelker, G.J. Dick, L.Maleki, "Ultra-Stable Hg^+ Trapped Ion Frequency Standard", *J. Mod. Opt.* **39**, No. 2, 221-232, (1992).

[6] N. F. Ramsey, *Molecular Beams* (Oxford University Press, Oxford, 1956).

[7] R.L. Tjoelker, J.D. Prestage, and L. Maleki, "Criteria for Ultra-Stable Operation of the Trapped Ion Frequency Standard," TDA Progress Report 42-108, Vol. October-December 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 19-30, February 15, 1992.

[8a] Smithsonian Astrophysical Observatory; Cambridge, Massachusetts 02138.

[8b] Jet Propulsion Laboratory; Pasadena, California 91109.

[9] J.D. Prestage, R.L. Tjoelker, G.J. Dick, and L. Maleki, "Doppler Sideband Spectra of Hg Ions in a Linear Ion Trap", Proc. 47th Ann. Symp. Freq. Control, June 1993.

[10] L. S. Cutler, R. P. Giffard, and M. D. McGuire, "Thermalization of ^{199}Hg Ion Macromotion by a Light Background Gas in an RF Quadrupole Trap", *Appl. Phys. B* **36**, 137--142, 1985.

[11] J. Vetter, M. Stuke, and E.W. Weber, "Hyperfine Density Shifts of $^{137}Ba+$ Ions in Noble Gas Buffers", *Z. Physik* **A 273**,no. 2, 129-135, 1975.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PROGRESS AT CSIRO AUSTRALIA TOWARDS A MICROWAVE FREQUENCY STANDARD BASED ON TRAPPED, LASER-COOLED $^{171}Yb^+$ IONS

P.T.H. Fisk, M.A. Lawn and C. Coles

CSIRO Division of Applied Physics
National Measurement Laboratory
P.O. Box 218
Lindfield NSW 2070 Australia

Abstract

We report preliminary results of precision microwave-optical double resonance spectroscopy on buffer gas cooled $^{171}Yb^+$ ions confined in a linear Paul trap. We also report laser cooling of trapped $^{171}Yb^+$ ions. Temperatures below 1K have been achieved, and evidence of a phase transition to a crystal-like state has been observed.

Introduction

The possibility of using trapped $^{171}Yb^+$ ions as a frequency standard in both the microwave and optical regimes has been under investigation by several research groups in recent years[1,2,3]. Part of the reason for this interest is that it has recently become feasible to build relatively inexpensive diode laser based sources for the laser light required to cool and detect the trapped ions[4]. However, most current work, including that reported here, still uses other types of lasers.

In this paper we report preliminary results of an investigation of the use of $^{171}Yb^+$ ions confined in a linear Paul trap as a microwave frequency standard at 12.6 GHz.

The linear Paul trap

The $^{171}Yb^+$ ions are confined in a linear Paul trap [5,6] shown in fig. 1.

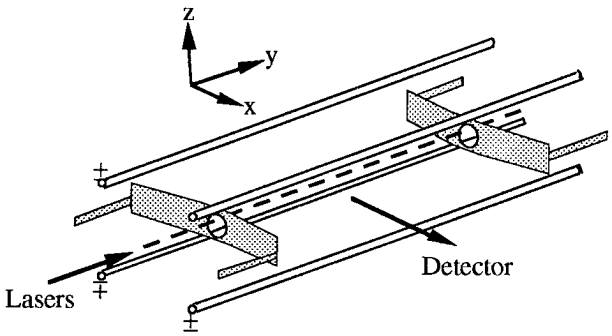


Figure 1: Schematic diagram of the ion trap electrode assembly. The DC end electrodes are shaded and the relative polarity of the RF electrodes is shown.

The RF rod electrodes are cut from 2 mm dia. Mo rod and are separated by 20 mm. The DC end electrodes are cut from 0.5 mm Cu sheet and are separated by 60 mm along the axis of the trap, indicated by the dashed line. Holes 6 mm in diameter are cut in the end electrodes for passage of the laser beams. The trap is mounted with its major axis horizontal in a cubic stainless steel vacuum chamber with outside edge lengths of 152 mm. The chamber is evacuated to a base pressure of $6x10^{-8}$ Pa using ion pumps and a titanium sublimation pump. Natural Yb or isotopically enriched ^{171}Yb (88 atomic %) is vaporised in an oven fabricated from 1.4 mm ID tantalum tube spot welded to a short length of tantalum foil through which an electric current is passed for heating to about 500°C. Electrons are generated by a heated tantalum filament located adjacent to the oven and biased to -24V.

The oven and filament are located 44 mm from the geometric centre of the ion trap. Ions are produced by collisions within the trap between the Yb atoms and electrons. A grounded plate with holes for the passage of the atom and electron beams is mounted between the ion trap and the oven and filament assembly to prevent the electric field from the biased filament perturbing the trapping potential during loading.

The trap is loaded with 380Vp-p at 350kHz applied to the main electrodes and 5V applied to the end electrodes. Loading is achieved by heating the oven for 40s and turning on the filament and its bias for 10s, starting 5s before the oven heating is turned off. The trap loading sequence is computer controlled for reproducibility.

Square Helmholtz coils with an edge length of 305 mm located around the outside of the vacuum chamber cancel most of the ambient magnetic field; however, no shielding is presently installed around the ion trap. Residual ferromagnetism of certain components of the vacuum system presently results in a rather unsatisfactory magnetic field variation of about 2 μT along the major trap axis (y direction) joining the two end electrodes.

The 369.5 nm and 609.1 nm laser beams propagate in the y direction along the major axis of the ion trap, passing

0-7803-0905-7/93 $3.00 © 1993 IEEE

through the holes in the end electrodes. The microwave field propagates in the x direction through a glass vacuum window from a horn located approximately 15cm from the centre of the trap. Both the laser and the microwave field (H vector) are linearly polarised in the z direction.

Fluorescence from the trapped ions emerging in the x direction is imaged onto a 0.1 cm slit by a cylindrical lens of 5 cm focal length, mounted 10 cm from the centre of the trap. The light passing through the slit and a bandpass filter (Newport UG11) is recorded by a photomultiplier. The net quantum efficiency of the detection system is approximately 0.1%.

Experiment

The energy levels of $^{171}Yb^{+}$ relevant to this work are shown in fig. 2.

The frequency of the $^{2}S_{1/2}$ F=0 - $^{2}P_{1/2}$ F=1 ground state hyperfine transition has recently been measured as 12,642,812,118.471±0.006 Hz[2]. The 609.1 nm transition is used in the laser cooling experiments to drain population accumulating in the metastable 4f145d $^{2}D_{3/2}$ F=1 level.

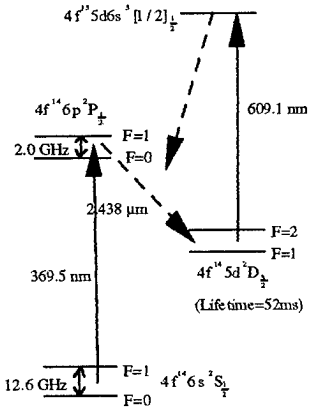


Figure 2: Partial energy level diagram of $^{171}Yb^{+}$ ions showing levels relevant to the present work. The solid arrowed lines indicate transitions driven by the applied laser or microwave fields, and the dashed lines indicate spontaneous transitions.

The experimental arrangement is shown in fig. 3. A titanium-sapphire laser with an intra-cavity frequency doubler is used to generate 369.5nm light. The beam was focussed to a diameter of 275 μm inside the trap, and the power measured at the entrance window of the trap was approximately 1 mW.

The 12.6 GHz microwave radiation is presently synthesised from the 5 MHz reference signal from a hydrogen maser. A commercial synthesiser generates a signal at 12.62 GHz, which is mixed with the output of a fine-tuneable local oscillator (Stanford Research DS345) tuned in the region of 22 812 120 Hz. The maximum available microwave power at 12.6 GHz at the feed to the microwave horn is -5 dBm.

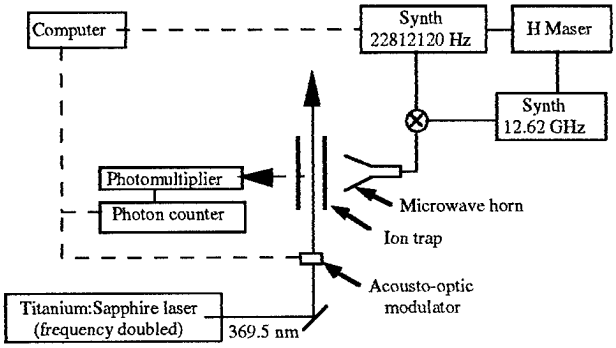


Figure 3: Simplified experimental arrangement

An excitation spectrum of a trapped ion cloud cooled by the introduction of 3 x 10^{-4} Pa He is shown in fig. 4.

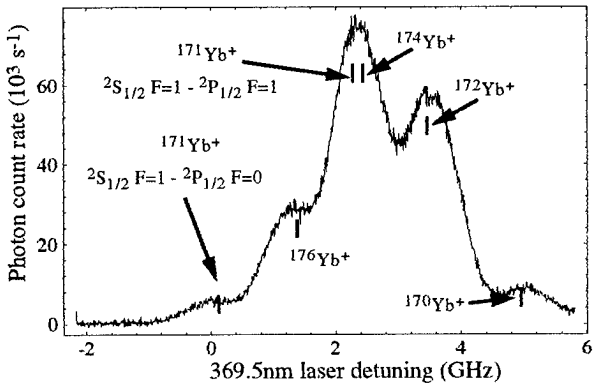


Figure 4: Excitation spectrum of a cloud of natural Yb^{+} cooled by 3 x 10^{-4} Pa He. The cloud was illuminated with approximately -5 dBm of 12.6 GHz microwave radiation. The approximate resonant frequencies of the even isotopes and $^{171}Yb^{+}$ are shown[7]. The peak due to the $^{2}S_{1/2}$ F=0 - $^{2}P_{1/2}$ F=1 transition is outside the range of this scan on the high detuning side.

Peaks due to the isotopes of Yb with even mass number are evident, as well as a peak due to $^{171}Yb^{+}$, resulting from illuminating the ion cloud with 12.6 GHz microwave radiation at the maximum available power. Without 12.6 GHz microwave illumination there is no signal from $^{171}Yb^{+}$, due to optical pumping of the ground state hyperfine levels. Fitting Gaussian profiles to the components of the excitation spectrum indicates an ion temperature of approximately 700 K.

Microwave-optical double resonance

For the microwave-optical double resonance experiments reported in this paper, the 369.5nm laser was tuned to resonance with the $^2S_{1/2}$ F=1 - $^2P_{1/2}$ F=1 transition in $^{171}Yb^+$ in a cloud of natural Yb^+ cooled by 3 x 10^{-4} Pa He.

The 609.1 nm laser was not used for these measurements because optical pumping into the metastable $^2D_{3/2}$ F=1 level was not found to be a problem, due to the short duty cycle of the 369.5 nm laser during these measurements.

The measurement sequence is as follows: The laser first illuminates the ions for about 500ms, which optically pumps nearly all of the population into the $^2S_{1/2}$ F=0 ground state hyperfine level. The laser is then blocked by the AO modulator. The ions are then illuminated by a pair of $\pi/2$ microwave pulses with Gaussian envelopes, separated by a delay time t_d. The laser then illuminates the ions again, and signal photons are counted for 10ms, during which time nearly all of the population driven by the microwave pulse sequence into the $^2S_{1/2}$ F=1 ground state hyperfine level is optically pumped back into the $^2S_{1/2}$ F=0 level. During the next 10ms, the photon counting system records mainly the background light level which results from resonant scatter of the 369.5nm laser light due mainly to $^{176}Yb^+$ and $^{174}Yb^+$, as well as stray scatter within the vacuum chamber. This background is subtracted from the signal. The laser continues to illuminate the ions for a further 480ms, to ensure that optical pumping into the $^2S_{1/2}$ F=0 ground state hyperfine level is complete. The cycle is then repeated, with a new microwave frequency.

Fig. 5 shows a Ramsey fringe pattern for a pulse delay t_d of 1s.

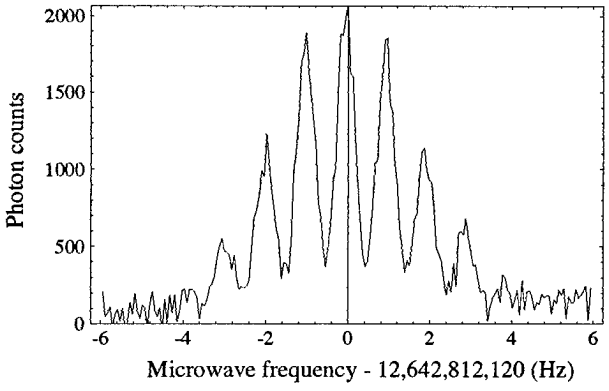


Figure 5: Ramsey fringes in $^{171}Yb^+$ generated using a microwave pulse separation t_d of 1 s. Each point is the sum of 10 pulse sequences. The trap was loaded with natural Yb^+.

The magnetic field at the centre of the trap was maintained at approximately 7.5 μT, with the field vector close to parallel with the polarisation directions of the laser and microwave fields. This magnetic field results in the frequency of the centre of the Ramsey pattern being higher than the recently measured value[2].

Laser cooling

Laser cooling of $^{171}Yb^+$ has also been demonstrated. For these experiments the trap was loaded with 88 atomic % isotopically enriched $^{171}Yb^+$ (which was not available to us when the microwave spectroscopy reported above was carried out). An excitation spectrum of a trapped ion cloud cooled by the introduction of 3 x 10^{-4} Pa He is shown in fig. 6.

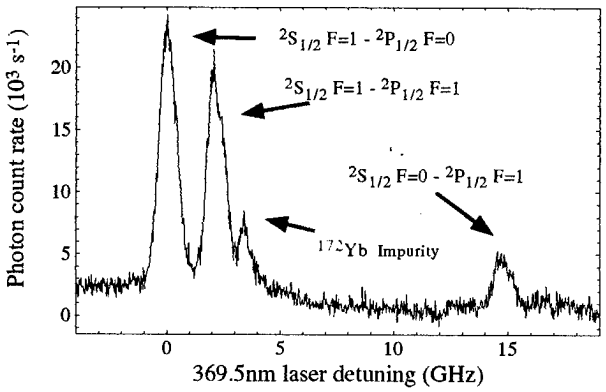


Figure 6: Excitation spectrum of a cloud of $^{171}Yb^+$ cooled by 3 x 10^{-4} Pa He. The cloud was illuminated with approximately -5 dBm of 12.6 GHz microwave radiation.

In order to achieve a rate of absorption and subsequent spontaneous emission processes sufficient for laser cooling, it was essential to prevent population accumulating in the metastable $^2D_{3/2}$ level and in the ground state hyperfine level out of resonance with the 369.5 nm laser light. A dye laser was therefore used to generate the 609.1 nm light, which passed through the ion cloud collinearly with the 369.5 nm laser light. The 609.1 nm laser power was approximately 45mW, with a beam diameter of 2.5mm inside the ion trap. For the cooling experiments the power of the 369.5 nm light was attenuated to approximately 100 μw.

Illumination with 12.6 GHz microwave radiation inhibited the ground state optical pumping process, which would otherwise prevent initial cooling of the ion cloud. Because the Rabi frequency of the 12.6 GHz ground state hyperfine transition was limited by the available microwave power to about 175 Hz, which is much less than the approximately 6 MHz Rabi frequency of the 369.5 nm interaction, cooling could only be achieved when the laser was tuned close to the $^2S_{1/2}$ F=1 - $^2P_{1/2}$ F=0 transition. Spontaneous emission from the $^2P_{1/2}$ F=0 excited state

level to the $^2S_{1/2}$ F=0 ground state level is dipole-forbidden, so that in this case population arrives in the $^2S_{1/2}$ F=0 ground state hyperfine level only via off-resonant optical excitation of the $^2S_{1/2}$ F=1 - $^2P_{1/2}$ F=1 transition and subsequent spontaneous emission, and via decay to the $^2D_{3/2}$ F=1, excitation by the 609.1 nm laser and then spontaneous emission to the ground state. These two processes result, in the absence of the microwave field, in approximately 1 s being required for almost complete optical pumping of the $^2S_{1/2}$ F=0 ground state hyperfine level to occur, as compared to a few milliseconds when the $^2S_{1/2}$ F=1 - $^2P_{1/2}$ F=1 optical transition is driven. The relatively weak microwave field is then sufficient to overcome this residual ground state optical pumping rate.

Cooling is achieved by reducing the confining RF and end electrode voltages after loading the trap[8] (this reduces RF heating of the ion cloud) and scanning the 369.5 nm laser upwards in frequency from about 200 MHz below resonance with the $^2S_{1/2}$ F=1 - $^2P_{1/2}$ F=0 transition over a period of about 3 minutes. No helium is added to the vacuum system for these experiments, and the system operates essentially at its base pressure. The laser directly cools only one degree of freedom of the ions' motion within the trap, and cooling of the other degrees of freedom occurs through Coulomb interaction between the ions[9]. A typical profile obtained in this way is shown in fig. 7.

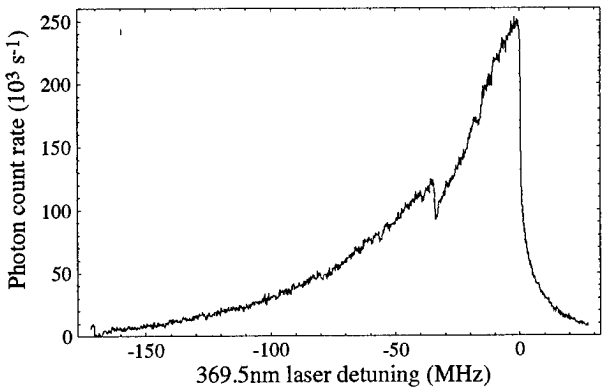


Figure 7: Excitation profile of cold $^{171}Yb^+$ ions. The sharp decay of the profile when the laser is above resonance with the optical transition (zero detuning on the x axis) is due to the resulting laser heating of the ion cloud.

The profile also exhibits a marked discontinuity at a detuning of -34 MHz. This type of discontinuity has been previously associated with a transition from a gaseous cloud phase to a crystalline arrangement of ions within the trap[10,11].

We estimate that between 10 and 100 ions are contributing to the signal shown in fig. 7. This estimate takes into account the photon count rate at the peak of the profile, the quantum efficiency of our detection system and the predictions of an approximate numerical model of the interaction of the 369.5 nm laser radiation and the 12.6 GHz microwave radiation with the four levels comprising the $^2S_{1/2}$ - $^2P_{1/2}$ optical transition. The unknown population remaining in the $^2D_{3/2}$ and other metastable levels[12] despite the presence of the 609.1 nm light is not taken into account, which means that our estimate may be lower than the actual number of cold ions.

The half width at half-maximum of the profile shown in fig. 7 corresponds to a temperature of approximately 0.7 K. Calculated in this way, the lowest temperature obtained to date is less than 0.2 K[8]. A temperature of 7 K corresponds approximately to a fractional second order Doppler shift of 10^{-14}.

In order to use the cold ions as a frequency standard, it will be necessary to maintain their low temperature with the 369.5 nm laser light blocked for at least a few seconds so that they may be probed by the microwave field under conditions free of light shifts and other perturbations introduced by the cooling laser. The time dependence of the ion temperature after the cooling laser is blocked will therefore need to be studied in detail. However, some preliminary observations are shown in fig. 8.

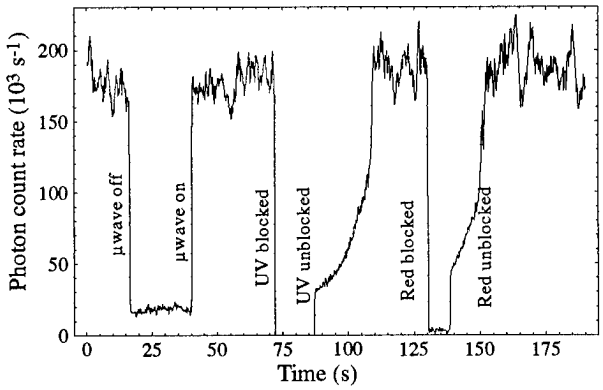


Figure 8: Behaviour of the 369.5nm fluorescence from an ion cloud cooled below 1K (see text).

In this case, a cold crystal is maintained with maximum microwave and 609.1 nm laser powers, and with the 369.5 nm laser tuned approximately 20 MHz below line centre. Other parameters are the same as for fig. 7. It was found that the microwave field can be turned off for more than 25 seconds, depending on the 369.5 nm laser detuning, without apparently destroying the crystalline phase. This is shown by the fact that the fluorescence count rate immediately returns to its original value when the microwave field is turned on again. However, after the 369.5 nm laser or the 609.1 nm laser is blocked for a shorter period, the fluorescence only gradually increases to the original value as the ion cloud cools to its original

temperature and returns to the crystalline phase. The increased noise in the photon count rate from the crystalline phase results from frequency fluctuations in the 369.5 nm light being transformed more effectively into photon count rate fluctuations by the steep slope of the line profile.

The fact that the microwave field can be turned off for such a relatively long period without apparently destroying the crystalline phase was an unexpected result. Furthermore, when the microwave field is absent the residual fluorescence count rate is significantly higher than when the 609.1 nm light is blocked.

There is therefore a process transferring population between the hyperfine levels in the ground-state, independently of the microwave field. The process most likely to be responsible is off-resonant excitation of the $^2S_{1/2}$ F=0 - $^2P_{1/2}$ F=1 optical transition by the 369.5 nm laser. Preliminary calculations show that the observed rate of population transfer is consistent with the predictions of the same simple numerical model used in the estimation of the number of cold ions.

Conclusion

We have presented preliminary microwave-optical double resonance spectroscopy on trapped $^{171}Yb^+$. Cooling of $^{171}Yb^+$ to <1 K has also been demonstrated. This temperature is sufficiently low to reduce the fractional second order Doppler shift of the microwave resonance to below 10^{-14}. The question of how long the ion cloud will stay cold when the cooling laser is blocked, so that microwave spectroscopy may be conducted without light shifts due to the laser light, remains to be answered.

References

[1] R. Casdorff, V. Enders, R. Blatt, W. Neuhauser, and P.E. Toschek, "A 12 GHz standard clock on trapped ytterbium ions," *Ann. Physik*, vol. 48, no. 7, 41, 1991.

[2] A. Bauch, D. Schnier, and C. Tamm, "Frequency measurement of the hyperfine splitting of $^{171}Yb^+$ ions stored in a Paul trap," in *7th European frequency and Time Forum*, 1993.

[3] A.S. Bell, P. Gill, H.A. Klein, A.P. Levick, and W.R.C. Rowley, "Precision measurement of the 2F7/2-2D5/2 3.43μm interval in trapped $^{172}Yb^+$," *J. Mod. Opt.*, vol. 39, no. 2, 381, 1992.

[4] C. Tamm, "A Tunable light source in the 370 nm range based on an optically stabilized, frequency-doubled semiconductor laser," 1993, To be published in App. Phys. B, 1993.

[5] J.D. Prestage, G.J. Dick, and L. Maleki, "New ion trap for frequency standard applications," *J. Appl. Phys.*, vol. 66, no. 3, 1013, August 1989.

[6] M.G. Raizen, J.M. Gilligan, J.C. Bergquist, W.M. Itano, and D.J. Wineland, "Linear trap for high-accuracy spectroscopy of stored ions," *J. Mod. Opt.*, vol. 39, no. 2, 233 , 1992.

[7] D.S. Gough, P. Hannaford, R.M. Lowe, and A.-M. Mårtensson-Pendrill, To be published, 1993.

[8] P.T.H. Fisk, M.A. Lawn, and C. Coles, "Laser cooling of $^{171}Yb^+$ ions in a linear Paul trap", To be published in Appl. Phys. B, 1993.

[9] I. Waki, S. Kassner, G. Birkl, and H. Walther, "Observation of ordered structres of laser-cooled ions in a quadrupole storage ring," *Phys. Rev. Lett.*, vol. 68, no. 13, 2007, March 1992.

[10] F. Diedrich, E. Peik, J.M. Chen, W. Quint, and H. Walther, "Observation of a phase transition of stored laser-cooled ions," *Phys. Rev. Lett.*, vol. 59, no. 26, 2931, December 1987.

[11] M.G. Raizen, J.M. Gilligan, J.C. Bergquist, W.M. Itano, and D.J. Wineland, "Ionic crystals in a linear Paul trap," *Phys. Rev. A*, vol. 45, no. 9, 6493, May 1992.

[12] A. Bauch, D. Schnier, and C. Tamm, "Collisional population trapping and optical deexcitation of ytterbium ions in a radiofrequency trap," *J. Mod. Opt.*, vol. 39, no. 2, 389, 1992.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Improved Linear Ion Trap Physics Package[1]

John D. Prestage, Robert L. Tjoelker, G. John Dick, and Lute Maleki

California Institute of Technology
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109

Abstract

This article describes an improvement in the architecture of the physics package used in the Linear Ion Trap (LIT) based frequency standard recently developed at JPL. This new design is based on the observation that ions can be moved along the axis of an LIT by applied dc voltages. The state selection/interrogation region can be separated from the more critical microwave resonance region where the multiplied local oscillator (LO) signal is compared to the stable atomic transition. This separation relaxes many of the design constraints of the present units. Improvements include increased frequency stability, and a substantial reduction in size, mass and cost of the final frequency standard.

Introduction

The primary reason that ions confined to electromagnetic traps are the basis for stable frequency standards is that in the containerless environment of a trap, processes that equalize atomic state populations and destroy coherence within the state prepared atomic ensemble are very weak. In past work we have measured coherence times in an ensemble of trapped Hg^+ ions of over 30 seconds on the 40.5 GHz transition. Such weak relaxation has permitted an atomic line-Q on the 40.5 GHz transition of $2x10^{12}$, the highest ever measured on a microwave atomic transition[1,2].

One disadvantage of a trapped ion based frequency standard is the relatively low density of the confined ions. This leads to low signal to noise ratio (SNR) in the detected atomic resonance and consequently limits clock stability. This situation was greatly improved by the use of the linear ion trap[2] where the number of trapped ions was increased by more than 10 over conventional hyperbolic traps. The stability reached with the improved SNR of the LIT together with the high line-Q has led to a 10 fold improvement in clock stability over conventional ion based clocks. Indeed, the present LIT based clock shows stability competitive with the best H-masers for averaging times less than 10,000 seconds and exceeds H-maser stability beyond 10,000 seconds [3] making it the most stable of all clocks for long term stability. The LIT clock is, however, relatively recent in its development and its configuration continues to evolve. The purpose of this article is to propose an improvement in the architecture of the current physics package which will lead to improved long term stability, and a substantial reduction in size, mass and cost of the final frequency standard.

Current LIT Operation

The present form of the LIT physics unit for frequency standard operation is shown in Figure 1. Ions are created inside the linear trap by an electron pulse which ionizes a weak vapor of parent neutral atoms introduced into the vacuum system from a heated HgO isotopic source. They are held in the four bar linear trap shown inside the region enclosed by mag-

[1] This work represents one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

0-7803-0905-7/93 $3.00 © 1993 IEEE

netic shields. Before the stable atomic frequency of the trapped ions can be compared to the multiplied output of an LO or flywheel oscillator a population difference between the hyperfine levels of the ground state must be created. This is done via optical pumping with UV light from a laser or a discharge lamp. As shown in Fig. 1, optical pumping is carried out by a ^{202}Hg lamp whose output light is collected and focussed onto the central portion of the ion cloud.

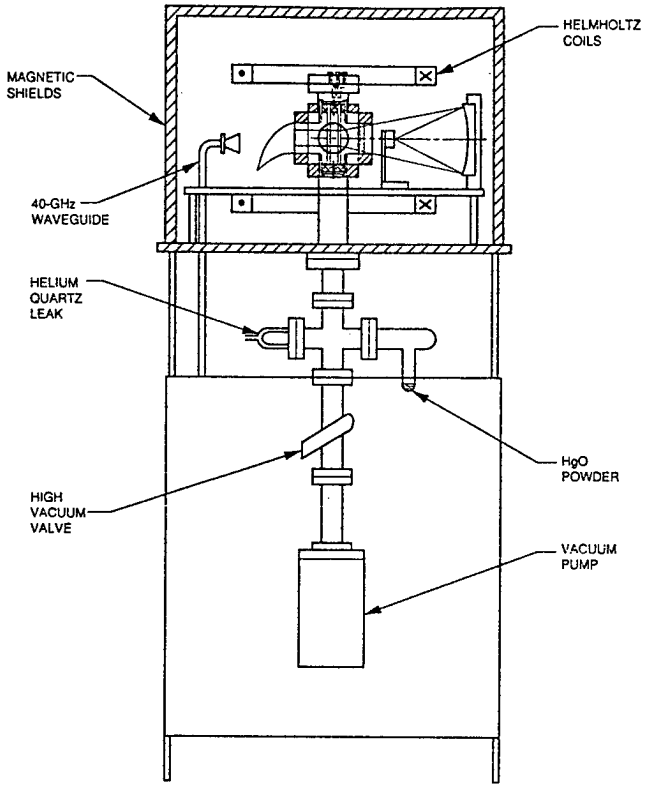


Figure 1: Present physics unit for the LIT based frequency standard. ^{199}Hg ions are state prepared and interrogated with the multiplied output of a local oscillator in the ion trap housed inside the magnetic shielded region. The apparatus is about 2 m high by 1 m on a side.

After state preparation, microwave radiation derived from the LO is passed onto the ions to make a frequency comparison with the stable atomic resonance. During this period it is crucial that the atomic resonance frequency not be perturbed by any changes in the trap environment. Such fluctuations will be transferred to the servoed LO thereby degrading clock stability. Because the state selecting light will shift the atomic clock resonance it is switched off during the microwave interrogation time. Following the microwave pulse, the lamp is turned on again to determine whether the microwave radiation has changed the population of the hyperfine levels of the atomic ion. Any frequency detuning of the LO from the reference atomic frequency will change the fluorescent light intensity measured when the lamp is turned on. These fluorescence changes are converted to a voltage and fed back to the frequency control port of the LO to keep it on frequency with the atomic oscillator.

It is apparent that the trap region is operated in two modes in the process of steering an LO. First, to prepare the atomic ions for microwave frequency comparison to the LO. In the second mode the atomic frequency is compared to the multiplied output of the LO. During the state preparation/interrogation mode there are no stringent requirements on environmental isolation or regulation whereas during the resonance mode great care must be exercised in the regulation of the atomic environment to insure stable operation.

Improved LIT Design

Figure 2 shows a configuration of an ion trap based clock where these modes of operation are carried out in two separate regions. The trap shown is a lengthened version of the the LIT of Fig. 1. In the new architecture of Fig. 2 the region labelled ion loading and fluorescence is somewhat smaller in size to the trap shown in Fig.1. A second linear trap has been added for a microwave resonance region.

A dc break forms a junction between these two regions to allow ions to pass from one region to another. The rf trapping voltage for transverse containment of the ions is continuous across this dc break. When the dc voltage level of all four trap rods are the same in both the upper and lower regions thermal motion of the ions will carry them through the junction with no change in axial velocity. When the four trap rods in the upper region are at negative dc voltage with respect to the lower region trap rods ions within a trap radius or so of the junction will be transported across the junction into the upper region. Only the ions near the junction will experience the electric field forcing them across the gap. Since each ion is in thermal motion along the axis of the trap it will reach the junction within a trap length transit time (typically about 1 millisecond) and then be pulled into the upper region emptying the lower region of ions. Similarly, when the upper region is dc biased positive with

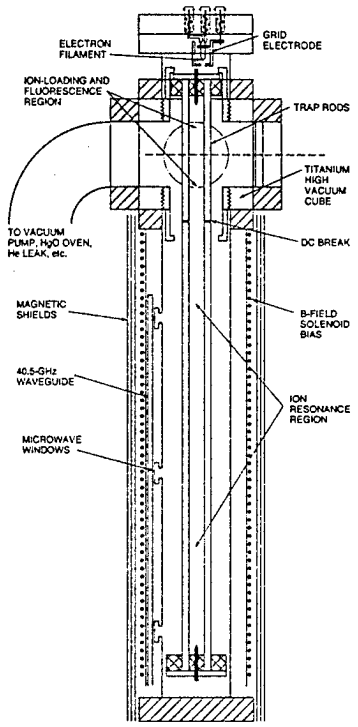


Figure 2: Improved physics unit for the trapped ion frequency standard.[199]Hg ions are created and state prepared in the upper region of the trap, then transported into the lower region, where the Hg^+ atomic resonance frequency is compared with the local oscillator frequency. The ions are then moved back to the state selection region to check for frequency detuning from the LO during the resonance comparison. Overall dimensions are 10 cm by 50 cm.

respect to the resonance region all ions will be transported from the fluorescence region to the ion resonance region. The helium buffer gas will provide the required energy damping with a 1/3 second time constant[4]. The dc potential along the length of the trap which controls which trap the ions occupy is shown in Fig. 3.

The separation of the clock resonance region from the optical/ion loading region relaxes many of the design constraints of the present units. The microwave resonance region can now be designed with no consideration of optical issues. A simple set of cylindrical shields and solenoid supply a very uniform and stable magnetic environment. In Figure 2 it is seen that the volume of the resonance region is now reduced by about 100 as the magnetic shields are 10 times smaller in diameter. Since the resonance region requires the most stringent temperature regulation of the entire physics package the size reduction enables simplification of thermal control of the standard.

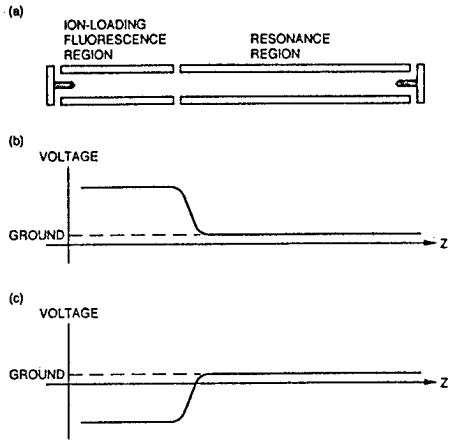


Figure 3: The improved LIT with separated state selection and microwave resonance regions: (a) the complete ion trap showing the dc junction; (b) a positive dc bias is applied to the loading and interrogation trap rods forcing the ions into the resonance region; and (c) ions are drawn back into the fluorescence region by applying a negative bias.

Similarly, the state selection optical design is now done with no worry of perturbing the atomic resonance. In practice, this means that the optical components are no longer required to be non-magnetic as in the previous design. The custom made Copper-Titanium flanged UV transmitting windows can be replaced with S-steel flanged windows which are each about 2 k$ less expensive. The 4 k$ non-magnetic UV light collecting horn is replaced by an S-steel elbow connecting the trap region to the vacuum pumps. The light source can be moved much closer to the ion cloud raising the possibility of a collimated UV source with no collection mirror thus further reducing the finalpackage size. Similar modifications could be made in the collection arms to reduce size. These design changes would not allow stable clock operation if the ions underwent microwave clock resonances inside the optical state selection region as in the previous design, because of the large magnetic shifts of the atomic levels induced by the close proximity of the light source and detectors. Another simplification gained by relaxing the non-magnetic requirement is in the electron gun assembly and its heater current supply. The electron source for ion creation is now remote from the atomic resonance region allowing use of the standard (magnetic) filament base and a floatable dc current drive. In the design of Fig. 1 ions are created inside the resonance region and require a floating audio frequency

filament heater current to prevent the residual dc magnetic field from the electron gun from shifting the atomic ion frequency.
Additionally, the largest frequency offset and potential instability can be reduced greatly by selecting a somewhat longer trapping length in the atomic resonance region. This frequency shift stems from the finite (i.e., non-zero) ion cloud diameter where ions spend time in regions of large rf trapping fields where thier motion results in frequency pulling via the second order doppler or relativistic time dilation effect. The magnitude of this offset depends only on the linear ion density, N/L, and is given by[5]

$$(\Delta f/f)_{2nd\ Doppler} = -(q^2/8\pi\epsilon_0 mc^2)N/L.$$

An increase in the resonance trapping length to about 200 mm would reduce the clock sensitivity to ion number variations by about a factor of 4 from that in the present system of Fig. 1. Since ion number stabilization to the 0.1% level has been demonstrated for 10^{-15} frequency standard operation in that system, an ion number induced instability noise floor below 2.5×10^{-16} should be achievable in this modified configuration.

Conclusions

An improved architecture for a linear ion trap based frequency standard is proposed. Its design is a natural extension of the original LIT and separates the state selection and ion creation region from the atomic resonance region. Many of the design constraints of the present configuration are eliminated and a smaller, cheaper more stable frequency standard will result.

References

[1] J. D. Prestage, R. L. Tjoelker, G. J. Dick and L. Maleki, "Ultra-stable Hg^+ trapped ion frequency standard," Journal of Modern Optics, vol. 39, pp. 221-232, 1992.

[2] J. D. Prestage, R. L. Tjoelker, R. T. Wang, G. J. Dick, and L. Maleki, "Hg^+ Trapped Ion Standard Performance with the Superconducting Cavity Maser Oscillator as L.O.," Proc. 1992 IEEE Frequency Control Symposium, pp. 58-63, 1992.

[3] R. L. Tjoelker, J. D. Prestage, G. J. Dick, and L. Maleki, "Long Term Stability of Hg^+ Trapped Ion Frequency Standards," Proc. 47th Ann. Symp. Freq. Control, June 1993.

[4] L. S. Cutler, R. P. Giffard, and M. D. McGuire, "Thermalization of ^{199}Hg Ion Macromotion by a Light Background Gas in an rf Quadrupole Trap," Appl. Phys. B36, pp. 137-142,1985.

[5] J. D. Prestage, G. J. Dick, and L. Maleki, "New Ion Trap for Frequency Standard Applications," J. Appl. Phys., vol. 66, pp. 1013-1017, 1989.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DOPPLER SIDEBAND SPECTRA FOR IONS IN A LINEAR TRAP[1]

John D. Prestage, Robert L. Tjoelker, G. John Dick, and Lute Maleki

California Institute of Technology
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109

Abstract

We describe a spectroscopic measurement of the temperature and linear density of Hg^+ ions held in a linear ion trap (LIT). The inferred temperature and number result from analysis of sidebands on the 40.5 GHz resonance line. The temperature of the ion cloud is determined by the Doppler broadened line when microwave radiation is propagated along the axis of the LIT. When propagation is perpendicular to the trap axis the microwave sidebands are displaced from the trap secular frequency by an amount dependent upon the ion cloud size and temperature. A Monte-Carlo simulation of the ion trajectories inside the cloud is used to model the position of these sidebands for a given cloud temperature and a comparison with measured sideband position is used to determine ion number.

Introduction

One of the largest frequency offsets in lamp based buffer gas cooled Mercury ion frequency standards is the second order Doppler shift from the thermal and micro-motion of the ions within the trap[1,2]. At room temperature with Hg^+ ions in a LIT this shift is $(1+2/3)3k_BT/(2mc^2)=3.5x10^{-13}$ for a small cloud where space charge screening of the time averaged trapping field is negligible. In this case a measurement of ion cloud temperature accurate to 10% would yield a room temperature Hg clock with accuracy at few parts in 10^{14}, approaching that of the best primary Cesium standards currently in operation. This paper describes temperature measurements and ion number determination for ion clouds held in a linear ion trap. Such measurements are necessary to assess clock frequency offset from the 2nd order Doppler shift and to determine clock accuracy[3]. Knowledge of cloud size and temperature also serves as a useful diagnostic of ion trap operation.

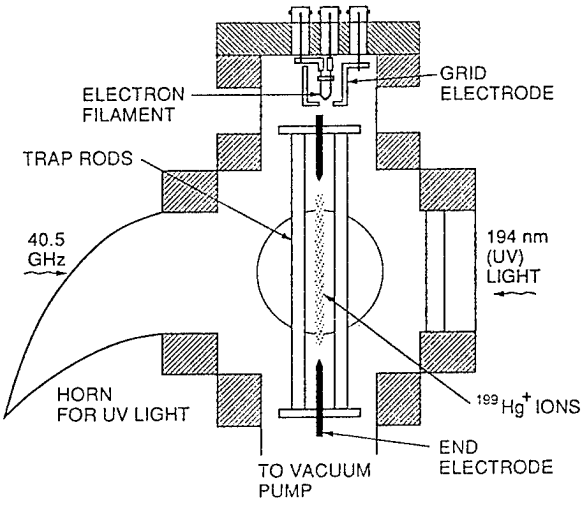


Figure 1: Linear ion trap residing in its high vacuum enclosure. State selection light from the ^{202}Hg discharge lamp is focused onto the central 1/3 of the of the ion cloud. Atomic fluorescence is collected normal to the page.

[1] This work represents one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

0-7803-0905-7/93 $3.00 © 1993 IEEE

Figure 1 shows the present LIT configuration as used in clock operation[4]. Ions are confined in the cylindrical cloud up to 75 mm in length by a few mm in diameter. An ion's thermal motion inside the cloud will lead to a high frequency secular motion corresponding to transit across the few mm cloud diameter and a much lower frequency corresponding to axial motion along the approximately 75 mm long cloud. When there are only a few ions in the trap the resultant forces on the ions are predominately from the trapping fields. The transverse motion is then harmonic and all ions undergo motion at the secular frequency of the trap. As the 40.5 GHz radiation propagates radially through the cloud an ion's motion in the trap will modulate the phase of the microwaves. This results in sidebands on the atomic absorption spectrum offset from the 40.5 GHz carrier by the trap secular frequency and its harmonics. As more ions are added to the trap the cumulative effect of the trapped charges begins to shield the time averaged trapping fields from the interior of the ion cloud. Ions moving inside the cloud will no longer feel the harmonic trapping well and the result is a broadened and downshifted Doppler sideband spectrum. We have computed the spectrum of microwave power seen by an ion moving through the cloud by a Monte-Carlo sampling of ion trajectories and the resulting phase variation. For a fixed cloud temperature the Doppler sideband shifts downward monotonically with increasing number of trapped ions. Once the ion temperature has been measured the cloud size is determined by adjusting ion number until the simulated and measured sideband frequencies match.

Ion Temperature Measurement

We measure ion temperature spectroscopically by propagating the 40.5 GHz resonance radiation along the 75 mm long axis of the ion cloud. In this configuration ion confinement is 10 times the wavelength of the 7.5 mm resonance radiation and the Doppler width of the 40.5 Ghz absorption line is used to determine the cloud temperature[5]. Because the 50 milli-gauss bias field is parallel to this axis of the trap and hence perpendicular to the oscillating magnetic field of the microwave radiation, the allowed transitions obey $\Delta m_F = \pm 1$. These transitions are from the F=0, $m_F=0$ state to the F=1, $m_F=\pm 1$ Zeeman states. Figure 2 shows the Doppler broadened microwave lines obtained by propagating microwaves along this direction. The width of the microwave transition is $\delta\nu = 2(\nu_0/c)(2k_BT \cdot \ln 2/m)^{1/2} = 2.056T^{1/2}$ kHz [6], where T is temperature in Kelvin. The 46 kHz wide spectral lines in Fig. 2 indicate an ion cloud temperature of 500 K. By use of this technique we have measured ion cloud temperature variations with helium buffer gas pressure. As shown in Figure 3 the ion temperature varies from 2610 K at 2×10^{-8} Torr(uncorrected) to 480 K at 7×10^{-6} Torr(uncorrected). For the clouds used in our clock measurements Helium pressures above 2×10^{-6} Torr(uncorrected) are necessary for minimum temperatures.

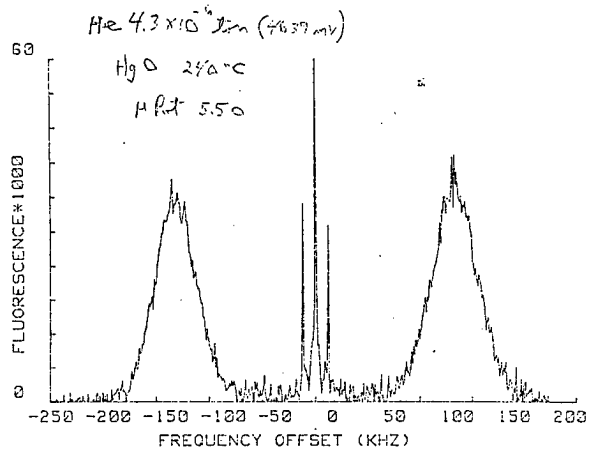


Figure 2: The Doppler broadened Zeeman hyperfine transitions F=0,m_F=0 to F=1,$m_F=\pm 1$ which are used to determine the Hg^+ temperature. For this measurement the 40.5 GHz radiation passed axially through the 75 mm long cloud.

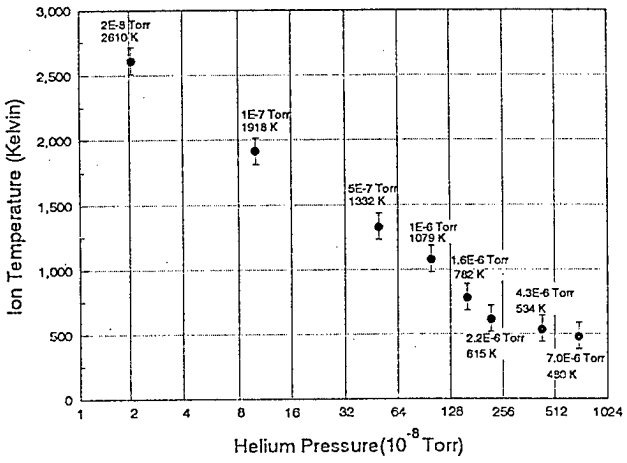


Figure 3: Hg^+ cloud temperature vs helium buffer gas pressure as determined from the doppler width of the 40.5 GHz transition.

Transverse Microwave Excitation of Ion Cloud

By propagating the 40.5 GHz resonance radiation perpendicular to the long axis of the ion cloud we can investigate the absorption sidebands which stem from the high frequency secular ion motion across the cloud diameter. The measured resonance structure is shown in Figure 4 (a) and (b). In Fig. 4(a) the helium pressure is $5x10^{-8}$ Torr(uncorrected) so ion temperature is over 2000 K. The calculated secular frequency at the operating rf voltage level (140 Volts rms @ 853 kHz) is 62 kHz and is in good agreement with the frequency of the transverse sidebands. This indicates low ion density since there is no significant screening of the trapping fields by the ion space charge. By contrast Fig. 4(b) shows that the sidebands have shifted closer to the 40.5 GHz carrier by about 10 kHz as the helium pressure was increased to $2x10^{-6}$ Torr(uncorrected). The signal size has also quadrupled since many more ions are trapped under these conditions.

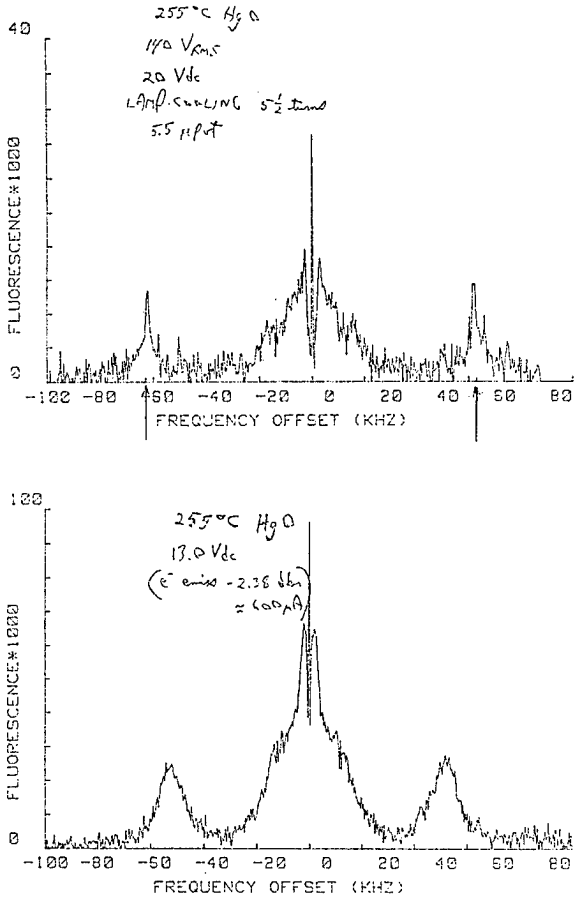


Figure 4: Sideband spectra obtained by transverse microwave illumination of the ion cloud as shown in Fig. 1. (a) Low helium buffer pressure and resulting ion temperature over 2000 K. Sidebands show very little shift from the 62 kHz trap secular frequency. (b) Sidebands shift by 10 kHz as the helium pressure is increased and ion temperature falls to 650 K.

Monte-Carlo calculation of Sideband Frequencies

The approach used in predicting the frequency of the secular sidebands as described above is to determine the microwave frequency spectrum seen by an ion inside the cloud[7]. A spherical wavefront is incident on the ion cloud from a distance R_0. The magnetic field seen by an ion at position r at time t is

$$B(r,t)=B_0\exp[i(k_{\mu w}r-\omega t)]/(k_{\mu w}r)$$

where $k_{\mu w}=2\pi/\lambda_{\mu w}$ and $\lambda_{\mu w}$ is the wavelength of the microwave radiation, $\approx$7.5 mm. The ions are in thermal motion inside the cloud so that r=r(t) and the spherical wave is modulated by the individual ion motions. This motion is determined by the velocity distribution and the shape of the ion cloud. The long thin cloud leads to microwave modulation frequencies of order v_{mp}/D and v_{mp}/L where v_{mp} is the most probable velocity for the ions at temperature T, D is the cloud diameter and L is the cloud length. The power spectral density of the 40.5 GHz radiation as seen by the ensemble of confined moving ions is related to the autocorrelation function of the microwave magnetic field through the cosine transform (Wiener-Khintchine Theorem)

$$P(\omega)=\int d\tau\psi(\tau)\cos\omega\tau$$

where $\psi(\tau)$ is the autocorrelation function

$$\psi(\tau)= B_0^2\mathrm{Re}[\exp(-ik_{\mu w}(r(\tau)-r(0))/(k_{\mu w}^2r(0)r(\tau))]$$

The averaging is to be carried out over the ionic trajectories through the cloud. To generate this average a set of initial conditions consistent with the velocity and position distribution functions is chosen with a trajectory being generated through the potentials seen inside the cloud. Ignoring collisions, an ion moves in response to the space charge potential of the other ions and the harmonic trap potential. The

starting point is calculation of the density profile for a given number of ions at temperature T in a cylindrically symmetric harmonic trapping well of secular frequency ω [8]. The ion cloud is assumed to be of uniform linear density up to the ends of the trap.

Boltzmann Density Profile

We assume that the ion cloud is in thermal equilibrium at temperature T so that the ion number density $n(\rho)$ satisfies a Boltzmann distribution [7]

$$n(\rho)=n(0)\exp[-\Phi_{total}(\rho)/k_BT]$$

where $\Phi_{total}(\rho)=m\omega^2\rho^2/2+q\varphi_{sc}(\rho)$ is the sum of the trapping field and ion cloud space charge potential energies, respectively and ρ is the radial distance from the trap axis. Poisson's equation ($\nabla^2\varphi_{sc}(\rho)=-qn(\rho)/\epsilon_0$) leads to a non-linear differential equation for $n(\rho)$

$$n''+n'/\rho-(n')^2/n+n(n_0-n)/(n_0\lambda_D^2)=0$$

where $n_0=2\epsilon_0m\omega^2/q^2$ is the pseudocharge or saturation density for ions in the trap. The Debye length $\lambda_D^2=k_BT\epsilon_0/(n_0q^2)=k_BT/(2m\omega^2)$ is a measure of the penetration length of the time averaged trapping fields into the ion cloud interior. This equation is numerically solved (subject to $2\pi\int\rho n(\rho)d\rho=N/L$) via Runge-Kutta methods to give density profiles as shown in Figure 5. The 62 kHz trap secular frequency leads to a saturation density of 34,840 ions per mm^3 which at 500 K ion temperature gives a Debye length 0.26 mm. From $4\pi n_0a^3/3=1$ we estimate an inter-ion spacing a=19 μm at saturation density. The density profile determines the total second order shift of the trapped ions as discussed in Appendix 1.

Ion Orbits

After the density equation has been solved the total potential seen by an ion is computed from $\Phi_{total}(\rho)=k_BT\cdot\ln(n(0)/n(\rho))$. An orbit in this 2-dimensional central force has two constants of motion [9], the angular momentum l_z around the z axis and the total energy in the (ρ,θ) plane E_{trans}, i.e., the kinetic plus potential energies. Following the procedures described in appendix 2 an initial position and velocity are generated consistent with the radial density profile for a set linear density N/L and the

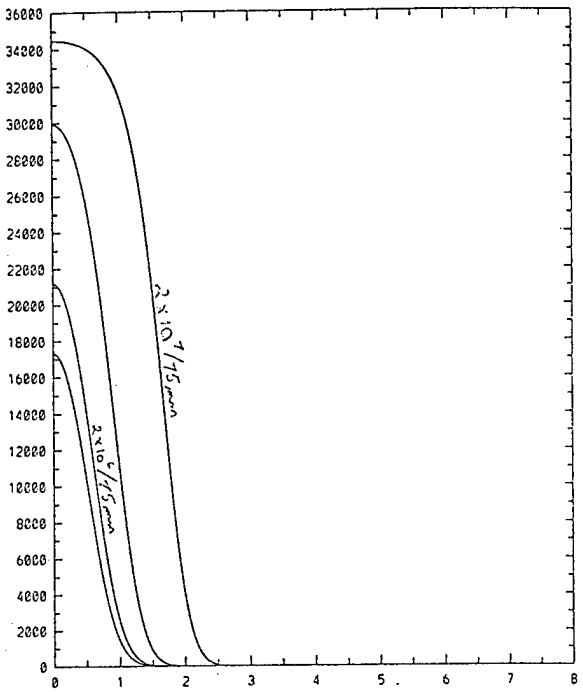


Figure 5: Calculated ion density profiles for 500 K ion temperature at 62 kHz trap secular frequency. The highest density shown is $2x10^7$ ions in 75 mm length.

Maxwell-Boltzmann velocity distribution at temperature T. l_z and E_{trans} are computed at this initial point and the radial turning points of the orbit are found by solving $E_{trans}-\Phi_{total}(\rho)-l_z^2/2m\rho^2=0$. The orbits are integrated to the radial turning points to find the angular turning points [9]

$$d\rho=(2(E_{trans}-\Phi_{total}(\rho)-l_z^2/2m\rho^2)/m)^{1/2}dt$$

$$d\Theta=(l_z/m\rho^2)dt.$$

Since the orbits are invariant under reflection about the turning points [9] the complete trajectory is generated by a number of successive reflections. The motion of an ion along the z-axis of the trap is assumed to be a uniform velocity (generated from the velocity distribution as described in Appendix 2) back and forth between the trap ends. For each initial condition a trajectory of duration $\approx$ 4 msec is generated. A time scale common to all the trajectories is generated so that the autocorrelation function at each time step τ_n is averaged over the 4,000 to 10,000 trajectories typically employed. This function is then Fourier transformed to give the power spectral density seen by the ensemble of moving atoms.

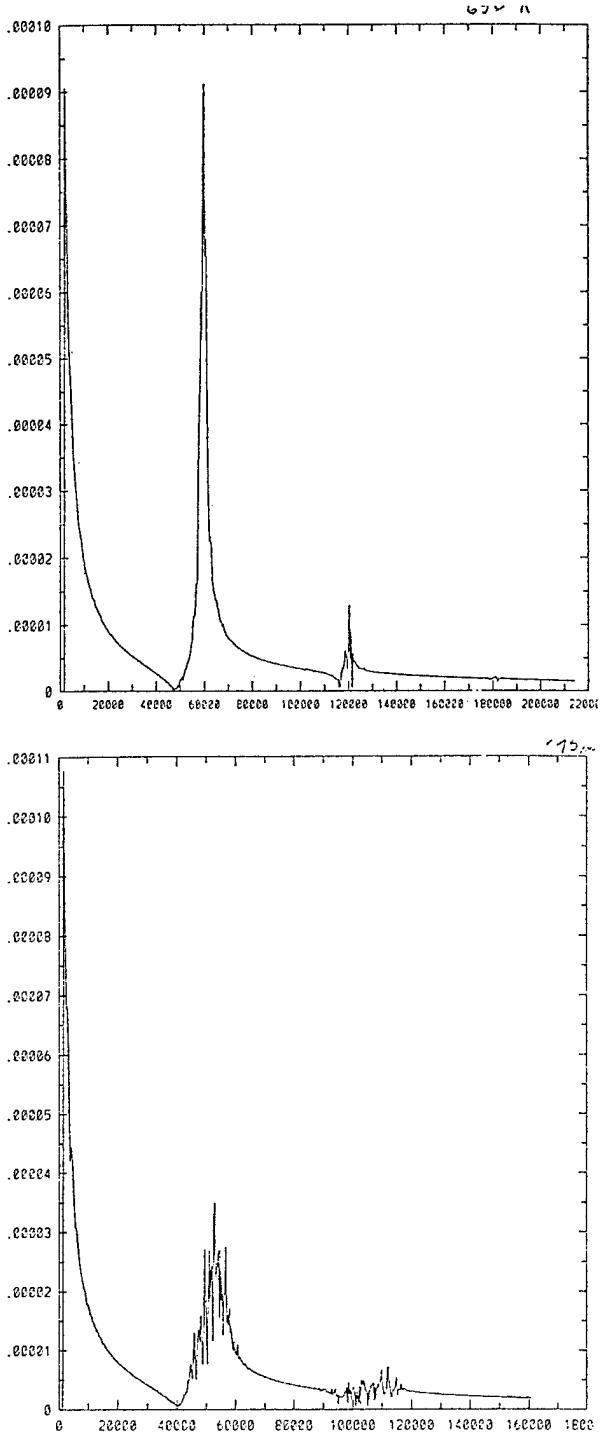


Figure 6: Simulated sideband structure for transverse propagation of 40.5 GHz radiation across ion cloud. The upper curve corresponds to 650 K with $5x10^5$ ions in 75 mm length. The lower curve at 650 K with $2.5x10^6$ ions in 75 mm has sidebands shifted in by 10 kHz from the 62 kHz trap secular frequency in agreement with the measured shift of Fig. 4(b).

Comparison with measured Sideband Frequency

Figure 6 shows the simulated sideband structure for a cloud of ions at 650 K of linear density $5x10^5$ ions per 75 mm at 62 kHz trap secular frequency. Under these conditions the sidebands are sharp and are slightly displaced inward from the 62 kHz secular frequency. As the linear density is increased to about $2.5x10^6$/75m-m the sidebands shift inward by about 10 kHz as shown in the measurement of Figure 4(b). The doppler frequency shift under these conditions is $-4.5x10^{-13}$ from thermal motion and $-4.3x10^{-13}$ from the micromotion generated by the trapping field. These are the approximate conditions for which the Hg^+ clock stability of $7x10^{-14}\tau^{-1/2}$ was obtained [4].

Conclusions

We have developed a spectroscopic method for measuring temperature, ion number and the resulting 2nd order Doppler shift of the 40.5 Ghz clock transition for Hg^+ ions in a linear ion trap. The ion cloud temperature is obtained from the Doppler broadened width of the 40.5 GHz Zeeman transition when the microwaves are propagated along the axis of the LIT. The ion number is then determined by modelling the frequency of the secular sidebands obtained when the 40.5 GHz radiation propagates across the ion cloud. In clock operation with $7x10^{-14}\tau^{-1/2}$ stability we have used this method to measure a $-8.8x10^{-13}$ frequency offset due to 2nd order Doppler shift at 650 K and linear ion density $2.5x10^6$/75 mm.

Appendix 1: Total Second-Order Doppler Shift for fixed T and N/L

The resonance frequency, f_0, of an ion moving with total velocity V will be shifted by Δf from that of a stationary ion by $\Delta f/f_0 = -V^2/2c^2$. For an ion in an rf trap the velocity is the sum of a driven motion V_{micro} and a random thermal motion V_{th}. The micromotion velocity V_{micro} is determined by an ion's position in the trap whereas the magnitude and direction of V_{th} is uncorrelated with an ion's position. The average second order doppler shift for an ensemble of ions at temperature T is

$$\overline{-(V_{micro}+V_{th})^2/2c^2} = -(V_{micro}^2+V_{th}^2)/2c^2$$

because the average over the product $V_{micro} \cdot V_{th}$ vanishes. The two sources can thus be computed separately. Since the average of V_{th}^2 over the velocity distribution

is $3k_BT/m$, the contribution from the thermal motion is $-3k_BT/2mc^2$.

The contribution from the micro-motion is the average of V_{micro}^2 over the density distribution $n(\rho)$ as determined numerically from eq (1). The time averaged micro-motion grows with radial distance ρ, $V_{micro}^2=\omega^2\cdot\rho^2$ [10] so that

$$\overline{V_{micro}^2}= \int n(\rho)\rho\omega^2\rho^2 d\rho / \int n(\rho)\rho d\rho.$$

In the limit of large λ_D, the solution to eq. (1) is $n(\rho)=n(0)\exp(-\rho^2/4\lambda_D^2)$ and the 2nd doppler from trap field generated ion motion is $-k_BT/mc^2$. This is 2/3 of the thermal contribution and stems from the fact that rf or ponderomotive forces provide the trapping in 2 of the 3 spatial dimensions. By contrast, when λ_D is very small compared to the ion cloud radius, the solution to eq. (1) is $n(\rho)=n_0$ out to the cloud radius $R_c=((N/L)/n_0\pi)^{1/2}$. In this limit, the trap field induced 2nd order doppler shift is $-(q^2/8\pi\epsilon_0 mc^2)N/L$.

Appendix 2: Selection of Initial Positions and Velocities

The selection of the six initial position and velocity coordinates for an ion ensemble at temperature T with density distribution $n(\rho)$ is at the heart of the averaging procedure used to compute the autocorrelation function. We follow the technique outlined in references [11,12] and will illustrate selection of the initial radial position coordinate. We compute the probability distribution for an ion's radial position from the computed density profile $n(\rho)$. The number of ions found between ρ and $\rho+d\rho$ is $2\pi n(\rho)\rho d\rho$ so that the probability distribution for radial position $p(\rho)$ is

$$p(\rho)=n(\rho)\rho/(\int_0^\infty n(\rho)\rho d\rho).$$

A random number, ξ, is selected in the interval (0,1) with the corresponding radial position ρ_ξ selected so that $\xi=\int_0^{\rho_\xi} p(\rho)d\rho$. The selection frequency of the radial position ρ is then in agreement with the distribution $p(\rho)$. All six of the initial values are chosen in this manner, the magnitude of the ion velocity from the Maxwell-Boltzmann distribution at temperature T, the angle of the ion velocity along the trap longitudinal direction (z-axis) from $\Theta_\xi=\cos^{-1}(1-2\xi)$, and so forth. For each of the six initial values an independent random number ξ is chosen so that there is no correlation. A representative histogram of initial velocity components along the z direction generated from the speed and direction sampling is shown in Figure A(1) for an ion temperature of 500 K.

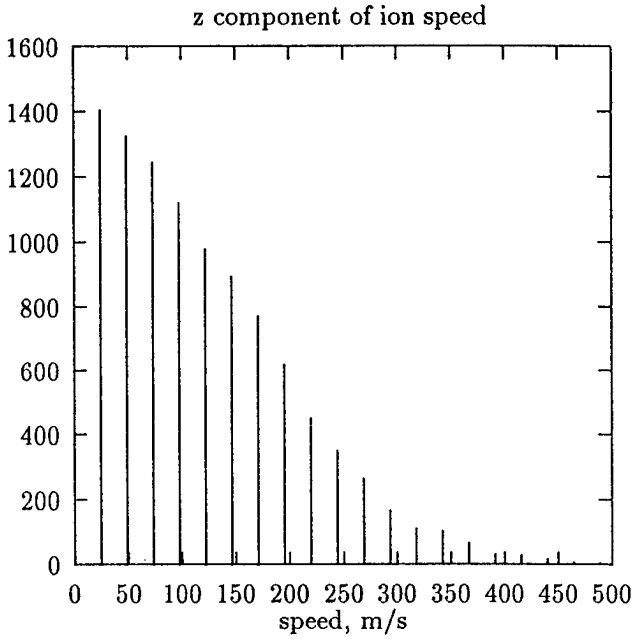


Figure A(1): The generated distribution of the z component of ion velocities. This histogram is created from two independently generated distributions- the Maxwell distribution of ion speeds and the velocity direction angle distribution with respect to the trap longitudinal axis.

References

[1] J. D. Prestage, R. L. Tjoelker, R. T. Wang, G. J. Dick, and L. Maleki, "Hg^+ Trapped Ion Standard with the Superconducting Cavity Maser Oscillator," IEEE Trans. Instr. Meas., vol. 42, No. 2, April 1993.

[2] L. S. Cutler, R. P. Giffard, P. J. Wheeler, and G. M. R. Winkler, "Initial Operational Experience with a Mercury Ion Storage Frequency Standard," Proc 41st Ann. Symp. Freq. Control, pp. 12-19, 1987.

[3] L. S. Cutler, R. P. Giffard, and M. D. McGuire, "Thermalization of ^{199}Hg Ion Macromotion by a Light Background Gas in an RF Quadrupole Trap," Appl. Phys. B 36, pp. 137-142, 1985.

[4] R. L. Tjoelker, J. D. Prestage, G. J. Dick, and L. Maleki, "Long Term Stability of Hg^+ Trapped Ion Frequency Standards," Proc. 47th Ann. Symp. Freq. Control, June 1993.

[5] R. H. Dicke, "The Effect of Collisions upon the Doppler Width of Spectral Lines," Phys. Rev. 89, pp. 472-473, Jan. 1953.

[6] A. Corney, Atomic and Laser Spectroscopy. Oxford: Clarendon Press, 1977, ch. 8, pp. 248-249.

[7] L. S. Cutler, C. A. Flory, R. P. Giffard, and M. D. McGuire, "Doppler Effects due to Thermal Macromotion of Ions in an RF Quadrupole Trap," Appl. Phys. B 39, pp. 251-259, 1986;
C. Meis, M. Desaintfuscien, and M. Jardino, "Analytical Calculation of the Space Charge Potential and the Temperature of Stored Ions in an rf Quadrupole Trap," Appl. Phys. B 45, pp. 59-64, 1988.

[8] G. R. Janik, J. D. Prestage, and L. Maleki, "Simple Analytic Potentials for Linear Ion Traps," J. Appl. Phys. 67, pp. 6050-6055, Jan. 1990.

[9] H. Goldstein, Classical Mechanics. Reading: Addison-Wesley, 1950, ch. 3, pp. 59-63.

[10] J. D. Prestage, G. J. Dick, and L. Maleki, "New Ion Trap for Frequency Standard Applications," J. Appl. Phys. 66(3), pp. 1013-1017, Aug. 1989.

[11] P. K. MacKeown and D. J. Newman, Computational Techniques in Physics. Bristol: Adam Hilger, 1987, ch. 7, pp. 141-145.

[12] S. E. Koonin, Computational Physics. Menlo Park: Benjamin/Cummings, 1986, ch. 8, pp. 191-195.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

TRACEABILITY IN ATOMIC FREQUENCY STANDARDS

John A. Kusters
Charles P. Little

Hewlett-Packard Co.
Santa Clara Division
Santa Clara, California 95052-8059

Abstract

In the manufacturing of cesium-based primary frequency standards, the cesium standard's normal operating frequency must be verified to be within the basic accuracy specifications set for that standard.

HP Santa Clara is the west coast reference station for the U.S. Naval Observatory. Daily comparison of time between HP and UTC(USNO) are made using common view GPS techniques. The data obtained permits continuous tracking of a bank of cesium standards. The bank of standards is used internally as a house reference standard. The production area is located several hundred meters from the house standard. In the production area, an active "real time" ensemble composed of three of the new HP5071A Primary Frequency Standards is used as a local reference. Continuous measurement of the house standard and up to 31 units under test are made relative to the local reference. This permits easy determination of the relative difference between UTC and the standards under test.

Introduction

Planning for the production of the HP5071A Primary Frequency Standard began about four years ago. At that time we realized, because of the accuracy and stability [1] of this new product, we would have to make a major upgrade to the house frequency standard system. The goal for the new system was to provide a clean, stable, accurate working reference, physically close to the test units in the production area. The job was complicated by the fact that the production line was located over 200 meters from the existing HP house standard, and the cables carrying the standard frequency signal passed through several environments in different parts of the factory causing unpredictable phase shifts in the standard frequency signal.

Because of budget constraints the new system had to use as much of the original configuration as possible. In addition, as part of a major upgrading of our production test facilities, an automated final test system for the new cesium standard was included in the plan.

In this paper we will discuss each of the elements in the measurement chain between our GPS reference receiver and the frequency standard units under test. Figure 1 shows a block diagram of the frequency reference chain and production test system.

HP Santa Clara Reference Standard

The house time and frequency reference standard is located in the Santa Clara Division Measurement Standards Lab. This lab is an official reference station for the U.S. Naval Observatory. The time and frequency reference standard is a "paper clock" ensemble consisting of four HP5061A/B cesium beam frequency standards. A block diagram of the system is shown in Figure 2. The most stable unit is selected as the master clock, and is continuously compared against each of the other ensemble units to monitor its stability and accuracy. From time-to-time the frequency of the master clock may be adjusted slightly to bring it into closer agreement with UTC. The master clock provides the ensemble's working output.

0-7803-0905-7/93 $3.00 © 1993 IEEE

A long term reference to UTC is maintained by use of a Global Positioning System (GPS) timing receiver. The receiver has an internal clock which operates from the 5 MHz output of the ensemble master clock. Time is compared through common view GPS to UTC(USNO) once per day by monitoring satellite PRN13 for a period while the satellite is at its highest point above the horizon. The link between the corrected UTC time and the ensemble master clock is provided by a time interval counter controlled by an automatic data logging system. On a daily basis, the system measures the time difference between the ensemble master clock and the updated clock in the GPS timing receiver. The resulting data is stored and processed to precisely determine the frequency difference between UTC and the ensemble master clock. A typical sample of this data is plotted in Figure 3. The plot shows a frequency difference of 9 parts in 10^{14} between UTC and the ensemble master clock for the month of September 1992.

A secondary source of timing information comes from Loran-C navigation system broadcasts. Timing data from four different Loran-C stations is compared against GPS time through the ensemble master clock. This data is used for two purposes: (1) it is collected by the USNO on a daily basis as a check on the monitored Loran-C stations, and (2), it is saved to provide the measurement standards lab with backup to the GPS data.

Real-Time Ensemble

Figure 4 shows a block diagram of the real-time ensemble. The term "Real-time" refers to the fact that the ensemble controller makes continuous, real time adjustments to the three frequency standards included in the ensemble. A key point is that the internal cesium loop operates unsteered, at its natural frequency. The output of each of the three frequency standards is precisely steerable in steps of 6.3 parts in 10^{15} by use of an internal high-resolution direct digital frequency synthesizer [4].

In the real-time ensemble, the output signal from each frequency standard is phase-compared with the ensemble output signal. This signal is derived from a power summer which combines the outputs of the three frequency standards [5]. The result of these comparisons is sent to the ensemble controller. The ensemble controller determines a frequency offset for each frequency standard output which will bring it in phase with the ensemble output. In order to produce a true average of the three undisturbed frequencies, the controller selects the minimum set of offsets which sum to zero. The three outputs are driven to operate in phase, and at the average frequency of the three standards in the ensemble.

A further advantage of power summing is that output frequency stability (Allan Variance) and phase noise are reduced by the square root of N and 10 log N (dB) respectively (where N is the number of sources in the ensemble). Finally, data kept by the ensemble controller allows a "three-cornered hat" measurement of the stability of each source. This data is continuously examined for instability and stored for future reference.

Even though the ensemble provides the working reference for production frequency measurements, it is not kept in a controlled environment. Extensive testing of the HP5071A by independent labs shows virtually no change in frequency due to environmental changes [2]. This environmental insensitivity reduces the cost of the test system because neither the ensemble nor the test units need to be environmentally controlled, even though the final test runs for a minimum of five consecutive days.

The output signal from the real-time ensemble is sent to the measurement system where it provides the reference for accuracy and stability measurements on the newly produced frequency standards.

Long-Term Stability Measurement System

The measurement system, shown in Figure 5, uses a time interval phase measurement technique to compare the frequency of the test units to that of the ensemble [3]. The time interval counter is used to measure the phase difference between the 5 MHz output from the ensemble and the unit under test.

Each unit under test is sampled every 12.5 seconds. Each sample represents the average of 1000 measurements of the time difference between the local reference and the unit under test. Data reduction from 80 consecutive samples produces an effective time difference for that 100 second period. Data is acquired on each unit for a total of 6 days, or 518 time difference measurements.

The data thus obtained yields a plot similar to Figures 6 and 7. The data in Figure 6 was derived from

an HP5071A which contains a high-performance cesium beam tube; Figure 7 data came from a standard cesium tube unit.

To precisely determine the relative output frequency of each test unit, a least-squares-fit line is computed from the time-difference data. The frequency offset of the test unit (with respect to the ensemble) is computed from the slope of that line.

In addition to the production frequency standards under test, the system also measures the signal from the HP house standard. This data is reduced in the same manner described above, and stored giving a continuous record of the frequency difference between the ensemble and the HP house standard. An automatic alarm in the measurement system warns the operator if there is a significant frequency change or jump between the frequency of the ensemble and that of the HP house reference standard.

Results

The data produced thus far must be further reduced to give frequency measurements traceable to UTC(USNO). Corrections for the offset between the ensemble and the HP house standard are computed from the data obtained by measuring the house standard on the measurement system. The frequency difference between the HP house standard and the ensemble is plotted in Figure 8. Also plotted is the offset data obtained via common view GPS showing the offset measured between the house reference standard and UTC(USNO). The figure shows the ensemble to be between 2.3 and 2.6 parts in 10^{13} low with respect to the HP house standard. Figure 8 also indicates clearly the problems with transmitting high precision frequency signals over coaxial cable, thorough several distribution amplifiers, and at least 4 different micro-climates. The environmentally-induced phase changes cause a significant spread in the measured frequency offset on the order of 1 part in 10^{13}. Frequency changes seen in the data were caused by switching master units in the House Standard, then making C-field corrections to syntonize it with UTC.

After all corrections are applied, the scattergram shown in Figure 9 was obtained. A histogram of the same data is shown in Figure 10. The data shown has a mean of 1.3 parts in 10^{14} and a standard deviation of 2.3 parts in 10^{13}. Also shown is that the maximum deviation seen on all units produced is less than $\pm 6 \times 10^{-13}$.

Accuracy

We have shown how production frequency measurements have a traceability chain from UTC to the measurement of new cesium beam frequency standards on the production line. The remaining issue is the accuracy with which the production units can be measured.

Sources of uncertainty in this measurement are the frequency determination from GPS, frequency transfer from the HP standards lab to the production area, real time ensemble frequency, and the production measurement system. Our best estimates of these uncertainties are as follows.

GPS

Monitoring one GPS satellite for long periods while using the master clock as a "flywheel" yields fairly good results. Over a four week period we are able to determine UTC with an uncertainty of around 100 ns. This translates to a frequency uncertainty of about 4 parts in 10^{14}.

Line Noise

The signal line from the from the HP house standard picks up noise and phase shifts as it passes through the factory to the cesium standard production area. These factors give this signal an uncertainty of about 7 parts in 10^{14}.

Measurement System

A system very similar to the production test system is described and characterized in [3]. Based on this data, and measurements of our production test system, we estimate its measurement uncertainty to be close to 3 parts in 10^{14}.

Local Reference

Finally, the uncertainty contribution of the real time ensemble, as described above, is the square root of 3 better than an individual ensemble unit. For a 5 day averaging time (the duration of the frequency measurement) the ensemble contributes an uncertainty of less than 1.5 parts in 10^{14}.

Using a square-root-of-the-sum-of-the-squares method the total system error is estimated at less than 1 part in 10^{13}. This level of accuracy is sustainable because of the number of cesium standards in the system, and the continuous cross checking between the system elements. As presently configured the system provides more than sufficient accuracy to guarantee our present specifications.

Conclusion

The data shown indicates that we can indeed measure the long-term accuracy of all units produced with sufficient precision to guarantee that all units meet a published accuracy specification of $\pm$ 1 x 10^{-12}, and that this accuracy level can be maintained in a normal production environment.

Acknowledgments

The authors thankfully acknowledge discussions with Robert Pitcock, manager of HP's Standards Lab, for details on how the house reference system is maintained, and for the data relating the house reference system to UTC(USNO).

References

[1] Hewlett-Packard model 5071A Primary Frequency Standard technical data sheet.

[2] Johnson, James L. and Kusters, John A., "A New Cesium Beam Frequency Standard Performance Data." Proceedings of the 1992 IEEE Frequency Control Symposium, page 143, IEEE catalog number 92CH3083-3.

[3] Kusters, John A., "Determination of the Aging Rates of High Stability Rubidium Frequency Standards." April 1989 Proceedings of the Second International Conference on Frequency Control and Synthesis.

[4] Giffard, Robin P. and Cutler, Leonard S., "A Low-Frequency, High Resolution Digital Synthesizer", Proceedings of the 1992 IEEE Frequency Control Symposium, page 188, IEEE Catalog number 92CH3083-3.

[5] Stern, A., "About Sum of Signals and Possible use in Time and Frequency Systems," Proceedings of the Forty-Fifth Annual Symposium on Frequency Control, pp 659-666, IEEE Catalog number 91CH2965-2

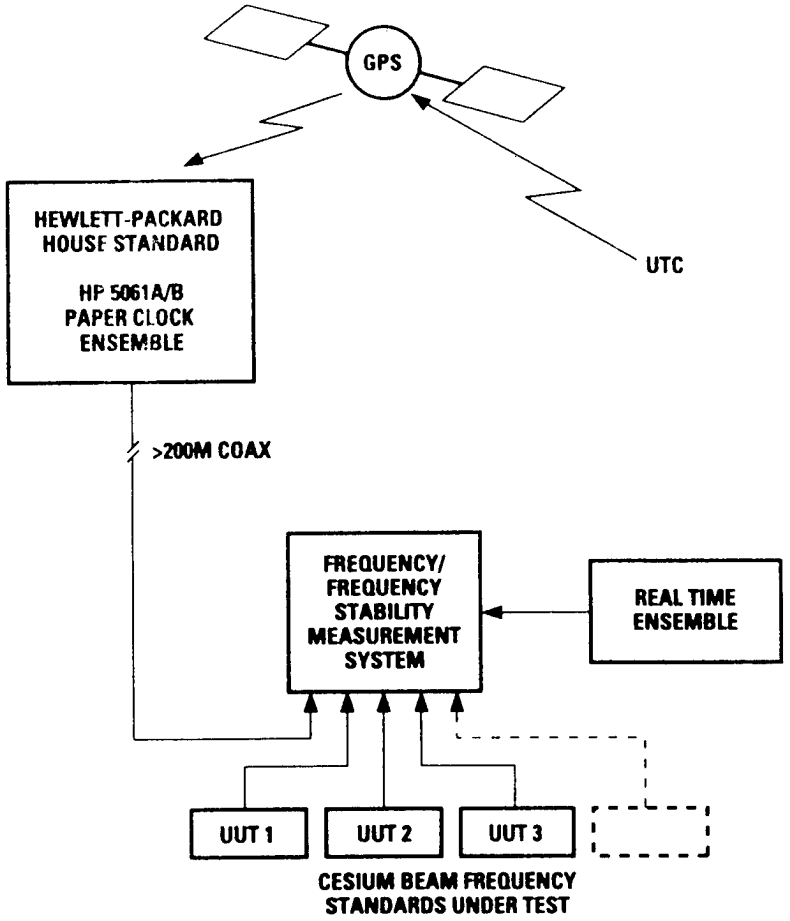


Figure 1. Block Diagram of Hewlett-Packard Frequency Standard/Measurement System

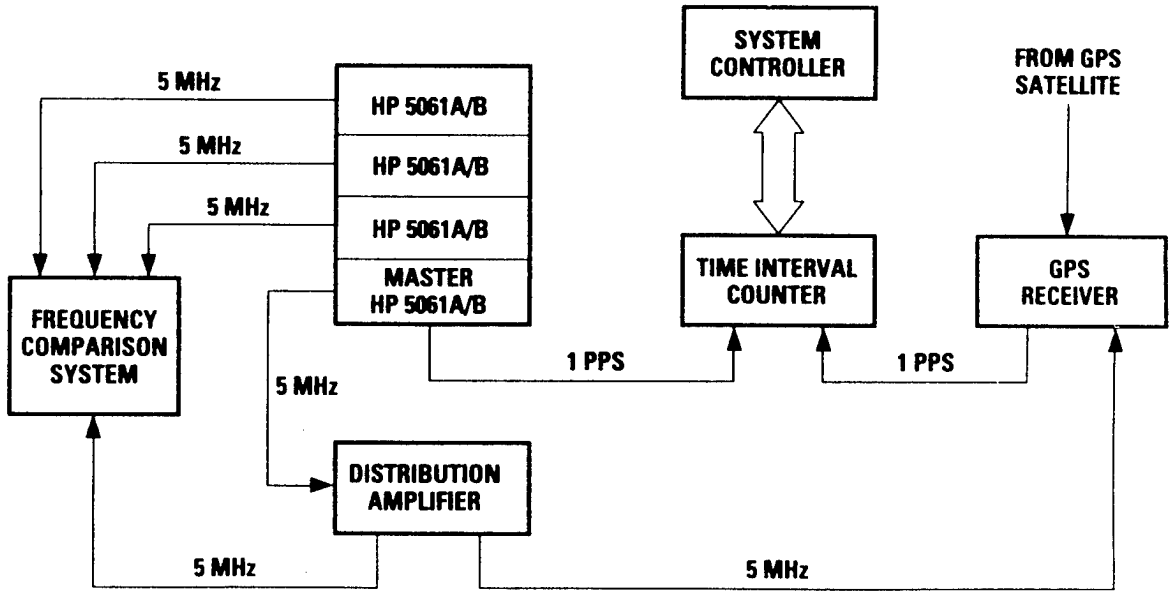


Figure 2. Hewlett-Packard Time and Frequency Reference Standard

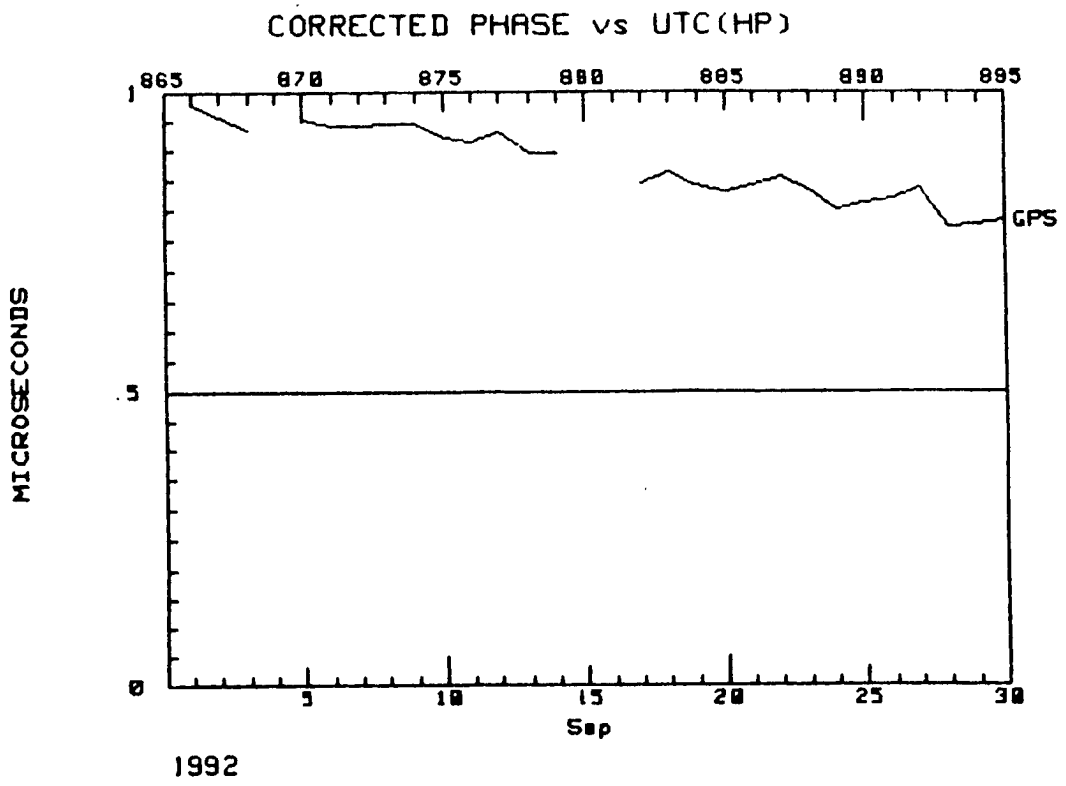


Figure 3. Ensemble master clock vs. UTC

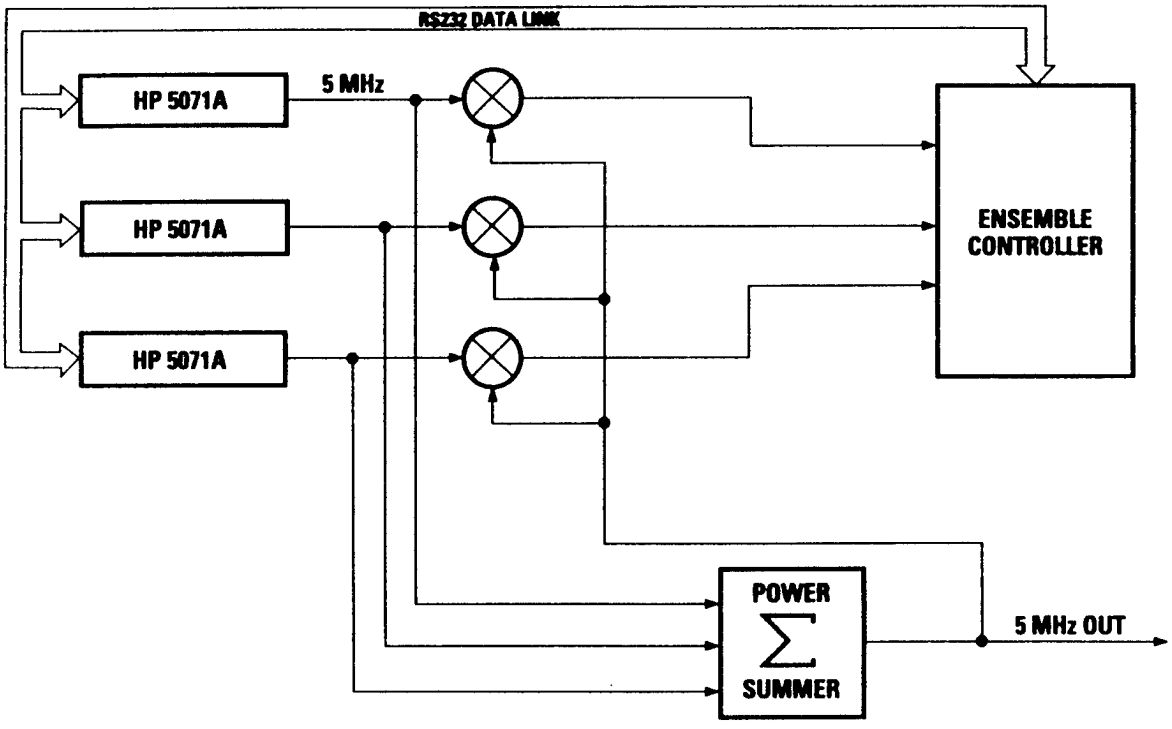


Figure 4. Block Diagram of Production Real-Time Ensemble

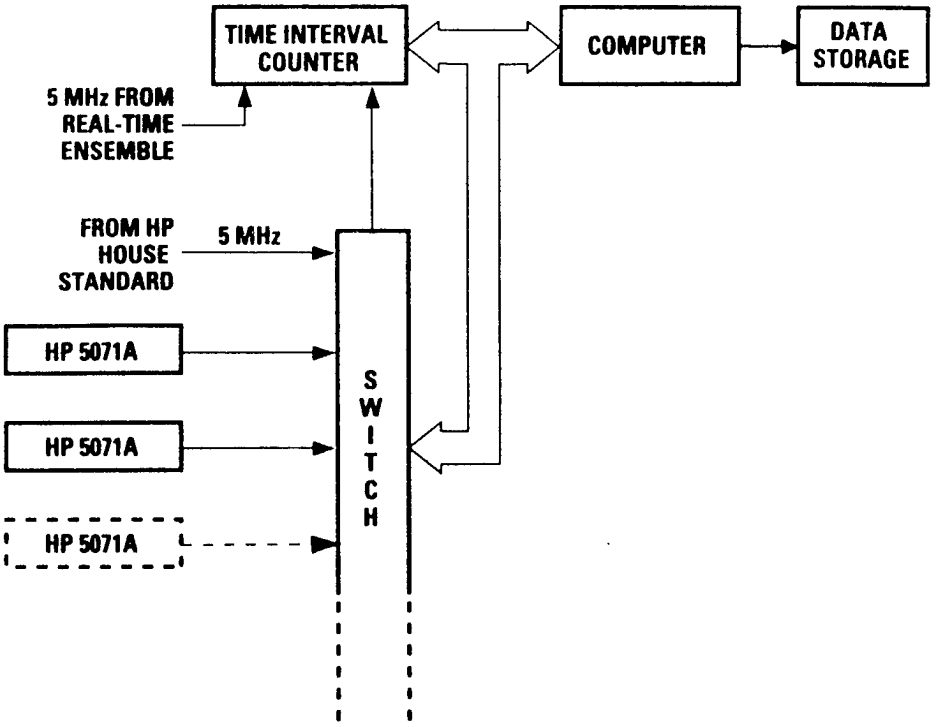


Figure 5. Frequency/Frequency Stability Measurement System

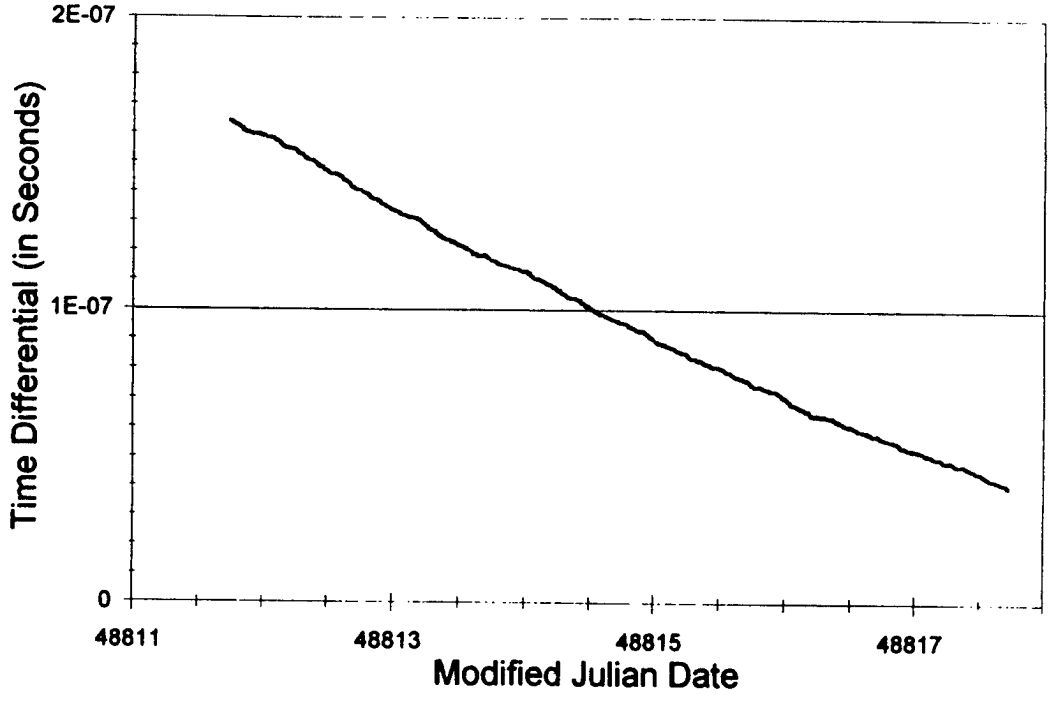


Figure 6. Continuous phase plot: HP 5071A with high performance cesium beam tube - 100 second data vs. Local Reference

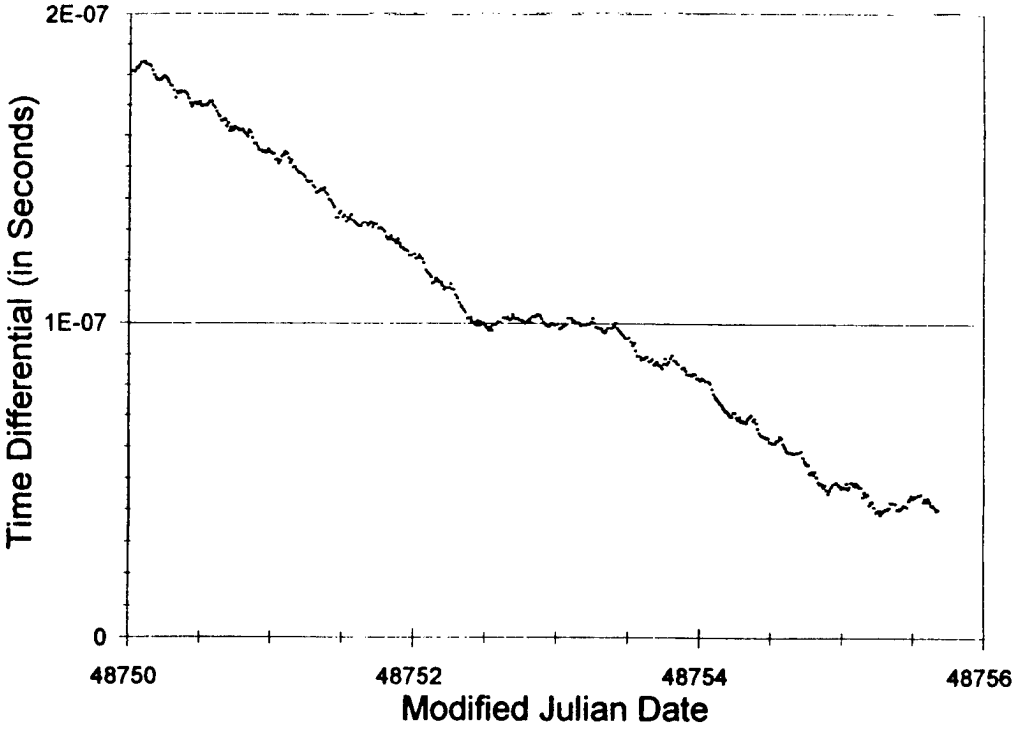


Figure 7. Continuous phase plot: HP 5071A with standard cesium beam tube - 100 second data vs. Local Reference

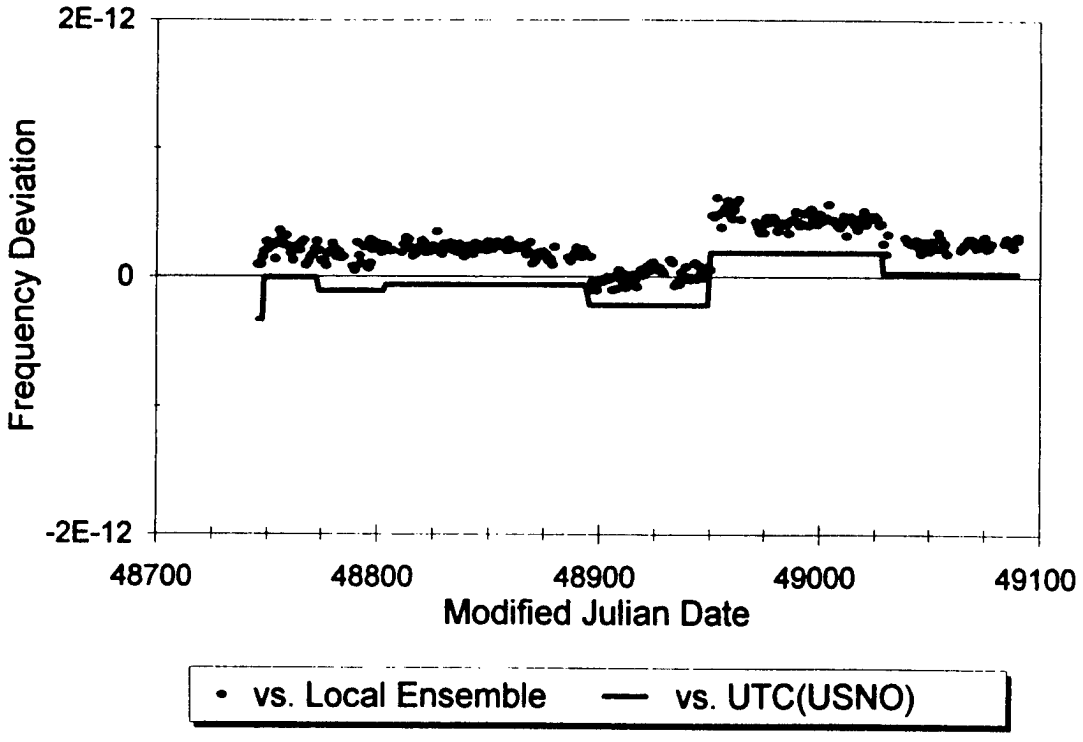


Figure 8. Frequency difference: HP house standard measurements Daily averages vs. Local Reference and UTC(USNO) via GPS

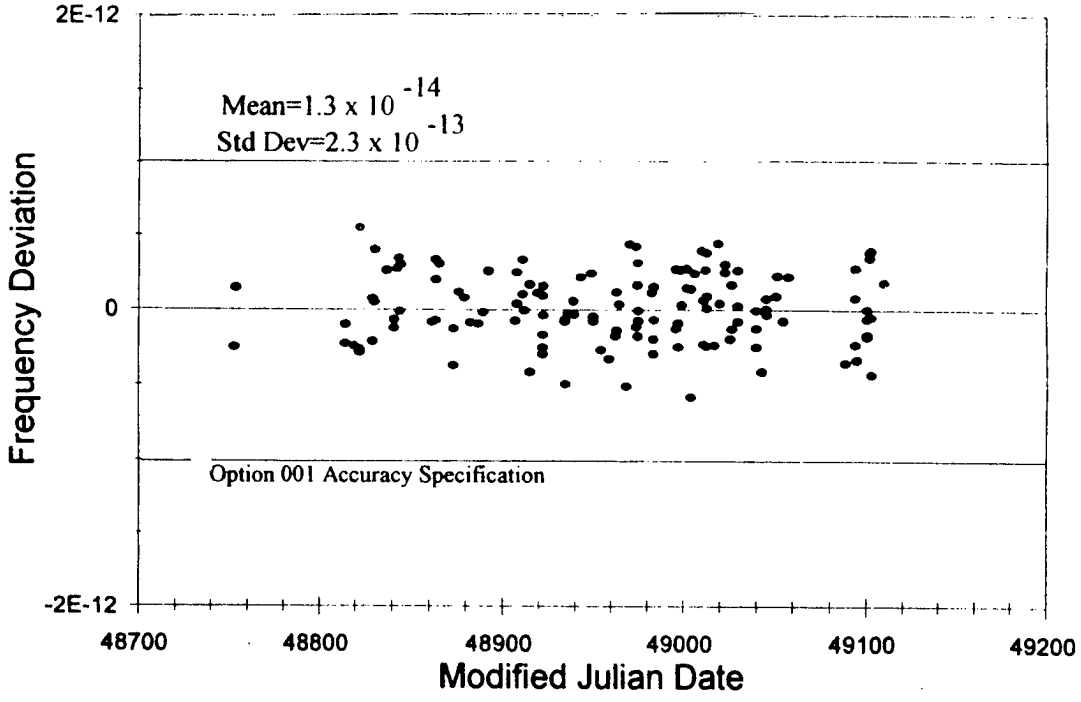


Figure 9. Test units vs. UTC(USNO) - 5 day averages

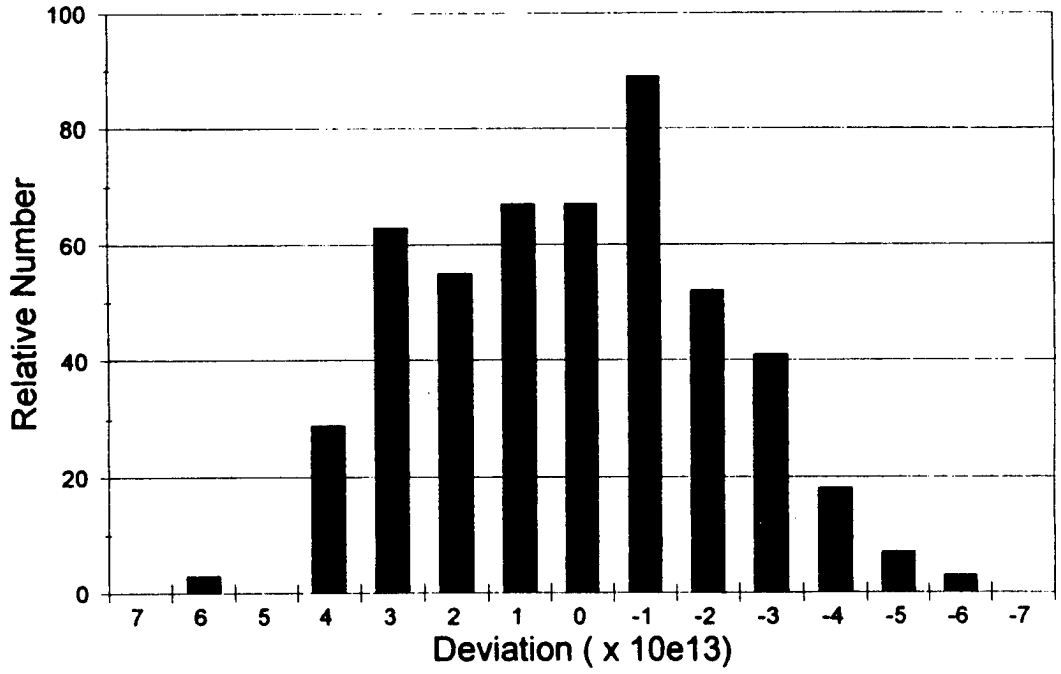


Figure 10. Histogram of test units vs. UTC(USNO) - 5 day averages

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

1993 IEEE FREQUENCY CONTROL SYMPOSIUM

SUBMINIATURE RUBIDIUM FREQUENCY STANDARD FOR COMMERCIAL APPLICATIONS

M. Bloch, I. Pascaru, C. Stone, T. McClelland
Frequency Electronics, Inc.
55 Charles Lindbergh Blvd.
Mitchel Field, NY 11553

Abstract

Frequency Electronics, Inc. has developed a new family of low cost Rb frequency standards for use in commercial applications. These Rb standards are designed to replace precision quartz oscillators, without any additional cost, but considerable improvement in performance. These devices are about a factor of two smaller than the smallest previously available commercial Rb standards, and dissipate about 25% less power; but with no compromise in performance capability. An important feature of these devices is an adaptable output interface which is easily tailored to different user demands. Any selected output interface is provided within the total package outline of 1.4 x 3 x 3 inches. The simplest (and least expensive) version provides a sine wave output at 50.255+ MHz (a sub-multiple of the atomic resonance). This option is ideal for users needing a stable clock reference for direct digital synthesis (DDS) systems in which the absolute frequency of the reference is not important. Another version incorporates a factory-set direct digital synthesizer in the output interface. This makes possible virtually any user specified output frequency between 1 Hz and 20 MHz, with the full stability of the Rb atomic system. A third version incorporates a digital communication link, in addition to the DDS output. This allows for remote adjustment of the output frequency by computer. In this paper a detailed description of the design concept and its implementation are provided. In addition performance test results are provided for Allan variance, aging, output phase noise and spurs, and frequency vs temperature.

Background

Applications, particularly in the telecommunications industry, which require atomic frequency control precision are growing dramatically. Often frequency standards are required which push the limits of capability of quartz crystal oscillators, especially when the constraints imposed by the operating environment are factored into the picture (e.g. temperature range, fast warm-up, vibration immunity). Rubidium vapor frequency standards are ideally suited to this type of application. Presently available Rb standards easily provide a factor of ten improvement in frequency drift compared to the best quartz oscillators, and typically achieve this level of performance after minimal warm-up (compared to quartz oscillators).

Most presently available Rb standards have been optimized in such a way as to push the performance capability of this (Rb vapor) technology. This approach has been driven almost exclusively by military requirements, which in general put a premium on performance and do not even consider cost. It has become apparent, however, that Rb standards which provide significant improvement over quartz technology, but at minimal cost, can be utilized with great benefit in many applications.

With this in mind, Frequency Electronics, Inc. has developed a new family of low cost Rb frequency standards for use in commercial applications.

Design

The goal of the design effort was to reduce the size, weight, power consumption, and cost of Rb

0-7803-0905-7/93 $3.00 © 1993 IEEE

technology as much as possible, while degrading the performance as little as possible. The result is the FE-5650A Rb Standard Series, a family of Rb standards targeted to meet this goal. A photograph of the FE-5650A, option -03, is shown in Figure 1. A block diagram of the FE-5650A series is shown in Figure 2.

Core Rb Standard

Conventional Rb standards use a quartz oscillator, at the desired output frequency, which is locked to the atomic frequency. This requires that a different quartz crystal and oscillator be employed each time a different output frequency is desired. This in turn necessitates re-packaging for different frequencies. This is avoided with the FE-5650A.

The FE-5650A series is based on a core Rb standard using a quartz crystal at a direct sub-multiple (50.255+ MHz) of the Rb atomic resonance frequency. This core Rb standard is the fundamental building block upon which all the various options of this family are built. (See Figure 2.) Because the crystal operates at a direct submultiple of the atomic frequency, no frequency synthesis is required in order to generate a comparison signal at the atomic resonance. This simplifies the design considerably, and also reduces undesirable interference effects from frequency mixing products. The core Rb standard is available as option -01 for those users able to accept the 50.255+ MHz output frequency. This is the simplest, and least expensive option.

Synthesizer Options

For those users unable to adapt to the 50.255+ MHz output frequency of the core Rb standard, additional options are provided, all of which utilize direct digital synthesis (DDS) to provide other output frequencies, referenced to the basic 50.255 MHz signal. These options are depicted in Figure 2 by the "synthesizer," and "digitally programmable synthesizer" boxes. In each of these cases, the 50.255+ MHz output of the core Rb standard is used as the clock reference input to the DDS circuitry.

Synthesizer: In this case, a single output frequency is provided. The output circuitry is optimized for the chosen frequency, and hence spurs and harmonics are minimized. Frequency adjustment to correct for aging is provided by a combination of c-field adjustment (resolution: 1×10^{-11}, range: 2×10^{-9}), and dip switches on the DDS (resolution: 1.2×10^{-9}, range: 2×10^{-7}).

Digitally Programmable Synthesizer: This is similar to the "synthesizer" output, except that adjustment is provided over a serial interface, instead of with dip switches. The circuitry is naturally somewhat more complex than for the dip switch version, and thus more costly. However, tremendous flexibility results. In this configuration, either a standard, or high resolution DDS implementation is available. With the high resolution implementation, a programmable frequency setting resolution of 2.7×10^{-12} Hz (2.7×10^{-18} at 1 MHz) is possible.

With either synthesizer, a wideband digital tuning option is available. In this case, the output frequency is provided directly from the DDS circuit, without any bandpass filtering. This results in an output which is a stepped approximation to a sine wave. With this option the output frequency can be selected remotely, via an RS-232 interface, to any value between 1 Hz and 20 MHz.

Packaging

In order to minimize the size of the FE-5650A, surface mount technology has been used for all circuitry. Figure 3 shows partially exposed views of the FE-5650A, in which the circuit assemblies are clearly visible.

The dimensions of the FE-5650A are 1.44 x 3.00 x 3.03 inches, making it possible to install the unit within a standard 1.5 inch height rack mount chassis. Input power (+15 VDC, and +5 VDC) is supplied via a 9 pin D-subminiature connector on the mounting baseplate. The RF output is supplied on an SMA connector, also on the baseplate. The optional RS-232 interface is provided on a separate RS-232 connector.

Performance Specifications

The basic performance specifications of the FE-5650A are as follows:

Warm-Up Time < 4 min. to lock (25°C)

Allan Variance $3 \times 10^{-11}/\sqrt{\tau}$ (1 to 1000 sec)

Drift	$<2 \times 10^{-11}$ / day $<2 \times 10^{-9}$ / year	
Phase Noise	10 Hz	-90 dBC
(10 MHz)	100	-125
	1000	-145
Spurs	<60 dBC	
Harmonics	<30 dBC	
Freq. vs Temp	$\pm 3 \times 10^{-10}$ (-5 to +55°C)	

Performance Test Results

Allan Variance

Allan variance data from several units is shown in Figure 4. As shown in the Figure, the actual performance is typically somewhat better than the specified performance.

Frequency vs Temperature

Data from a typical unit is shown in Figure 5. Because the C-field is adjusted only over a very limited range (long term frequency aging is corrected with the DDS), this performance does not change with time, as is the case with conventional Rb standards.

Phase Noise

Phase noise data is shown in Figures 6 and 7. Figure 6 shows the performance for the core Rb standard, at an output frequency of 50.255+ MHz. Figure 7 shows the performance for the -03 option, which provides a 5 MHz output. Note the improvement of the phase noise at 5 MHz, corresponding to the frequency division by ~10.

Spurious, and Harmonic Outputs

In general, the spur / harmonic content of the output is sensitive to the exact configuration, since these differ in the type of output filtering provided. Also spurs generated by the DDS depend on the output frequency. Typical data for an output frequency of 10 MHz is shown in Figures 8 through 11.

Summary

A new family of low cost Rubidium vapor frequency standards has been developed. These standards are extremely small, low power, atomic frequency standards designed for use in a variety of demanding commercial applications. These devices achieve low cost and small size with the use of surface mount technology, and provide a wide range of frequency outputs by utilizing direct digital synthesis technology (DDS) for the output interface.



Figure 1
Photograph of the FE-5650A, option -03 (5MHz output)

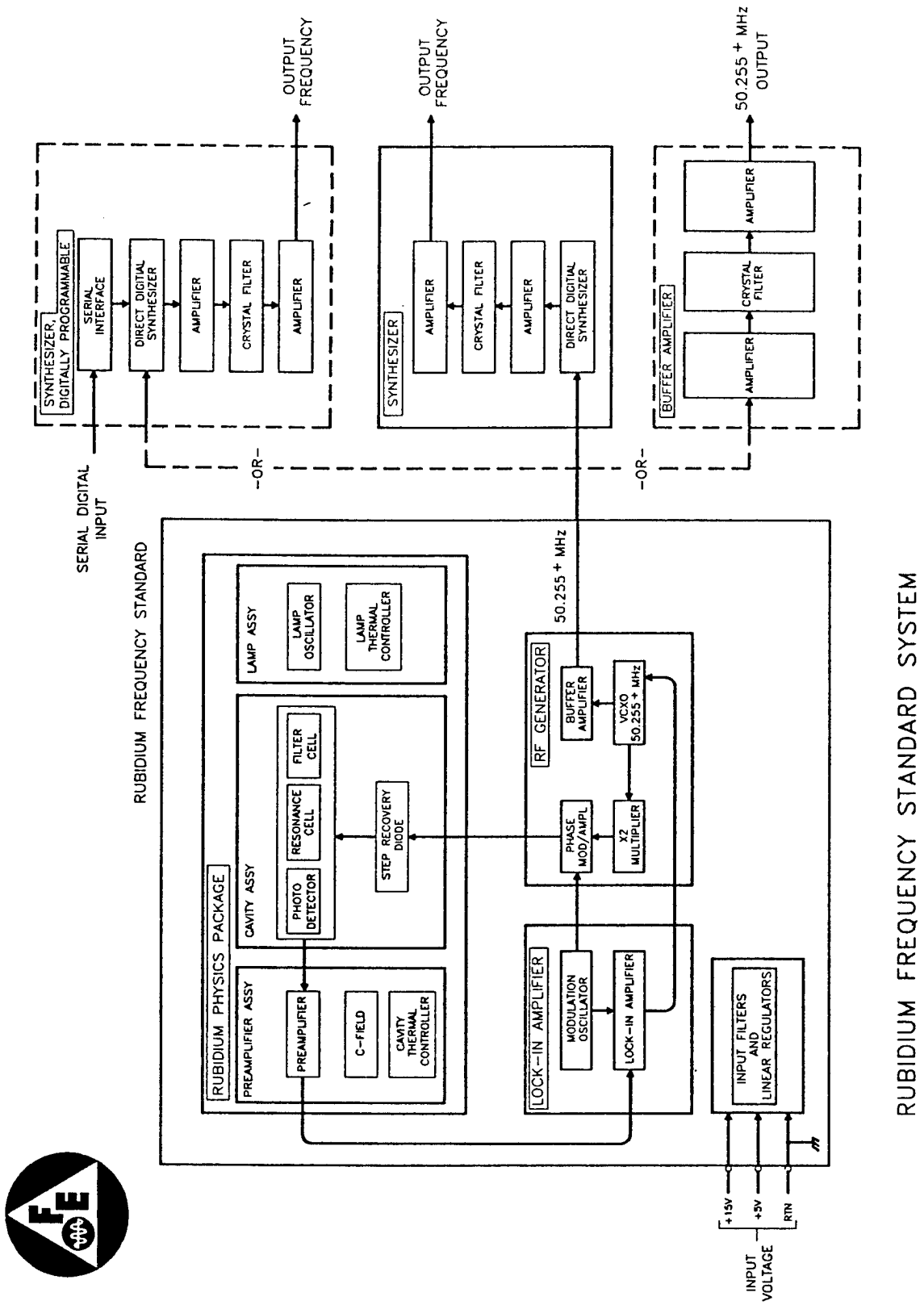
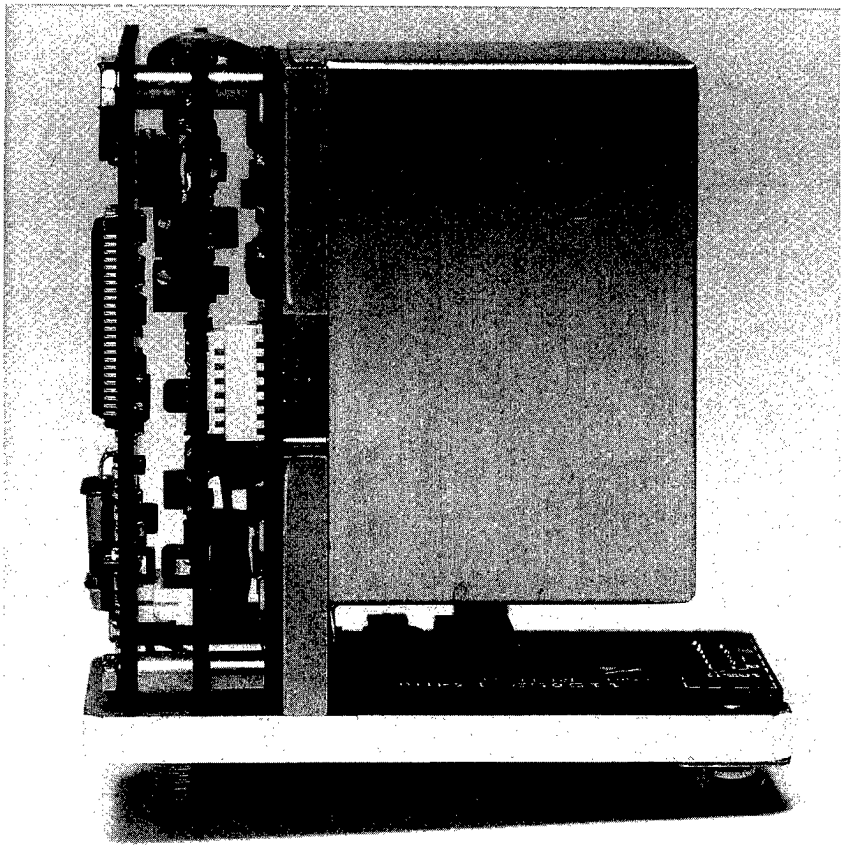
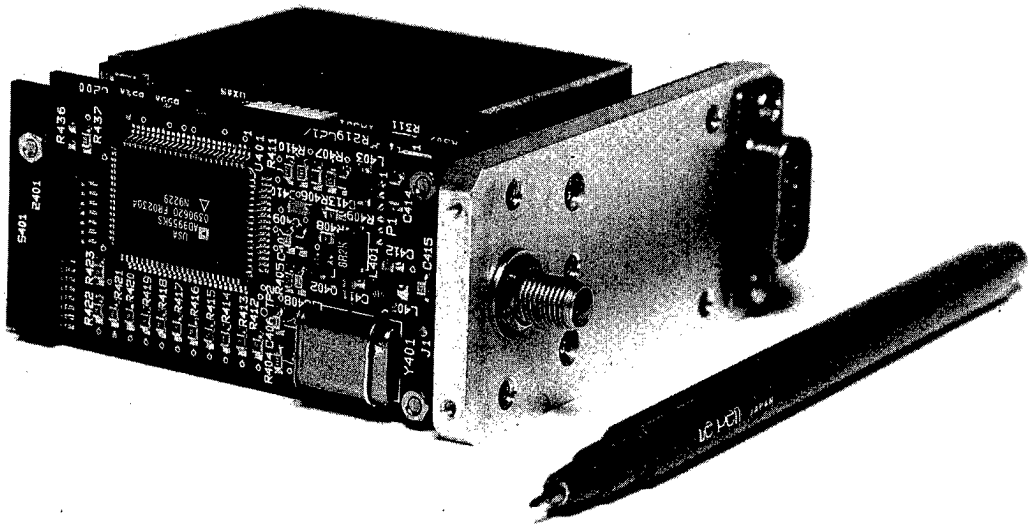


Figure 2
Block diagram of the FE-5650A series. Optional output blocks are shown at the right.



a.) Side view showing three PC assemblies (edge-on), and the physics package (shielded module)



b.) Isometric view showing the DDS PC assembly

Figure 3
Photographs of the FE-5650A, option -03, partially dis-assembled

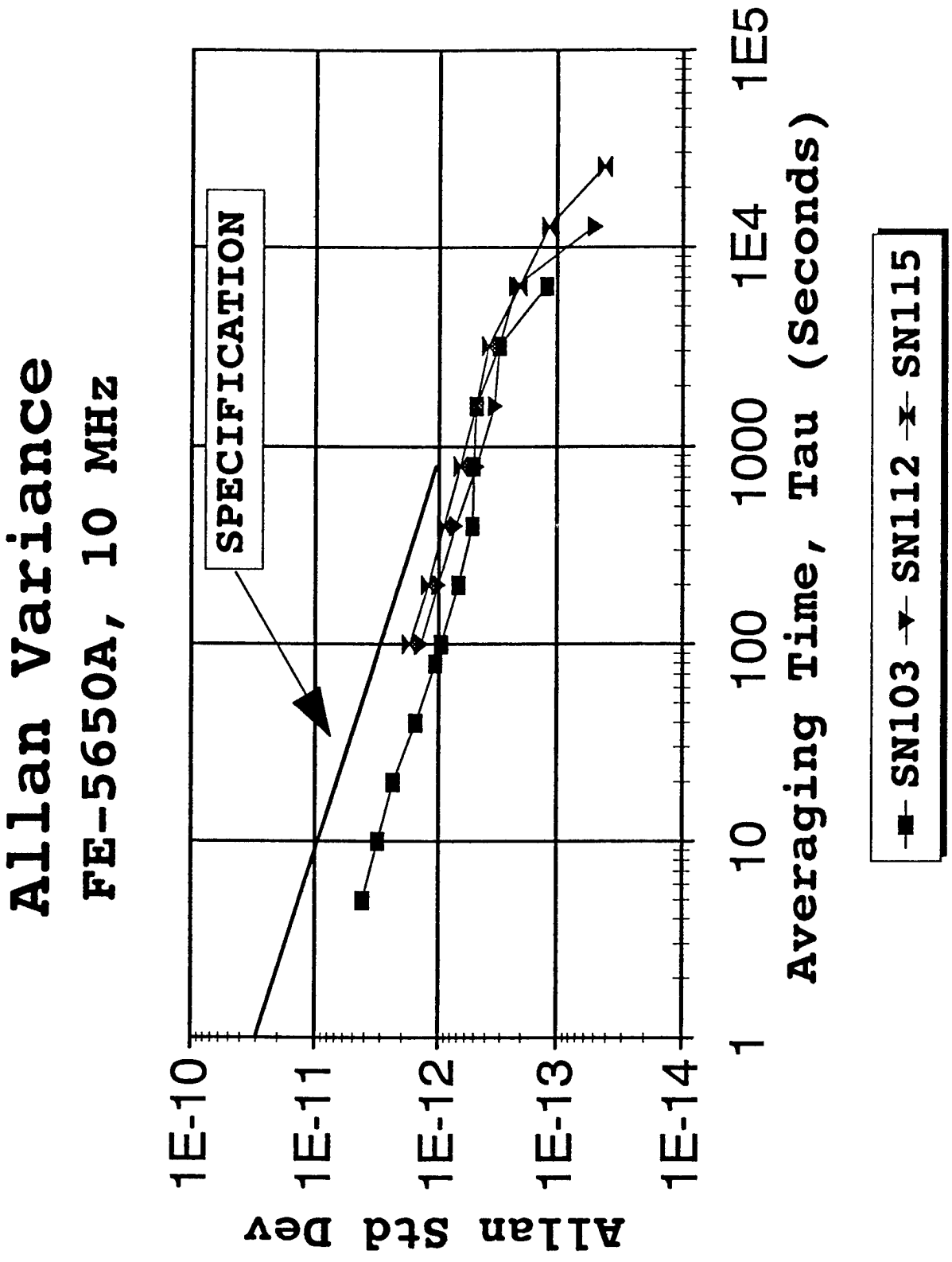


Figure 4
Allan variance data for three different units

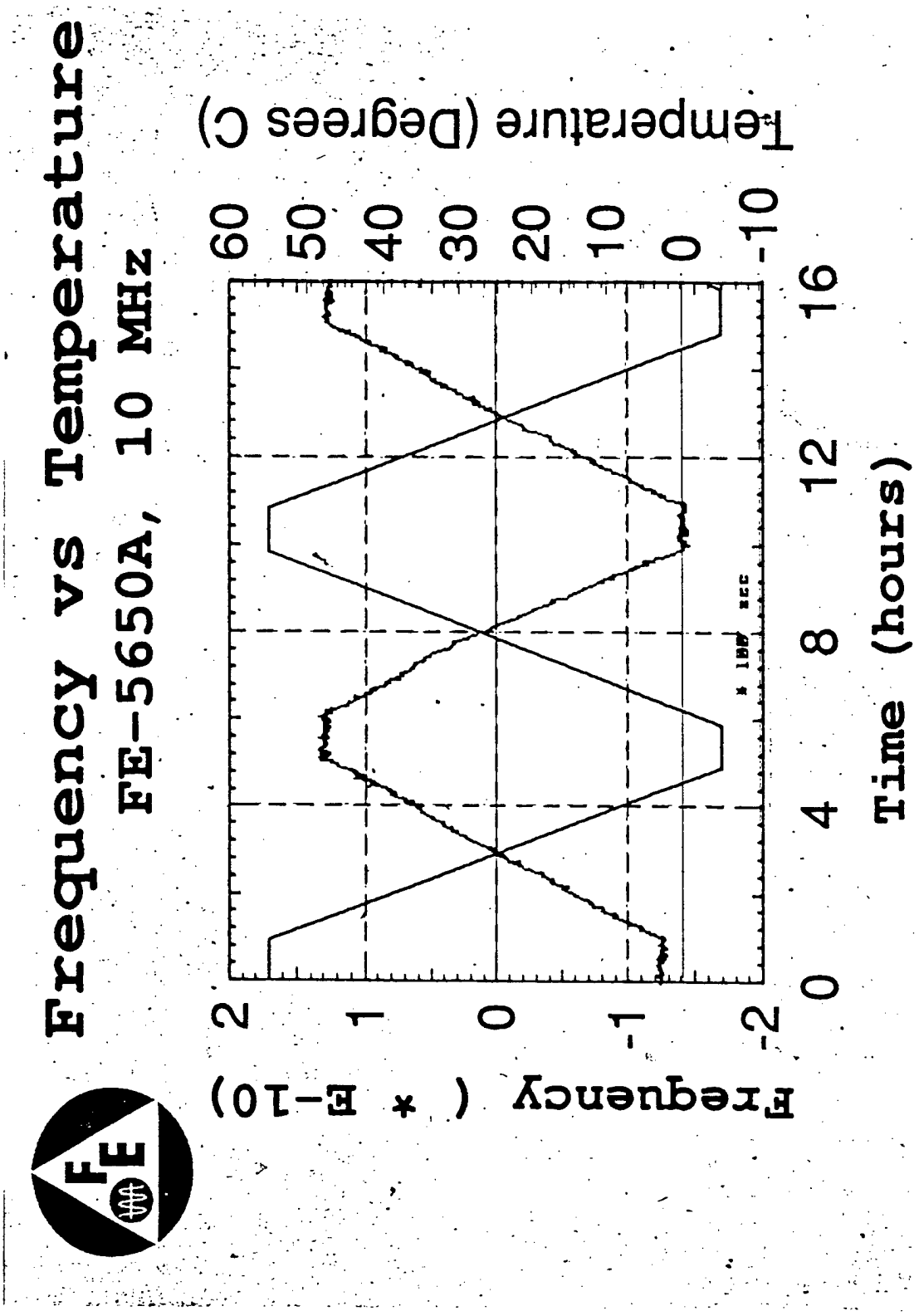


Figure 5
Frequency vs temperature data. The jagged line is the frequency output, the bold line is the temperature.

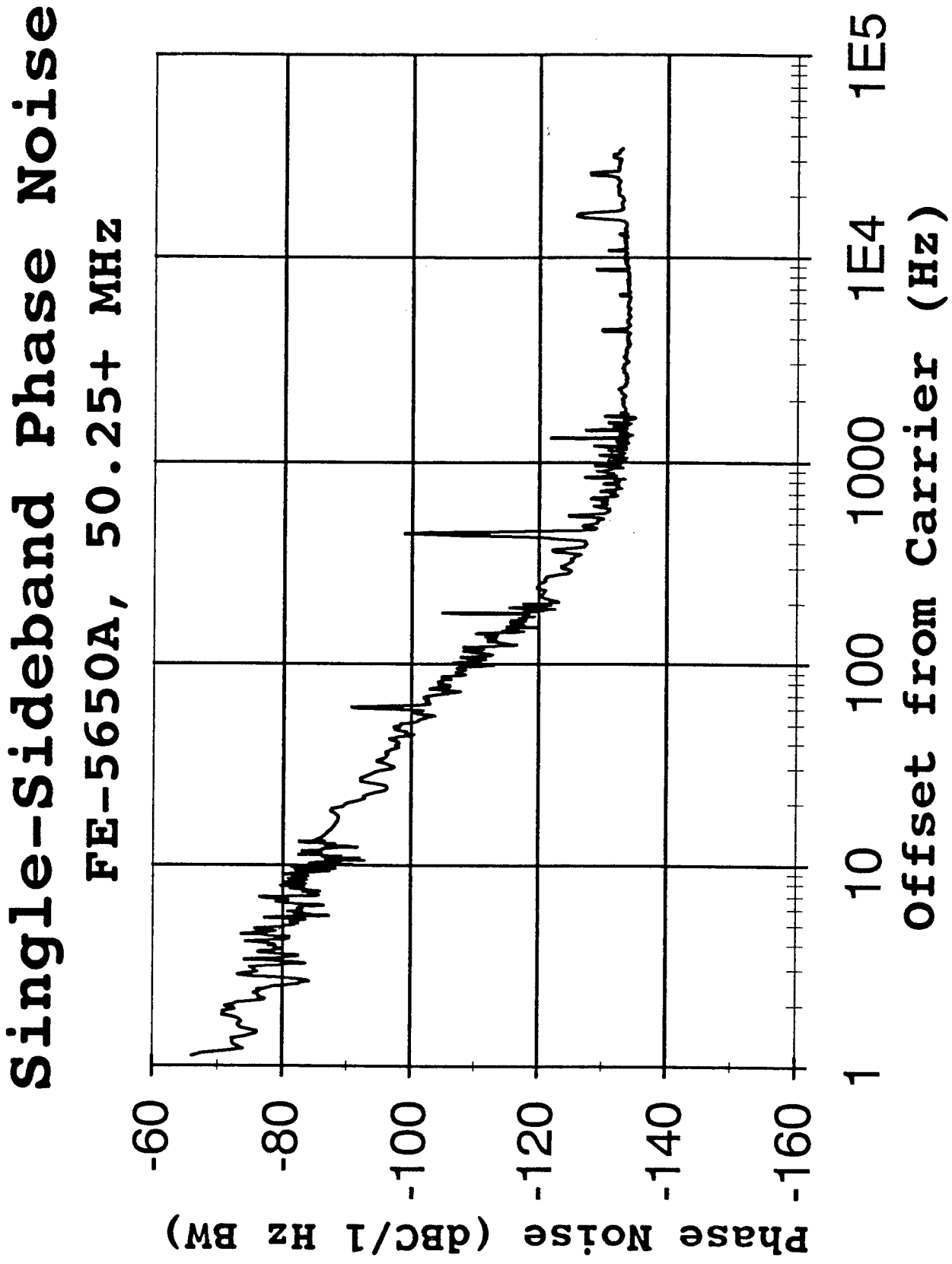


Figure 6
Phase noise data, 50.255+ MHz output frequency

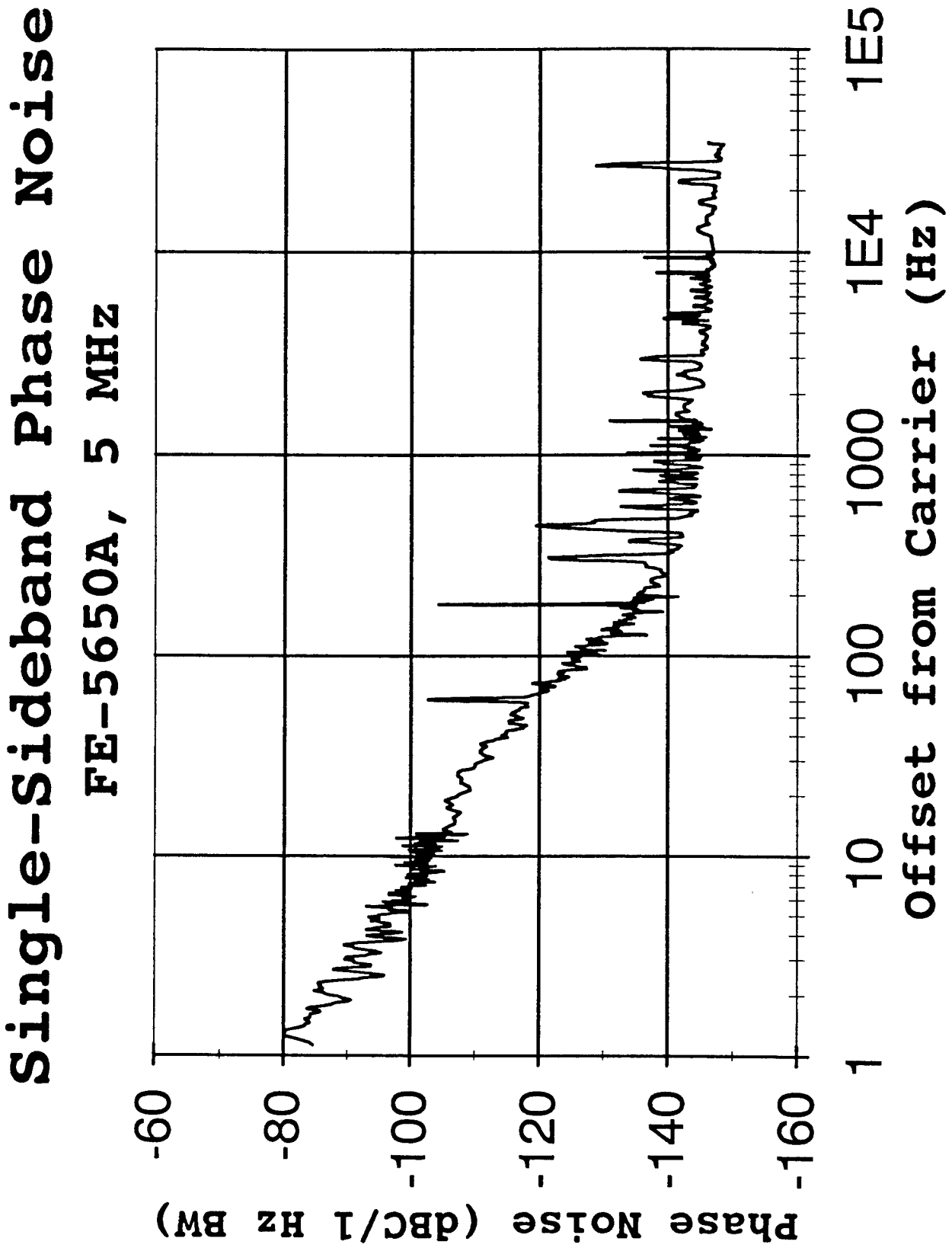


Figure 7
Phase noise data, 5 MHz output frequency

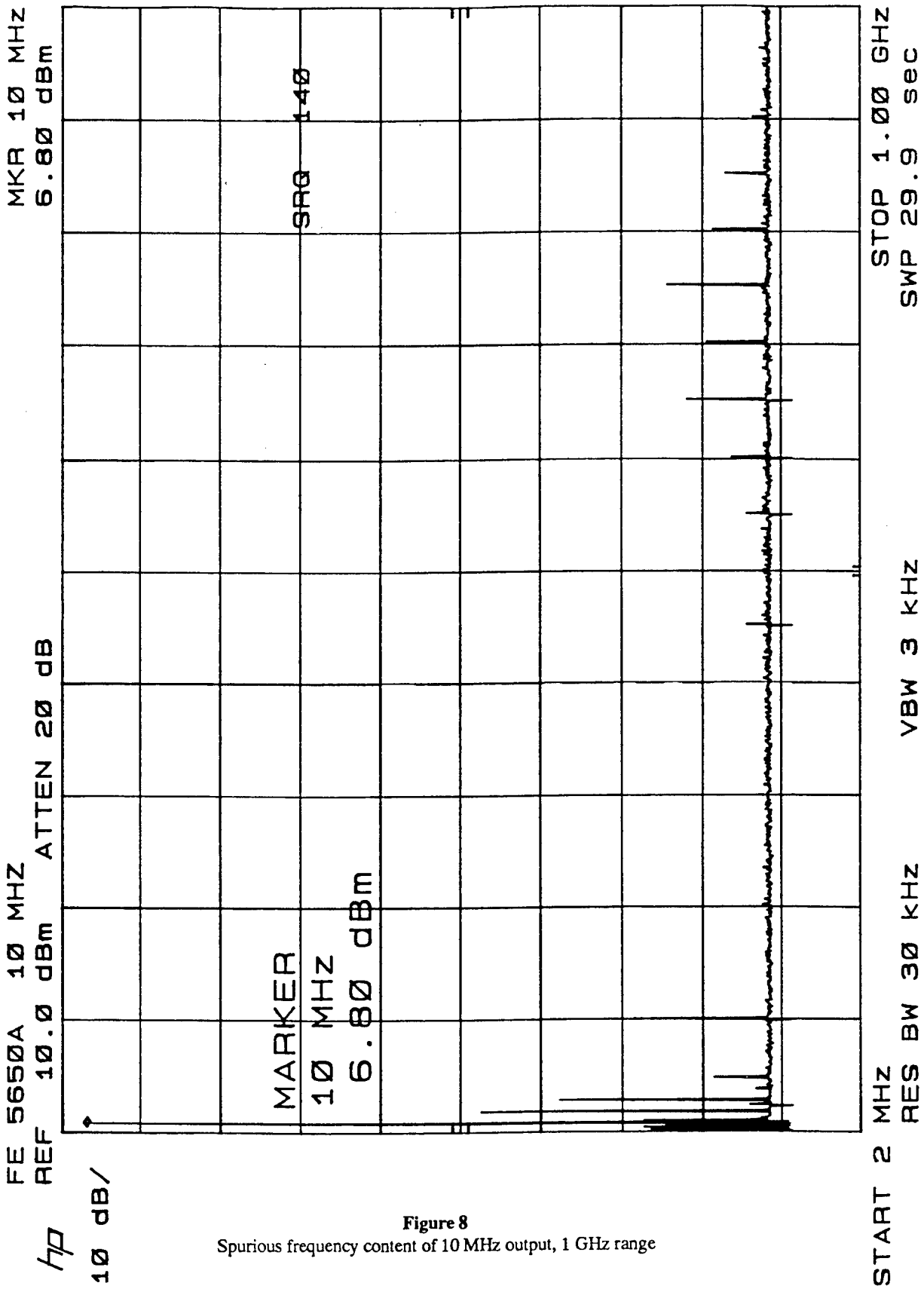


Figure 8
Spurious frequency content of 10 MHz output, 1 GHz range

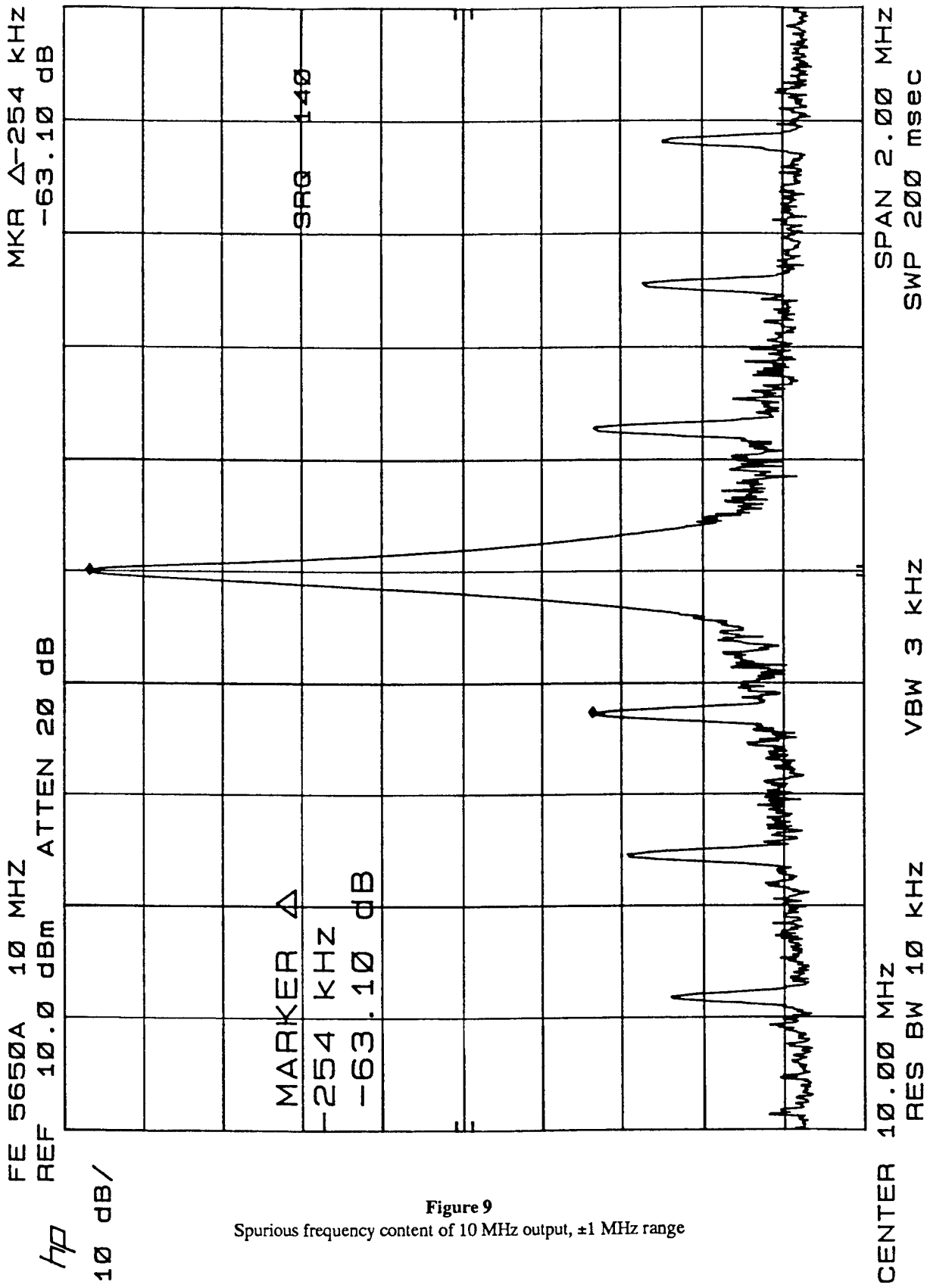


Figure 9
Spurious frequency content of 10 MHz output, ±1 MHz range

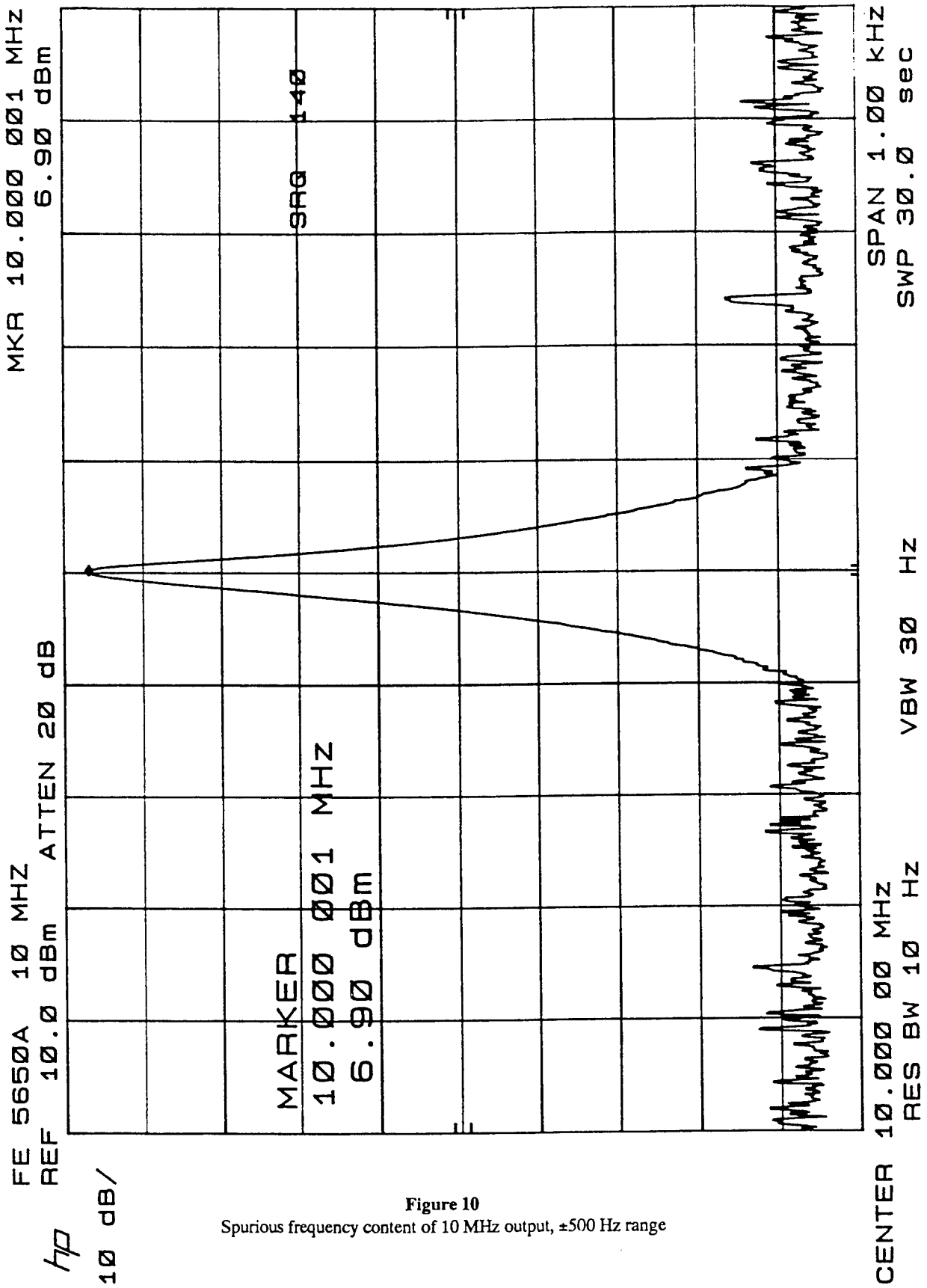


Figure 10
Spurious frequency content of 10 MHz output, ±500 Hz range

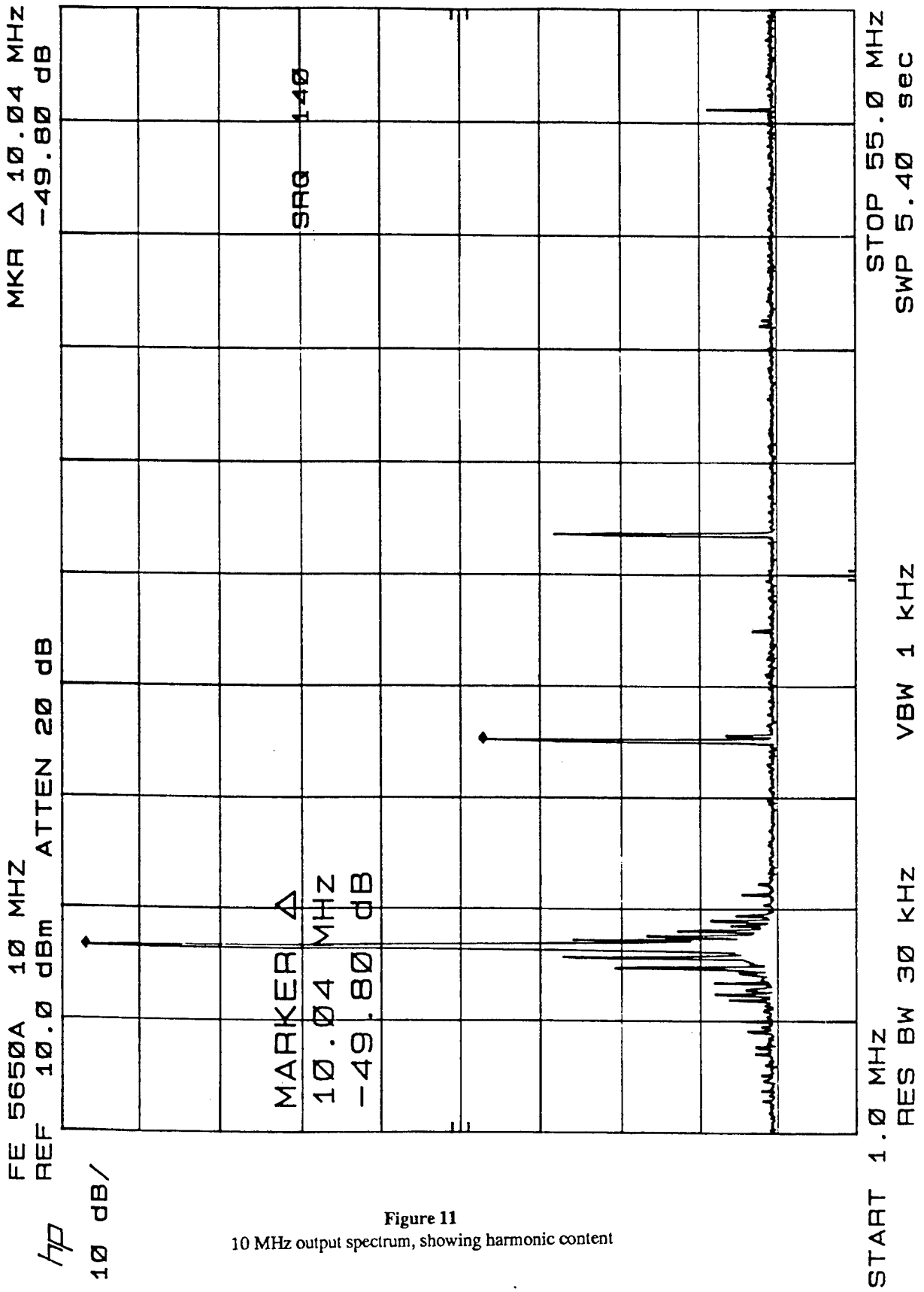


Figure 11
10 MHz output spectrum, showing harmonic content

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DEPENDENCE OF SAW RESONATOR *1/f* NOISE ON DEVICE SIZE

Thomas E. Parker and Donald Andres

Raytheon Company
Research Division
131 Spring Street
Lexington, MA 02173 USA

Abstract

Experiments were conducted with eight 450 MHz two-port SAW resonators of four different designs which demonstrate that a resonator's 1/f noise depends approximately inversely on the active acoustic area of the device. The ion etched groove resonators were all fabricated from a single piece of quartz. The active acoustic areas of these devices ranged from approximately $1.7\text{x}10^4$ square wavelengths ($8.4\text{x}10^{-3}$ cm^2) to $7.0\text{x}10^4$ square wavelengths ($3.4\text{x}10^{-2}$ cm^2). The measured flicker noise levels varied from $S_y(f=1Hz) = 3.4\text{x}10^{-23}$ to $6.4\text{x}10^{-24}$/Hz, with the smallest area devices exhibiting the highest noise levels. This observation is consistent with a proposed theory that 1/f noise in acoustic resonators can be caused by localized, independent, velocity (or dimensional) fluctuations in the quartz. The fractional velocity fluctuations would have to have a magnitude close to $S_V(f=1Hz) \approx 2\text{x}10^{-19}$/Hz in order to cause the observed level of frequency fluctuations. Information from additional experimental data in the literature also suggests that the average distance between these localized fluctuations is on the order of 10 microns.

Introduction

A number of studies have been carried out over the years in an attempt to characterize the nature of 1/f noise (often called flicker noise) in both Surface Acoustic Wave (SAW) and Bulk Acoustic Wave (BAW) resonators [1]-[5]. These efforts to understand 1/f noise have been made more difficult by strong evidence that there is more than one source of flicker noise in many acoustic devices. To minimize this complication, generally only devices having an unloaded Q close to the material limit have been studied in detail. Recently, a theory was proposed that states that the 1/f noise in a BAW resonator [6] is proportional to the active acoustic volume, Vol, of the device and inversely proportional to Q_u^4, as shown in Eq. 1.

$$S_y(f) = \frac{\beta}{f}\left(\frac{Vol}{Q_u^4}\right) \qquad (1)$$

Here $S_y(f)$ is the spectral density of fractional frequency fluctuations, f is the noise frequency, Q_u is the unloaded Q of the resonator, and β is a constant close to 1. A set of experiments has been carried out to determine if this theory applies to SAW devices by evaluating the dependence of 1/f noise on SAW device size. This paper reports an observed decrease in 1/f noise with increasing SAW device size which is clearly not in agreement with Eq. 1.

Experiment

Four different 450 MHz SAW resonator designs were evaluated to determine the dependence of 1/f frequency fluctuations on the size of the resonator. All of the designs were two-port resonators with 49 quarter-wavelength wide, recessed fingers in each unapodized transducer. Two different acoustic apertures of 90 wavelengths (λ) and 180λ were used along with two different effective cavity lengths [7] of 190λ and 390λ. The 390λ cavity length designs have a space of 250λ between the two transducers, and the 190λ designs have a space of 50λ between the transducers as illustrated in Fig. 1. All devices were fabricated on 36.7 degree rotated Y-cut quartz (ST-cut family) and had ion etched reflective gratings with 1000 grooves in each reflector. The groove depth and transducer metal thickness were both 1 percent of an acoustic wavelength. The transducers were fabricated with ~ .05 atomic percent copper doped aluminum and the devices were hermetically sealed under high vacuum as "All Quartz Packages" [8].

0-7803-0905-7/93 $3.00 © 1993 IEEE

Table 1

SUMMARY OF DEVICE DESIGN PARAMETERS AND PERFORMANCE DATA

F= 449.75 MHz±50 kHz λ = 7.01 microns

Design Ident.	Acoustic Aperture	Effective Cavity Length	Active Acoustic Area (λ^2)	Insertion Loss(dB)	Q_L	Q_u
NS	90λ	190λ	1.71×10^4	6.8	11,800	21,950
WS	180λ	190λ	3.42×10^4	5.1	6,375	14,600
NL	90λ	390λ	3.51×10^4	10.5	16,200	22,300
WL	180λ	390λ	7.02×10^4	7.4	11,400	20,850

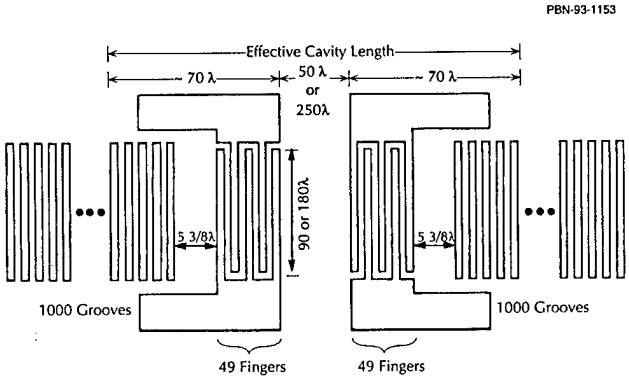


Figure 1. Illustration of the four different two-port SAW resonator designs.

A total of eight devices were fabricated, consisting of two of each design. All of the devices were fabricated at the same time on small rectangular plates cut from one large Premium Q, Pure Z quartz substrate. The locations of each type of device were randomly distributed about the large substrate. The resonant frequency, minimum insertion loss, loaded Q, Q_L, and the unloaded Q, Q_u, were measured on each of the sealed devices and the average values, along with the device design parameters, are shown in Table 1. All of the devices had a center frequency, F, of 449.75 MHz ±50 kHz. Each different design is identified in Table 1 by a pair of letters (for example NS, which signifies "Narrow-Short"). All of the resonators, except for the WS (Wide-Short) devices, had unloaded Q's greater than 89% of the material limit, which is 23,300 at 450 MHz. The WS devices exhibit a very low loaded Q as compared to the other three designs because the combination of a wide aperture and a short cavity results in stronger coupling to the higher order transverse modes. This allows energy from the lowest mode to leak into the other transverse modes which results in a low Q.

All of the devices were subjected to a high power RF burn-in to ensure that the lowest possible noise levels were achieved [9]. The burn-in power levels were adjusted for each design so that all devices were subjected to equal peak acoustic stress levels of approximately 2.3×10^8 N/m^2. Flicker noise measurements were made on all of the resonators, both before and after the burn-in, with an HP3048 Phase Noise Measurement System using the phase bridge technique [10]. In addition to the phase bridge measurements, each resonator was also successively mounted (after burn-in) on an oscillator board (containing an Avantek UTO-509 loop amplifier) and noise measurements were made on each oscillator. These measurements were made at noise frequencies from 1 Hz to 100 Hz using an HP8662A Synthesizer as a reference. $\mathcal{L}(f)$ was measured on each oscillator with an HP5390A Frequency Stability Analyzer using the modified Hadamard variance and the noise measurement was also repeated on two of the oscillators with the HP3048. Both types of oscillator measurement techniques gave results that agreed within 1 dB. Oscillator noise measurements were made at several values in the range from 8 volts to 15 volts for the DC bias to the UTO-509 loop amplifier to observe the dependence of the flicker noise on incident RF power level. No dependence was present.

Figure 2 shows the measured values of $\mathcal{L}(f=10Hz)$ for each of the oscillators as a function of the SAW resonator active acoustic area (in square wavelengths). The open circles represent the as measured oscillator data, and the solid circles represent the oscillator noise due only to the resonator. The latter data was obtained by removing the

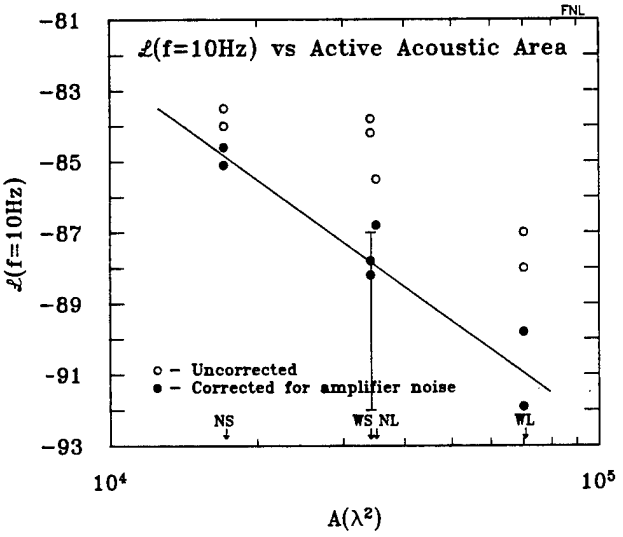


Figure 2. Measured phase noise of the SAW oscillators as a function of the resonator active acoustic area. The solid line illustrates a 1/A dependence. NS, WS, NL, and WL identify the different resonator designs.

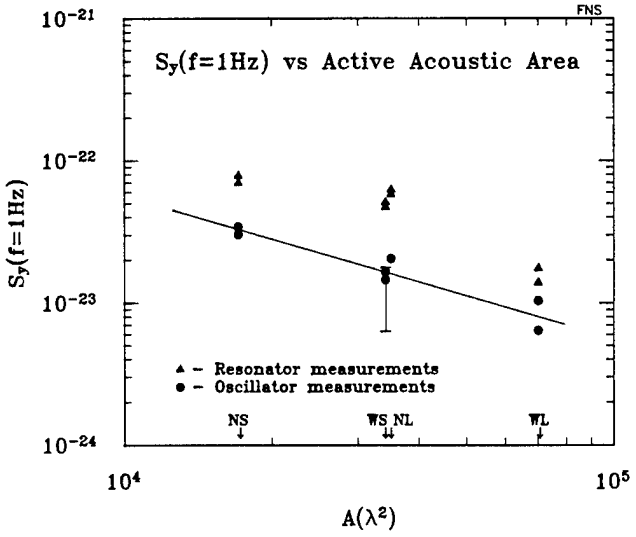


Figure 3. Measured values of S_y(f=1Hz) from the resonator residual noise (solid triangles) and from the oscillator noise (solid circles) as a function of SAW resonator size. All data is corrected for either amplifier noise or measurement system noise.

contribution of the loop amplifier to the oscillator noise. (The residual noise of the amplifier was measured to be $\mathcal{L}'$(f=1Hz) = -146 dBc/Hz.) The solid line in Fig. 2 illustrates a 1/A dependence. Figure 2 shows a large error bar for the corrected WS data because these low Q resonators had residual noise levels very close to the amplifier noise level, and consequently there was a large correction factor which was very sensitive to the particular value used for the amplifier noise level. The phase noise level of the reference synthesizer was 13 dB below the noise level of the quietest oscillator at $\mathcal{L}$(f=10Hz) and therefore was not a significant factor. f = 10 Hz was chosen as the best offset frequency for the data in Fig. 2 since the data at 10 Hz was least corrupted by frequency drift (which the Hadamard variance is susceptible to) and the SAW oscillator noise was still well above the reference oscillator noise level. The phase noise level of one of the NL oscillators was very unstable with time so it was not used. Therefore, there are only seven data points for the oscillator data. However, the phase bridge data on the resonator used in that oscillator was of good quality and is included in the data in Fig. 3.

Figure 3 shows the values of S_y(f=1Hz) as a function of resonator size as determined from the phase bridge measurements on the resonators alone (solid triangles), and from the oscillator measurements (solid circles). The oscillator data is derived from the corrected data in Fig. 2 and extrapolated to 1 Hz. The phase bridge data was corrected for the noise contribution of the measurement system which was approximately $\mathcal{L}$(f=1Hz) = -146 dBc/Hz. Both the resonator and oscillator data show the same decrease in noise level with increasing resonator size, but the measured noise levels from the two measurement techniques disagree by about 4 dB. At this time there is no explanation for the discrepancy.

Figure 4 shows the measured values for the fractional frequency deviations (Allan deviations) of the oscillators. Again the decreased noise level with increased resonator size is clearly evident. It should be noted that the values for $\sigma_y(\tau)$ are more consistent with the resonator measurements in Fig. 3 than the oscillator measurements, even though $\sigma_y(\tau)$ was measured on the oscillators.

It is clear from the data in Figs. 2, 3 and 4 that the device flicker noise does not increase with increased area, but decreases with size at a rate of approximately 6 dB for a factor of four increase in device area. Also there is no strong dependence on unloaded Q, since the WS devices exhibited a very low loaded Q, but were not exceptionally noisy. The data is definitely not consistent with Eq. 1.

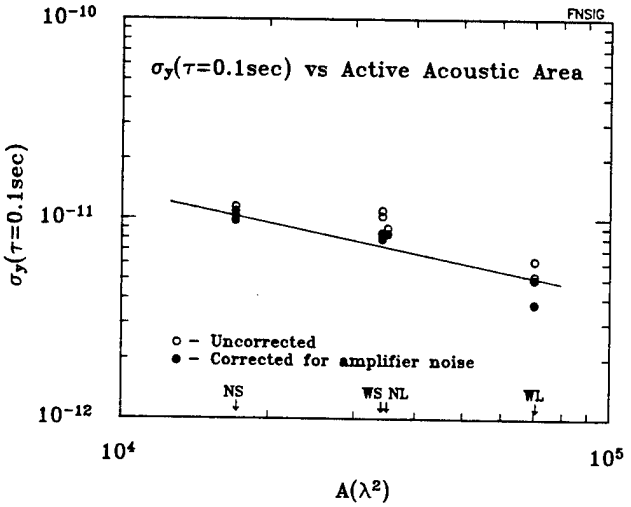


Figure 4. Fractional frequency deviations of the SAW oscillators as a function of SAW resonator size. The solid line represents a $(1/A)^{1/2}$ dependence.

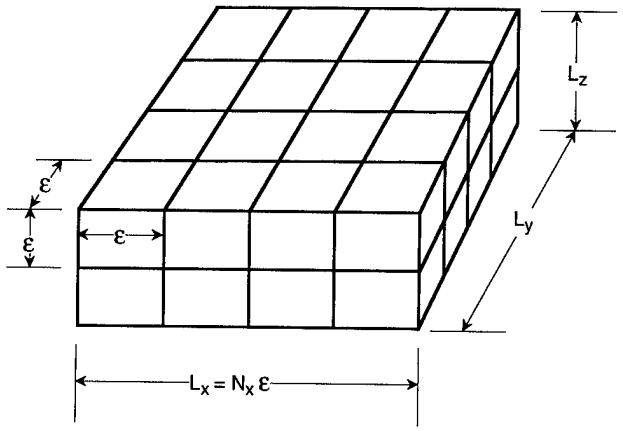


Figure 5. The SAW substrate is divided into small elements of dimension ε which represents the dimension over which the velocity fluctuations are coherent.

Theory

A possible explanation for the observed size dependence is that 1/f noise in SAW devices can be caused by localized, independent velocity fluctuations with a 1/f power spectral density. These fluctuations could be caused by mobile impurities or defects in the quartz, or possibly by a fundamental process involving phonon-phonon interactions. Similar phenomena have been proposed as an explanation for the observed 1/f characteristic of resistance fluctuations present in metal films and in semiconductors [11], [12]. No matter what the cause of the fluctuations, the observed size dependence will be present if the average spacing between the sources of independent velocity fluctuations is small compared to the dimensions of the SAW device in the plane of the substrate. This is illustrated in Fig. 5. The SAW substrate is divided into small cells of linear dimension ε, which is defined as the average separation between localized, independent velocity fluctuations. Consider a surface acoustic wave propagating in the x-direction, with the z-axis normal to the substrate surface. A small change in velocity in one of the elements in Fig. 5 will cause a small change in transit time across this element. Since a fractional change in velocity is related to a fractional change in delay time by the equation, $\Delta v/v = -\Delta t/t$, we can write

$$\Delta v_\varepsilon = -v\frac{\Delta t_\varepsilon}{t_\varepsilon} \qquad (2)$$

where v is the acoustic velocity in one of the elements, Δv_ε is the velocity fluctuation in this element, t_ε is the transit time for the element, and Δt_ε is the fluctuation in the transit time in this element. If the distance L_X includes N_X elements, then the average change in velocity over the distance L_X due to a fluctuation in one element is

$$\Delta v_A = -v\frac{\Delta t_\varepsilon}{N_x t_\varepsilon} = \frac{1}{N_x}\Delta v_\varepsilon. \qquad (3)$$

If more than one element is fluctuating, and the fluctuations are not coherent between elements, the total result of all of the velocity fluctuations along the x-direction, Δv^2_{Ax}, is obtained by summing the squares of the individual fluctuations.

$$\Delta v^2_{Ax} = \frac{1}{N_x^2}\sum_0^{N_x}\Delta v_\varepsilon^2 = \frac{1}{N_x}\Delta v_\varepsilon^2 = \frac{\varepsilon}{L_x}\Delta v_\varepsilon^2 \qquad (4)$$

The assumption has been made that the magnitude of the fluctuations is the same in all elements.

Similar arguments can be used in summing over the y- and z-directions, with the final result being

$$\Delta v^2_{Axyz} = \frac{\varepsilon^3}{L_x L_y L_z}\Delta v_\varepsilon^2. \qquad (5)$$

Since $\Delta F/F = \Delta v/v$, where F is the resonator frequency, we can write

$$\frac{\Delta F^2}{F^2} = \frac{1}{v^2}\frac{\varepsilon^3}{L_x L_y L_z}\Delta v_\varepsilon^2. \qquad (6)$$

Expressed as spectral densities this becomes

$$S_y(f) = \frac{1}{v^2}\frac{\varepsilon^3}{L_x L_y L_z} S_{\Delta v}(f) = \frac{1}{v^2}\frac{\varepsilon^3}{Vol} S_{\Delta v}(f) = \frac{\varepsilon^3}{Vol} S_v(f) \qquad (7)$$

where Vol is the physical volume occupied by the acoustic energy in the resonator, $S_{\Delta V}(f)$ is the spectral density of the velocity fluctuations and $S_V(f)$ is the spectral density of the fractional velocity fluctuations. In a SAW resonator L_X is the effective cavity length, L_Y is the effective acoustic aperture and L_Z is approximately half an acoustic wave length.

Equation 7 shows that a factor of two increase in L_X and L_Y results in a factor of four decrease in a resonator's flicker noise. This is in good agreement with the experimental data in Fig. 3, considering the small number of devices that were measured. The size dependence indicates that ε is much smaller than the smallest in-plane resonator dimension, and therefore ε << 630 microns. However, there are other data in the literature that this theory can also be tested against. Specifically, it has been reported that $S_y(f)$ in SAW resonators has an approximate F^2 dependence [4]. Since L_X, L_Y and L_Z scale inversely (at least approximately for L_X and L_Y) with frequency in most SAW designs, Eq. 7 shows that $S_y(f)$ should increase roughly as F^3. This is clearly not in agreement with the observed F^2 dependence. However, L_Z, which is approximately half an acoustic wavelength for a surface wave, is much smaller than either L_X or L_Y. If L_Z is smaller than the average spacing, ε, of the velocity fluctuations, the correct frequency dependence will be obtained since a change in L_Z results in no change in averaging. Therefore, ε would also have to be larger than half an acoustic wavelength for frequencies above approximately 300 MHz. Combining these two results yields

$$\sim 5\ \mu m < \varepsilon < \sim 200\ \mu m \qquad (8)$$

Equation 7 then becomes

$$S_y(f) = \frac{1}{v^2}\frac{\varepsilon^2}{L_x L_y} S_{\Delta v}(f) = \frac{1}{v^2}\frac{\varepsilon^2}{A} S_{\Delta v}(f) = \frac{\varepsilon^2}{A} S_v(f) \qquad (9)$$

where A is the active acoustic area of the SAW resonator. Equation 9 is in good agreement with both the current data on size dependence and the existing data on SAW resonator frequency dependence and applies when $\lambda/2 < \varepsilon$. It should be noted that the effect of velocity fluctuations on resonator frequency cannot be distinguished from dimensional fluctuations. Therefore, the observed 1/f noise could also be caused by localized dimensional fluctuations, or both.

If the frequency fluctuations in SAW devices are caused by velocity (or dimensional) fluctuations in the quartz itself, then one should also expect to see similar fluctuations in BAW resonators. However, it must be noted that the acoustic wavelength of a 5 MHz BAW device is approximately 90 times larger than that of the 450 MHz SAW devices, and therefore L_Z in the BAW resonator will be larger than ε. Thus Eq. 7 should be used for BAW calculations rather than Eq. 9. The measured flicker noise level of the best 5 MHz, fifth overtone, SC-cut BAW resonators (Vol = 0.06 cm^3) is approximately $S_y(f=1Hz) \approx 3 \times 10^{-27}/Hz$ [6],[13]. From this information and Eq. 7, and the SAW oscillator data in Fig. 3 and Eq. 9, one can calculate approximate values for $S_V(f)$ and ε. The results are

$$S_V(f=1Hz) \approx 2 \times 10^{-19}/Hz \qquad (10)$$

$$\varepsilon \approx 10\ \text{microns} \qquad (11)$$

Note that the above value of ε falls in the range estimated from the SAW data alone (Eq. 8). This is good evidence that similar processes may be taking place in both BAW and SAW devices. The value used for L_Y for the SAW resonators in the above calculations was 0.7 times the transducer acoustic aperture to account for the approximate transverse mode shape in a SAW resonator [7].

Proposed model for 1/f noise

The observed size dependence of the SAW resonator flicker noise could be caused by either a surface or a bulk effect. However, the observed dependence on resonator frequency reported in the literature [4] is not consistent

with a surface effect that is confined to a depth of substantially less than an acoustic wavelength. This is evident by noting that if the velocity fluctuations were caused by a surface effect, one would have $\Delta v \propto 1/\lambda \propto F$, since a higher acoustic frequency (hence a shorter wavelength) would concentrate more of the acoustic energy in the fluctuation region. Thus the influence of a surface velocity fluctuation would increase proportionally with the acoustic frequency. Therefore, including the size dependence, a surface effect would result in $S_y(f)$ being proportional to F^4. This is clearly not the case, so surface effects less than an acoustic wavelength in depth can be ruled out. Given the careful preparation of SAW substrates, it is very unlikely that surface effects could penetrate into the quartz by anything close to an acoustic wavelength.

Another interesting observation regarding the size dependence is that an increase in either dimension in the plane of the substrate appears to have the same effect in reducing the 1/f noise. This reduces the probability that the transducers are the source of the noise since doubling the length of the cavity involves no change in the transducers and would result in a 6 dB noise reduction if the velocity fluctuations were localized to the transducers. This is clearly not the case. The fact that the effect of the increase in device size is essentially the same for both in-plane directions strongly suggests that the quartz is the source of the 1/f noise in these devices.

Therefore, the proposed model for one possible source of 1/f noise in acoustic resonators is that 1/f frequency fluctuations are caused by localized, independent fractional velocity (or dimensional) fluctuations in the quartz, $S_V(f)$, that have a 1/f dependence. The average separation between these independent noise sources is ε. The fluctuations in the quartz would appear to be independent of the frequency and power level (for moderate power levels) of the acoustic wave. Equation 9 should be used for SAW resonators where $\lambda/2 < \varepsilon$, and Eq. 7 should be used for BAW resonators as long as the resonator thickness is significantly larger than ε. Approximate values for $S_V(f)$ and ε are given in Eqs. 10 and 11. The concept of flicker noise being caused by localized, independent noise sources is now generally accepted as a fact in most metal and semiconductor systems [11], and therefore should not be unexpected for acoustic devices.

It should be noted that the model presented above is also entirely consistent with the observation of Driscoll [14] that the use of two BAW resonators in series or parallel results in a 3 dB reduction in flicker noise as compared to a single resonator. Combining two resonators doubles the volume of the acoustic energy.

Discussion and speculation

In order to firmly establish the proposed localized velocity fluctuations in the quartz as a source of flicker noise in SAW and BAW resonators it will be necessary to identify a physical process (or processes) that can cause fractional velocity fluctuations on the order of 0.5 parts per billion with an average separation, or coherence length, of approximately 10 microns. Given the experience of investigators of 1/f noise in other materials such as metal films and semiconductors, this will no doubt be a difficult task. As with metal films, a reasonable starting place is with mobile impurities or defects. Considerable information is available on defects and impurities in quartz, and it is known that both can be mobile [15]. However, a brief survey of this information has not yielded any useful results. The density of dislocations in high quality synthetic quartz is about $100/cm^2$, and is therefore too low to give an effective spacing on the order of 10 microns. On the other hand, the concentration of hydrogen impurities is too high (250 to 2500 atomic PPM relative to silicon) and gives an average distance between impurities of 10 to 100 Å. Many other impurities may also be present in quartz at lower concentrations than hydrogen, but even their concentrations are too high [15] to give a mean separation on the order of 10 microns. If the velocity fluctuations are related to mobile impurities, it is necessary that only a small fraction of the available impurity atoms be involved in the fluctuation process.

A theory for 1/f frequency fluctuations in BAW resonators based on the diffusion of mobile defects in quartz has been proposed by Chaban [16],[17] and the expression for $S_y(f)$ in this theory includes a term equivalent to ε^3/Vol in Eq. 7. Furthermore, Chaban's estimate for ε (r_0 in ref. 16) is on the order of 10 microns. However, there are some problems with this theory. In addition to the ε^3/Vol term in ref. 16, there is also a multiplicative Q^{-3} term which gives a much stronger dependence of the flicker noise on the resonator frequency ($S_y(f)$ approximately proportional to F^5) than is observed. (Note that the acoustic volume of typical BAW resonators tends to decrease with increased resonator frequency approximately as F^{-2}.) Also, it is not clear to the authors how the value of 10 microns for r_0 in ref. 16 was derived.

An additional class of data that any theory for 1/f noise in acoustic resonators must be consistent with is the observed temperature dependence of flicker noise [3] in BAW resonators. Generally, the decrease in 1/f noise at low temperatures is explained by the accompanying increase in unloaded Q (see Eq. 1). However, Eq. 7 in the present paper is based on localized velocity

fluctuations and does not explicitly contain Q. But it would not be unrealistic to expect $S_V(f)$ to vary inversely with the Q·F product of quartz and therefore could bring in an appropriate temperature dependence.

Since the publication of ref. 6, more noise data on a variety of BAW resonators in the frequency range of 40 MHz to 160 MHz has been published [18] and the results have been compared to Eq. 1 and Eq. 7. Neither theory fits this data particularly well, with both approaches having an average error of between 4 and 5 dB. If the theory based on localized velocity fluctuations is applied to the data in ref. 6, a poor fit is obtained with an average error between 7 and 8 dB, and a worst case error of 20 dB.

It is clear that much more work is still needed to explain 1/f noise in acoustic resonators, and one of the tasks must be to test some of the predictions of the proposed model (Eqs. 7 and 9). More data over a larger range of SAW device sizes would be very useful, as well as more data at different resonator frequencies, since much of the data on the F dependence in ref. 4 comes from oscillator data where the contribution of the loop amplifier is not known. If the proposed theory is true for SAW resonators, it should also hold for SAW delay lines. Estimates of residual noise levels and a prediction of the dependence on delay time can be tested. Unlike SAW resonators, delay lines can be operated at harmonics of the fundamental frequency and very useful information can be gathered by measuring the 1/f noise at different operating frequencies on the same device.

BAW resonators can, in principle, be operated at several harmonics, and noise measurements at several different operating frequencies of the same device would be very useful for determining the importance of unloaded Q in 1/f noise levels (Eq. 1 or Eq. 7). Also it would be very useful to have noise measurement results for BAW resonators over a range of resonator sizes at a constant operating frequency .

Conclusions

Experiments have been carried out to evaluate the dependence of SAW resonator flicker noise on device size and it has been found that the flicker noise decreases as the size increases. A theory based on localized, independent velocity or dimensional fluctuations in the quartz has been proposed which gives the correct dependence of $S_y(f)$ on resonator size and frequency if an average separation between noise sources of not less than approximately 5 microns and not more than about 200 microns is assumed. The theory also gives a good estimate for the flicker noise level in state-of-the-art 5 MHz BAW resonators if an average spacing, ε, of approximately 10 microns is used along with a value of $S_V(f=1Hz) \approx 2x10^{-19}/Hz$. Given the high degree of variation in flicker noise levels that occurs in otherwise "identical" devices, it will be necessary to test this theory with measurements on many more devices and with of a variety of SAW and BAW designs before it can truly be verified.

Acknowledgments

The authors would like to acknowledge G. Montress and J. Greer for valuable discussions, M. Hall for making the residual phase noise measurements on the SAW resonators, and B. Howard for fabricating the SAW devices.

References

[1] J. J. Gagnepain, "Fundamental Noise Studies of Quartz Crystal Resonators", in *Proc. of the 30th Annual Symp. on Freq. Control*, 1976, pp. 84-91.

[2] P.H. Handel, "Nature of 1/f Frequency Fluctuations in Quartz Crystal Resonators", *Solid State Electronics*," vol. 22, 1979, pp. 875-876.

[3] J. J. Gagnepain, J. Ubersfeld, G. Goujon, and P. Handel, "Relation Between 1/f Noise and Q-Factor in Quartz Resonators at Room and Low Temperatures", *Proc. of the 35th Annual Symp. on Frequency Control*, 1981, pp. 476-483.

[4] T. E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators", in *Proc. of the 41st Annual Symp. on Freq. Control*, 1987, pp. 99-110.

[5] V. F. Kroupa, "The State-of-the-Art of the Flicker Frequency Noise in BAW and SAW Quartz Resonators", *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 35, no. 3, pp. 406-420, May 1988.

[6] F. L. Walls, P.H. Handel, R. Besson and J. J. Gagnepain, "A New Model of 1/F Noise in BAW Quartz Resonators", in *Proc. of the 1992 IEEE Freq. Control Symp.*, pp. 327-333.

[7] W. R. Shreve and P. S. Cross, "Surface Acoustic Waves and Resonators", in *Precision Frequency Control: Acoustic Resonators and Filters*, Vol. 1, E. A. Gerber and A. Ballato, Eds., New York: Academic Press, 1985, pp. 118-145.

[8] T. E. Parker, J. Callerame and G. K. Montress, "A New All Quartz Package for SAW Devices", in *Proc. of the 39th Annual Symp. on Frequency Control*, 1985, pp. 519-525.

[9] G. K. Montress, T. E. Parker, M. J. Loboda, and J. A. Greer, "Extremely Low-Phase-Noise SAW Resonators and Oscillators: Design and Performance", *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. 35, No. 6, pp. 657-667, November 1988.

[10] F. L. Walls and A. E. Wainright, "Measurement of the Short-Term Stability of Quartz Crystal Resonators and the Implications for Crystal Oscillator Design and Applications", IEEE Transactions on Inst. and Meas., Vol. IM-24, pp. 15-20, 1975.

[11] M. B. Weissman, "*1/f* Noise and Other Slow, Non exponential Kinetics in Condensed Matter", *Reviews of Modern Physics*, Vol. 60, No. 2, pp. 537-571, April 1988.

[12] A. Van der Ziel, "Unified Presentation of *1/f* Noise in Electronic Devices: Fundamental *1/f* Noise Sources", *Proc. of the IEEE*, Vol. 76, No. 3, pp. 233-258, March 1988.

[13] T. E. Parker and G. K. Montress, "Spectral Purity of Acoustic Resonator Oscillators", in *Proc. of the 1992 IEEE Freq. Control Symp.*, pp. 340-348.

[14] M. Driscoll, "Reduction of Quartz Crystal Oscillator Flicker-of-Frequency and White Phase Noise (Floor) Levels and Acceleration Sensitivity via use of Multiple Resonators", in *Proc. of the 1992 IEEE Freq. Control Symp.*, pp. 340-348.

[15] J. C. Brice, "Crystals for Quartz Resonators", *Reviews of Modern Physics*, Vol. 57, No. 1, pp. 105-146, January 1985.

[16] I. A. Chaban, "Flicker Noise in Quartz Resonators", *Sov. Phys. Acoust.*, Vol 36, No. 6, pp. 614-617, November 1990.

[17] I. A. Chaban, "On the Nature of Flicker Noise", *Sov. Journal of Communications Tech. and Elec.*, Vol 33, No. 5, pp. 65-72, May 1988.

[18] M. M. Driscoll and W. P. Hanson, "Measured vs. Volume Model-Predicted Flicker-of-Frequency Instability in VHF Quartz Crystal Resonators", (in this Proceedings)

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MEASURED VS. VOLUME MODEL-PREDICTED FLICKER-OF-FREQUENCY INSTABILITY IN VHF QUARTZ CRYSTAL RESONATORS

*M. M. Driscoll and **W. P. Hanson

*Westinghouse Electronic Systems Group, Baltimore, MD
**Piezo Crystal Co., Carlisle, PA

Abstract

Levels of flicker-of-frequency instability have been measured for several types of low noise, VHF, quartz crystal resonators. The results have been compared to those predicted using a formula proposed in a paper presented by Fred Walls at last year's Frequency Control Symposium [1]. In that paper, a β_e (electroded crystal) data-fitting value of 4 was found to accurately describe best-measured crystal flicker-of-frequency instability. The formula suggests crystal Sy(f) flicker-of-frequency instability that varies as the (electroded) volume, and as the inverse of Q^4.

The data reported in this paper are the result of both passive, in-bridge as well as in-oscillator measurements for AT, SC, and BT-cut, overtone-mode resonators operating at 40, 80, 100, and 160MHz. Measurements were made using four-crystal oscillators as low noise reference signal generators [2]. The data indicates poor agreement between measured and model-predicted resonator stability, especially for 5th overtone AT and SC-cut resonators operating at 40 and 160MHz.

Measured data was also compared to that predicted using an alternative model (formula) recently proposed by T. Parker [3]. Using the Parker model, relatively poor agreement was found for the case of the BT-cut resonators. The data suggests that further refinement of resonator short-term stability models is necessary.

Introduction

The signal phase noise spectral characteristics of quartz crystal-controlled oscillators can be divided into two distinct regions: (1) a near-carrier region where the signal spectrum is dominated by flicker-of-frequency instability, and (2) a white phase noise (floor) region. For many years, prediction of crystal oscillator output signal spectral performance was made based on calculation and/or measurement of sustaining stage open loop phase noise characteristics, with subsequent prediction of resultant oscillator closed loop behavior made by considering the conversion of phase-to-frequency instability in accordance with the resonator in-circuit group delay [2]. Using this model, oscillator signal flicker-of-frequency noise was often assumed to be the result of sustaining stage open loop, flicker-of-phase noise.

Beginning in the mid 1970's, measurement of quartz crystal resonator short-term frequency stability by Walls and others showed that quartz crystal (as well as surface acoustic wave) resonator flicker-of-frequency noise, rather than sustaining stage flicker-of-phase noise, is usually the primary contributor to oscillator signal near-carrier spectral performance [4-8]. For the purpose of simplification, resonator short-term frequency instability will be referred to as "self-noise." Substantial effort has been undertaken in the frequency control community in the past several years in an effort to reduce resonator self-noise levels in order to improve the output signal spectral performance of quartz crystal and SAW oscillators. A good portion of this effort has involved determination, often via trial-and-error and educated-guess experimentation, of resonator processing modifications resulting in improved yields of low noise resonators.

On the analytical side, a number of theories have been advanced and calculations performed in an effort to accurately account for quartz crystal self-noise. Included are: (1) efforts to define the actual mechanisms involved (i.e., surface contaminant molecular motion, 1/f quantum noise, etc.) and (2) related efforts aimed at fitting measured data to a model that would indicate the relationship of various, controllable, resonator parameters to resultant device self-noise level [8-11].

The results of these efforts is such that the dominant causes of quartz crystal resonator self-noise remain unclear. One reason for this is the likelihood of several, distinct contributors to resonator self-noise. We know, for example, that resonator surface condition is critical. [High-noise resonators have been found to exhibit noise levels comparable to that of the lowest noise level devices after the surface was stripped and electrodes re-plated]. Therefore, attributing measured device self-noise levels to a single mechanism appears inappropriate. Another problem involves the non-independent relationship between resonator parameters used to develop formulas intended to accurately model noise behavior. For example, low noise is consistently obtained in high Q resonators. High Q values are

0-7803-0905-7/93 $3.00 © 1993 IEEE

normally associated with large active volume (i.e., lower resonant frequency and/or overtone operation. However, high Q value is also usually associated with high motional impedance, or a smaller relative (electroded) volume. In addition, much of the analysis is based on data where additional factors that might influence measured noise level are unrecorded.

At the 1992 IEEE Frequency Control Symposium, F. L. Walls proposed a new model for describing lowest attainable flicker-of-frequency noise levels in quartz crystal resonators. The model is described by the equation:

$$Sy(f) = (ß)(\text{electrode volume, in } cm^3)(Q^{-4})f \quad (1)$$

Using best (lowest noise level) measured data for plated electrode and non-plated electrode (BVA) crystals, an excellent data fit was obtained, using a value $ß_e = 4$ for the plated electrode crystals and $ß_b = 1.4$ for the BVA devices.

This paper compares published, flicker-of-frequency data obtained for low noise, VHF crystals purchased by Westinghouse from Piezo Crystal Co. to that reported by Walls in terms of the model-predicted (equation 1) behavior.

Measurement of Quartz Crystal Flicker-of-Frequency Noise

For the results reported here, standard, in-bridge and in-oscillator techniques for measurement of quartz crystal self-noise level were used. Simplified block diagrams for each of the two measurement techniques are shown in figures 1 and 2. The only novel aspect of these measurement techniques involves the use of ultra-low noise oscillators employing four (lowest available noise level) crystals. The resultant oscillator output signal flicker-of-frequency noise is typically 6dB below that exhibited by each of the individual oscillator crystals. Using these oscillators, measurement of the self noise of individual UUT crystals was possible, and the necessity for trial-and-error measurements using various combinations of crystal pairs in order to locate the lowest noise crystal for use as a test reference unit was eliminated. For all measurements reported here, comparative resonator/oscillator noise evaluation was made at 100Hz carrier offset frequency. Three reference oscillators were available for use, each employing four, lowest-available noise crystals, as well as varactor diode frequency tuning circuitry.

For the 40MHz (5th overtone, AT-cut crystal) oscillator:

$$£(f=100Hz) = -157dBc/Hz,\ Sy(f=100Hz) = 2.5X10^{-27} \quad (2)$$

For the 80MHz (3rd overtone, SC-cut crystal) oscillator:

$$£(f=100Hz) = -143dBc/Hz,\ Sy(f=100Hz) = 1.6X10^{-26} \quad (3)$$

For the 100MHz (3rd overtone, SC-cut crystal) oscillator:

$$£(f=100Hz) = -143dBc/Hz,\ Sy(f=100Hz) = 1.0X10^{-26} \quad (4)$$

Measurement Results

Figure 3 shows typical plotted results obtained for in-bridge measurement of individual crystal self-noise level. Similarly, figure 4 shows a typical in-oscillator crystal self-noise measurement result. Table 1 provides a summary of the best self-noise measurement results obtained for each of the evaluated crystal resonator types, together with associated resonator parameters applicable to equation (1). As shown in the table, the equation data fit (in terms of actual vs equation-predicted stability) is rather poor, especially for the case of the 5th overtone, 160MHz SC-cut crystals. For these crystals, the measured [Sy(f=100Hz)] stability is a factor of 13 times better than predicted by the formula.

Table 2 compares the same measured data to that predicted using an alternative formula proposed by Tom Parker given by:

$$Sy(f) = K/(V^2 \cdot f \cdot Vol) \quad (5)$$

where V = acoustic velocity in meters/sec, Vol = resonator acoustic energy volume in cubic meters, and $K = 2X10^{-27}$. The formula proposed by Parker is substantially different than the Walls formula in that there is no Q dependency, and the volume term appears in the denominator rather than the numerator [3]. As shown in table 2, the measured data fit to this formula is also not good, especially for the case of the 100MHz BT-cut resonators (25:1 discrepancy) and 40MHz, AT-cut resonators (45:1 discrepancy). For both the table 1 and table 2 calculations of model-predicted stability, the effective resonator volume was calculated as total electroded volume.

Measurement Results Using Actual (Acoustic Energy) Resonator Volume

Calculations of actual acoustic energy resonator volume (as opposed to the electroded volume) are based on a set of equations developed by Stevens and Tiersten describing the mode pattern and inharmonic frequencies for doubly-rotated, contoured resonators [12]. Measured data presented by Stevens and Tiersten for 5MHz, SC-cut resonators closely matches predicted behavior. Agreement between predicted and measured behavior has recently also been confirmed for contoured 10MHz, 3rd overtone, SC-cut devices as well [13].

The vibrating portion of the resonator is a potential source of noise. The mode patterns vary with orientation. The main 100, 300, and 500 modes in the AT, SC, and BT cut have central regions of large displacement with decreasing displacement near the edges. Noise sources inside of or on the surface of the resonator may depend upon the magnitude of the displacement where they are located.

For this work, resonators operating in the frequency range 40MHz to 160MHz were analyzed. These resonators all nearly flat. Residual asymmetrical stresses in the resonator blanks make the calculation of resonator acoustic volume more difficult. The stresses spread from the mounting structure and vary from resonator to resonator. It is possible however, to make contour estimates, based on measurement of the resonator inharmonic response spectrum.

The acoustic energy volume was calculated using Stevens and Tiersten's equations for doubly rotated contoured quartz crystal resonators [12]. Equation (6) has been extrapolated to nearly flat blanks. The inharmonic frequencies were measured, and a "best fit" contour was determined from the inharmonic frequency spacing. The electrode material and thickness were included in the calculations. The inharmonic frequencies were determined from Stevens and Tiersten's equation 4.6 repeated here as equation (6):

$$\omega^2_{mnp} = \frac{\pi^2 \bar{c}^{(1)}}{4 h_0^2 \rho}\left[1+\frac{1}{m\pi}\left(\frac{2h_0}{R}\right)^{1/2}\left(\sqrt{\frac{M_n'}{\bar{c}^{(1)}}}(2m+1)+\sqrt{\frac{P_n'}{\bar{c}^{(1)}}}(2p+1)\right)\right] \quad (6)$$

The acoustic volume was calculated from the normalized mode shape and thickness. A MathCad spreadsheet was used to calculate the volume using equation 4.4 in Stevens and Tiersten's paper, and the calculated thickness determined from the Christoffel solution. Figure 5 is the set of calculations for the 160MHz, 5th overtone SC-cut crystal. The MathCad spreadsheet solved the Christoffel problem, including temperature effects, for the elastic, piezoelectric, dielectric, and thermal expansion coefficients. The rotated constants, eigenvectors, and eigenvalues were then used in Stevens and Tiersten's equations to determine the mode shape and inharmonic frequencies. The contour was varied until a reasonable match was found.

The calculated mode volumes are shown in tables 3 and 4. The Q of the resonator is known to be limited by the viscosity (fluency), mounting configuration, and, in some cases, the atmospheric pressure. The model for the Q in a resonator is limited by the material viscosity for the orientation and mode in question. Actual resonator Q values fall at or below this material limit. The models for phase noise may be limited in a similar way.

Comparison between predicted and measured performance using the acoustic energy volume in the Wall's formula (table 3) indicates somewhat better agreement than was obtained using electroded volume, except for the case of the 100MHz BT-cut resonators. Using the acoustic energy volume a factor of 12 difference between predicted and measured stability was obtained for the BT-cut devices.

Using acoustic energy volumes in the Parker formula (table 4) resulted in poor agreement between predicted and actual results for the 100MHz BT-cut and 40MHz AT-cut resonators. The discrepancy between predicted and actual results is reduced however, when active volume rather than electroded volume is used.

Conclusions

Short-term frequency stability measurements have been made at Westinghouse for low noise, VHF AT, BT, and SC-cut crystals.

Using the best (lowest noise) data, relatively poor agreement exists between measured resonator stability and that predicted using the electroded volume in formulas proposed by either Walls and Parker for certain of the resonator types.

When actual acoustic energy volume (rather than electroded volume) is used, the data fit is better for certain resonator types and worse for others. Using either electroded or active volume in either formula, discrepancies of more than 10:1 exist between predicted and measured performance for certain resonator types.

Formula data fit inaccuracy is likely result of the fact that resonator short-term frequency instability is related to several, distinct mechanisms, the parameters of which are not fully included in the formulas.

In addition, the contrary (volume) implications of the two formulas is indicative of continuing uncertainty as to the mechanisms involved in determining resonator short-term frequency instability.

References

[1] F. Walls, "A New Model of 1/f Noise in BAW Quartz Resonators", Proc. 1992 IEEE Freq. Contr. Symp., May 1992, pp. 327-333.

[2] D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum", Pro. IEEE, Vol. 54, No. 2, Feb. 1966, pp. 329-330.

[3] Correspondence with T. Parker (Raytheon Research Labs)

[4] F. Walls, et al., "Measurement of the Short-Term Stability of Quartz Resonators and the Implication for Crystal Oscillator Design", IEEE Trans. Instr. and Meas., Vol. IM-24, No. 1, March 1975, pp. 15-20.

[5] J. Gagnepain, et al., "Excess Noise in Quartz Crystal Resonators", Proc. 37th Freq. Contr. Symp., May 1983, pp. 218.

[6] S. Eliot, et al., "Direct Phase Noise Measurements of SAW Resonators", Proc. 1984 IEEE Ultras. Symp., Oct. 1984, pp. 180-185.

[7] M. M. Driscoll and B. W. Kramer, "Spectral Degradation in VHF Crystal-Controlled Oscillators Due to Short-Term Instability in the Quartz Resonator", Proc. 1985 IEEE Ultras. Symp., Oct. 1985, pp. 340-345.

[8] T. Parker, "1/f Fluctuations in Acoustic and Other Stable Oscillators", Proc. 39th Freq. Contr. Symp., May 1985, pp. 97-106.

[9] J. Gagnepain, "Fundamental Noise Studies of Quartz Crystal Resonators", Proc. 30th Freq. Contr. Symp., May 1976, pp. 84-91.

[10] J. Gagnepain, et al., "Relationship Between 1/f Noise and Q Factor in Quartz Resonators at Room and Low Temperatures, First Theoretical Interpretation", Proc. 35th Freq. Contr. Symp., May 1981, pp. 476-483.

[11] Y. Yong and J. Vig, "Resonator Surface Contamination - A Cause of Frequency Fluctuations?", Proc. 42nd Freq. Contr. Symp., May 1988, pp. 397-403.

[12] D. S. Stevens and H. F. Tiersten, "An Analysis of Doubly-Rotated Contoured Quartz Crystal Resonators", Proc. 39th Freq. Contr. Symp., May 1985, pp. 436-447.

[13] W. P. Hanson, "Nondestructive Evaluation Methods for Vibrating Quartz Resonators Using Synchrotron X-Ray Topography and Laser Interferometry", 1993 Thesis, The Johns Hopkins University, Baltimore, MD.

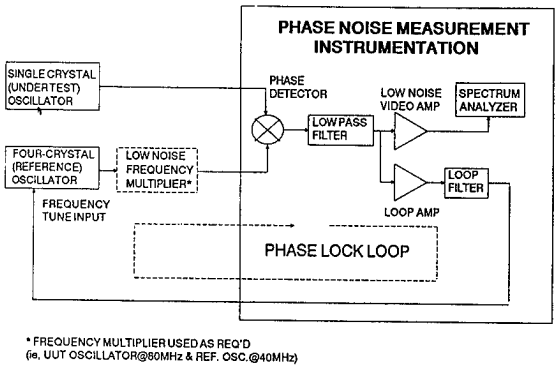


Figure 1. Block Diagram for In-Oscillator Measurement of Crystal Resonator Flicker-of-Frequency Noise

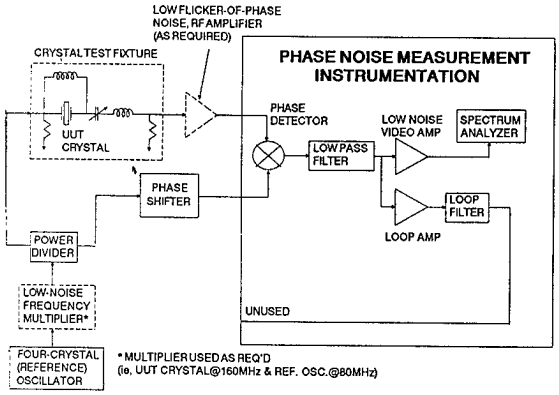


Figure 2. Block Diagram for In-Bridge Measurement of Crystal Resonator Flicker-of-Frequency Noise

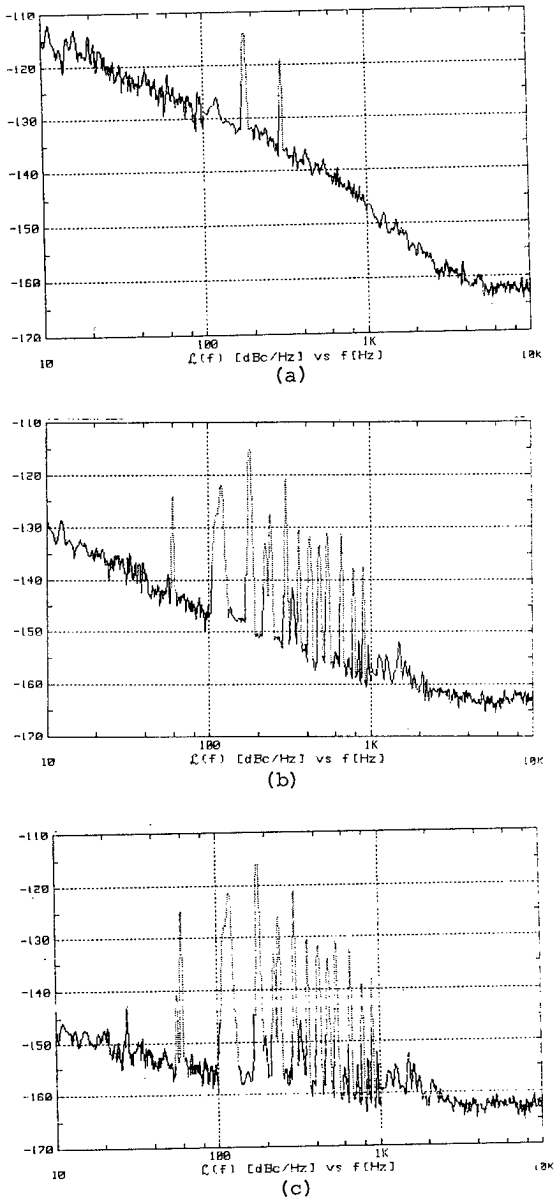


Figure 3. Typical In-Bridge Crystal Noise Measurement Results for (a) High Noise, (b) Moderate Noise, and (c) Very Low Noise 100MHz, SC-Cut Crystals. In-Bridge Delay = 160usec.

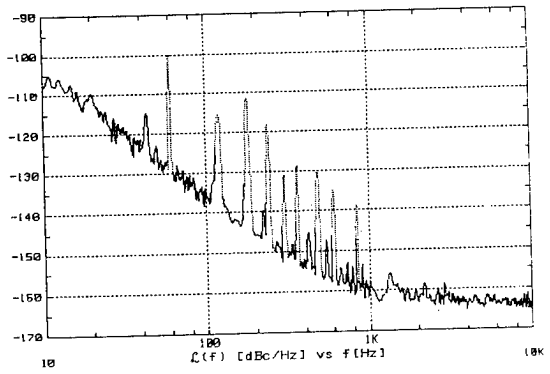


Figure 4. Typical In-Oscillator Measurement Results for 40MHz, Four-Crystal Oscillator-Doubler vs. Single (UUT) Crystal Oscillator

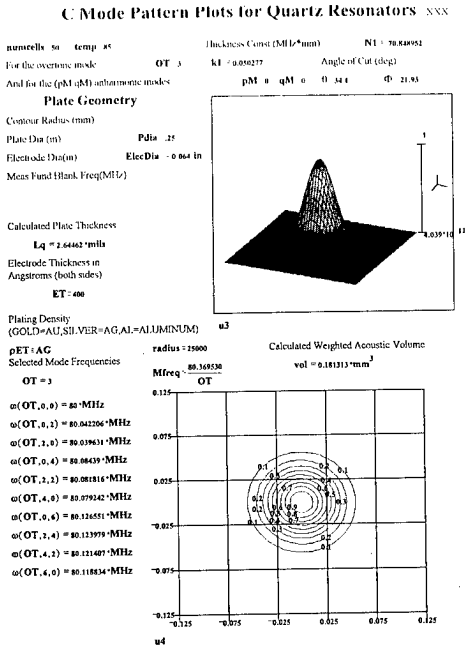


Figure 5. Mode Pattern Results for a 5th Overtone, 80MHz, SC-Cut Resonator

Table 1. VHF Crystal Flicker-of-Frequency Noise: Best Measured Results in Comparison to Walls' Volume Model Predictions Using Electroded Volume

CRYSTAL CUT	SC	SC	SC	BT	AT
FREQUENCY	80MHz	100MHz	160MHz	100MHz	40MHz
OVERTONE	3RD	3RD	5TH	5TH	5TH
UNLOADED Q	125K	119K	75K	300K	250K
BLANK THICKNESS (in.)	.00267	.00214	.00223	.005	.0083
ELECTRODE DIAMETER (in.)	.064	.055	.100	.125	.144
ELECTRODED VOLUME (cu. cm.)	.00014	.000083	.00028	.001	.0022
BEST- MEASURED SY(f=100Hz)*	$7.4X10^{-26}$	$3.6X10^{-26}$	$4.1X10^{-26}$	$3.0X10^{-26}$	$1.25X10^{-26}$
RESONATOR Sy(f=100Hz)**	$6.4X10^{-26}$	$2.6X10^{-26}$	$2.5X10^{-26}$	$2.0X10^{-26}$	$1.0X10^{-26}$
MODEL-PREDICTED Sy(f=100Hz) [using β_e=4 in eq. 1]	$2.3X10^{-26}$	$1.7X10^{-26}$	$\underline{3.3X10^{-25}}$	$5.0X10^{-27}$	$2.2X10^{-26}$

* Including contribution of 4-crystal reference oscillator
** Removing contribution of 4-crystal reference oscillator

Table 2. VHF Crystal Flicker-of-Frequency Noise: Best Measured Results in Comparison to Parker's Formula Predictions Using Electroded Volume

CRYSTAL CUT	SC	SC	SC	BT	AT
FREQUENCY	80MHz	100MHz	160MHz	100MHz	40MHz
OVERTONE	3RD	3RD	5TH	5TH	5TH
UNLOADED Q	125K	119K	75K	300K	250K
BLANK THICKNESS (in.)	.00267	.00214	.00223	.005	.0083
ELECTRODE DIAMETER (in.)	.064	.055	.100	.125	.144
ELECTRODED VOLUME (cu. cm.)	.00014	.000083	.00028	.001	.0022
BEST- MEASURED SY(f=100Hz)*	$7.4X10^{-26}$	$3.6X10^{-26}$	$4.1X10^{-26}$	$3.0X10^{-26}$	$1.25X10^{-26}$
RESONATOR Sy(f=100Hz)**	$6.4X10^{-26}$	$2.6X10^{-26}$	$2.5X10^{-26}$	$2.0X10^{-26}$	$1.0X10^{-26}$
MODEL-PREDICTED Sy(f=100Hz) [using eq. 2]	$1.1X10^{-26}$	$1.9X10^{-26}$	$5.5X10^{-27}$	$\underline{8x10^{-28}}$	$\underline{2.2X10^{-28}}$

* Including contribution of 4-crystal reference oscillator
** Removing contribution of 4-crystal reference oscillator

Table 3. VHF Crystal Flicker-of-Frequency Noise: Best Measured Results in Comparison to Walls' Volume Model Predictions Using Acoustic Energy Volume

CRYSTAL CUT	SC	SC	SC	BT	AT
FREQUENCY	80MHz	100MHz	160MHz	100MHz	40MHz
OVERTONE	3RD	3RD	5TH	5TH	5TH
UNLOADED Q	125K	119K	75K	300K	250K
BLANK THICKNESS (in.)	.00267	.00214	.00223	.005	.0083
ELECTRODE DIAMETER (in.)	.064	.055	.100	.125	.144
ELECTRODED VOLUME (cu. cm.)	.00018	.000056	.000054	.00034	.0012
BEST- MEASURED SY(f=100Hz)*	$7.4X10^{-26}$	$3.6X10^{-26}$	$4.1X10^{-26}$	$3.0X10^{-26}$	$1.25X10^{-26}$
RESONATOR Sy(f=100Hz)**	$6.4X10^{-26}$	$2.6X10^{-26}$	$2.5X10^{-26}$	$2.0X10^{-26}$	$1.0X10^{-26}$
MODEL-PREDICTED Sy(f=100Hz) [using β_e=4 in eq. 1]	$2.9X10^{-26}$	$1.1X10^{-26}$	$6.8X10^{-26}$	$1.6X10^{-27}$	$1.2X10^{-26}$

Table 4. VHF Crystal Flicker-of-Frequency Noise: Best Measured Results in Comparison to Parker's Formula Predictions Using Acoustic Energy Volume

CRYSTAL CUT	SC	SC	SC	BT	AT
FREQUENCY	80MHz	100MHz	160MHz	100MHz	40MHz
OVERTONE	3RD	3RD	5TH	5TH	5TH
UNLOADED Q	125K	119K	75K	300K	250K
BLANK THICKNESS (in.)	.00267	.00214	.00223	.005	.0083
ELECTRODE DIAMETER (in.)	.064	.055	.100	.125	.144
ELECTRODED VOLUME (cu. cm.)	.00018	.000056	.000054	.00034	.0012
BEST- MEASURED SY(f=100Hz)*	$7.4X10^{-26}$	$3.6X10^{-26}$	$4.1X10^{-26}$	$3.0X10^{-26}$	$1.2X10^{-26}$
RESONATOR Sy(f=100Hz)**	$6.4X10^{-26}$	$2.6X10^{-26}$	$2.5X10^{-26}$	$2.0X10^{-26}$	$1.0X10^{-26}$
MODEL-PREDICTED Sy(f=100Hz) [using eq. 2]	$8.5X10^{-27}$	$2.7X10^{-26}$	$2.8X10^{-26}$	$2.3X10^{-27}$	$3.8X10^{-28}$

* Including contribution of 4-crystal reference oscillator
** Removing contribution of 4-crystal reference oscillator

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A THERMODYNAMIC RESONANCE IN PIEZOELECTRIC CRYSTAL PLATES OF THICKNESS-SHEAR VIBRATIONS

Yu.I.Yevdokimenko *, Yu.S.Shmaly **

* Research Laboratory, ** Department of Radio-Transmitting and Radio-Receving Devices, Kharkov Military University, 228 Klochkovskaya Street, Kharkov, Ukraine, 310165

ABSTRACT

While investigating the thermodynamic behavior of thickness-shear vibrations of a disc quartz piezoelectric plate of singly or doubly rotated orientations it has been show that in the space of at least three parameters - an ambient temperature variation frequency, a heat transfer coefficient of a quartz resonator to the medium, and a diameter/thickness relation,- the function has a particularly pronounced resonance called a thermodynamic resonance. The top of the resonance in the frequency range of less than 1 Hz for different quartz resonators designs and its character is responsible for a power spectral density burst of OCXO fluctuations with extremely small turning away from the carrier frequency.

Practical recomendations have been given as to the choice of the quartz resonator parameters providing a minimum thermo-dynamic frequency instability effect in quartz crystal resonators.

INTRODUCTION

A characteristic property of the behavior of quartz-crystal resonators of thickness-shear vibrations in a nonstationary fluctuating field of the medium is the existence of the effect of a spectral density splash of frequency fluctuations in a generator spectral line range[1]. In 1965 L.Ye.Ivlev[2] drew attention to the temperature nature of this phenomenon. He introduced the thermodynamic coefficient of frequency (TDCF)[2,3], which is numerically equal to the ratio of relative frequency instability to amplitude of sinusoidal medium temperature variation.

First theoretical thermodynamics studies of an infinite plane quartz-crystal plate were carried out by R.Holland[4,5] in 1974, and in 1978 A.Ballato introduced a concept of the thermodynamic coefficient $\tilde{a}$[6] analogous to TDCF.Further stadies of thermodynamic characteristics of resonators were continued in[7-10], a noticeble step in this direction being taken by J.P.Valentin who published the results of the temperature gradients studies in quartz-crystal resonators[11-13]. A common shortcoming of these works and others devoted to thermodynamics studies of quartz-crystal resonators consists in their trend to solve particular problems. As a result the authors failed to develop practical recommendations with respect to improvement of quartz-crystal resonators constructions.

In the present paper the problems of a statistic thermodynamic frequency instability of quartz-crystal resonators are considered from a position of the theory of full dynamic characteristics[14]. Such approach allows to consider TDCF and $\tilde{a}$ to be particular dynamic characteristics. The study of full transient,impulse and amplitude phase-frequency characteristics allows to detect the effect of a quartz plate thermodynamic resonance, the account of which is an important practical task as it will be shown later.

THERMODYNAMIC MODEL OF QUARTZ CRYSTAL PLATE

Pursuing the main object of the paper - to provide physical and mathematical grounds for a thermodynamic resonance in quartz-crystal resonators, - we shall study only an autonomous unelectroded disc quartz plate with central symmetry, diameter D and thickness H, in order to simplify the task. In the general case it is assumed that the plate has singly or doubly rotated orientations. Figure 1 gives a resonator model where rotated crystallographic axes correspond to singly x',y' and doubly x'',y'' rotated orientations.

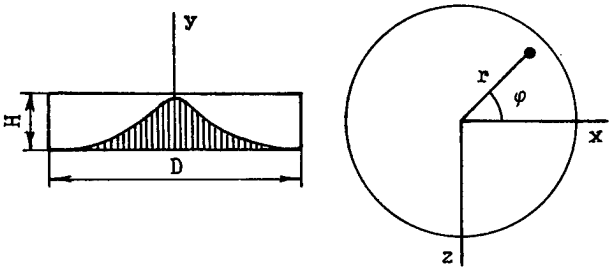


Figure 1. Model of an autonomous quartz crystal plate

In the general case the plate thermodynamic state is described by the equation of heat condaction

$$c\rho\frac{du}{dt}-\mathrm{div}(\kappa\ \mathrm{grad}u)=F(r,\varphi,y,t) \qquad (1)$$

where c and ρ are the specific thermal conductivity and quartz density, κ is the tensor of thermal conduction, u is the temperature at the point of the plate volume, t is the current time,$F(r,\varphi,y,t)$ is the nonstationary function of the internal heart sources distribution proportional to the normalized function of the mechanical vibrations amplitudes distribution[15].

0-7803-0905-7/93 $3.00 © 1993 IEEE

$$e_{npq}=e^{-\alpha_n x_1^2/2}H_p(\sqrt{\alpha_n}x_1)e^{-\beta_n x_3^2/2}H_q(\sqrt{\beta_n}x_3) \qquad (2)$$

where $H_p(X)$ and $H_q(X)$ are Hermite polynomials, α_n and β_n are the functions defined in [15,16], $x_1 \in x', x''$, $x_3 \in z', z''$, n=1,3,5,..., p,q=0,2,4,....

During the piezoelectric crystal plate excitation at the fundamental frequency of vibrations (p, q=0) the amplitude distribution of mechanical vibrations in the plate volume in view of can be written in the form of [2]:

$$E_{n00}(r,\varphi,y)=Ae^{-0,5r^2(a\cos^2\varphi+b\sin^2\varphi)}\times$$

$$\times\cos[\pi y(2n-1)/H] \qquad (3)$$

where $a=1/R_x^2$, $b=1/R_z^2$, R_x and R_z are spot dimensions of the mechanical vibrations field in the directions of the axis x and z, A is the amplitude of mechanical vibrations proportional to the amplitude of the piezoelectric current $I_q(t)$. Taking into consideration that the power scattered in the plate and transformed into heat is defined as $W(t)=R_q I_q^2(t)/2$, where R_q is the dynamic resistance of the resonator the distribution density of heat sources in the resonator volume is described with (3) in the form of

$$F(r,\varphi,y,t)=\frac{R_q I_q^2(t)}{2N}e^{-r^2(a\cos^2\varphi+b\sin^2\varphi)}\times$$

$$\times\cos^2[y\pi(2n-1)/H] \qquad (4)$$

where $N_n=A^{-2}\int_v E_{n00}(r,\varphi,y)drd\varphi dy=\pi H^2/2\sqrt{ab}$.

After transformation of κ into cylindrical system of coordinates of the plate plane and taking the symmetry of the received matrix into account (1) is transformed to the form of

$$c\rho\frac{du}{dt}-\kappa_{rr}\frac{\partial}{r\partial r}(r\frac{\partial u}{\partial r})-\frac{\partial}{r^2\partial\varphi}(\kappa_{\varphi\varphi}\frac{\partial u}{\partial\varphi})-\kappa_{yy}\frac{\partial^2 u}{\partial y^2}-$$

$$-2\kappa_{r\varphi}\frac{\partial^2 u}{\partial r\partial\varphi}-\kappa_{ry}\frac{\partial^2 u}{\partial r\partial y}-\kappa_{\varphi y}\frac{\partial^2 u}{\partial\varphi\partial y}-\frac{1}{r}(\frac{\partial\kappa_{r\varphi}}{\partial\varphi}\frac{\partial u}{\partial r}+$$

$$+\frac{\partial\kappa_{\varphi y}}{\partial\varphi}\frac{\partial u}{\partial y})=F(r,\varphi,y,t) \qquad (5)$$

where κ_{ij} are the tensor κ components in cylindrical coordinates. In view of the object of the paper we confine ourselves to the study of the approximation $\forall H\ll 2R_o \exists \kappa_{rr}=\kappa_{\varphi\varphi}=\kappa_1, \kappa_{yy}=\kappa_3; \forall i,j\in r,\varphi,y; i\neq j \exists\kappa_{ij}=\kappa_{ji}=0$, a=b, since according to [12,17] anisotropy of κ_{ij} has weak influence on the space-time distribution of temperature in the plate volume and is taken into account only at a rather small temperature gradient In this case (5) can be defined as

$$c\rho\frac{du}{dt}-\kappa_1\frac{\partial^2 u}{\partial r^2}-\kappa_1\frac{\partial u}{r\partial r}-\kappa_3\frac{\partial^2 u}{\partial y^2}=F(r,y,t) \qquad (6)$$

adding to (6) zero initial $F(r,y,t)$ and boundary conditional of the form

$$-\kappa_{1,3}\frac{\partial u}{\partial n}\Big|_b=h(u|_b-\tilde{u}) \qquad (7)$$

where h is the coefficient of the heat exchange taken as a sum over all types of heat exchange and heat transfer; $\partial u/\partial n|_b$ is the derivative from u at the boundary point in the direction of the external normal to it; $\tilde{u}$ is the external temperature.

From the coordinate y the eq.(6) is brought into the equation $\partial^2\mathcal{L}_y/\partial y^2=\nu^2\mathcal{L}_y$, and from the radius r - into Bessel equation $\partial^2\mathcal{K}_r/\partial r^2+\partial\mathcal{K}_r/r\partial r+\xi^2\mathcal{K}_r=0$ [12,17]. The solutions of (6) with account of u (7) from the coordinates r and y by the Fourier method it is possible to bring it into nonstationary differential equations with fluctuating parameters of piezoelectric current and ambient temperature.

PIEZOELECTRIC CURRENT AMPLITUDE VARIATION

Ambient temperature is assumed to be constant and piezoelectric current amplitude is varying according to the law

$$I_q(t)=\bar{I}_q(t)[1+\alpha_q(t)]=\bar{I}_q(t)+\tilde{I}_q(t) \qquad (8)$$

where $\bar{I}_q(t)$ is the deterministic function of the piezoelectric current amplitude variation, $\alpha_q(t)$ is the function of small relative amplitude instability.

Transforming (6) by the Fourier method [17] and taking the assumption $I_q^2\cong\bar{I}_q^2(1+2\alpha_q)$, following from (8), into account we bring (6) into the form

$$\frac{d\hat{u}_{kl}}{dt}=-a\hat{u}_{kl}+b\bar{I}_q^2+2b\bar{I}_q^2\alpha_q(t)=f(\hat{u}_{kl},t) \qquad (9)$$

where $a\equiv a_{kl}=\kappa_1(\nu_k^2+\xi_l^2)/c\rho$, $b\equiv b_{kl}=\gamma\mathfrak{X}_{kl}Q_{kn}W_l/c\rho P_k L_l$,

$$Q_{kn}=\frac{1}{\nu_k}\frac{(\nu_k H)^2-2\pi^2(2n-1)^2}{(\nu_k H)^2-4\pi^2(2n-1)^2}[1+\frac{\kappa_1^2\nu_k^2+h^2}{\kappa_1^2\nu_k^2-h^2}\cos(\nu_k H)],$$

$$P_k=0,5H(\kappa_1\nu_k^2/h^2+1)+\kappa_1/h,\ W_l\approx\frac{1}{2a}\exp(-\xi_l^2/4aR^2),$$

$$L_l=R^2[J_o^2(\xi_l R)-\xi_l^2 J_1^2(\xi_l R)]/2, \mathfrak{X}_{kl}=[\frac{\kappa_3\nu_k}{h}\cos(\nu_k y)+$$

$\sin(\nu_k y)]\ J_o(\xi_1 r)$, $\hat{u}_{kl}$ is the partial temperature, $J_o(X)$ is the Bessel functions, ν_k and ξ_l are defined by solutions of transcendental equations

$$tg(\nu H)=2h\kappa_3\nu/(\kappa_3^2\nu^2-h^2) \qquad (10)$$

$$R\xi\kappa_1/(2\kappa_1+Rh)=J_o(\xi R)/J_1(\xi R) \qquad (11)$$

In relation to slow temperature variations in the piezoelectric plate the random process $\alpha_q(t)$ has white noice properties.

Since for (8) the condition of Lipshits[18] is satisfied $|f(\hat{u}'_{kl},t)-f(\hat{u}''_{kl},t)|\le|(\partial/\partial\hat{u}_{kl})f(\hat{u}'_{kl},t)|\times|\hat{u}'_{kl}-\hat{u}''_{kl}|$ then according to the theorem of Dub[18], the random process $\alpha_q(t)$ is Marcovian. In this case the coefficient of drift and diffusion of the process (9) can be found in the form of

$$K_1(\hat{u}_{kl})=-a\hat{u}_{kl}+b\bar{I}_q^2(t) \qquad (12)$$

$$K_2(\hat{u}_{kl})=2b^2N_{oq}\bar{I}_q^4(t) \qquad (13)$$

where $N_{oq}/2=S_q$ is the spectral density of the process $\alpha_q(t)$, and now we can get on to the equation of Fokker-Plank

$$\frac{d}{dt}W(\hat{u}_{kl},t)+\frac{\partial}{\partial\hat{u}_{kl}}(-\hat{u}_{kl}+b\bar{I}_q^2)W(\hat{u}_{kl},t)-$$

$$-b^2\bar{I}_q^4N_{oq}\frac{\partial^2}{\partial\hat{u}_{kl}^2}W(\hat{u}_{kl},t)=0 \qquad (14)$$

Since the coefficient $K_2(\hat{u}_{kl},t)$ is constant and the coefficient $K_1(\hat{u}_{kl},t)$ is linearly dependent on $\hat{u}_{kl}$, the Marcovian process ordinate $\hat{u}_{kl}$ under zero initial conditions in accordance with (12)-(14) obeys the normal distribution law

$$W(\hat{u}_{kl},t)=b\bar{I}_q^2\sqrt{a/[2\pi N_{oq}(1-e^{-2at})}\times$$

$$\times\exp\left\{-\frac{[\hat{u}_{kl}-b\bar{I}_q^2(1-e^{-at})/a-T_{oo}a^{-at}]^2}{2b^2\bar{I}_q^4N_{oq}(1-e^{-2at})/a}\right\} \qquad (15)$$

where the expectation is a function of the partial temperature regular variation

$$\bar{\hat{u}}_{kl}(t)=b\bar{I}_q^2(1-e^{-at})/a \qquad (16)$$

and the dispersion is found in the form of

$$\sigma_u^2(t)=b^2\bar{I}_q^4N_{oq}(1-e^{-2at})/a \qquad (17)$$

The resultant temperature variation at the plate arbitrary point according to the Fourier method with account of (16) and (9) has the form of

$$u=T_{oo}+\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}b_{kl}\bar{I}_q^2(1-e^{-a_{kl}t})/a_{kl}+$$

$$+2\bar{I}_q^2\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}b_{kl}e^{-a_{kl}t}\int_o^t e^{a_{kl}\lambda}\alpha_q(t)d\lambda \qquad (18)$$

where T_{oo} is the initial value of the plate temperature.

Regular temperature variations after averaging (18) is found in the form of

$$\bar{u}=T_{oo}+\bar{I}_q^2\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}b_{kl}(1-e^{-a_{kl}t})/a_{kl} \qquad (19)$$

In order to find the process dispersion (18) during the response to a piezoelectric current fluctuation with the accepted white noise and the correlation function $K_q(t_2,t_1)=0,5N_{oq}\delta(t_2-t_1)$ we write down

$$\sigma_{uI}^2=4\bar{I}_q^4\sum_i\sum_j b_ib_je^{-(a_i+a_j)t}\times$$

$$\times\int_o^{t_1}\int_o^{t_2}e^{a_it'}e^{a_jt''}\langle\alpha_q(t'')\alpha_q(t')\rangle dt'dt'' \qquad (20)$$

where $i\in k,l$, $j\in k,l$, $\langle\rangle$ is a sign of the time averaging. After the operation of integration with the use of the filtering property of δ-function we finally write down

$$\sigma_{uI}^2=2N_{oq}\bar{I}_q^4\sum_i\sum_j b_ib_j[1-e^{-(a_i+a_j)t}]/(a_i+a_j) \qquad (21)$$

Functions (15)-(21) adequately fully describe temperature statistic dynamics at an arbitrary point of a piezoelectric plate volume during its response to the piezoelectric current amplitude.

AMBIENT TEMPERATURE VARIATION

To solve eq.(6) during the plate response to to ambient temperature variation it is assumed that

$$u(r,\varphi,y,t)=\bar{u}(r,\varphi,y,t)[1+\alpha_t(t)]=\bar{u}(r,\varphi,y,t)+\mathring{u}(t) \qquad (22)$$

where $u(r,\varphi,y,t)$ is the initial function of the plate internal temperature, $\mathring{u}(t)$ is the ambient temperature which is fluctuating in the general case. The piezoelectric current amplitude is taken as constant.

Taking (22),(7) into account and having performed the transform in terms of the Fourier method[17] we obtain

$$\frac{d\hat{u}_{kl}}{dt}=-a\hat{u}_{kl}+b\bar{I}_q^2+c\frac{d\mathring{u}}{dt} \qquad (23)$$

where $c=V_kD_l\mathfrak{X}_{kl}/L_l$. ν_k and ξ_l are found from (10) and (11), $V_k=\left\{\frac{(\kappa_э\nu_k)^2+h^2}{(\kappa_э\nu_k)^2-h^2}\cos(\nu_kH)+1\right\}/$

$$/\left\{\frac{\nu_kH}{2h^2}[(\kappa_э\nu_k)^2+h^2]+\frac{\kappa_э\nu_k}{h}\right\},D_l=hRJ_o(\xi_lR)/(\kappa_1\xi_l^2).$$

The first theorem of Dub[18] is not applicable to (23), that is why the partial temperature process $u_{kl}(t)$ cannot be considered as Marcovian. Taking into account that in the general case $\mathring{u}(t)$ has arbitrary time of correlation with the constant of temperature transient time and that the temperature instability is small, we shall use the method of statistical linearization to evaluate statistic properties of $u_{kl}(t)$ process.

The general solution of eq.(23) under initial conditions $u_{kl}(t=o)=b\bar{I}_q^2/a$ has the form of

$$\hat{u}_{kl}=\bar{u}_{kl}+\tilde{u}_{kl}=\frac{b\bar{I}_{qo}^2}{a}+T_{oo}ce^{-at}\int_o^t e^{at}\frac{d[(1+\bar{\alpha})(1+\alpha_T)]}{dt}dt \qquad (24)$$

where $\bar{u}_{kl}$ and $\tilde{u}_{kl}$ are regular and random components of partial temperature regular $\bar{\alpha}$ and random α_T modulating processes are connected with the ambient temperature by the relation $\hat{u}=T_{oo}(1+\bar{\alpha})(1+\alpha_T)$.

Temperature fluctuations dispersion $\tilde{u}_{kl}=$ $=T_{oo}ce^{-at}\int_c^t e^{at}(\alpha_t d\bar{\alpha}/dt+d\alpha_T/dt+\bar{\alpha}d\alpha_T/dt)dt$ for small fluctuations $\alpha_T d\alpha/dt \ll d\alpha_T/dt$, $\bar{\alpha}d\bar{\alpha}_T/dt \ll$ $\ll d\alpha_T/dt$ we write down in the form of

$$\sigma_{uT}^2=T_{oo}^2c^2e^{-2at}\iint_o^t\left\langle\frac{d\alpha_T(t')}{dt'}\frac{d\alpha_T(t'')}{dt''}\right\rangle e^{at'}e^{at''}dt'dt'' \qquad (25)$$

Taking into account that $\left\langle\frac{d}{dt}\alpha_T(t'')\frac{d}{dt}\alpha_T(t')\right\rangle =$ $=-\frac{\partial^2}{\partial\tau^2}\langle\alpha_T(t')\alpha_T(t'')\rangle$ where $\tau=t''-t'$ for delta-like noise we obtain

$$\sigma_{uT}^2=T_{oo}\frac{N_{oT}}{2}c^2e^{-2at}\int_o^t\int_o^t\delta''(t''-t')e^{at'}e^{at''}dt'dt'' \qquad (26)$$

With account of the filtering property of the second derivative of δ-function we pass from (26) to the relation

$$\sigma_{uT}^2=\frac{1}{4}T_{oo}N_{oT}c^2a(1-e^{-2at})\Big|_{t=\infty}=\frac{1}{4}N_{oT}c^2aT_{oo} \qquad (27)$$

To find statistical estimations of the resultant temperature in terms of the Fourier method we proceed from (24) to the relation

$$u=T_{oo}(1+\bar{\alpha})(1+\alpha_T)+$$
$$+\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\left(\frac{b_{kl}}{a_{kl}}\bar{I}_q^2+c_{kl}e^{-a_{kl}t}\int_o^t e^{a_{kl}t}d\hat{u}\right) \qquad (28)$$

Expectation of (28) has the form of

$$\bar{u}=T_{oo}(1+\bar{\alpha})+\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\left(\frac{b_{kl}}{a_{kl}}\bar{I}_q^2+T_{oo}c_{kl}e^{-a_{kl}t}\int_o^t e^{a_{kl}t}d\bar{\alpha}\right) \qquad (29)$$

and its dispersion with account of (25)-(27) is found as

$$\sigma_u^2=T_{oo}\frac{N_{oT}}{2}\left[\delta(0)+\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}c_{kl}-\right.$$
$$\left.-\sum_i\sum_j\frac{c_ic_ja_j^2}{a_i+a_j}[1-e^{-(a_i+a_j)t}]\right] \qquad (30)$$

both in the case of partial temperatures and in the case of the resultant temperature at an arbitrary point of the plate volume (29),(30) the character of random processes obeys the normal law of probability density distribution so far as the medium fluctuations near the stationary temperature point most probably have the properties of white noise in relation to the temperature transient time of the plate

FULL THERMODYNAMIC CHARACTERISTICS

Temperature Frequency Transforms

To evaluate frequency changes caused by non-uniformity of temperature fields reliefs (19) and (29) we take into consideration that partial relative shift of the plate oscillations frequency caused by the force $P(\varphi)$ acting along the diameter D at its edges, is found as

$$\delta f(\varphi,\Psi,\Phi,t)=a_fK_f(\varphi,\Psi,\Phi)P(\varphi,t) \qquad (31)$$

where $K_f(\varphi,\Psi,\Phi)$ is the Rataiski coefficient[19], $a_f=f_o\eta/Dn$, f_o is the nintial frequency of the plate oscillations, n is the harmonic order of an acoustic wave.

Increment of the internal stresses tensor in the plate elementary volume at the expense of its temperature increment is written down in the form of

$$\bar{t}_{ij}(r,\varphi,y)=\bar{d}_{ij}(r,\varphi,y)\Delta T(r,\varphi,y) \qquad (32)$$

where $\bar{d}_{rr}(r,\varphi,y)=c_{ijkl}(r,\varphi,y)\alpha_{ik}(r,\varphi,y)$, c_{ijkl} and α_{ik} are the elastic constants and crystal thermal expansion tensors, $\Delta T(r,\varphi,y)$ is the temperature difference of the plate elementary volume. For a plate construction with central symmetry the tensor radial component is found as

$$\bar{t}_{rr}=\bar{d}_{rr}(\varphi)\int_o^R\Delta T(r,\varphi,y)dr=\bar{d}_{rr}(\varphi)[u_R(\varphi)-u_o] \qquad (33)$$

where $u_R(\varphi)$ and u_o are temperature at the edges and in the center of the plate.

To obtain from (33) the resultant stress acting along the plate diameter and the sign of the force action we shall obtain a transform operator of the nonuniformity of the temperature field relief into the relative instability of oscillations frequency[20]

$$L_\varphi\{\chi\}=\delta f(\varphi)=0.5HDa_f\int_o^\pi K_f(\varphi)\bar{d}_{rr}(\varphi)\chi(\varphi)d\varphi \qquad (34)$$

where $\chi(\varphi)=u_o-u_R(\varphi)$. All the functions from φ in (34) are written as a generalized series of the type

$$\omega(\varphi)=\sum_i a_i\cos(2i\varphi)+\sum_j b_j\sin(2j\varphi) \qquad (35)$$

Taking (35) into account, we transform (34) into the form

$$L_\varphi\{\chi\}=0.5\pi HDa_f\langle K_f(\varphi)\rangle\langle\bar{d}_{rr}(\varphi)\rangle\langle\chi(\varphi)\rangle \qquad (36)$$

where $\langle\chi(\varphi)\rangle=u_o-\langle u_R(\varphi)\rangle, \langle K_f(\varphi)\rangle, \langle d^-_{rr}(\varphi)\rangle, \langle u_R(\varphi)\rangle$ are the average angular values of corresponding functions. In [20] it is shown that for the majority of angles of the cut of thickness-shear vibrations (36) can be brought into the form

$$L_\varphi\{\chi\}=3.85\times10^6 HDa_f\langle K_f(\varphi)\rangle\langle\chi(\varphi)\rangle \qquad (37)$$

from which it is seen now that the main source of thermodynamic frequency variations is the average angular temperature difference between the center and the edges of a piezoelectric plate.

Response to Piezo-Current Amplitude Variation

To find full thermodynamic characteristics it is assumed that $I_q(t)$ is changing stepurse and that there are no fluctuations sources.In this case (18) can be written in the form of

$$u=T_{oo}+\sum_{k=1}^{\infty}\sum_{l=1}^{\infty} b_{kl}\bar{I}_q^2(1-e^{-a_{kl}t})/a_{kl} \qquad (38)$$

Now we apply the operator (37) to (38) and find the transient characteristic

$$h_{fI}(t)=\frac{\delta f(t)-\delta f(0)}{\delta f(\infty)-\delta f(0)}\ , \quad t\in\overline{0,\infty} \qquad (39)$$

where $\delta f(t)=\delta f_o(T)+L_\varphi\{\chi\}$, $T=\langle u(r,y,t)\rangle_{r,y}$,

$$\chi=\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}[M_{kl}(0)-M_{kl}(R)](1-e^{-a_{kl}t})/a_{kl}\ ,$$

$$M_{kl}=\bar{I}_{qo}^2\gamma\mathfrak{X}_{kl}\hat{Q}_{kn}\hat{W}_l/c\rho\ ,$$

$$\delta f_o(t)=a_o(T-T_o)+b_o(T-T_o)^2+c_o(T-T_o)^3 \qquad (40)$$

are the frequency-temperature characteristics, a_o and b_o and c_o are the temperature coefficients, T_o is reference temperature.

It can be shown that main physical processes taking place in the during the transient mode, are well described with account of the series terms with indexes kl=11, kl=12 in (38). The second term of the series describes the initial response process of the plate to the action of an internal heat source due to the stepwise variation of the piezoelectric current amplitude. After the central part of the plate having been heated, a slow redistribution of the heat to the plate edges is occurring which is taken into account by the series term with the index kl=11. The rest terms of the series have large decrements, small amplitudes with an alternate sign, that is why their influence is insignificant. With kl=11 and kl=12 the processes are described by the relation

$$u=T_{oo}+\Delta T[1-e^{-a_1t}+b(1-e^{-a_2t})] \qquad (41)$$

where $\Delta T=\bar{I}_q^2\gamma Q_{1n}\bar{W}_2J_o(\xi_1r)/\kappa_1P_1(\xi_1^2+\nu_1^2)$,

$a_1=\kappa_1(\xi_1^2+\nu_1^2)/c\rho$, $a=\kappa(\xi_2^2+\nu^2)/c\rho$,

$b=\hat{W}_2J_o(\xi_2r)(\xi_1^2+\nu_1^2)/\hat{W}_1J_o(\xi_1r)(\xi_2^2+\nu_1^2)$.

Functions of (41) and (38) are given in Figure 2.

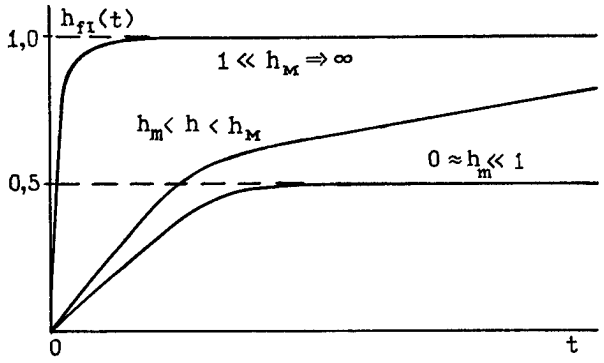


Figure 2. Turn-on transient characteristics for a sudden rise of a piezo current amplitude.

Proceeding from (39) to impulse characteristic

$$g_{fI}(t)=\frac{d[\delta f(t)]/dt}{\delta f(\infty)-\delta f(0)}=\frac{1}{\Delta f_p}d[\delta f(t)]/dt \qquad (42)$$

and using the Fourier transform

$$K_{fI}(j\Omega)=\int_{-\infty}^{\infty}g_{fI}(t)e^{-j\Omega t}dt \qquad (43)$$

we pass to the transient function of the plate $K_{fI}(j\Omega)=K_{fI}(\Omega)e^{j\varphi_{fI}(\Omega)}$,where for (41)the amplitude -frequency and phase-frequency characteristics are found:

$$K_{fI}(\Omega)=f_o\sqrt{L_\varphi^2\{K_1^A\}+L_\varphi^2\{K_1^R\}}\ /\Delta f_p \qquad (44)$$

$$\varphi_{fI}(\Omega)=\mathrm{arctg}(L_\varphi\{K_1^R\}/L_\varphi\{K_1^A\}) \qquad (45)$$

where $K_1^A(\Omega)=[a_1^2a_2^2(b_1+b_2)+\Omega^2(a_1^2b_1+a_2^2b_2)]/$

$$/(a_1^2+\Omega^2)(a_2^2+\Omega^2) \qquad (46)$$

$$K_1^R(\Omega)=\Omega[a_1a_2(a_2b_1+a_1b_2)+\Omega^2(a_1b_1+a_2b_2)]/$$

$$/(a_1^2+\Omega^2)(a_2^2+\Omega^2) \qquad (47)$$

Respectively the spectral density of the plate frequency fluctuations power, due to the amplitude fluctuations of piezo-current is found

$$S_{fI}(f_I)=|K_{fI}(f)|^2S_I(f_I)\ , \quad f_I=2\pi\Omega \qquad (48)$$

where f_I is the amplitude fluctuations frequency, $S_I(f_I)$ is the spectral density of fluctuations I_q.

Figure 3 gives functions (44) and (45) for various parameters of AT-cut plate. It follows from the analysis that (48) has two characteristic regions in conventionally close and distant areas of the spectrum. With fixed H and D and with increase of h the spectrum (48) is expanding in the close area of $10^{-4}\ldots10^{-2}$Hz (Figure 3a). And with increase of h the spectrum variations are becoming

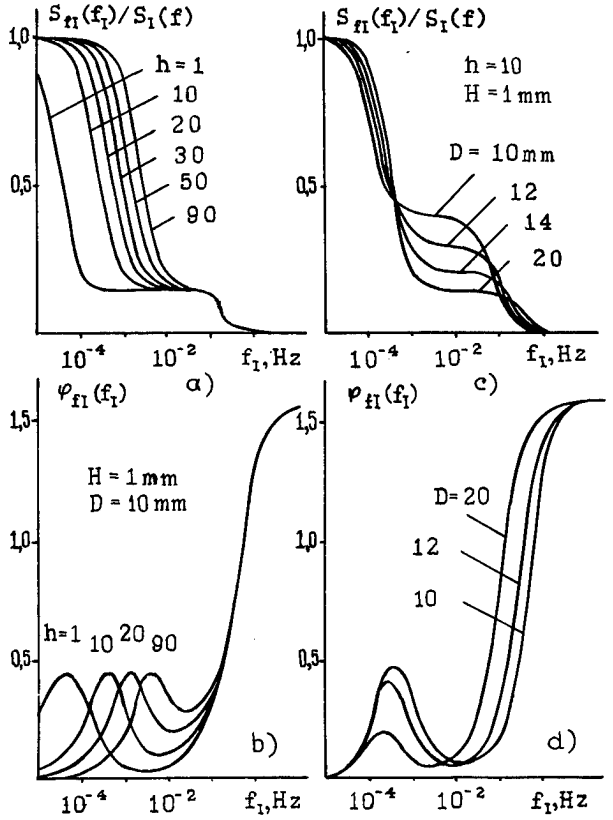


Figure 3. Amplitude-frequency (a,c) and phase-frequency (b,d) characteristics for an influence of a piezo current amplitude.

less noticeable. With fixed h and H increase of D causes increase of (48)in the frequency region 10^{-4} ... 10^{-1} without the spectrum expansion (Fig.3c). The corresponding phase characteristics variations are shown in Fig.3c,d.

From the analysis of Fig.3 it is possible to make a conclusion that decrease of the influence of piezoelectric current amplitude fluctuations is achieved by the decrease of h and D to possibly small quantities.

Thermodynamic Resonance

Now we shall consider thermodynamic characteristics of the plate during its response to stepwise variation of the ambient temperature by the quantity of ΔT. Similar to (38) external fluctuations are assumed to be absent and (29) can be put into the form

$$u=T_{oo}+\Delta T(t)+\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}(b_{kl}\bar{I}_q^2/a_{kl}+$$

$$+c_{kl}e^{-a_{kl}t}\int_c^t e^{a_{kl}t}\frac{d\Delta T}{dt}dt) \qquad (49)$$

Let us assume now, that $\Delta T(t)=\Delta T(1-e^{-\theta t})$, where θ is a rather big number.The solution of (49) under such a condition has the form of

$$u=T_{oo}+\Delta T(1-e^{-\theta t})+\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\left[b_{kl}\bar{I}_q^2/a_{kl}+\right.$$

$$\left.+\Delta T\theta c_{kl}e^{-(a_{kl}+\theta)t}/(a_{kl}+\theta)\right] \qquad (50)$$

Let θ go to infinity in (50), then we obtain

$$u=T_{oo}+\Delta T+\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\left[b_{kl}\bar{I}_q^2/a_{kl}-\Delta Tc_{kl}e^{-a_{kl}t}\right] \qquad (51)$$

During the process of the temperature T change the plate temperature field relief (51)goes through three specific stages (Fig.4) causing thermodynamic effects[17].The first stage (Fig.4a) is characterized by large warming up of the plate center and practically uniform temperature distribution between the center and the edges[12,17].The second stage (Fig.4b) is characterized by zero temperature gradient connected with warming up of the plate edges after the ambient temperature rise by ΔT. The third stage (Fig.4c) is characterized by maximum warming up of the plate edges relative to its center. With time the plate passes over to the second and the first stages. With this, however, the plate integral temperature becomes equal to $T+\Delta T$. In electroded plates the temperature field reliefs are somewhat modified[12], but the stages of the reliefs change with the temperature rise by remain the same.

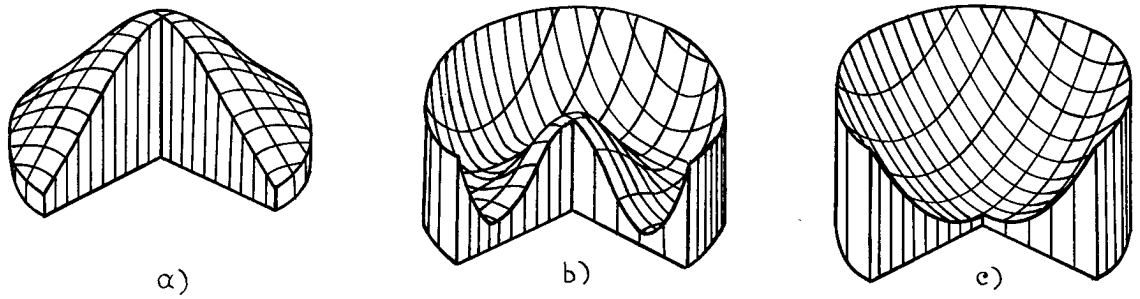


Figure 4. Reliefs of themperature distributions in a quartz crystal plate for variouse time moments after sudden rise in a temperature

The account of the series terms with indexes kl=11,12 with h=1...1000 in (51) allows to stady thermodynamic processes with the error not exceeding 10%. The account of the series terms with indexes kl=14,...,110 allows to define more exactly the result in the 2nd-order infinitesimal with which the error is not over 1%.

The transient process function is defined by (39) with account of (40), where

$$\chi=\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\left[\left[M_{kl}(0)-M_{kl}(R)\right]/a_{kl}-\right.$$

$$\left.-\left[N_{kl}(0)-N_{kl}(R)\right]e^{-a_{kl}t}\right] \qquad (52)$$

where $N_{kl}=\Delta T V_k \dot{D}_{kl} \Im_{kl}$. With h<1 the transient process goes rather slowly and has a monotonous character. The increase of h leads to the decrease of the transient process time with simultaneus appearance of a negative frequency overshoot which can reach 10^{-6} at the point of the temperature control in AT-cut plates. The decrease of D causes the decrease of the transient process time, and the decrease of H causes the decrease of its overshoot. Typical transient process character for a AT-cut plate is given in Fig.5.

Amplitude- and phase-frequency characteristics with account of (52),(42),(43) are defined by (44),(45) where similarly to (46),(47)

$$K^{A}_{fT}(\Omega)=\nu\sum_{p}\rho_1 a_p-\mu\sum_{p}\sum_{i}\rho_2(a_p+a_i)+ \qquad (53)$$

$$+3c_o\sum_{i}\sum_{j}\sum_{p}\rho_3(a_p+a_i+a_j)-L_{\varphi}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\rho_4 a_{kl}$$

$$K^{P}_{fT}(\Omega)=-\Omega\left[\nu\sum_{p}\rho_1+\mu\sum_{p}\sum_{i}\rho_2+\right.$$

$$\left.+3c_o\sum_{i}\sum_{j}\sum_{p}\rho_3-L_{\varphi}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\rho_4\right] \qquad (54)$$

where $\rho_2=\bar{N}_p\bar{N}_i a_p(a_p+a_i)/[(a_p+a_i)^2+\Omega^2]$,

$\rho_3=\bar{N}_i\bar{N}_j\bar{N}_p a_p(a_p+a_i+a_j)/[(a_p+a_i+a_j)^2+\Omega^2]$,

$\rho_4=[N_{kl}(0)-N_{kl}(R)]\,a_{kl}^2/(a_{kl}^2+\Omega^2)$,

$\rho_1=\bar{N}_p a_p^2/(a_p^2+\Omega^2)$, $\nu=a_o+2b_o(u_c-T_o)+3c_o(u_c-T_o)^2$,

$\mu=2b_o+6c_o(u_c-T_o)$, $u_c=T_o+\Delta T+\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\bar{M}_{kl}/a_{kl}$,

$\bar{M}_{kl}=\langle M_{kl}\rangle_{r,\varphi,y}$, $\bar{N}_{kl}=\langle N_{kl}\rangle_{r,\varphi,y}$.

With account of (53),(54) investigation of the functions

$$K_{fT}(\Omega)=\sqrt{\left[K^{A}_{fT}(\Omega)\right]^2+\left[K^{R}_{fT}(\Omega)\right]^2} \qquad (55)$$

$$\varphi_{fT}(\Omega)=\operatorname{arctg}\left[K^{R}_{fT}(\Omega)/K^{A}_{fT}(\Omega)\right] \qquad (56)$$

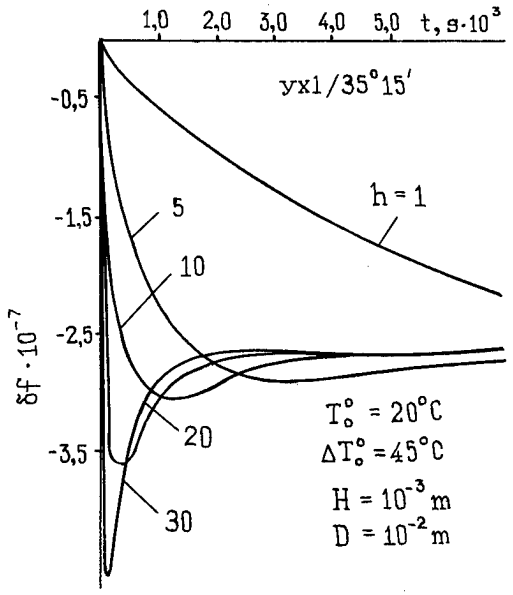


Figure 5. Warm-up for AT-cut.

and their influence on the spectral density

$$S_{fT}(\Omega)=|K_{fT}(\Omega)|^2 S_T(\Omega) \qquad (57)$$

where $S_T(\Omega)$ is the spectral density of the temperature fluctuations power, has shown that (57) is characterized by the effect which is called a thermodynamic resonance. The essence of the effect is that in the space of at least three parameters $-f_T=2\pi\Omega$, D/H, h the function (55) has a particularly pronounced splash. Its physical essence is explained by the fact that the ambient temperature effect leads to rather nonuniform warming up of the plate center and edges with different rates and in different moments of time (Fig.4).

The essence of the thermodynamic resonance effect is explained by Figs.6 and 7.

With fixed H,D and h the ambient temperature rise f_T leads to the change of $K_{fT}(\Omega)$ so that its splash is observed in some region f_T. Maximum amplitude of the latter with H=const takes with a diameter D and at the borders of theoretical values D=0 and D=∞ the splach is not observed. A simular resonance phenomenon is observed with the change of h (Fig.7), and maximum dynamic increase of $K_{fT}(\Omega)$ can exceed its static value by some orders of magnitude.

The influence of the plate parameters h and D is explained by Fig.8.

With D/H=const and h<3 the resonance is not observed. The increase of h leads to the increase of $K_{fT}(\Omega)$, and the maximum resonance is shifted to higher frequencies f_T (Fig.8a,b). The increase of D leads to the shift of the resonance characteristic (55) to lower frequencies. At this (55) goes through the resonance. With the diameter D change of 10 - 12 mm the splash $|K_{fT}(\Omega)|^2$ (57) in the frequency range of $10^{-1}...10^{-3}$ is increased by a factor of 3 (Fig.6c,d).

CONCLUSION

The conducted investigation of statistic thermodynamics of a quartz-crystal plate of thickness-shear vibrations and in particular, the study of full thermodynamic characteristics of the plate during its response to the piezoelectric current amplitude and ambient temperature variations enabled to reveal and to ground a thermodynamic resonance effect. Practical recommendations for taking the investigated effect into account come to the fact that for real quartz-crystal resonators constructions it is necessary to calculate nomograms of the construction parameters choice (Figs. 6b,7d) on the basis of which to select geometrical dimensions of the plate, degree of vacuum and thermal insulation from the enviroment so as the spectral density (57) has a minimum thermodynamic overshoot. Calculation of nomograms should be carried out by solving (5) completing it with boundary conditions, for exemple, with account of[12,21]

REFERENCES

1. E.J.Baghdady, R.N.Lincoln, B.D.Nelin. Short-term frequency stability:Characterization, Theory and Measurement. Proc.IEEE, v.53,p.704-722, 1965.
2. Л. Е. Ивлев. О температурно-динамическом коэффициенте частоты резонаторов. Вопросы радиоэлектроники. Серия 3, Детали и компоненты аппаратуры. вып. 2. 1965.
3. Л. Е. Ивлев. Температурно-динамический коэффициент частоты и температурно-динамические характеристики прецизионных кварцевых резонаторов. Электронная техника. Серия 9, Радиокомпоненты, вып. 4, с. 20-28, 1967.
4. R.Holland. Non uniformly heated anisotropic plates: I Mechanical distorsion and relaxation. IEEE Trans., SU-21, n.3, p.171, 1974.
5. R.Holland. Non uniformly heated anisotropic plates: II Frequency transients in AT and BT quartz plates. Proc.of IEEE Ultrasonics Symp., IEEE Cat, CHO 896-1SU, p.592, 1974.
6. A.Ballato. Static and dynamic behaviour of quartz resonators. IEEE Trans., vol.SU-26, n.4, p.299-306, 1979.

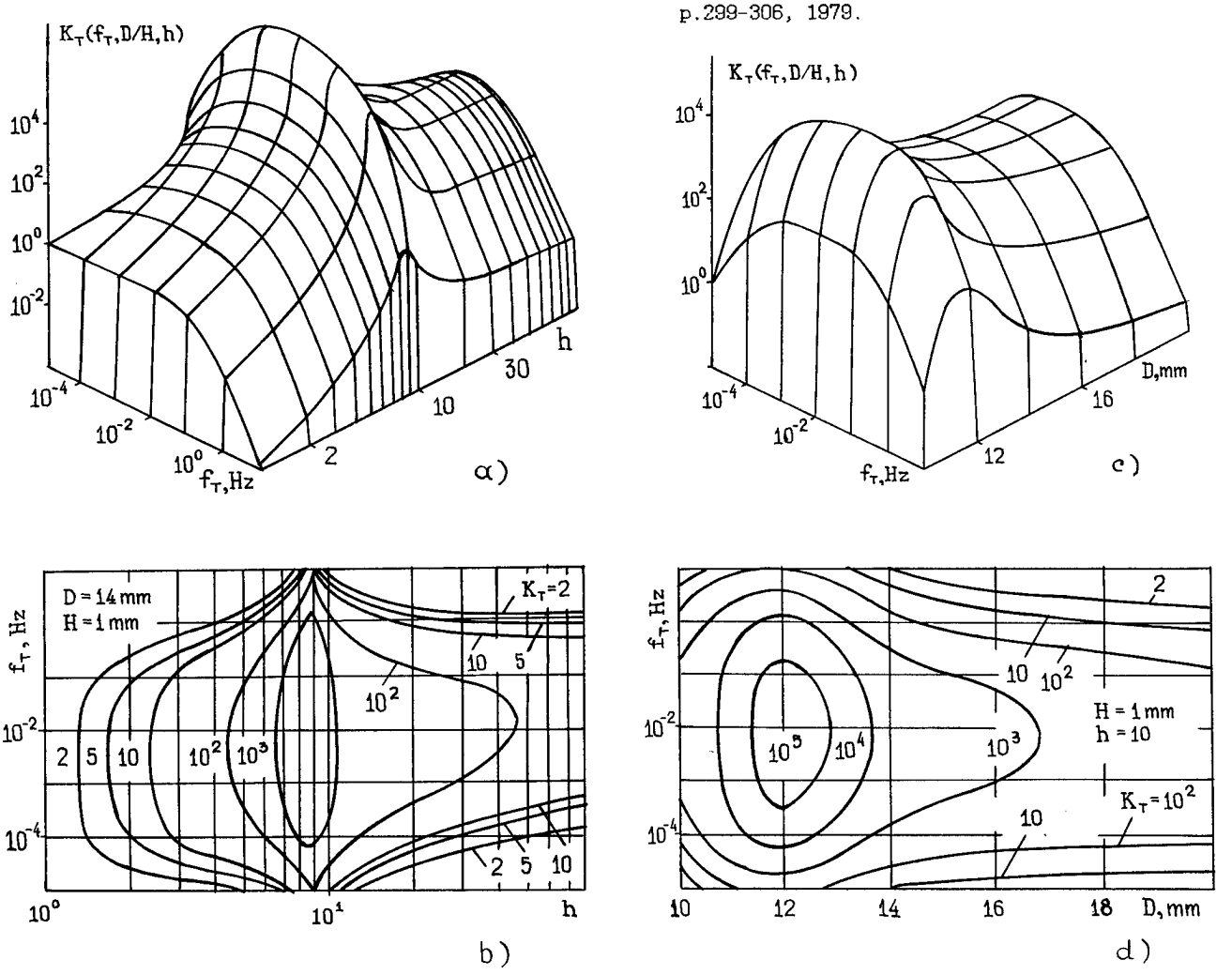


Figure 6. Reliefs of spectrum density (a,c) equal density line (b,d)
The THERMODYNAMIC RESONANSE.

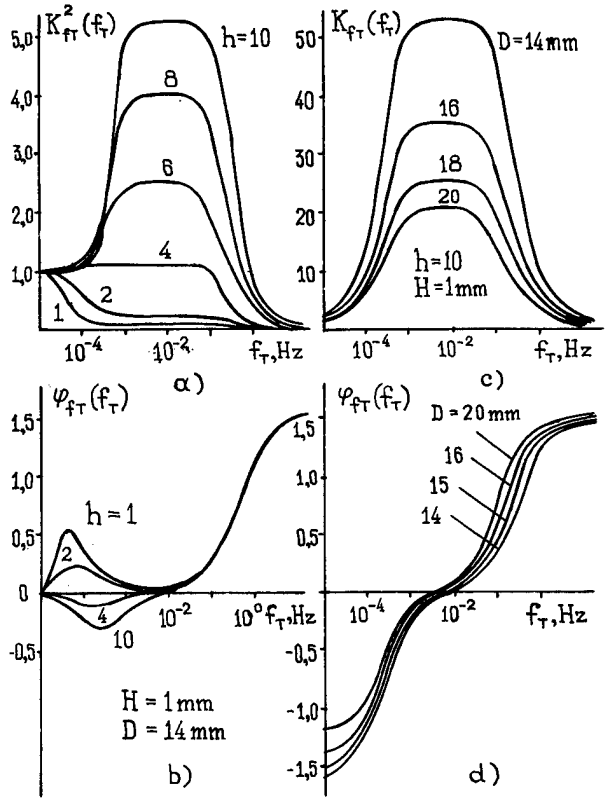


Figure 7. Amplitude-frequency (a,c) and phase-frequency (b,d) characteristics for an influence of temperature.

7. В. Я. Баржин, Э. Б. Крутофалов, И. И. Тартаковский. Исследование нестационарного температурного поля пьезоэлемента. Электронная техника, серия 5, Радиодетали и радиокомпоненты, вып. 5, с. 56-58. 1975.
8. D.S.Stevens, H.F.Tiersten. Transient thermally induced frequency excursions in AT- and SC-cut quartz trapped energy resonators. Proc. of 34th AFCS, p.208-217, 1980.
9. B.K.Sihna, H.F.Tiersten. Transient thermally induced frequency excursions in doubly-rotated quartz thickness-mode resonators. Proc. of 34th AFCS, p.393-402, 1980.
10. D.Janiaud, R.Besson, J.J.Gagnepain, M.Valdois. Quartz resonator thermal transient due to crys tal support. Proc. of 35th AFCS,p.340-344, 1981.
11. J.P.Valentin. Doctoral thesis, n.178, University of Besancon, France, 1983.
12. J.P.Valentin,G.Theobald,J.J.Gagnepain. Frequency shifts arising from in-plane temperature gradient distribution in quartz resonators. Proc.of 38th AFCS, p.157-163, 1984.
13. J.P.Valentin. Thermal gradient distributions in trapped energy quartz resonator. J.Appl.Phys, p.128, 1984.
14. Ю. С. Шмалий. Динамические искажения в частотно-модулируемых автоколебательных системах. Изв. вузов. Радиоэлектроника, т. 29, N12, с. 40-44, 1986.
15. H.F.Tiersten, R.S.Smythe. An analysis of contoured crystal resonators operating in overtones of coupled thickness shear and thickness twist. J.Acoust.Coc.Am., v.65, p.1455, 1979.
16. D.S.Stevens,H.F.Tiersten. An analysis of doubly-rotated contoured quartz crystal resonators. Proc. of 39th AFCS, p.436-447, 1985.
17. Ю. И. Евдокименко, Ю. С. Шмалий. Термодинамическая нестабильность частоты кварцевых автогенераторов. Изв. вузов. Радиоэлектроника, N5, с. 49-55, 1991.
18. В. И. Тихонов, М. А. Миронов. Марковские процессы. Москва: Советское радио, 488с., 1977.
19. J.M.Rataiski. Force-frequency coefficient of singly rotated vibrating quartz crystals. IBM J.Res.Dev., V.12, N.1, p.92-99, 1968.
20. Ю. И. Евдокименко, Ю. С. Шмалий. Термодинамическая нестабильность частоты объемных акустических колебаний кварцевой пьезопластины. Акустический журнал, т. 38, вып. 2, с. 283-289, 1992.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PREDICTION OF ACOUSTIC RESONATOR-STABILIZED OSCILLATOR SUSTAINING STAGE PHASE NOISE CHARACTERISTICS FROM LOW Q RESONATOR OSCILLATOR NOISE MEASUREMENTS

M. M. Driscoll

Westinghouse Electronic Systems Group
P.O. Box 746, MS 85, Baltimore, MD

Abstract

Accurate prediction of quartz crystal resonator and surface acoustic wave resonator (SAWR) oscillator output signal phase noise characteristics due to the contribution of the non-resonator portion of the circuit can be facilitated via phase noise measurement of the oscillator circuit with the resonator(s) replaced by resistive/reactive components that simulate resonator operating point impedances.

In the case of oscillators incorporating 50 ohm modular amplifiers and mode-selection, L-C tuned circuits (used to prevent oscillation at undesired crystal resonant modes), crystal replacement with fixed, resistive/reactive components results in a low Q, L-C oscillator operating at the same signal level and (approximately) same frequency. The operating point and closed loop group delay of the low Q oscillator circuit is easily determined via open loop, transmission response measurements. For oscillator sustaining stage designs (i.e., SAWR oscillators) not requiring mode-selection circuit selectivity, the acoustic resonator itself can be replaced by a known low Q value, R-L-C resonant circuit whose resistance value is equivalent to that of the resonator.

Measurement of the phase noise characteristics for the known, low Q value, L-C oscillator allow accurate prediction of crystal/SAWR-controlled oscillator near-carrier and noise floor performance exclusive of that resulting from resonator short-term frequency instability. The low Q oscillator white frequency noise levels (which are used to predict crystal oscillator noise floor level) are well above the noise contribution of all auxiliary measurement equipment. Excellent agreement between predicted and actual oscillator performance has been demonstrated using this technique for evaluation of oscillator configurations where resonator out-of-band impedance effects (on output signal spectral performance) is negligible. It should prove useful in connection with evaluation of the effects of sustaining stage amplifier gain compression and ALC diode bias circuit configurations on resultant oscillator flicker-of-frequency noise and noise floor performance.

Introduction

The phase noise spectral characteristics of quartz crystal and other high Q acoustic resonator-stabilized oscillators can be divided into two distinct regions: (1) a near-carrier region where the output signal spectrum is dominated by flicker-of-frequency instability, and (2) a white phase noise (floor) region. For many years, prediction of quartz crystal oscillator output signal spectral performance was made based on calculation and/or measurement of sustaining stage open loop phase noise characteristics, with subsequent prediction of resultant oscillator closed loop behavior made by considering the conversion of phase-to-frequency instability in accordance with the resonator in-circuit group delay [1]. Using this model, oscillator signal flicker-of-frequency noise was often assumed to be the result of sustaining stage open loop, flicker-of-phase noise.

Beginning in the mid-1970s, measurement of quartz crystal resonator short-term frequency characteristics by Walls and others showed that acoustic resonator flicker-of-frequency instability, rather than sustaining stage flicker-of-phase noise, is often the primary contributor of oscillator signal near-carrier spectral performance [2-6]. Accurate measurement of oscillator/resonator flicker-of-frequency noise is possible using commercially available test equipment via use of: (1) pairs of phaselocked oscillators, and/or (2) pairs of resonators installed in a passive phase bridge.

Surface acoustic wave resonator processing refinements, the development of the Stress-Compensated (SC) cut quartz crystal, and the use of multiple resonators have resulted in low noise oscillators employing relatively high resonator drive levels. As a result, it is not uncommon for oscillators to be designed for which output signal phase noise floor levels near or below -180 dBc/Hz are predicted. These low noise floor oscillator configurations often employ self-contained, 50-ohm modular amplifiers as the sustaining stage gain element. The benefits of such configurations include ease of measurement of open loop transmission and phase noise characteristics, and the availability of amplifiers exhibiting unconditional stability and very low levels of flicker-of-phase noise [7-9].

0-7803-0905-7/93 $3.00 © 1993 IEEE

The most common method used for measurement of oscillator phase noise performance involves spectral analysis of the amplified video signal recovered from a Schottky diode, double balanced mixer serving as the phase detector in a phase lock loop in which the oscillator under test is locked to a similar (or known, lower phase noise) oscillator (figure 1). Verification of exceptionally low oscillator phase noise floor level can be difficult, however. As shown in figure 2 and table 1, this is because the noise floor of associated measurement equipment is not appreciably lower than that of the oscillator(s) under test. Accurate determination of oscillator flicker-of-frequency noise, on the other hand, is possible because the contribution of the measurement equipment is negligible in the (near-carrier) region. Often, additional attempts are made to reduce the effects of measurement instrumentation noise via signal frequency multiplication and/or level amplification prior to measurement (phase detection). These techniques suffer from the same drawback; namely, that the cumulative phase noise floor level of the multipliers/amplifiers is not significantly less than that of the oscillator.

Methods exist for determination of the effective noise floor levels associated with phase noise measurement equipment itself, as well any auxiliary multiplier/amplifier circuitry, including driving pairs of the circuits under test from a common source. In either case, measurement inaccuracies and/or actual changes in non-oscillator hardware phase noise floor levels occurring from test-to-test due to differences in test conditions can have a large impact on accurate interpretation of measured results.

Oscillator Noise Measurements Using A Low Q, L-C Resonator

When the primary oscillator frequency-determining element is a low Q resonant circuit, the conversion of carrier signal open-loop phase noise to closed-loop frequency noise begins at a high carrier offset frequency, The effect, as shown in figure 3, is that the oscillator output signal spectrum includes a significant offset frequency band where the dominant spectral characteristic is that of white frequency (20 dB/decade) noise. The signal phase noise sideband level in this region is well above that of associated phase noise measurement equipment and is directly related to the oscillator circuit white phase noise (floor) level and closed-loop group delay. In addition, the oscillator white frequency noise level in many cases is substantially higher than that of available synthesized signal generators like the HP8662A that include adequately wide frequency deviation FM capability and can be phaselocked to the (low Q resonator) oscillator as part of the phase noise measurement procedure.

Table 1. Measurement Errors Introduced by Auxiliary Circuit and Measurement Equipment Additive Noise

Ratio of Total Measured Noise to All Non-Oscillator Measurement Hardware Noise	Corresponding Ratio of Total Measured Noise to Oscillator Noise
1.0 db	6.9 dB
1.5 dB	5.3 dB
2.0 dB	4.3 dB
2.5 dB	3.6 dB
3.0 dB	3.0 dB
4.0 dB	2.2 dB
5.0 dB	1.6 dB
6.0 dB	1.2 dB

VHF Quartz Crystal Oscillator Measurements

A large percentage of crystal oscillator circuits employ, in addition to the crystal resonator itself, additional L-C tuned circuits in order to prevent oscillation at unwanted crystal resonant modes. Normally, the tuned circuit loaded Q is made as low as possible in order to minimize oscillator frequency sensitivity to tuned circuit element value tolerance/drift. For oscillators designed to operate on an overtone mode and/or employing SC-cut crystals exhibiting an undesired, b mode response, use of a single-tuned, L-C circuit having loaded Q value in the range 5 to 10 is typical. Figure 4 shows simplified block diagrams for VHF crystal oscillators previously described by the author [8,9]. As shown in the figure, the circuits utilize a 50-ohm, modular amplifier as the sustaining RF gain element. In oscillator circuit topologies such as those of figure 4, the resonator out-of-band impedance has negligible effect on sustaining stage noise floor level. When the crystal resonators in the figure 4 circuits are replaced by resistors equal in value to the crystal series resistance, the oscillators operate at approximately the same frequency and quiescent signal level. Measurement of modified oscillator open loop group delay (usually made by breaking the loop at the sustaining stage amplifier input) is easily accomplished using a network analyzer. Open loop circuit characterization is made at drive levels that duplicate oscillator (closed loop) steady-state conditions.

Measurement Results: 80 and 100 MHz Crystal Oscillators

Using on-hand, 80 and 100 MHz crystal oscillators configured as shown in figure 4, measurements of oscillator open loop group delay and

closed loop phase noise were made with resistors substituted for the crystal resonators. Figure 5 shows the open loop transmission response measurement results obtained for an oscillator-frequency doubler normally employing a 3rd overtone, 80 MHz, AT-cut crystal, where the crystal has been replaced by a resistor equal in value to the crystal series resistance. As shown in the figure, the open loop group delay (primarily due to the L-C mode selection circuit) is 22 nsec. This implies the conversion of open loop, white phase noise to closed loop, white frequency noise begins at $f_c = 1/2\pi\tau = 7.2$ MHz.

Figure 6 shows the phase noise measurement results for the oscillator/doubler. The measurement was made by phase locking the output of an HP8662A synthesized signal generator, operated in the DCFM mode (50 KHz/volt sensitivity) to the oscillator under test. As shown in the figure the spectrum includes a wide region where the noise characteristic is white frequency (20 dB/decade) in nature. At a carrier offset frequency fm = 100 KHz, the phase noise level is -130 dBc/Hz at the doubler output, or -136 dBc/Hz at the (80 MHz) oscillator frequency. Therefore, the oscillator floor level is equal to:

$$-136 \text{ dBc/Hz} + 20 \log(2\pi\tau \text{ fm}) = -136 \text{ dBc/Hz} - 37 \text{ dB} = -173 \text{ dBc/Hz} \quad (1)$$

The phase noise sideband level at 100 Hz offset frequency is −60 dBc/Hz (-66 dBc/Hz at the oscillator output). This implies a sustaining stage flicker-of-phase noise level of -66 dBc/Hz -97 dB or -163 dBc/Hz, which is in good agreement with that separately measured for the (QBit QBH-125) sustaining stage amplifier used. Figure 7 shows the results of measurement of actual crystal-controlled oscillator-doubler output signal phase noise. For this test, a voltage controlled crystal oscillator (VCXO) plus doubler exhibiting very low noise floor performance (approximately -173 dBc/Hz at 160 MHz) and employing four, series-connected, SC-cut crystals was used as the comparison unit [8]. The results indicate a noise floor level of -165 dBc/Hz at 160 MHz, or -171 dBc/Hz referred to the oscillator frequency. The 2 dB discrepancy between calculated and measured results is due to the additive noise floor contributions of the VCXO, frequency doublers, and phase noise test equipment.

Figure 8 shows phase noise measurement results for a pair of 100 MHz oscillators, each employing two BT-cut crystals [8]. As shown in the figure, the (two-oscillator comparison) noise floor level is -172 dBc/Hz, or -175 dBc/Hz on a per-oscillator basis. Figure 9 shows the result of measurement of oscillator phase noise (again, using the HP8662A, DCFM mode) with the crystals replaced by resistors. In this case, the measured oscillator open loop group delay was 25 nsec ($fc = 1/2\pi\tau = 6.4$ MHz). As shown in the figure, the white frequency noise level at fm = 10 KHz is -120 dBc/Hz. Therefore the phase noise floor level is:

$$-120 \text{ dBc/Hz} + 20 \log(2\pi\tau \text{ fm}) = -120 \text{ dBc/Hz} - 56 \text{ dB} = -176 \text{ dBc/Hz} \quad (2)$$

This implies only 1 dB error exists for the figure 8 data.

Measurement Results: 320 MHz SAWR Oscillators

Figure 10 shows the block diagram for a voltage-controlled, 320 MHz SAWR oscillator. As shown in the figure, a pair of parallel sustaining stage amplifiers, and a pair of series-connected SAWRs were used to double oscillator/resonator drive level capability in an effort to improve output signal noise floor level. Figure 11 shows phase noise measurement results obtained at 640 MHz for a pair of oscillator-doublers. As shown in the figure, the measured noise floor level was -169 dBc/Hz, or -178 dBc/Hz on a per-device basis, referred to the oscillator frequency.

Figure 12 shows the open loop amplitude and phase response for the oscillator, obtained after replacing the SAWRs with an L-C tuned circuit/resistive pad combination having the same loss as the two, series-connected SAWR circuit. As shown in the figure, the group delay is 32 nsec (12.6°/MHz). Therefore fc = 4.6 MHz. Figure 13 shows the measured (L-C) oscillator-doubler phase noise spectrum. For the figure 13 measurement the 640 MHz comparison signal was derived from a low noise, 80 MHz crystal oscillator, and with the phase shifter incorporated in the SAWR oscillator (with SAWR replaced by the L-C tuned circuit) providing requisite, voltage-frequency tuning capability for oscillator phaselock.

As shown in figure 13, the white frequency noise level for the oscillator doubler at fm = 10 KHz is -123 dBc/Hz, or -129 dBc/Hz, referred to the oscillator frequency. Therefore the oscillator floor level is:

$$-129 \text{ dBc/Hz} + 20 \log(2\pi\tau \text{ fm}) = -129 \text{ dBc/Hz} - 53 \text{ dB} = -182 \text{ dBc/Hz} \quad (3)$$

In this case, the implied discrepancy between measured and actual oscillator phase noise floor level is 4 dB. Again, the "error" is due to the extremely low noise floor level exhibited by the oscillator plus the additive effects of doubler circuit and test set phase noise floor levels. Spectral "irregularities" in the region fm < 300 Hz are due to the effect of low level, benchtop vibration on the trimmer capacitor used in the tuned circuit.

Conclusions

Accurate determination of low oscillator phase

noise floor levels in crystal and SAWR oscillators employing high resonator drive level can be difficult, due to the relative contribution of circuitry following the oscillator and/or phase noise measurement test equipment.

Measurement of the flicker-of-frequency and white frequency noise of crystal and SAWR oscillators with the resonators replaced with a resistor and the oscillator operating with a known value, low Q, tuned circuit, provides a useful method for accurate determination of sustaining stage flicker-of-phase and white phase noise (floor) level.

This technique should be used for oscillator configurations, such as those employing modular amplifier sustaining stages, where resonator out-of-band impedance does not play a role in determination of oscillator signal noise floor performance.

References

[1] D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," Proc. IEEE, vol. 54, no. 2, Feb. 1966, pp. 329-330.

[2] F. Walls, et al., "Measurement of the Short-Term Stability of Quartz Crystal Resonators and the Implication for Crystal Oscillator Design," IEEE Trans. Instr. and Meas., vol. IM-24, no. 1, March 1975, pp. 15-20.

[3] J. Gagnepain, et al., "Excess Noise in Quartz Crystal Resonators," Proc. 37th Freq. Contr. Symp., May 1983, pp. 218.

[4] S. Eliot, et al., "Direct Phase Noise Measurements of SAW Resonators," Proc. 1984 IEEE Ultras. Symp., Oct. 1984, pp. 180-185.

[5] T. Parker, "1/f Frequency Fluctuations in Acoustic and Other Stable Oscillators," Proc. 39th Freq. Contr. Symp., May 1985, pp. 97-106.

[6] M. M. Driscoll and B. W. Kramer, "Spectral Degradation in VHF Crystal-Controlled Oscillators Due to Short-Term Frequency Instability in the Quartz Resonator," Proc. 1985 IEEE Ultras. Symp., Oct. 1985, pp. 340-345.

[7] G. K. Montress, et al., "Extremely Low Phase Noise SAW Resonator Oscillator Design and Performance," Proc. 1987 IEEE Ultras. Symp., Oct. 1987, pp. 47-52.

[8] M. M. Driscoll, "Reduction of Quartz Crystal Oscillator Flicker-of-Frequency Noise and White Phase Noise (Floor) Levels and Acceleration Sensitivity Via Use of Multiple Resonators," Proc. 1992 IEEE Freq. Contr. Symp., May, 1992, pp. 334-339.

[9] M. M. Driscoll, "Low Noise Crystal Oscillators Using 50-Ohm Modular Amplifier Sustaining Stages," Proc. 40th Freq. Contr. Symp., May 1986, pp. 329-335.

[10] M. M. Driscoll and R. W. Weinert, "Low Noise, Microwave Signal Generation using Cryogenic, Sapphire Dielectric Resonators: An Update," Proc. 1992 Freq. Contr. Symp., May, 1992, pp. 157-162.

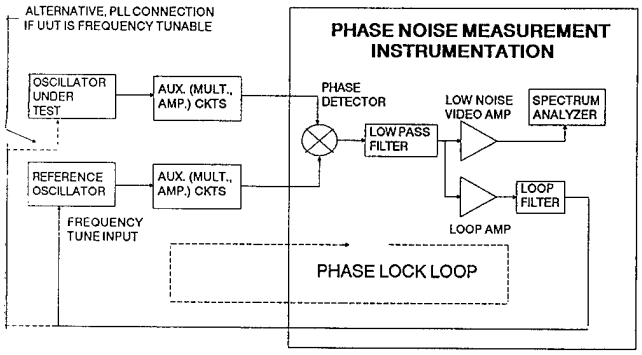


Figure 1. Two-Oscillator Phase Noise Measurement Block Diagram

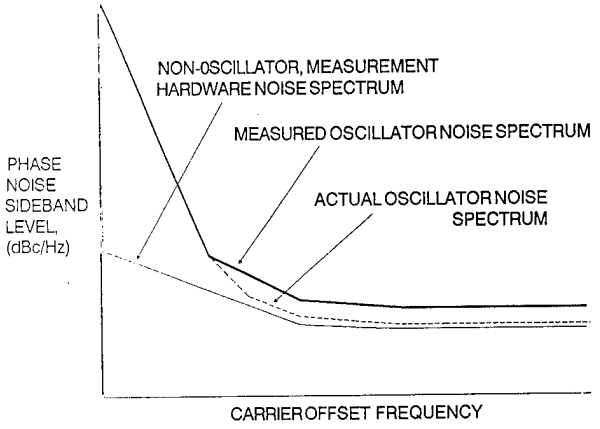


Figure 2. Typical, Comparative Phase Noise Spectra for a Low Noise Crystal Oscillator Signal and Associated Non-Oscillator Circuitry, Including Phase Noise Measurement Hardware

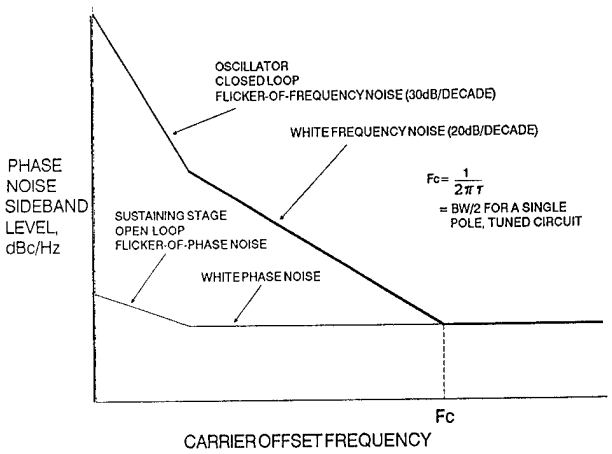
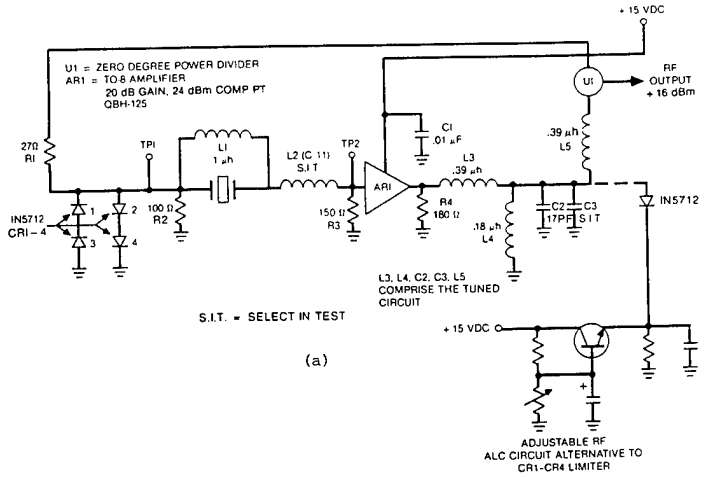
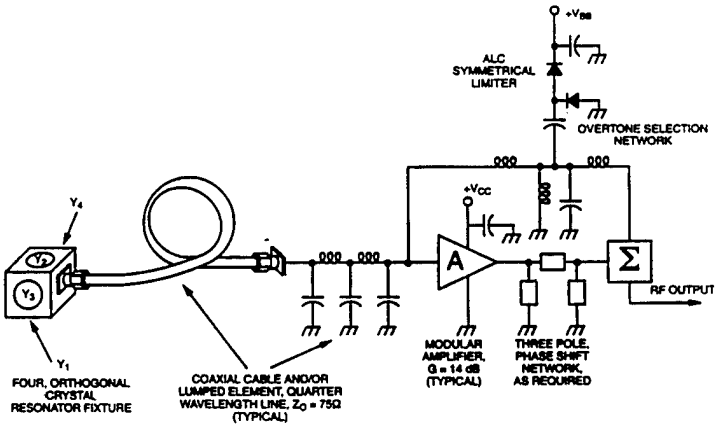


Figure 3. Typical Signal Phase Noise Spectrum for a Low Q, L-C Resonator-Controlled Oscillator



(a)



(b)

Figure 4. Crystal Oscillators Employing 50-Ohm, Modular Amplifier Sustaining Stages

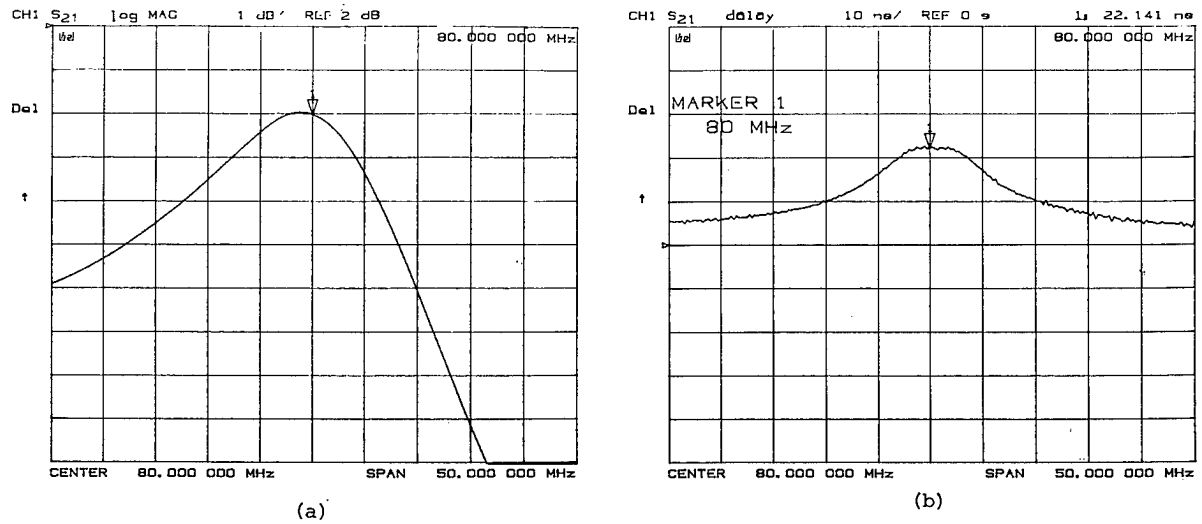


Figure 5. Measured Open Loop Transmission Response for an 80 MHz Crystal Oscillator with the Crystal Replaced by a Resistor
(a) Amplitude (b) Group Delay

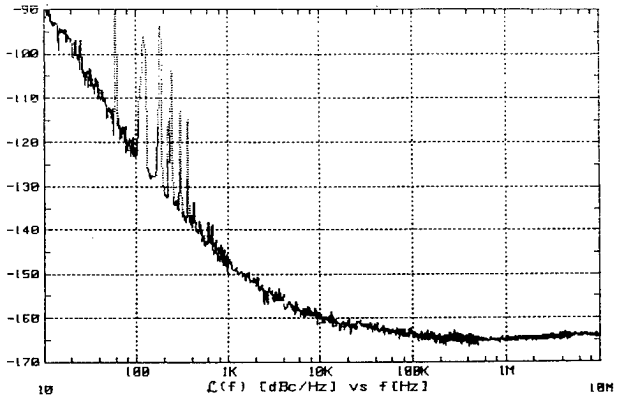


Figure 7. Phase Noise Measurement Results for the 80 MHz Crystal Oscillator-Doubler with the Oscillator Crystal Reinstalled

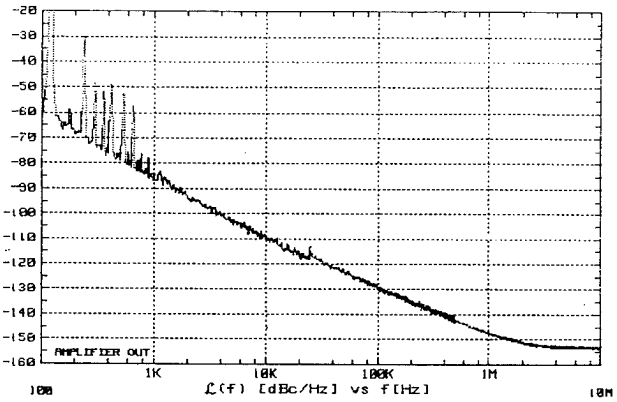


Figure 6. Phase Noise Measurement Results for the Figure 5 Oscillator plus Output Frequency Doubler

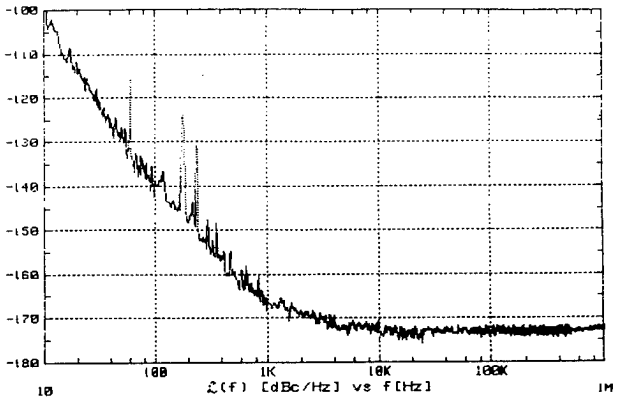


Figure 8. Phase Noise Measurement Results for a Pair of 100 MHz Crystal Oscillators, Each Employing Two BT-Cut Crystals

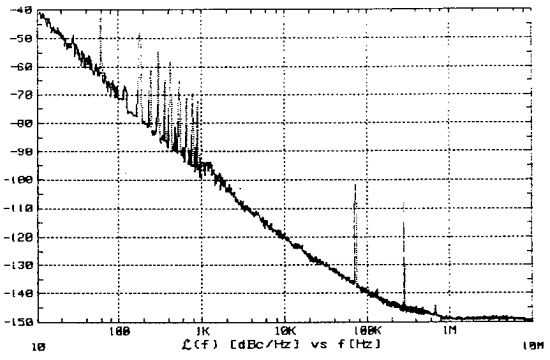


Figure 9. Phase Noise Measurement Results for the 100 MHz Crystal Oscillator with the Crystals Replaced by Resistors

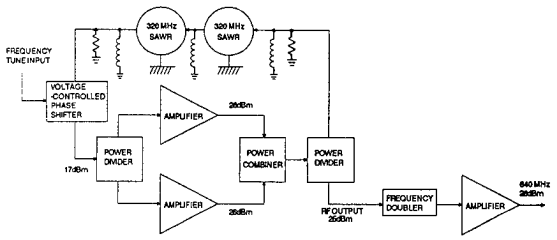


Figure 10. 320 MHz Voltage-Controlled, SAWR Oscillator Block Diagram

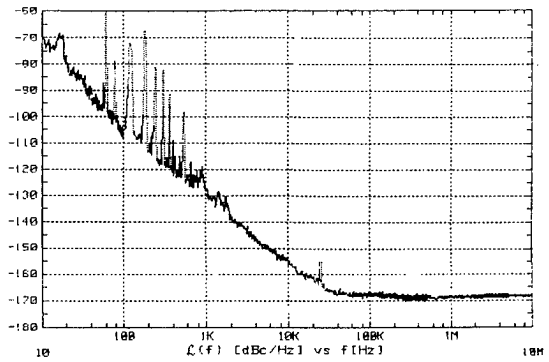


Figure 11. Phase Noise Measurement Results for a Pair of 320 MHz SAWR Oscillator-Doublers (Output Frequency = 640 MHz)

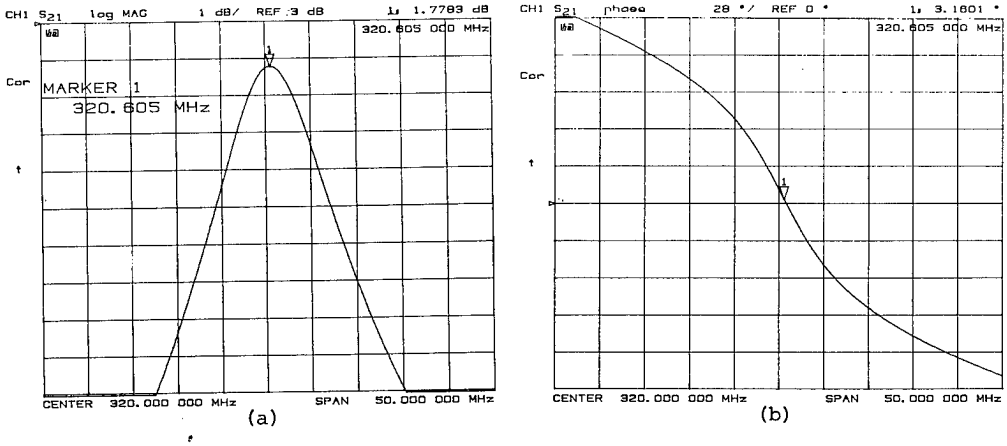


Figure 12. Measured, 320 MHz SAWR Oscillator Open Loop Transmission Response with the SAWRs Replaced by a Low Q, Tuned Circuit
(a) Amplitude (b) Phase

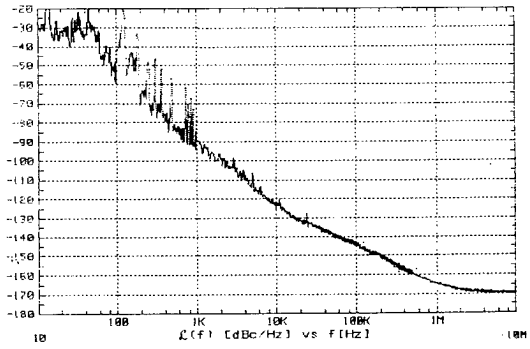


Figure 13. Phase Noise Measurement Results for the SAWR Oscillator-Doublers with the SAWRs Replaced by the Low Q, Tuned Circuit

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A HIGH PERFORMANCE VHF OSCILLATOR WITH OPTIMIZED CRYSTAL DRIVE POWER

Vaughn H. Estrick

Hughes Aircraft Co.
Sensors and Communications Systems Division
1901 West Malvern Avenue
Fullerton, CA 92634 USA

Abstract

The common method to predict the phase noise for crystal controlled oscillators is to use Leeson's model. Unfortunately, many oscillator circuits do not give results consistent with this simple model because of the interaction between the active devices and the frequency control element. A model has been formulated that takes into account the effects of the resonator impedance on the noise levels occurring in an emitter coupled oscillator and its associated buffer circuit. This model uses the standard transistor noise sources to predict the output noise level of a circuit with the resonator connected to the transistor emitter. The model was also used to analyze the performance of a high performance crystal oscillator that exhibits a low phase noise floor and minimal frequency chirp after power is applied. The drive level needs to be kept as low as possible to prevent frequency pulling caused by heating of the crystal. A transformer is used in the oscillating transistor emitter to control the drive power to the resonator while optimizing the loaded Q of the circuit. A wideband noise floor below -172 dBc/Hz has been achieved while dissipating less than one-half milliwatt of power in the resonator. Higher crystal drive levels provide correspondingly lower noise floors.

Introduction

Recent improvements in VHF oscillator design, resonator technology and phase noise measurements have resulted in very low (-180 dB/Hz) wideband phase noise floors [1], [2]. However, the shape of the phase noise curve and the noise floor is often different from that predicted with simple models. Leeson's model shown in Figure 1 has works well for oscillators that use amplifiers whose parameters are not sensitive to variations in load [2],[3]. For instance, if the noise level of the sustaining stage amplifier varies with frequency shown by the dotted line in the figure, then the output phase noise will be lower. Some oscillator designs can take advantage of this condition and produce lower noise than those using load insensitive devices [4]. Likewise, the use of a narrowband filter behind the oscillator (such as with a crystal filter) can make the noise floor decrease.

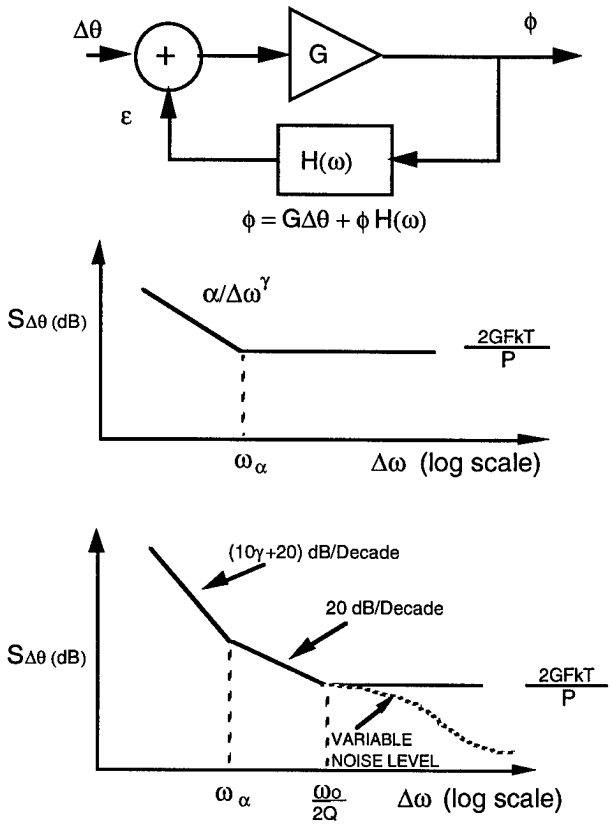


Figure 1. Leeson's Model for Predicting Oscillator Phase Noise.

0-7803-0905-7/93 $3.00 © 1993 IEEE

One such oscillator design reported by Driscoll used this effect to reduce the wideband noise floor [5]. A simplified schematic circuit is shown in Figure 2. This 80 MHz design had a noise floor of -176 dBc/Hz while dissipating about 6 mw (7.8 dBm) in the crystal resonator. An advantage of this circuit is that the resonator acts as a filter and attenuates the wide band noise from the oscillator stage that is fed into the common base buffer. The wideband noise is then primarily set by the buffer amplifier.

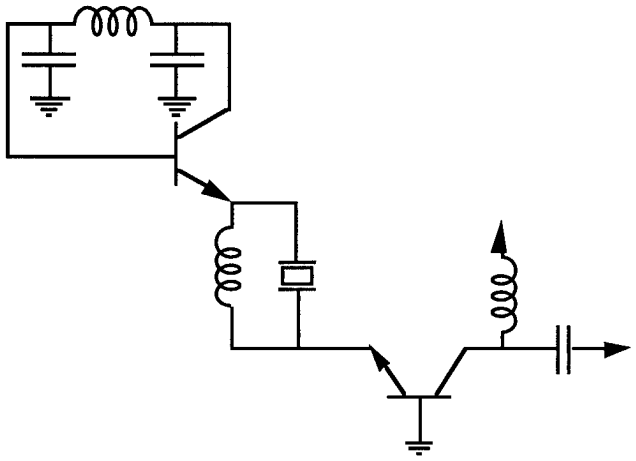


Figure 2. Emitter Coupled Oscillator with Common Base Buffer Amplifier Stage.

This configuration proved particularly attractive for an application that required a moderately low phase noise floor of -170 dBc/Hz. An additional requirement was that the oscillator reaches a stable frequency quickly after cold power-up. Self heating of the crystal after turn on causes excessive frequency drift. The RF power dissipated in the crystal needed to be less than a milliwatt to keep drift within acceptable limits. This limits the achievable level of the phase noise floor forcing a careful design of the oscillator circuit. A transformer adjusts the drive to the crystal and transforms the crystal impedance to achieve the desired loaded Q. This is shown in Figure 3 where the resonator and transformer are attached to the input of a common base amplifier. A 9:1 impedance transformation converts the 135 ohms series resistance of the crystal to 15 ohms at resonance.

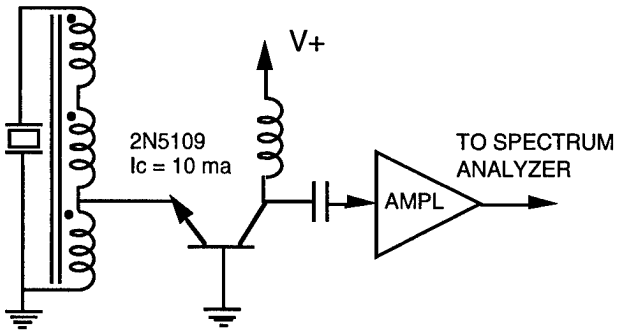


Figure 3. Common Base Amplifier with Transformed Resonator Load on the Input.

The unloaded Q of the crystal used for this set up was 80,000. A 2N5109 biased at 10 ma provided the grounded base stage. The output uses an L-transformation with the signal fed into a spectrum analyzer after suitable gain. The measured output noise of this configuration is shown in Figure 4. The interaction of the transformed crystal impedance and the internal noise sources of the transistor determine the shape of the output noise spectrum. The constant amplitude trace is the noise when a 50 ohm termination is substituted for the transformer and crystal.

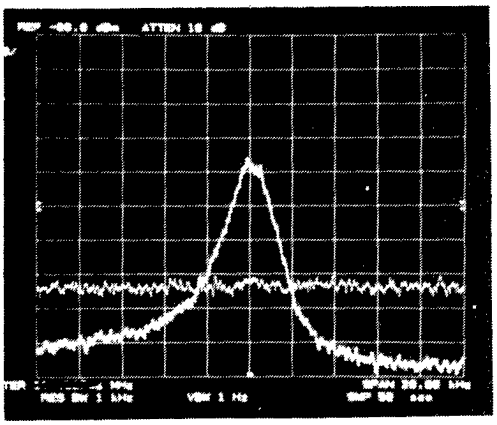


Figure 4. Output Noise of Common Base Amplifier with Crystal Connected to Emitter. Scales are 2 dB/div and 2 kHz/div.

Transistor Noise Modeling

A model of the circuit in Figure 3 was developed to include the effects of crystal load variation transistor and resistive noise sources.

The crystal modeled as a simple series RLC circuit is shown in Figure 5. Note that this model does not account for the shunt capacitance of the crystal or any asymmetry caused by other modes. The R in the circuit can include any external resistance so that the loaded Q can be determined from the unloaded Q (Q_u).

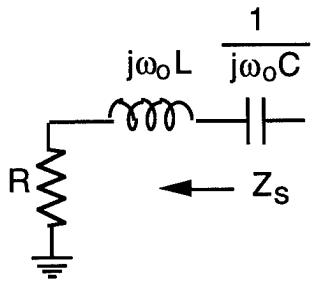


Figure 5. Simplified Model of Grounded Crystal Resonator.

The Q of this circuit is

$$Q = \frac{\omega_0 L}{R} = \frac{1}{\omega_0 CR} \qquad (1)$$

The impedance Z_S is

$$Z_S = R + j\omega L + 1/j\omega C \qquad (2)$$

$$= R(1 + \frac{j\omega Q}{\omega_0} - \frac{j\omega_0 Q}{\omega}).$$

For $\omega = \omega_0 + \Delta\omega$

$$Z_S = R[1 + \frac{j(\omega_0+\Delta\omega)Q}{\omega_0} - \frac{j\omega_0 Q}{(\omega_0+\Delta\omega)}]$$

$$Z_S = R[1 + jQ(\frac{\omega_0+\Delta\omega}{\omega_0} - \frac{\omega_0}{\omega_0+\Delta\omega})]. \qquad (3)$$

Setting

$$\Omega = \frac{\omega_0+\Delta\omega}{\omega_0} - \frac{\omega_0}{\omega_0+\Delta\omega}$$

gives

$$Z_S = R[1 + jQ\Omega] = R + j RQ\Omega. \qquad (4)$$

Figure 6 gives a convenient model for a grounded base transistor including noise sources with the crystal connected to the emitter [6]. There are four noise sources, with three voltage sources and one current source. Each voltage source has an associated resistance. The source resistance Rs is the crystal series resistance with an associated squared voltage noise source of $4KTR_S$. The resistor $r_b = r_b + \frac{r_e}{2}$ is the sum of the real base resistance r_b and the collector shot noise referred to the base [7]. The value of $4KTr_b$ is the voltage squared noise of this resistor combination in a one Hz bandwidth. The base shot noise is a squared current $\overline{i_n}^2$ that is equal to 2qIB. The noise from the output load resistance is a voltage generator with a voltage squared value equal to $4kTR_L$.

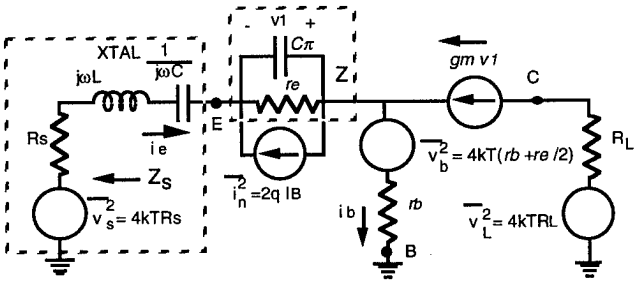


Figure 6. Transistor Grounded Base Model with Resonator Connected to the Emitter.

From the diagram,

$$i_b = i_e + g_m v_1$$

and

$$v_1 = -(i_e + \overline{i}_n) Z$$

so

$$i_e = -\frac{v_1}{Z} - \overline{i}_n \qquad (5)$$

where Z is the parallel combination of *re* and Cπ. Writing the loop equation gives

$$\overline{v}_b + \overline{v}_s = i_e Z_S - v_1 + i_b r_b \qquad (6)$$

$$\bar{v}_b + \bar{v}_s = (-\frac{v_1}{Z} - \bar{i}_n)Z_S - v_1 + (i_e + g_m v_1) r_b$$

$$= -\frac{v_1}{Z} - \bar{i}_n Z_S - v_1 - \frac{v_1}{Z} r_b - \bar{i}_n r_b + g_m v_1 r_b \,.$$

Rearranging terms and solving for v_1 gives

$$v_1 = \frac{Z[\bar{v}_b + \bar{v}_s + \bar{i}_n(Z_S + r_b)]}{Z_S + r_b + Z - g_m Z r_b} \,. \qquad (7)$$

Since the noise sources are independent their powers add, so

$$v_1^2 = \frac{|Z|^2[|\bar{v}_b|^2 + |\bar{v}_s|^2 + |\bar{i}_n(Z_S + r_b)|^2]}{|Z_S + r_b + Z - g_m Z r_b|^2}$$

Substituting the values for v_b, v_s and $\bar{i}_n$ from Figure 6 gives

$$v_1^2 = \frac{|Z|^2[4KT(r_b + R_S) + 2qI_B|r_b + Z_S|^2]}{|Z_S + r_b + Z - g_m Z r_b|^2} \,. \qquad (8)$$

The noise power contributed from the load resistor is $4KTR_L$ when referenced back to v_1 becomes

$$\frac{4KTR_L}{g_m^2 R_L^2} = \frac{4KT}{g_m^2 R_L} \,. \qquad (9)$$

since $g_m R_L$ is the voltage gain of the transistor circuit. If the transistor is followed by a second stage with a noise figure of 10 log F, where F is the noise factor the effective noise referred back to the input is

$$v_{12}^2 = \frac{4KTF}{g_m^2 R_L} \qquad (10)$$

where v_{12} is the noise from the second stage.

The total noise in a one Hz bandwidth is the sum of the powers since the noise sources can be assumed to be independent or

$$v_1^2 = \frac{|Z|^2[4KT(r_b + R_S) + 2qI_B|r_b + Z_S|^2]}{|Z_S + r_b + Z - g_m Z r_b|^2}$$

$$+ \frac{4KTF}{g_m^2 R_L} \qquad (11)$$

It is left now to express the noise in more familiar transistor and circuit terms such as transistor β, f_T, crystal unloaded Q, etc. As stated earlier, Z is the parallel combination of r_e and $C\pi$, or

$$Z = \frac{r_e \frac{1}{j\omega C\pi}}{r_e + \frac{1}{j\omega C\pi}} = \frac{r_e}{1 + j\omega r_e C\pi}$$

Since Z will not change noticeably over the resonator frequency band, ω can be assumed constant and equal to ω_0. Therefore,

$$Z \approx \frac{r_e}{1 + j\omega_0 r_e C\pi} \,.$$

However, the current gain bandwidth product f_T of the transistor is also defined as $\frac{\beta}{2\pi r_\pi C_\pi}$ where $r\pi = r_e(1 + \beta)$.

Therefore,

$$Z = \frac{r_e}{1 + j\beta_M} \qquad (12)$$

where

$$\beta_M = \frac{\beta f_0}{(\beta + 1) f_T} \,.$$

Z also can be expressed as

$$Z = \frac{r_e(1 - j\beta_M)}{1 + \beta_M^2}$$

and

$$|Z|^2 = \frac{r_e^2(1 + \beta_M^2)}{(1 + \beta_M^2)^2} = \frac{r_e^2}{1 + \beta_M^2} \,. \qquad (13)$$

The next step reduces the denominator term from equation 8 into its real and imaginary parts, such that

$$r_b + Z + Z_S - g_m Z r_b = r_b + \frac{r_e}{1 + \beta_M^2} + j\frac{\beta_M r_e}{1 + \beta_M^2}$$

$$+ Z_S - \frac{g_m r_e r_b}{1 + \beta_M^2} - j\frac{g_m r_e r_b \beta_M}{1 + \beta_M^2}$$

The term $r_b + Z_S$ can be rewritten using equation 4 as

$$r_b + Z_S = r_b + R_S + j(R_S Q_u \Omega)^2$$

and

$$| r_b + Z_S |^2 = (r_b + R_S)^2 + (R_S Q_u \Omega)^2 .$$

Finally, after substituting into equation 11, the noise referred to the input in a one Hz bandwidth can now be expressed as

$$v_1^2 = \frac{\frac{r_e^2}{1 + \beta_M^2}}{R^2 + X^2} \{4KT(r_b + R_S)$$

$$+ 2q I_B[(r_b + R_S)^2 + (R_S Q_u \Omega)^2]\} + \frac{4KTF}{g_m^2 R_L} \qquad (14)$$

where

$$R = r_b + R_S + \frac{r_e}{1 + \beta_M^2} (1 - g_m r_b) \qquad (15)$$

and

$$X = R_S Q_u \Omega - \frac{r_e \beta_M}{1 + \beta_M^2} (1 - g_m r_b) . \qquad (16)$$

Comparison of Measurements with Model

A spreadsheet program was developed to compare the measurements of Figure 4 to values calculated using equation 14. For this condition, a 2N5109 transistor was biased at 10 ma. The *rb* of the transistor was determined to be 40 ohms with β = 100 and f_T = 1200 MHz. The load resistor was 600 ohms and the second stage noise figure was 3 dB.

Figure 7 shows how the measured data compares with values calculated using the previous analysis. The heavy line is the calculated response and the lighter lines are from the bandwidth corrected spectrum analyzer response photo of Figure 4. As shown, the calculated curve matches the measured response close to resonance. Some of the difference is due to circuit parasitics and from not knowing all parameters exactly. More careful determination of circuit characteristics would likely give more exact results.

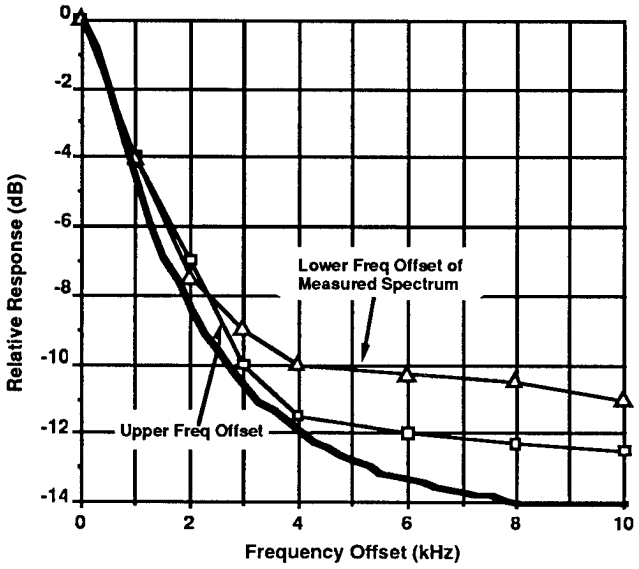


Figure 7. Comparison of Calculated and Measured Noise of Common Base Amplifier with Crystal Input Loading.

Oscillator Performance

Several oscillator circuits that used emitter coupling to a common base buffer were constructed and evaluated for phase noise (See Figure 8).

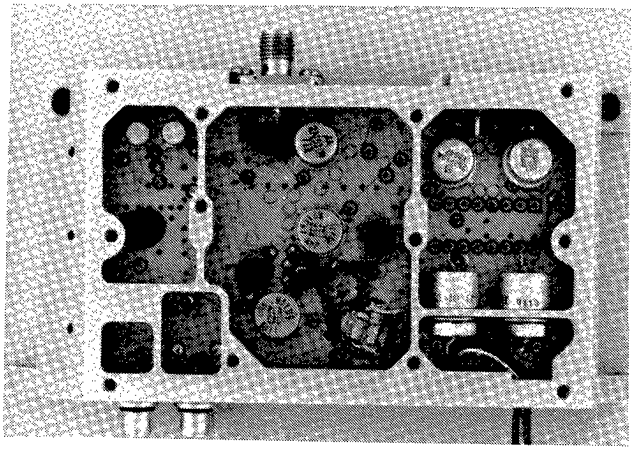


Figure 8. Low Phase Noise Oscillator with Optimized Crystal Drive.

A simplified schematic of the oscillator and buffer stage is shown in Figure 9. The oscillator stage uses a transformer to feedback from collector to base. The pi-attenuator, in conjunction with L1 and C1, adjust the level and phase of the feedback to sustain oscillations. Transformer T2 sets the desired crystal drive power and loaded Q. The network of Z1 helps reduce the tendency of the buffer stage to oscillate. A 100 MHz SC cut crystal was used with an unloaded Q of about 80,000.

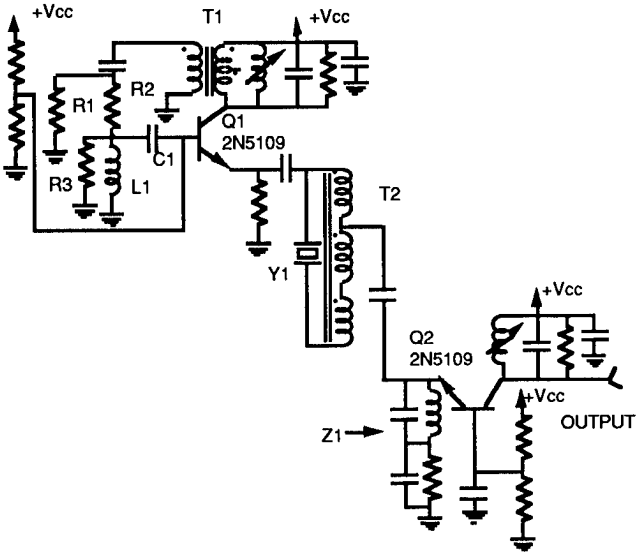


Figure 9. Schematic of Emitter Coupled Oscillator with a Common Base Buffer Amplifier.

The performance of the oscillator with a crystal dissipation of -3 dBm is shown in Figure 10. This configuration had a noise floor of -172 dBc/Hz.

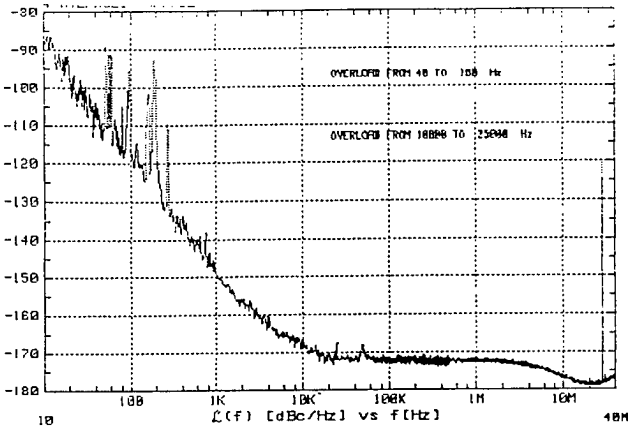


Figure 9. Phase Noise of Emitter Coupled Oscillator with -3 dBm of Power Dissipation in the Crystal.

The crystal drive power was increased 11 dB to +7 dBm giving phase noise performance shown in Figure 11. This oscillator used an SC-cut crystal different from the one that gave the above data. (The spurious that occur above 20 kHz are caused by RFI from the CRT in the phase noise test set.) The measured floor for this second oscillator was -180 dBc/Hz. This was only about 3 dB above the limit of the test set. It is likely that the actual floor for an individual oscillator was around -183 dBm/Hz.

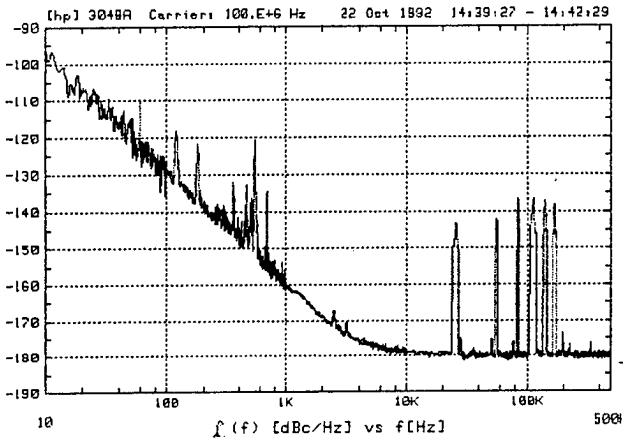


Figure 11. Phase Noise of Emitter Coupled Oscillator with +7 dBm of Power Dissipation in the Crystal.

Even lower noise floor can be obtained at higher drive powers, and SC-cut crystals can handle levels above 20 dBm. However, it is difficult to sustain this level through subsequent amplification and multiplication so often these latter stages often limit the overall noise floor. The close to carrier noise appears to be limited by both flicker noise of the transistors and by excess noise from the resonator. Additional reduction of the close in phase noise from this configuration will probably have to come from improving the flicker noise from these components.

Conclusions

The advantageous effect of using the loading of the crystal on a common base oscillator configuration has been shown to improve the phase noise floor. A model has been developed that predicts the noise performance of the buffer stage. By using this approach, the phase noise of the oscillator can be reduced without dissipating excessive power in the resonator.

Acknowledgments

The author would like to thank John Safford and Sam Hernandez for the design and testing of the oscillator circuits. Also thanks to Richard Docter for his support of this effort.

References

[1] R.A. Docter, et. al., "Low Noise Frequency Exciter Study," Final Report US Army LABCOM, SLECT-TR-87-0765-F, October 1990.

[2] T.E. Parker and G. K. Montross, "Spectral Purity of Acoustic Resonator Oscillators," *Proc. of the 1992 IEEE Frequency Control Symposium*,1992, pp.340-447.

[3] D.B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proc. of the IEEE*, Vol. 54, No. 2, Feb. 1966, pp. 329-330.

[4] M.M. Driscoll, "Low Moise, Microwave Signal Generation Using Bulk and Surface Acoustic Wave Resonators", *Proc. 42nd Annual Symposium on Frequency Control*, 1988, pp. 369 -377.

[5] M.M. Driscoll, "Low Noise, VHF Crystal - controlled Oscillator Utilizing Dual, SC-cut Resonators", *Proc. 38 nd Annual Symposium on Frequency Control*, 1985, pp.197-210.

[6] P.E. Gray and R Meyer, *Analysis and Design of Analog Integrated Circuits,* Wiley, New York, 1984, Chapter 11.

[7] Y. Netzer, "The Design of Low Noise Amplifiers", *Proc. of IEEE*, Vol. 69, No. 6, June 1981, pp. 728-742.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A HIGH ISOLATION LOW NOISE AMPLIFIER WITH NEAR UNITY GAIN UP TO 100 MHz

A. De Marchi, F. Mussino, M. Siccardi

Dip. di Elettronica, Politecnico di Torino.
C.so Duca degli Abruzzi 24, Torino, Italy.

Abstract

Increasing needs at ever higher frequencies in high resolution frequency metrology pose increasing demands on electronics. Available low phase noise isolation amplifiers may well be inadequate for the coming era. Designing amplifiers suitable for this purpose is no easy task. They must have high output-to-input isolation, low noise, wide bandwidth with near unity gain, and the capability of handling with low distortion relatively high power levels (up to 13 dBm). In this paper the design philosophy of a new type of isolation amplifier is illustrated, and the results obtained with a small number of prototypes are reported. Highlights of these are: unity gain with a 3 dB bandwidth of 350 MHz, isolation in excess of 140 dB up to 100 MHz, phase noise floor at −168 dBc/Hz with corner frequency about 300 Hz (at 13 dBm drive level), delay stability in excess of a few ps/°C at 100 MHz, and very good input and output matching to 50 Ω.

1 Introduction

Future needs of high resolution frequency metrology require high isolation between oscillators to prevent injection locking. For example, preventing a quartz oscillator in the 100 MHz range from injection locking to external signals of the same level with a 10^{-16} offset requires the availability of isolation amplifiers providing 140 dB of isolation if its own output buffer guarantees only 80 dB, which is a reasonable figure.

Designing an amplifier to meet these specifications represents a problem, particularly in view of the low phase noise and the high linearity required in most frequency metrology applications.

In this paper we report on results obtained with a circuit in which three cascaded common base BJT stages have short base leads directly connected to ground. Alternating NPN and PNP transistors biased with a bipolar supply was necessary for this purpose. In this way the bases act as shields within the device, which prevent any direct capacitive collector-to-emitter coupling, and the only reverse path for the signal is represented by the small internal base resistances R_b, as can be seen by inspection in fig.1. Backward isolation in excess of 40 dB per transistor was obtained, up to more than 100 MHz, with 4 GHz gain-bandwidth product devices. Other stage to stage coupling paths were cut by shielding and filtering on the power leads.

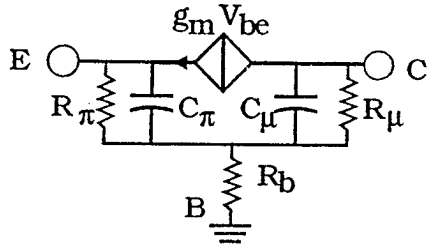


Figure 1: Giacoletto's high frequency small signal transistor model. The model is still valid at the high power level driving the circuit because linearity is guaranteed by the current operating mode.

Distortion problems may arise in this design from the nonlinearity of the first transistor's input characteristic. These were solved by current

0-7803-0905-7/93 $3.00 © 1993 IEEE

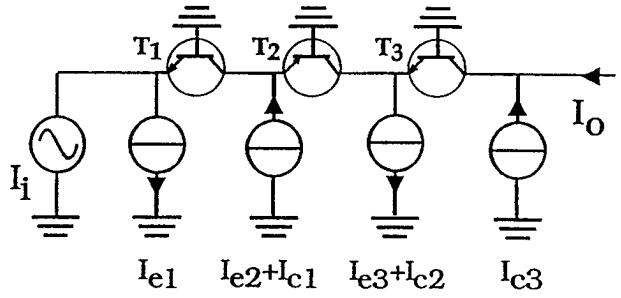


Figure 2: Three transistor NPN-PNP cascade configuration of the isolation amplifier.

source driving the first emitter. A high quiescent current (around 50 mA) is imposed by the requirement of remaining in linearity with the 60 mA_{pp} current swing induced by a 13 dBm drive level.

2 Design guidelines

The design guidelines which were followed to meet the specifications loosely referred to in the introduction are:

- Use of high frequency BJT transistors (as opposed to FETs) in a common base configuration, with directly grounded short leads on the bases, as already justified.
- Use of local feedback only, in order to avoid the existance of reverse paths for the signal. The sole feedback used is the intrinsic transistor's current feedback, which yields $I_c = I_e\, h_{fe}/(h_{fe}+1)$.
- Use of DC coupling everywhere, with the obvious exception of input and output connections, in order to reduce the number of noise producing components.
- Current operation of all stages. This produces various beneficial effects: 1) It guarantees high linear dinamic range, as the transistors are kept at an almost constant V_{ce}. 2) It allows good broadband input matching without signal power attenuation (this is necessary for low noise) by simple voltage to current conversion on a 47 Ω series resistor. 3) It lowers the impedance levels of the circuit, since the emitter impedance of the common base stages is then $Z_{ie} = V_T/I_{ee}$ <0.5 Ω at the operation current. 4) This in turn allows current sourcing with 100 Ω resistors, reducing in this way the number of components for low noise purposes.
- Use of current drive also for the output coupling emitter follower stage, after the signal is reconverted to voltage on a resistor; this is realized with an active load in order to safeguard the high linear dynamic range.

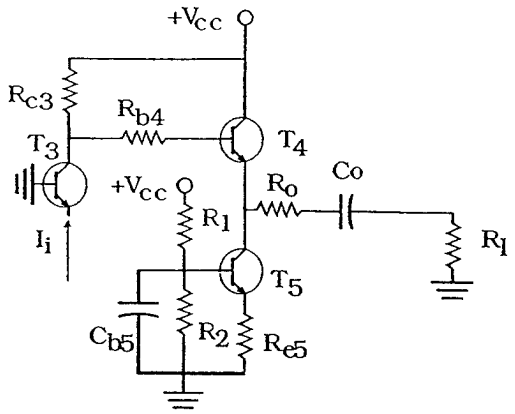


Figure 3: Schematics of the output coupling emitter follower stage, with active load and decoupling base resistor.

3 Circuit design and realization

The effective circuit which was designed and realized is shown schematically in fig.2, in which the bias current sources were implemented with 100 Ω resistors (which matches well the need to feed the designed current from a reasonable supply), and the driving signal current generator was obtained by voltage to current conversion on a 47 Ω series resistor. Computer assisted analysis of this three stage isolation amplifier yields an isolation in excess of 160 dB at 100 MHz. The used surface mount transistors were a BFG-194 for the

PNP type, with f_T=4.5 GHz and h_{fe}= 50, and four BFG-196 for the NPN type, with f_T=7 GHz and h_{fe}=90. Both types can stand a maximum current of 100 mA.

In order to increase the V_{ce} of the first two transistors, which would otherwise be of the order of 1.5 V only, a small series resistor was introduced between each collector and the emitter of the following stage. These resistors are also useful in increasing slightly the isolation.

The required output power of about 13 dBm is delivered to the 50 Ω load by the output emitter follower stage shown in fig.3, which is equipped with an active load for linear dynamic range reasons. In order to prevent undesired oscillations in this emitter follower stage, a 220 Ω series resistor was inserted on the base. The amplifier output impedance is 50 Ω.

Surface mount components were used throughout for the realization. Attention was paid to the printed circuit layout, allowing for the maximum extension of ground planes in both sides, using passing holes and preventing undesired r.f. current flow from one stage to the others. The r.f. current flow is confined in one side, while the d.c. supply leads are obtained in the opposite side of the printed circuit. Stage to stage coupling paths through the supply leads were also cut by filtering. No shielding by separating septs was instead necessary between stages.

4 Measured performances

The measured forward gain of the amplifier is shown in fig.4, where it can be seen that the 3 dB bandwidth is wider than 350 MHz. Beyond that frequency there are still some problems of overshoot which are not shown, but should be relatively easy to eliminate in the next iteration of the circuit.

In table.1 the relevant performances of the isolation amplifier are summarized. Careful positioning of ferrite beads close to some of the emitters is generally necessary in order to reach the quoted isolation.

The reverse isolation is also shown graphically

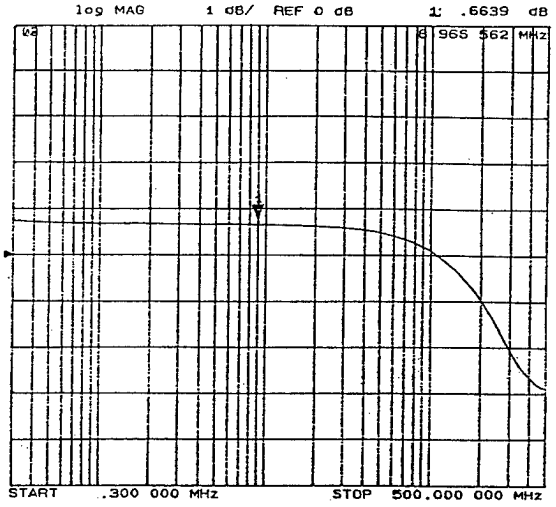


Figure 4: Measured direct gain of the amplifier.

** Forward gain:*
0 dB +/- 1dB
(10 MHz to 200 MHz)

** Reverse gain (isolation):*
<-150 dB at 10 MHz
-140 dB at 100 MHz
-135 dB at 150 MHz
-130 dB at 200 MHz
-125 dB at 300 MHz
-120 dB at 400 MHz
-115 dB at 500 MHz

** Noise : < 5 nV/$\sqrt{Hz}$*

** Non linear (harmonic) distorsions at an output level of 13 dBm :*

2nd order < -45 dB up to 100 MHz
< -40 dB up to 250 MHz
3rd order < -52 dB up to 100 MHz
< -44 dB up to 250 MHz

Table 1: Measured performances

in fig.5, where it can be noticed that the logarithmic slope of isolation vs carrier frequency is 40 dB/dec.

Measurements were also taken of added phase noise vs Fourier frequency, and of delay stability vs temperature, on one realization of the amplifier. These were taken at NIST in Boulder with the usual bridge configuration. The phase noise results, which were the same for 5 MHz and 100 MHz carrier frequency, are shown in fig.6 for a drive level of 13 dBm.

The delay stability vs temperature is also in-

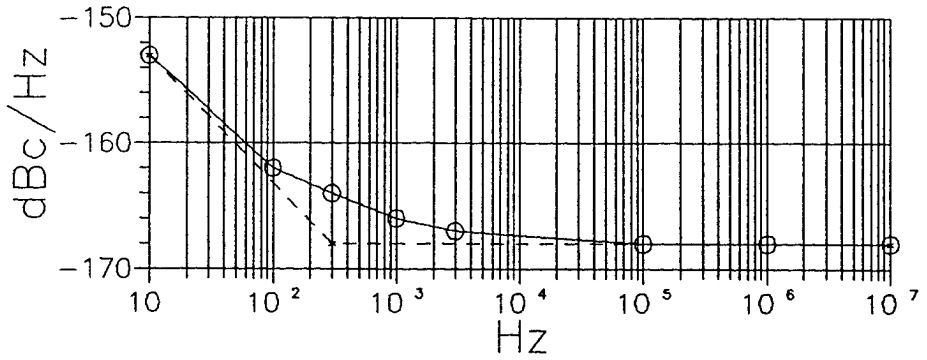


Figure 6: Measured phase noise vs Fourier frequency. The same results were obtained for 5 MHz and for 100 MHz carrier frequency.

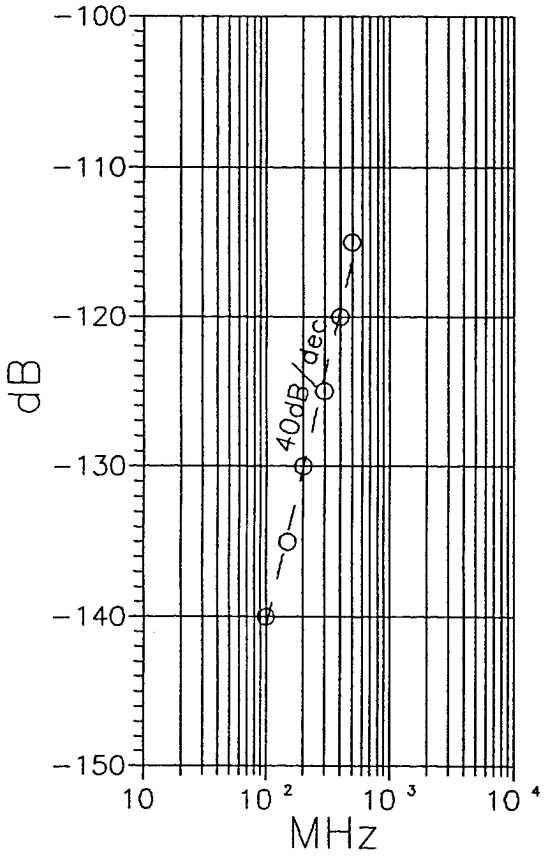


Figure 5: Typical output to input isolation as a function of frequency.

teresting, as it does not exceed a few ps/°C at frequencies which are within the 3 dB bandwidth of the amplifier.

Acknowledgements The authors wish to acknowledge the active help of Fred Walls in the evaluation of the performances of the amplifier.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A NEW HIGH-ISOLATION, ENVIRONMENTALLY INSENSITIVE, WIDEBAND DISTRIBUTION AMPLIFIER

Richard T. Lee and Kenneth Ho

Hewlett-Packard Company
Santa Clara, CA 95052

Abstract

This paper describes a wideband distribution amplifier that has a high degree of automatic gain control, low environmental sensitivity, and very good isolation between channels. It performs within design specifications from 100 KHz to 10 MHz without tuning or operator adjustment. Interchannel isolation of better than 100 dB is combined with low phase noise and low harmonic distortion. The amplifier has been designed to be operated in a VXIbus environment to allow its inclusion in larger distribution systems. The amplifier is described both in terms of its block diagram as well as particular details, and measured performance data are presented. Also described is a novel method of measuring interchannel isolation.

Introduction

A wideband distribution amplifier has been implemented in a single-input, six-output configuration for 50 ohm systems, covering a frequency range of 100 KHz to 10 MHz without tuning. The amplifier accepts a sinusoidal signal at +7 dBm to +19 dBm, levels it by means of a highly effective automatic gain control (AGC) circuit, and distributes it to identical, well-isolated channels having output levels of +13 dBm. The output signals have low harmonic distortion, low phase noise, and are isolated from each other by more than 100 dB. In addition, the distribution amplifier has characteristics that are very insensitive to temperature variations over the range of 0° C to +55° C.

Since knowledge of the operating status of the amplifier is essential for its use in a distribution system, the instrument status is indicated both by light-emitting-diode (LED) front panel indicators, and by having a status register read over a VXIbus interface.

Amplifier Block Diagram Description

Figure 1 is a block diagram of the analog portion of the distribution amplifier. Key to this design is an input AGC loop preceded by a variable attenuator. The attenuator is a resistive divider whose shunt element is a photoconductor having resistance controlled by feedback around the AGC loop. This photoconductor resistance is indicated by "R_{PH}" in the diagram. The attenuated input signal is amplified by a wideband, low-noise operational amplifier, whose output ($V_{leveled}$) drives the six output channels. A precision level detector measures the amplitude of the gain-controlled signal, and its DC output is summed with a reference DC voltage by an integrating amplifier stage. The integrator output controls the current flowing in a light-emitting diode. Since the LED is intimately packaged with the photoconductor, the current through the LED affects the photoconductor resistance in a manner that achieves AGC action.

Figure 2 shows the input stage of the AGC loop. $V_{leveled}$ equals the input attenuation $R_{PH} / (R_{PH} + R_1)$ multiplied by the gain of the non-inverting amplifier stage. If R_{PH} is much less than R_1 (which it is in this design), then $V_{leveled}$ can be held constant if changes in V_{IN} are balanced by inverse changes in R_{PH}.

Figure 3 is a typical resistance versus current curve for the optodiode/photoconductor in the AGC circuit. This curve has the desired inverse characteristic. On this graph are drawn two horizontal limit lines representing the required R_{PH} for the minimum and maximum specified signal input levels. Notice that large portions of the curve extend above and below the limit lines. This large dynamic range guarantees that the AGC loop will work satisfactorily for temperature changes, component aging, and for different photoconductor units.

The input operational amplifier has been chosen on the basis of several important considerations. First, it has to have a high enough bandwidth and slew rate to minimize waveform distortion at 10 MHz. It also has to have input noise less than the thermal noise associated with the input attenuator. Finally, the operational amplifier has to possess a high enough output capability to drive six output channels.

0-7803-0905-7/93 $3.00 © 1993 IEEE

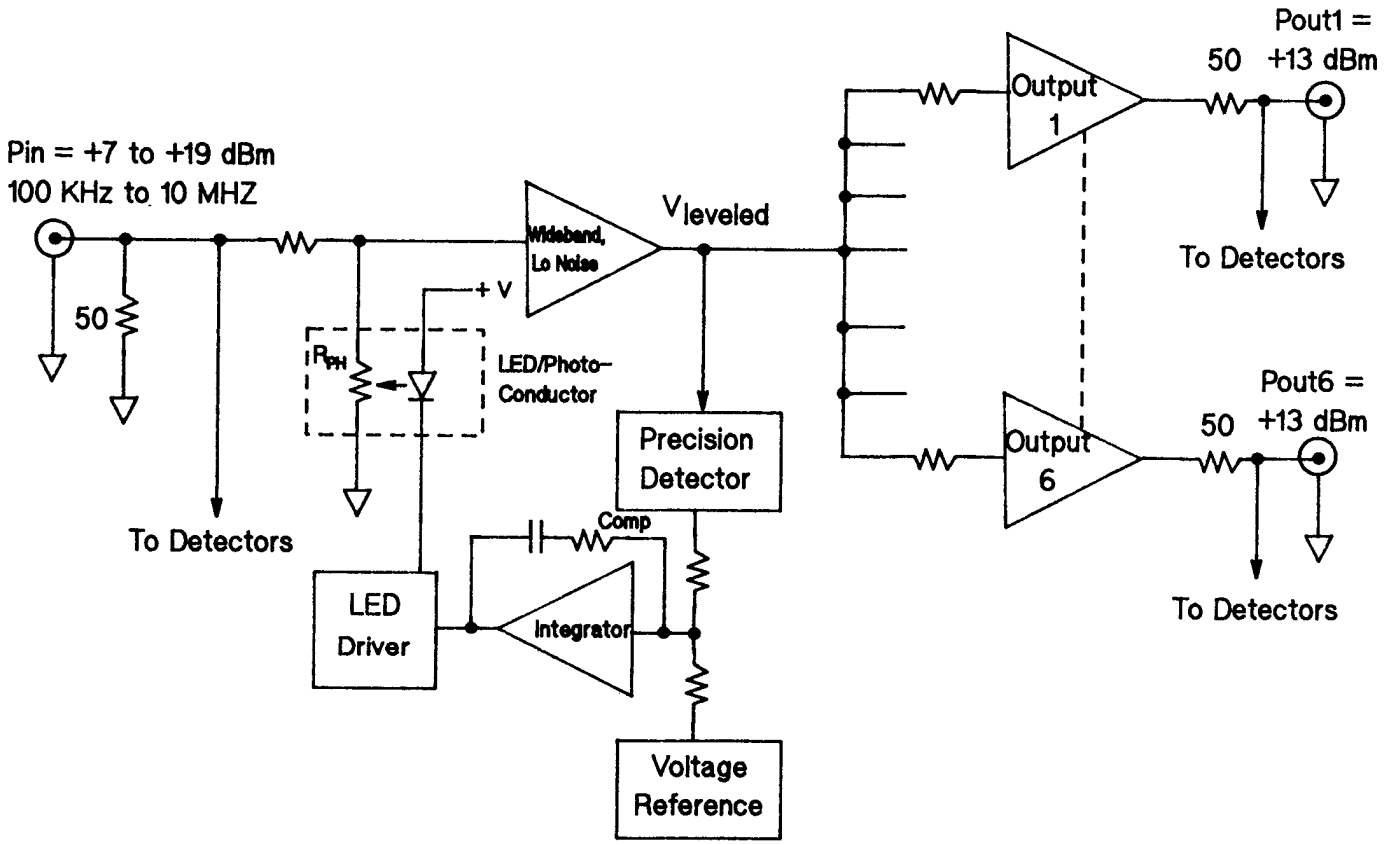
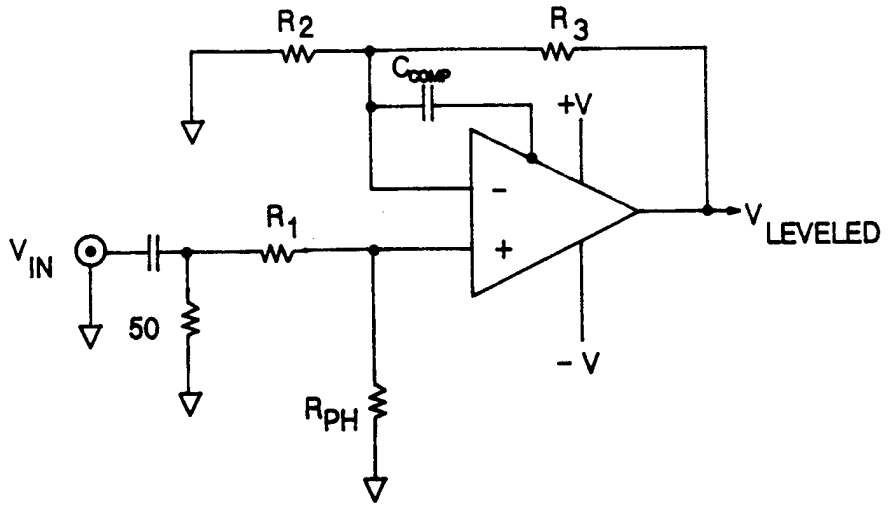


Figure 1. Block diagram of input attenuator, AGC loop and output channels



$$V_{LEVELED} = [R_{PH}/(R_{PH}+R_1)][1+R_3/R_2][V_{IN}]$$

Figure 2. Input stage of AGC loop

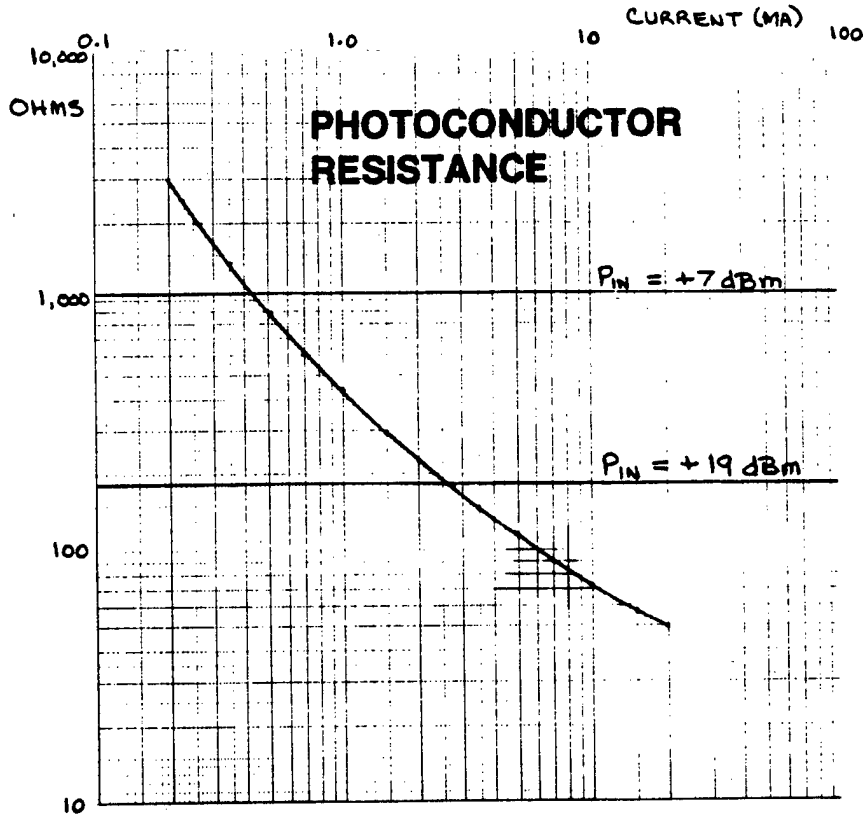


Figure 3. Resistance characteristic of typical LED/photoconductor element

The AGC loop can be viewed as a control system that has the DC voltage reference as its input and the leveled AC signal as its output. Because loop gain is quite high within the loop bandwidth, the wideband precision detector in the feedback path becomes the dominant transfer element. It is responsible for distribution amplifier properties such as flat frequency response and good amplitude stability versus temperature. Variations of P_{IN} are just disturbances in the forward path of the control system and are attenuated by the loop gain, resulting in automatic gain control of the signal. Loop dynamics are well-behaved, with the compensation resistor in the integrating amplifier stage chosen for critical damping at a small-signal loop bandwidth of 40 Hz. For large-signal transients, such as the sudden application of input power, the loop reaches equilibrium within 8 msec.

The output channels consist of IC operational amplifiers having wideband, low-noise, low-distortion characteristics. Output stage amplification is very stable, being set by gain resistors. The desired output impedance is achieved with 50 ohm series resistors.

Performance Data for the Distribution Amplifier

Measurement data will now be presented. Currently, very repeatable performance data has been obtained on twelve amplifier units. Performance has been verified under conditions of extreme temperature and humidity, as well as mechanical stress. The design appears to be a robust one.

The effectiveness of the AGC circuit was determined by measuring amplifier output power as input power was varied from below the specified minimum of +7 dBm to above the specified maximum of +19 dBm. P_{OUT} was flat at a typical value of +13.05 dBm for P_{IN} between -1 dBm and +22 dBm. (Higher input power was not applied.) For P_{IN} below -1 dBm the AGC was no longer effective, and the output followed the input.

The frequency response of the distribution amplifier is shown in Figure 4. From 100 KHz to 10 MHz the output amplitude was flat within ±0.25 dB, a variation well within the ±1 dB specification for output power level.

Harmonic distortion for the distribution amplifier was typically -60 dBc at frequencies at and below 1 MHz, increasing to -54 dBc at 5 MHz, and -48 dBc at 10 MHz. This increase was due to the signal slope becoming a larger fraction of the input operational amplifier slew rate. However, it is important to note that harmonic distortion is independent of input signal amplitude because the AGC maintains a constant signal amplitude at the operational amplifier input. Also, the photoconductor is a very linear component, with distortion more than 80 dB below the signal level.

The distribution amplifier has very stable characteristics with temperature. Over a temperature range from 0° C to +55° C the output amplitude changed 0.17 dB typically, compared with an instrument specification of 0.50 dB. The temperature coefficient of the input to output phase difference varied between 0.004 degrees/°C (or 1.1 psec/°C in time) at 10 MHz and 0.028 degrees/°C at 100 KHz, figures well within a specification of 0.050 degrees/°C.

The residual phase noise of the distribution amplifier is largely determined by the noise of the attenuator and amplifier. The amplifiers in the output channels contribute only slightly, since they are low-noise units (-150 dBc/Hz at 1 KHz) with noise effectively reduced by the gain (11 dB) of the input stage. The noise density of an operational amplifier referred to its input is given by the expression[1]

$$e_t = [e_n^2 + 4kTR_{eq} + (i_n R_{eq})^2]^{1/2} \text{ nV}/\sqrt{\text{Hz}}$$

where e_n, i_n = input voltage, current noise density
R_{eq} = total equivalent input resistance
k = Boltzmann's constant
T = temperature, degrees Kelvin

For the input resistors and operational amplifier used, the noise referred to the input is

$$e_t = [2.89 + 12.80 + 1.37]^{1/2} = 4.13 \text{ nV}/\sqrt{\text{Hz}}$$

In this case, the major contribution to e_t, at a sideband frequency of 1 KHz, is the thermal noise of the attenuator resistors and operational amplifier gain resistors. This dominant contribution shows up as the large middle term in the calculation. Since $V_{leveled}$ referred to the input is 0.189 vrms, the input noise power divided by the input signal power in a 1 Hz bandwidth, expressed in decibels, is

$$20 \log(4.13 \times 10^{-9} / 0.189) = -152 \text{ dBc/Hz}$$

However, as noise sidebands can be separated into equal phase and amplitude modulation components[2], only half of the noise power is attributable to phase noise. Therefore, 3 dB must be subtracted from the ratio.

Calculated phase noise (1 KHz) = -155 dBc/Hz

Figure 5 shows the phase noise of a distribution amplifier at a carrier frequency of 5 MHz. The phase noise was measured using a phase-quadrature arrangement at the input of a phase noise measurement system. The measured phase noise at a 1 KHz sideband was -155 dBc/Hz, which agrees well with the calculated value. The phase noise at sideband frequencies below 100 Hz was due

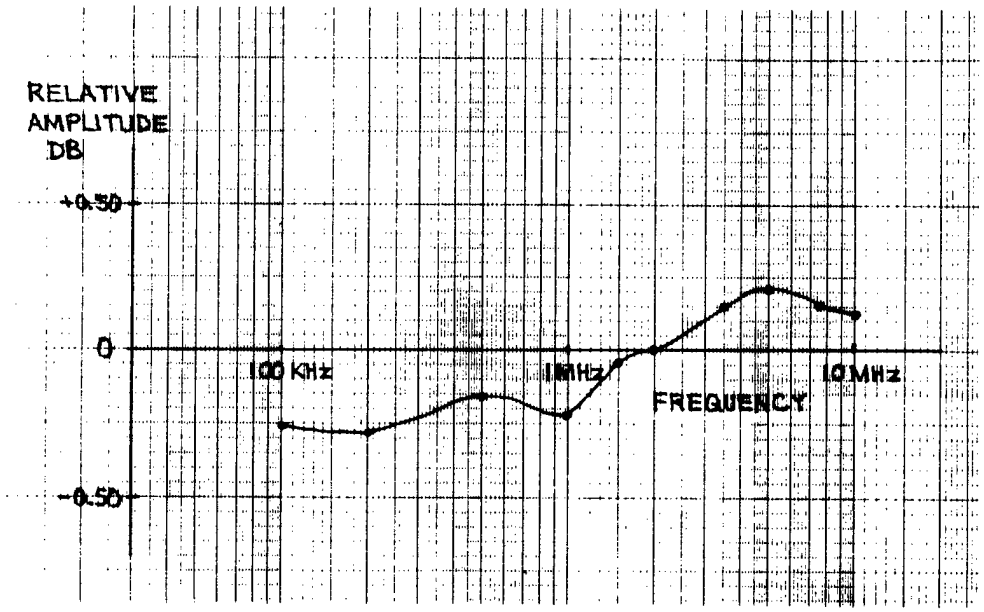


Figure 4. Output frequency response

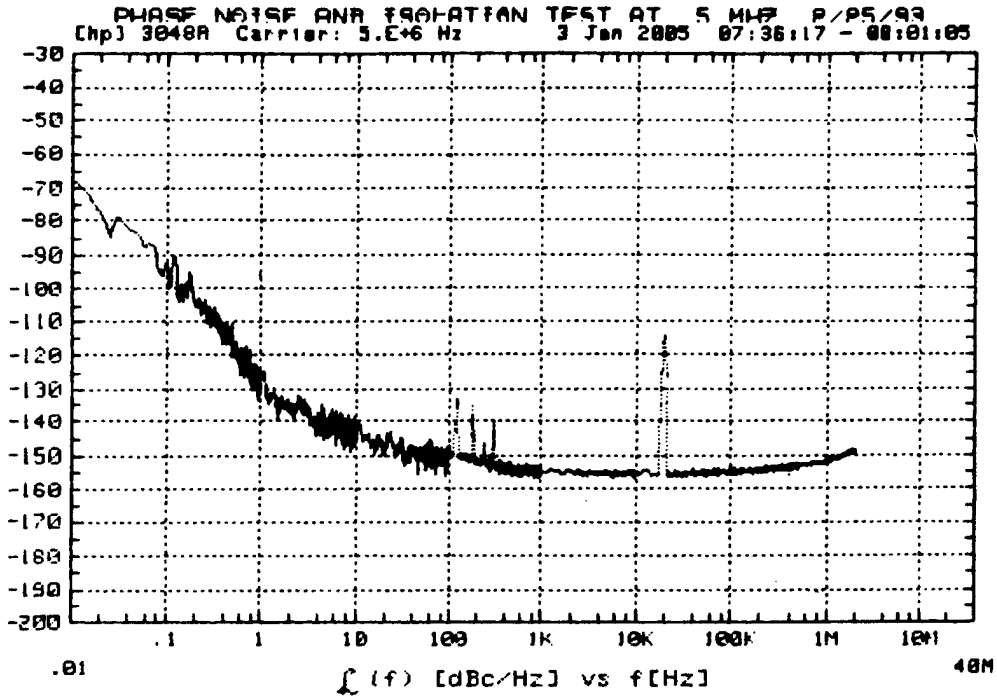


Figure 5. Phase noise for a 5 MHz signal

largely to the operational amplifier input voltage and current noise densities rather than R_{eq}. (The upturn in the noise plot at Fourier frequencies above 100 KHz is due to decorrelation of the signal source noise in the measurement setup. It is not an indication of distribution amplifier phase noise.)

Phase noise was also measured at a carrier frequency of 10 MHz. The results averaged around -151 dBc/Hz, an apparent increase of 4 dB over the 5 MHz case, but still significantly below the instrument specification of -145 dBc/Hz at a 1 KHz sideband frequency. This difference may have been measurement system related.

A very important property of a distribution amplifier is that the outputs be well isolated electrically from each other. A disturbance on one output caused by switching or removal of equipment, for instance, should have extremely little effect on the signal appearing at any other output. The distribution amplifier has been designed for a minimum of 100 dB of isolation between outputs.

Several steps have been taken to insure this degree of isolation. First of all, the output amplifiers have been chosen to have good reverse isolation at 10 MHz. In addition, isolation resistors at the front of each output channel further attenuate signals coming back through the output stages. Ground plane shielding has been carefully placed on the printed circuit board between output amplifiers, and also inside the board between output transmission lines. Finally, the printed circuit board assembly is placed inside a shielded enclosure.

Since isolation is often specified in terms of the effect that applying an open or short circuit to one output has on another, a test technique has been adopted that utilizes this kind of load disturbance. This technique was developed by one of our colleagues at Hewlett-Packard, Richard Karlquist[3].

As shown in Figure 6, a signal can be reflected back into any output channel by switching the load between a matched 50 ohms and either an open or a short circuit. The degree of isolation between this port (Output 1) and any target port (Output 2) determines the amount of phase and amplitude disturbance appearing on the latter. If switching of loads is done in a periodic fashion, then Output 2 is amplitude and phase modulated. This effect can be maximized by varying the electrical distance to the switched load with a line stretcher.

A practical switched load can be set up by toggling a PIN diode at an audio rate between open and short circuit conditions. The phase modulation disturbances at Output 2 can be measured with a phase noise measurement system. Using a two step procedure even eliminates the line stretcher. First, sideband levels are observed with the switched PIN load connected directly to Output 1. Next, the PIN load is moved one-eighth of a wavelength away so that the output is now switched between purely capacitive and inductive loads, and sideband readings are taken again. The two sideband measurements are then root-mean-square summed to give a worst-case value. Finally, the interchannel isolation is obtained by subtracting a 2 dB correction from this sum. This correction accounts for the following factors: the sideband amplitude is only half of the peak phase disturbance (add 6 dB); the sinusoidal fundamental component is 2 dB greater than the peak value of the square wave modulation (subtract 2 dB); and the load is being switched between total mismatches of opposite polarity (subtract 6 dB).

Figure 7 is a phase noise plot of the modulation sidebands with the PIN diode connected directly to Output 1. The carrier frequency in this case is 10 MHz, and the switching rate is 2 KHz. Superimposed on the phase noise curve are harmonics of 2 KHz with a fundamental amplitude of -110 dBc. Figure 8 shows the phase modulation sidebands with the PIN diode moved an eighth wavelength away. The magnitude of the 2 KHz fundamental is -123 dBc. Adding the two sidebands in an rms manner results in a net amplitude of -110 dBc. When the 2 dB correction factor is applied, the interchannel isolation is determined to be 112 dB. Isolation measurements on twelve distribution amplifiers have given results in the range of 105 to 116 dB.

VXI Interface and Status Indicators

The distribution amplifier was built on a VXI (VMEbus extensions for Instrumentation) platform. VXI is a well defined, modular system based on the VMEbus[4]. Since the VXI hardware and software interfaces are well defined, this platform allows the distribution amplifier to be easily integrated into the user's environment. With the wide use of VXI and VME, the amplifier is ready for use in many metrology, communications, and other field applications.

A typical configuration for the distribution amplifier is in distributing the outputs of a primary frequency source (e.g. HP 5071A cesium standard) within a site. In this type of system the VXI mainframe would contain a command module that monitors the system status and flags the user if there are any problems. In particular, monitoring the status of the signal health is important in a distribution system. If a cable is disconnected or an output is burned out, then the system should be able to locate the fault.

Figure 9 is a block diagram of the status indicators, fault logic, and VXI interface. The instrument input signal and each of the output signals are detected to produce DC

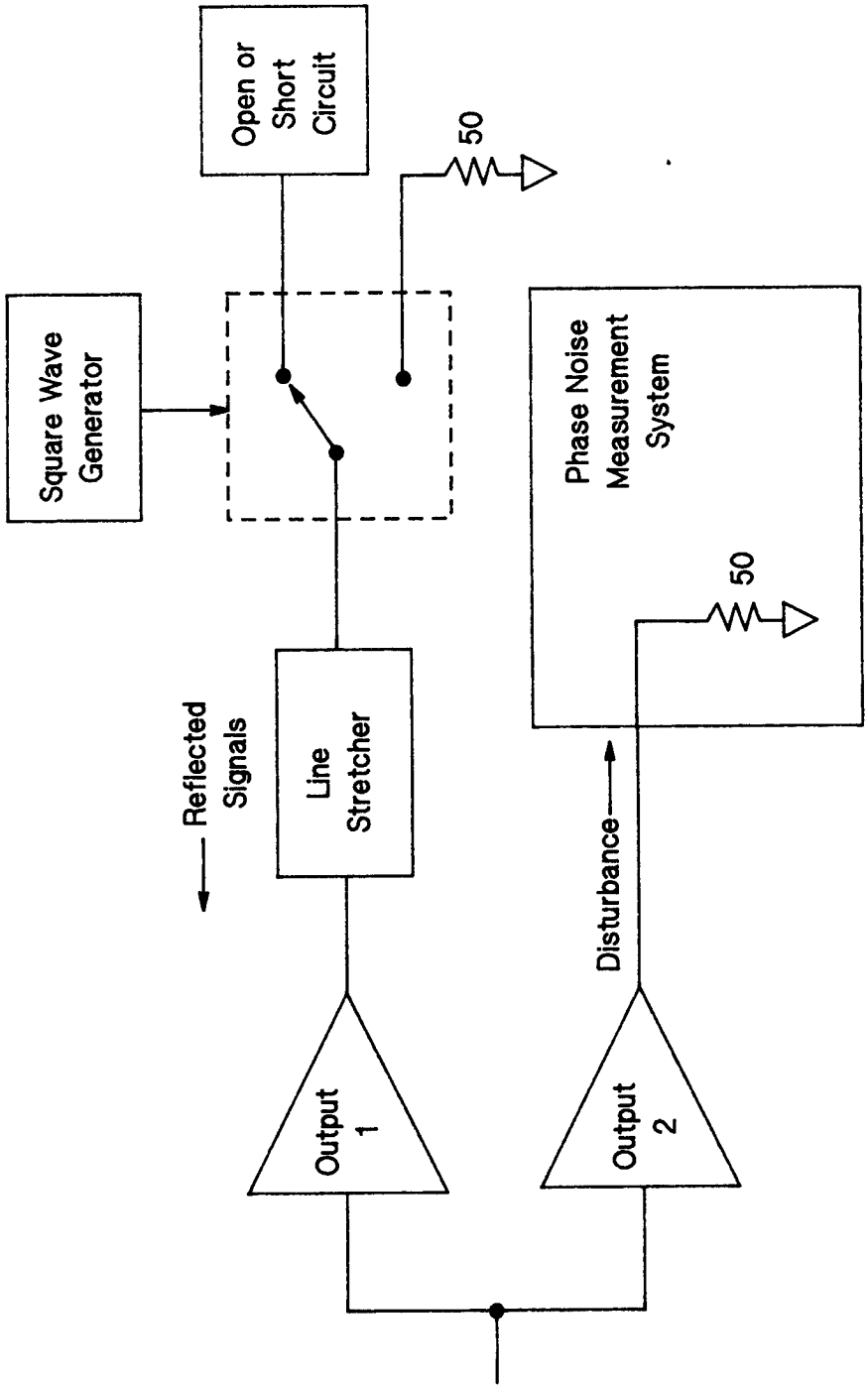


Figure 6. Measurement of isolation between channels

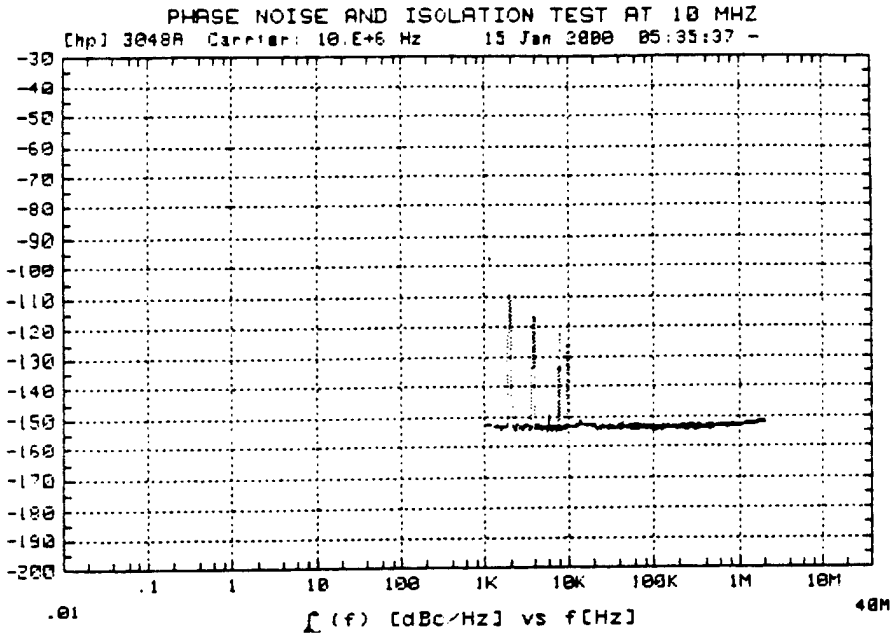


Figure 7. Isolation sidebands for a 10 MHz signal.
PIN diode load at Output 1.

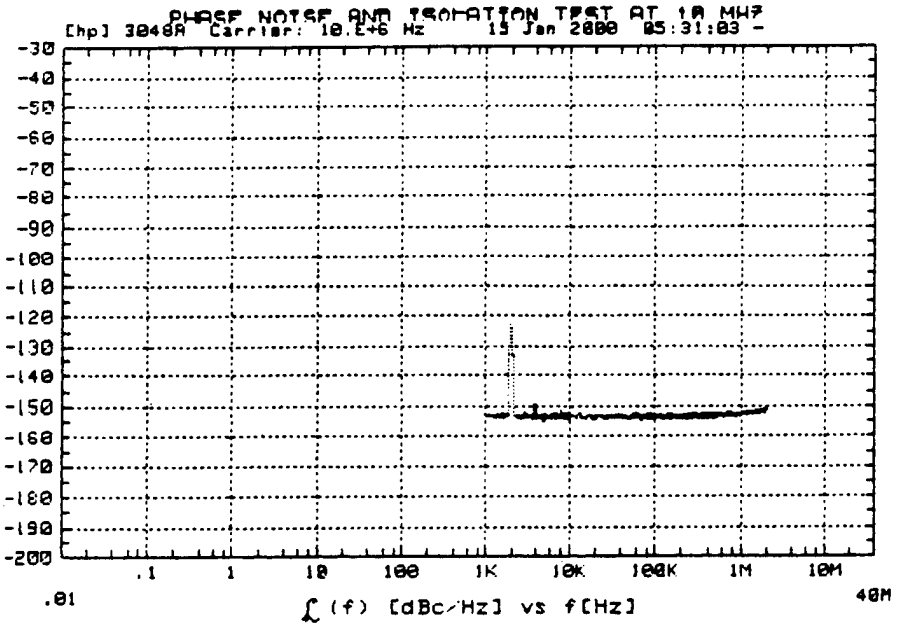


Figure 8. Isolation sidebands for a 10 Mhz signal.
PIN diode load λ/8 away from Output 1.

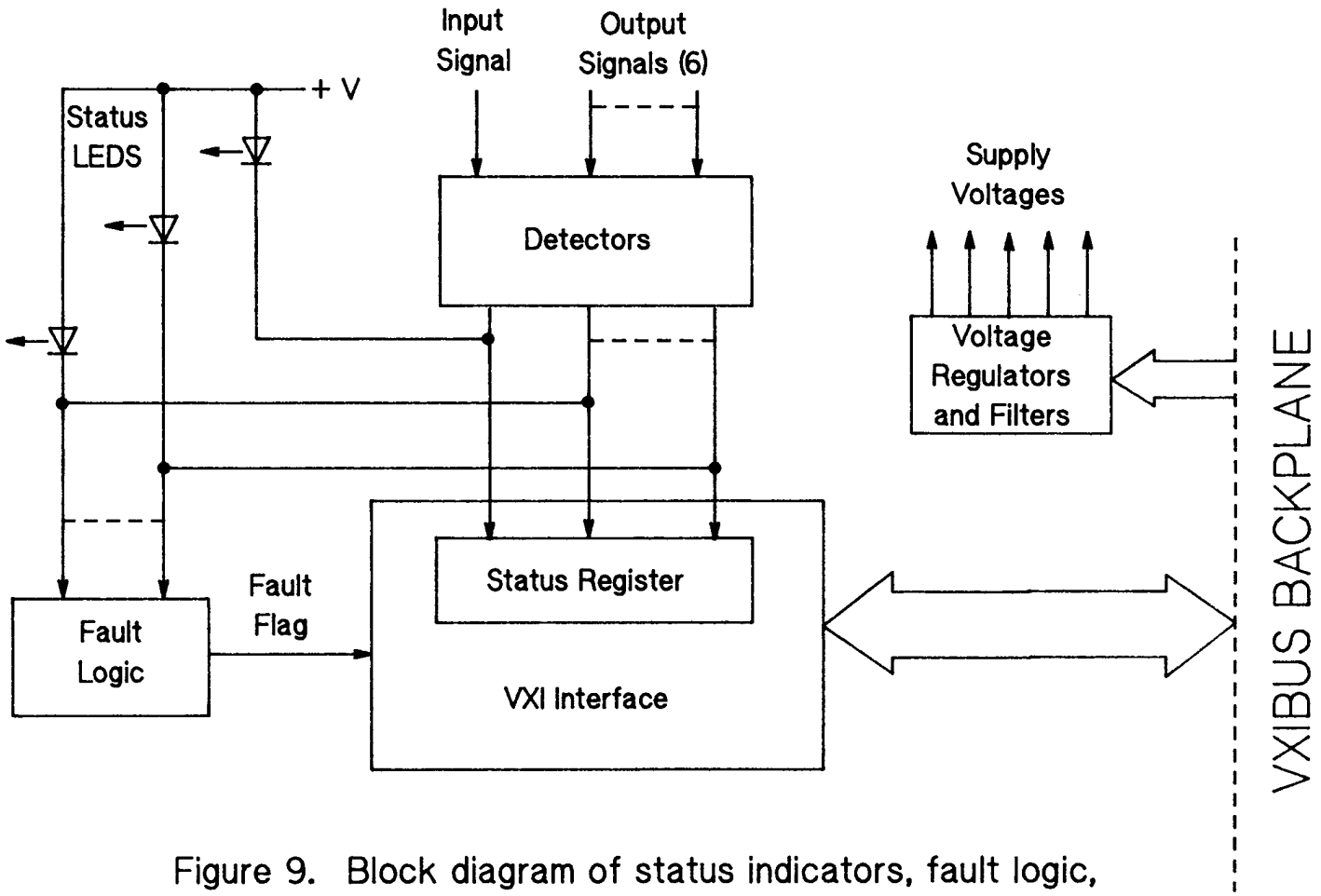


Figure 9. Block diagram of status indicators, fault logic, VXI interface, and voltage regulators

voltages that turn on the front panel status LEDs and also write data into a status register in the VXI interface, which is used mainly for signal status determination. The input and each output port of the distribution amplifier is given a "bit" in the VXI status register, so that the absence of a signal can be detected over the VXI backplane by reading this register. Polling the status register is the simplest method of monitoring the signal status, but a more efficient method is to use interrupts. Using interrupts is particularly effective when the VXI mainframe has many modules, and the command module is busy performing other tasks. When a signal is lost at the input or any of the outputs (e.g. someone shorts one of the outputs), a fault flag is sent to the VXI interface, which in turn transmits an interrupt request to the command module via the VXI backplane. The VXI command module can then service that interrupt and locate the fault.

Conclusion

A wideband distribution amplifier has been developed that the authors believe has many excellent performance characteristics. It combines good automatic gain control with low distortion, low phase noise, high interchannel isolation, and insensitivity to environmental changes. Ease of use is a distinguishing feature, as no tuning or operator adjustments are required. In addition, it readily lends itself to system applications because of its modular VXIbus implementation.

References

[1] C. D. Motchenbacher and F. C. Fitchen, Low-Noise Electronic Design. New York: John Wiley & Sons, 1973, ch. 2, pp. 30-31.

[2] William F. Egan, Frequency Synthesis by Phase Lock. New York: John Wiley & Sons, 1981, ch. 4, pp. 65-67.

[3] Richard Karlquist, "Isolation Specification Definition and Measurement." Hewlett-Packard Company unpublished technical article, April 2, 1992.

[4] VXIbus Consortium, Inc., VMEbus Extensions for Instrumentation System Specification, Revision 1.3. Vancouver, WA: VXIbus Consortium, Inc., July 14, 1989.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SPUR REDUCTION TECHNIQUES IN DIRECT DIGITAL SYNTHESIZERS

Victor S. Reinhardt
Hughes Space and Communications Company
P. O. Box 92919, Los Angeles, California 90009

Abstract

This paper reviews spur reduction techniques used in direct digital synthesizers (DDSs) or numerically controlled oscillators (NCOs). First, the classification and operation of conventional DDSs are reviewed. Covered are the pulse output DDS, sine output DDS, fractional divider, and phase interpolation DDS. It is shown that DDSs produce spurs as well as the desired output frequency due to the aliasing of harmonic imperfections in the generated waveform. Next, spur reduction techniques which reduce spurs by destroying the coherence of the aliasing process are discussed. Architectures described are the spurless fractional divider, the Wheatley jitter injection DDS, the randomized DAC DDS, and the nonuniform clock DDS. The spur reduction and phase jitter properties of each architecture are also discussed.

Introduction/Review of DDSs

Direct digital synthesizers (DDSs) or numerically controlled oscillators (NCOs) directly synthesize frequency waveforms from a fixed reference frequency using digital waveform generation techniques. DDSs can be classified as the pulse output DDS, sine output DDS, fractional divider, phase interpolation DDS, and other minor DDS variations [1]. The major classifications are summarized below. For more detailed descriptions, see References 1, 2, 3, 4, 5, and 6.

Pulse Output DDS

Figure 1 shows a block diagram of the pulse output DDS and outlines its operation. The pulse output DDS merely consists of an N-bit accumulator set up to add a frequency word K every clock cycle t_c or

$$R_{n+1} = \text{Mod}(R_n + K, 2^N) \quad (1)$$

where R_n is the N-bit register value in the accumulator after the nth clock cycle. One can take the frequency output f_o from the carry output as a pulse or from the most significant bit (MSB) of the accumulator as a "square" wave. From the figure, one can see that on average

$$f_o = F_o f_c \quad (2)$$

where F_o the fractional frequency is

$$F_o = \frac{K}{2^N} \quad (3)$$

and where the clock frequency f_c is $1/t_c$.

The main limitation of the pulse output DDS is the high degree of time or phase jitter it exhibits. One can see from Figure 1, that the time error between the actual zero crossings of the output and that of an ideal frequency source varies approximately uniformly from 0 to t_c. This produces a time jitter of

$$\sigma_t \cong \frac{t_c}{2\sqrt{3}} \quad (4)$$

Let us define a fractional register value

$$r_{n+1} = r_n + F_o \quad (5)$$

which is related to the accumulator register word R_n by

$$\text{Fract}(r_n) = \frac{R_n}{2^N} \quad (6)$$

One can see from Figure 1 that, at sample time nt_c, r_n is equal to the cycle count elapsed in an ideal oscillator with frequency f_o. Thus the elapsed phase is

$$\phi = 2\pi r_n \quad (7)$$

0-7803-0905-7/93 $3.00 © 1993 IEEE

and the value of R_n at a carry is proportional to the phase error of the output pulse. (This will be important in other DDS designs.) Finally, from (4) and Figure 1, one can see that the phase jitter of the unfiltered output is given by

$$\sigma_\phi \cong \frac{\pi F_o}{\sqrt{3}} \tag{8}$$

The filtered jitter is a completely different matter. Figure 2 shows a typical plot of the spectrum from a pulse output DDS. Notice the dense collection of large spurs around the carrier due to aliasing. The character and position of these spurs change drastically as f_0 is changed. Because large aliasing spurs, which contain large amounts of jitter energy, can appear very close to the carrier at certain frequencies, one cannot reliably improve the phase jitter of a pulse output DDS by filtering. This problem is corrected by the spur reduction techniques discussed later.

Sine Output DDS

The sine output DDS reduces the phase jitter of the pulse output DDS by adding a sine look-up table and digital to analog converter (DAC or D/A converter) to the N-bit accumulator as shown in Figure 3. In this DDS, The N-bit accumulator register provides a W-bit phase word for a J-bit sine look-up table, whose output is converted to an analog voltage by an M-bit DAC. An example of the resultant stepped DDS output is also shown in Figure 3.

If the sine output voltage is perfect, using the sampling theorem [7], one can show that a pure sine wave at f_0 will be recovered when the output is low-pass filtered with a cut-off frequency below $f_C/2$. When the look-up table and DAC have finite resolution, spurs will be generated [1,3,9,10]. Generally, DAC quantization is the limiting factor in sine output DDSs. A rough order of magnitude for the size of spur levels (power) is 2^{-2M}, but other analysis has shown that the dependence on DAC resolution M is faster than this when the sine table has a much higher resolution than the DAC [9,10]. Figure 4 shows typical sine output DDS frequency spectrums with 5-bit and 11-bit DACs.

Fractional Divider and Phase Interpolation DDS

The fractional divider or pulse swallower shown in Figure 5 is similar in operation to the pulse output DDS. In the fractional divider, a programmable divide by n/n+1 counter normally divides clock cycles by n. At each divide output, an N-bit accumulator is clocked to add the frequency word K to itself. Whenever there is a carry, the next divide cycle is then set for divide by n+1. One can show that this produces an output whose average frequency is given by [1]

$$f_o = \frac{f_c}{n + F_o} \tag{9}$$

The fractional divider, like the pulse output DDS, produces a high degree of jitter. The phase interpolation DDS shown in Figure 6, like the sine output DDS, reduces this jitter. As shown in the figure, the fractional divider is used in a phase lock loop. This inverts the fractional division process, so the output is

$$f_o = (n + F_o) f_c \tag{10}$$

Since the register value R_n is proportional to the phase error of the fractional divider output [1], its value is used to provide a phase correction to the loop through a DAC. Like the sine output DDS, the spurs are roughly given by 2^{-2M} for an M-bit DAC. The main disadvantage of the phase interpolation DDS is that M-bit linearity and accuracy are also required of the phase detector.

Theory of Conventional DDS Spurs

To simplify the derivations that follow, let us use normalized time and frequency units given by

$$F = f/f_C = f\, t_C \tag{11}$$

$$T = t/t_C = t\, f_C \tag{12}$$

Figure 7 shows the DDS model we will use. The first block in the model is a look-up table v(r). This look-up table uses the normalized register value r to generate a voltage v. An important property of v(r) is that it is periodic in r with a period of one; that is, the look-up table only utilizes the fractional part of r to generate voltage values.

The next block in our model is an accumulator which generates a sequence of normalized register values r_n and samples of v(r) given in the time domain by

$$v_s(T) = \sum_n v(r_n)\delta(T-n) \quad (12)$$

The final block in our model is a hold function given by

$$h(t) = \begin{cases} 1 & \text{if } -1/2<T<1/2 \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

which generates the stepped DDS output given by

$$v_h(T) = \int h(T-T')\, v_s(T')\, dT' \quad (14)$$

This model can be applied to both pulse output and sine output DDSs. For the pulse output DDS, v(r) is a square wave, and for the sine output DDS, v(r) is a quantized sine wave. The fractional divider and phase interpolation DDSs, can also be understood by suitably modifying the model.

Fourier Transform of DDS Output.

The fourier transform of the DDS output $V_h(F)$ is then given in terms of the fourier transform of the sampled signal $V_s(F)$ and the fourier transform of the hold function H(F)

$$V_h(F) = V_s(F)\cdot H(F) \quad (15)$$

where

$$V_h(F) = \int_{-\infty}^{\infty} v_h(T) e^{-j2\pi FT} dT \quad (16)$$

$$H(F) = \mathrm{Sinc}(\pi F) \quad (17)$$

and

$$V_s(F) = \sum_n v(r_n) e^{-j2\pi nF} \quad (18)$$

Because of the normalized units, note that all fourier transforms and spectral densities are per f_c Hertz. Thus to obtain per Hz formulas, one must divide frequency domain formulas in this paper by f_c and substitute f_o/f_c and f/f_c for F_o and F respectively.

Since v(r) is periodic with an r-period of one, we can expand v(r) as a fourier series [10]

$$v(r) = \sum_m a_m e^{j2\pi mr} \quad (19)$$

Thus (18) becomes

$$V_s(F) = \sum_{m,n} a_m e^{j2\pi(mr_n - nF)} \quad (20)$$

Spurs in a Conventional DDS

In a conventional DDS

$$r_n = F_o n \quad (21)$$

so from (20), V_s is given by

$$V_s(F) = \sum_{m,n} a_m e^{j2\pi F_m n} \quad (22)$$

where

$$F_m = mF_o - F \quad (23)$$

Utilizing the well known aliasing relationship for sampled signals [1,7]

$$\sum_{n=-\infty}^{\infty} e^{j2\pi nF} = \sum_{m'=-\infty}^{\infty} \delta(F-m') \quad (24)$$

we obtain

$$V_s(F) = \sum_m a_m \sum_{m'(m)} \delta(F - F_{m,m'}) \quad (25)$$

where

$$F_{m,m'} = mF_o - m' \quad (26)$$

Thus one obtains the well known fact that sampling produces aliasing [7]; that is, the harmonics of $v(F_o t)$ at $f=mf_o$ are aliased down to spurs by multiples of the clock frequency

$$f_{m,m'} = mf_o - m'f_c \quad (27)$$

The frequencies of DDS spurs can be predicted using (27). It can be shown [1] that the spur frequencies $F_{m,m'}$ are given by a permutation of 1/b, 2/b, ..., (b-1)/b where F_0 is written as the simplified fraction a/b (a and b relatively prime to each other). The size of these spurs can also be predicted by calculating the a_m. For a pulse output DDS, the a_m are the harmonics of a square wave [3,7]. For a sine output DDS, the evaluations of a_m for the quantized sine wave have also been published using analytical techniques [8] and computer simulations [9].

Spur Reduction Techniques

As described in the previous section, spurs occur in a DDS because of aliasing from the uniformly stepped periodic (period = $N't_C$) sequences generated. In the following sections we will discuss methods of reducing these spurs by destroying the periodicity of these sequences. Table 1 summarizes spurs reduction techniques in the literature. The last technique by Nicholas and Samueli [16] does not involve destroying the periodicity of the DDS steps and will not be discussed.

Table 1. Spur reduction techniques.

Technique	Refer-ence	Method
Spurless Fractional Divider	[11]	Totally Random Output
Wheatley Jitter Injection	[12] [13]	Word Jitter Injection
Randomized DAC DDS	[14]	Word Jitter Injection
Nonuniform Clock DDS	[15]	Nonuniform Sampling
Nicholas & Samueli DDS	[16]	Force K to Be Odd

Spurless Fractional Divider

Figure 8 shows a block diagram of a spurless fractional divider [11]. This consists of a programmable divide by n/n+1 counter, an N-bit random or pseudorandom number generator, and an N-bit word comparator. At every output of the divider, the random number generator produces a new random word P_n, and the comparator compares this word with the frequency word K. If $P_n<K$, the counter is set to divide by n+1 for the next cycle. On average, the frequency will be given by (9) just as for the conventional fractional divider, but the sequence of divide by n's and n+1's will be totally random, so no spurs will be generated. However, this process generates frequency jitter, so the spurless fractional divider has a $1/f^2$ phase noise spectrum near the carrier.

Wheatley Jitter Injection DDS

Figure 9 shows one form of the Wheatley jitter injection DDS [12,13]. In this variation on a pulse output DDS, a random word k_n, which can vary randomly from 0 to K-1, is generated each clock cycle and added to the accumulator register value R_n. The sequence of carries from this addition then becomes the pulse output. In the original Wheatley circuit, this output is also divided by two to produce a "square" wave. Figure 10 shows the spectrum of a pulse output DDS with and without Wheatley jitter injection. Notice that the Wheatley DDS quite effectively removes most spurs, but trades these spurs for a high degree of broadband phase noise S_ϕ. In essence, the phase jitter of the pulse output DDS is still there, but now smeared out into broadband noise. This broadband noise is more easily filtered out than spurs; now the filtered phase jitter is just $\sqrt{S\phi\, B}$ where B is the filter noise bandwidth (B assumed to be small).

One can show that the spur power at $F_{m,m'}$ in the Wheatley jitter injection DDS is smaller than that in a pulse output DDS by a factor of [1]

$$|M|^2 = \mathrm{Sinc}^2(\pi m F_0) \qquad (28)$$

Notice that only m, the order of the original harmonic expansion of v(r), matters in reducing spurs, not the final frequency $F_{m,m'}$. Thus spurs which come from low order harmonics are not completely eliminated by the jitter injection process.

Several formulas have been published for the spectral density of the noise floor [1,12,13]. These have the same dependence on the basic parameters f_0 and f_C, but have different coefficients.

Randomized DAC DDS

The comparable jitter injection technique for sine output and phase interpolation DDSs is the randomized DAC DDS [14]. Figure 11 shows its sine output embodiment. Here a random word from 0 to $2^{J-M}-1$ is added to the J bit output of the sine look-up table before the sum is truncated to M bits for the DAC.

Figure 12 shows the spectrum of a randomized DAC DDS with a 5-bit DAC and those of a conventional sine output DDS with 5-bit and 11-bit DACs. Comparing spur levels, one can see that the randomized DAC DDS levels are at least 10 dB lower than those of the conventional DDS with the same resolution DAC, and has spur levels comparable to those of the conventional DDS with an 11-bit DAC. Especially note that the persistent spur nearest carrier in both sine output DDS plots is gone in the Randomized DAC DDS plot.

A final note is that the f_o in Figure 12 was chosen to provide a worst case example. Since f_o is approximately 1/3 f_c, many of the spurs near f_o are aliases of low order harmonics (m = ±2, ±4, etc.). These are are difficult to remove with jitter injection. At other values of f_o, the experimental results were even better [14].

Nonuniform Clock DDS

Another way to reduce the periodicity which produces spurs is to randomly vary the time interval between DDS steps. Figure 13 shows a DDS which accomplishes this by stepping a sine output nonuniformly in time [15]. To achieve this nonuniform stepping, a pseudorandom clock generator produces a sequence of pseudorandomly varying time intervals Δt_n and generates alternate odd and even clock pulses based on this sequence. A ΔR calculator next generates an accumulator increment

$$\Delta R_n = K\Delta t_n \qquad (29)$$

and an accumulator adds this to itself to produce a sequence of register values R_n. These R_n values are then utilized by a sine table to produce a sequence of sine values. At this point the sequence is broken up into odd and even values. These are stored in a buffer and sent individually to odd and even DACs. Finally, the DACs are clocked by the odd and even clock pulses and a Ping-Pong switch combines both DAC outputs into a single stepped output.

The separation into odd and even DACs is required to give the DACs sufficient time to settle. To eliminate spurs, the sequence Δt_n must uniformly vary from 0 to $2t_c$ [15], so some values of Δt_n will be virtually zero. The odd and even separation guarantees that each DAC will have a time of at least t_c to settle.

At first glance, this nonuniform clock method should produce no broadband noise because the exact phase increment for the sine table is properly calculated. However, the method does produce a broadband noise floor [15].

Conclusions

The randomization techniques described are an effective way of reducing spurs with minimum added complexity. Of the methods discussed, the jitter injection techniques--the Wheatley technique for the pulse output DDS and the randomized DAC DDS for the sine output and phase interpolation DDS--offer the best combination of manageable added phase noise and minimum added complexity. These techniques are also useful in reducing spurs in the face of technological component limitations such as limited DAC resolution and excessive settling time at high speeds.

Acknowledgment

The author acknowledges J. Damir, A. Strodtbeck, and M. Fashano of Hughes Space and Communications Company for generating simulations of the Wheatley jitter injection DDS.

References

[1] V. S. Reinhardt, "Direct Digital Synthesizers," Proceedings of the 17th Annual Precise Time and Time Interval Planning Meeting (NASA/DOS). Washington, D. C., December 3-5, 1985.

[2] J. Gorsky-Popiel (ed), Frequency Synthesis and Applications, IEEE Press 1975.

[3] J. Noordanus, "Frequency Synthesizers--A Survey of Techniques," IEEE Trans. Comm. com-17, #2, April, 1969.

[4] V. S. Reinhardt, K. V. Gould, and K. M. McNab, "A short Survey of Frequency Synthesizer Techniques

[5] V. S. Reinhardt, "Frequency Synthesizer Basics," Tutorial for the 1993 IEEE International Frequency Control Symposium, to be published.

[6] U. L. Rhode, Digital Frequency Synthesizers: Theory and Design, Prentice-Hall, 1963.

[7] A. V. Oppenheim, Digital Signal Processing, Section 1.7, Prentice-Hall, 1975.

[8] J. Garvey and D. Babitch, "An Exact Spectral Analysis of a Number Controlled Oscillator Based Synthesizer," 44th Frequency Control Symposium, 1990.

[9] R. E. Lundgren, V. S. Reinhardt, and K. W. Martin, "Designs and Architectures for EW/Communications Direct Digital Synthesizers," Research and Development Technical Report, SLCET-TR-0424-F, U. S. Army Laboratory Command, August, 1987.

[10] S. M. Selby, ed., CRC Standard Mathematical Tables, 22nd Edition, CRC Press, 1974.

[11] V. S. Reinhardt and I. Shahriary, "Spurless Fractional Divider Direct Digital Synthesizer and Method," U. S. Patent 4,815,018, March 21, 1989.

[12] C. E. Wheatley, III and D. E. Phillips, "Spurious Suppression in Direct Digital Synthesizers," Proceedings of the 35th Frequency Control Symposium, May, 1981.

[13] C. E. Wheatley III and D. E. Phillips, "Digital Frequency Synthesizer with Random Jittering for Reducing Discrete Spectral Spurs," U. S. Patent 4,410,954, October 18, 1983.

[14] V. S. Reinhardt, K. V. Gould, and K. M. McNab, "Randomized Digital/Analog Converter Direct Digital Synthesizer," U. S. Patent 5,014,231, May 7, 1991.

[15] V. S. Reinhardt, "Method and Apparatus for Reduced Aliasing in Signal Processing," U. S. Patent 4,890,248, December 26, 1989.

[16] H. T. Nicholas III and H. Samueli, "An Analysis of the Output of Direct Digital Frequency Synthesizers in the Presence of Finite Word Length Effects, 42nd Frequency Control Symposium, 1987.

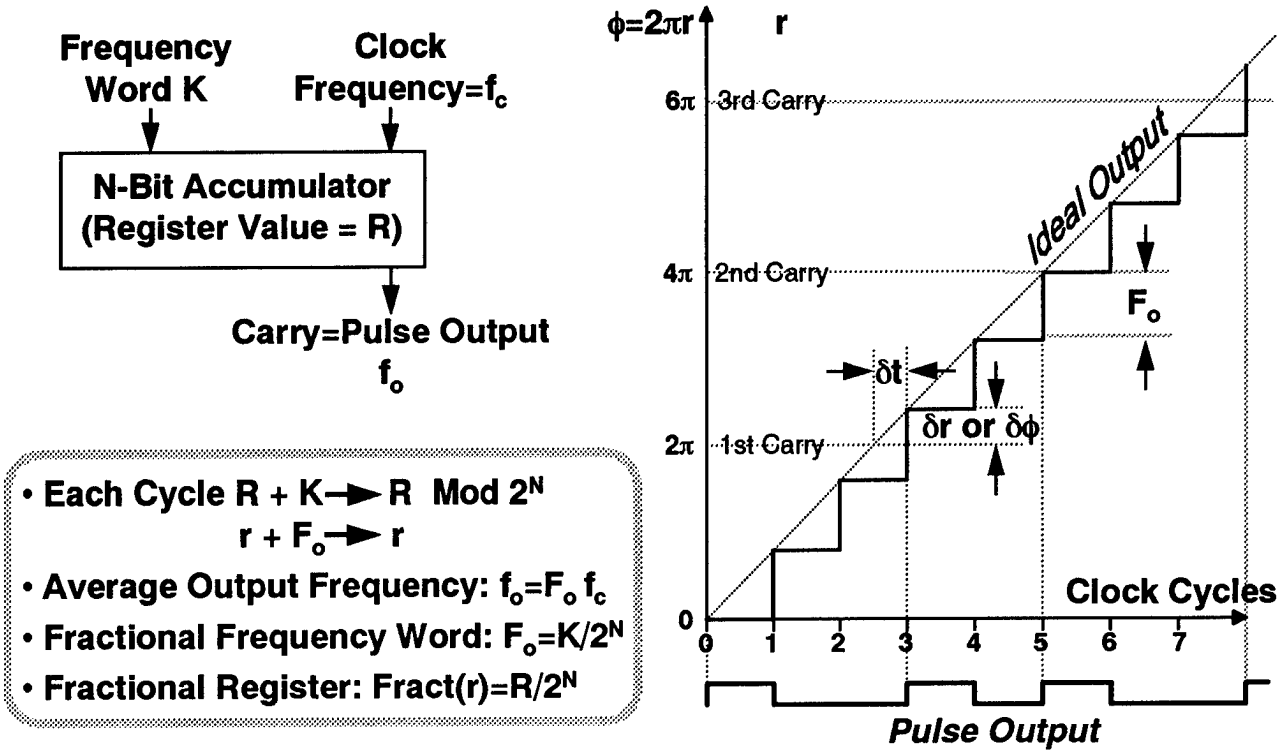


Figure 1. Pulse output DDS block diagram and operation.

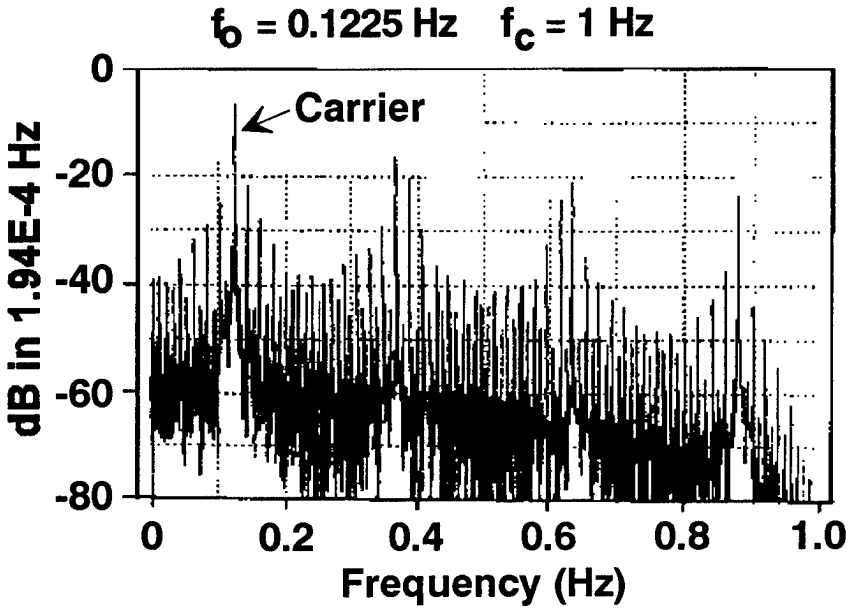


Figure 2. Typical pulse output DDS frequency spectrum.

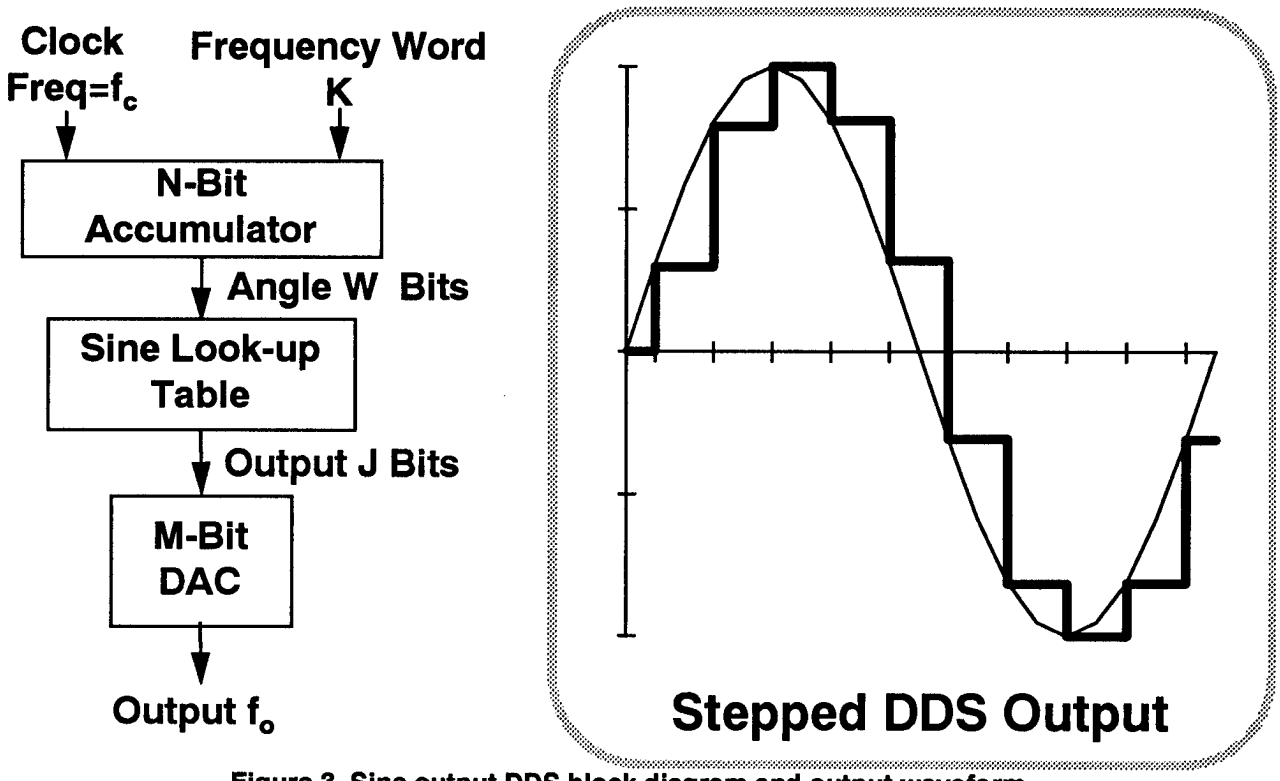


Figure 3. Sine output DDS block diagram and output waveform.

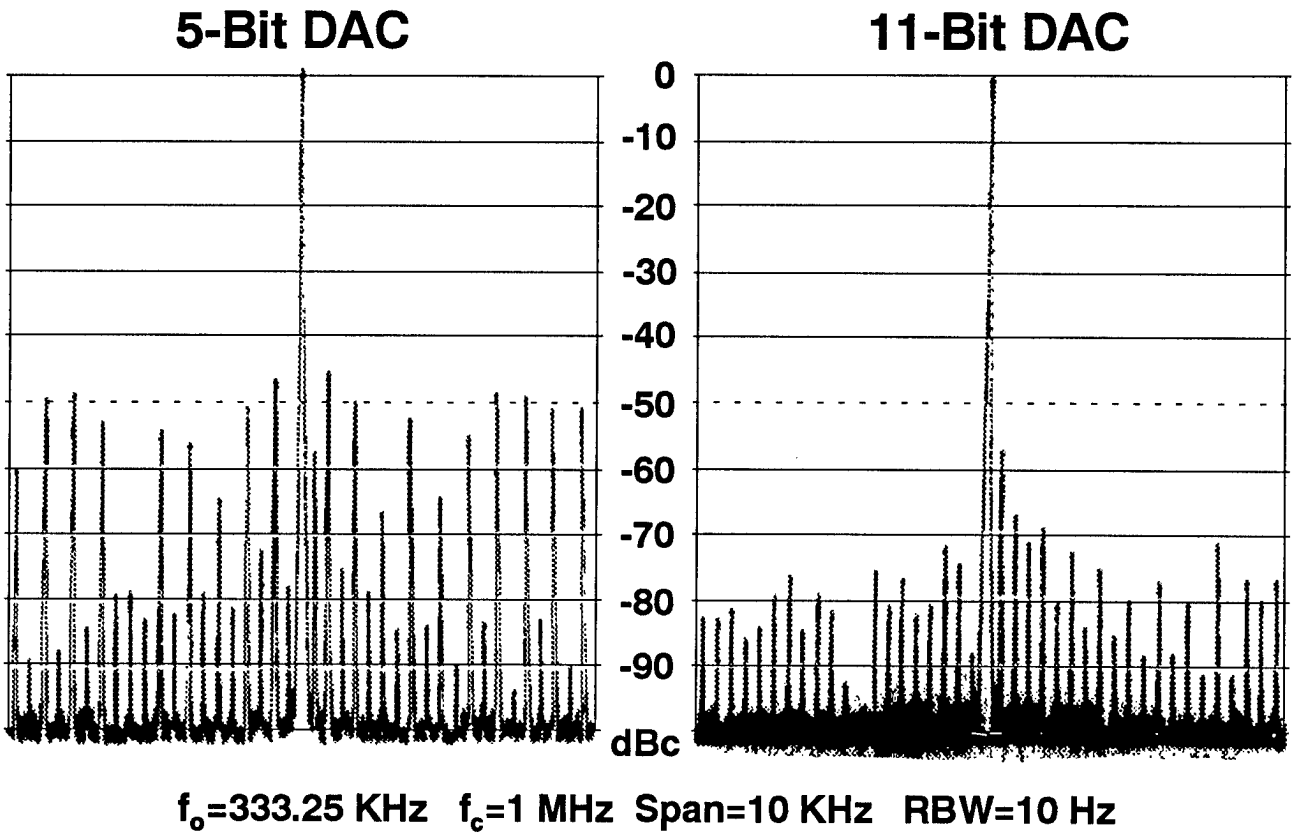


Figure 4. Typical sine output DDS frequency spectrum for 5-bit and 11-bit DACs

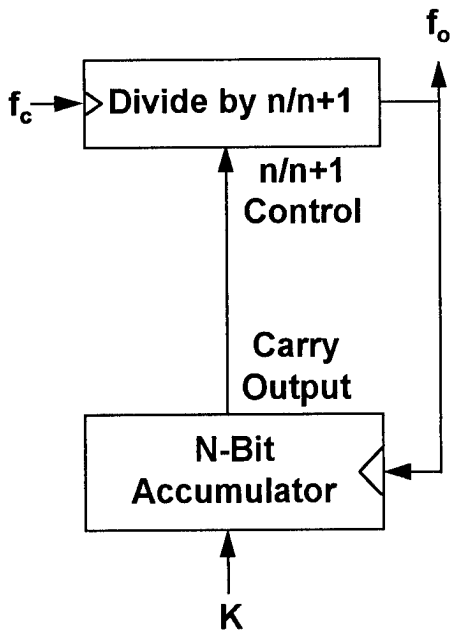


Figure 5. Fractional divider block diagram.

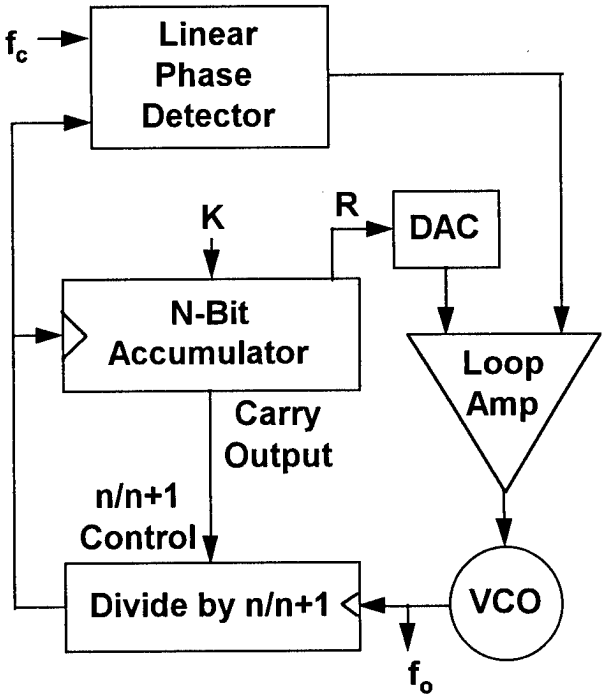


Figure 6. Phase interpolation DDS block diagram

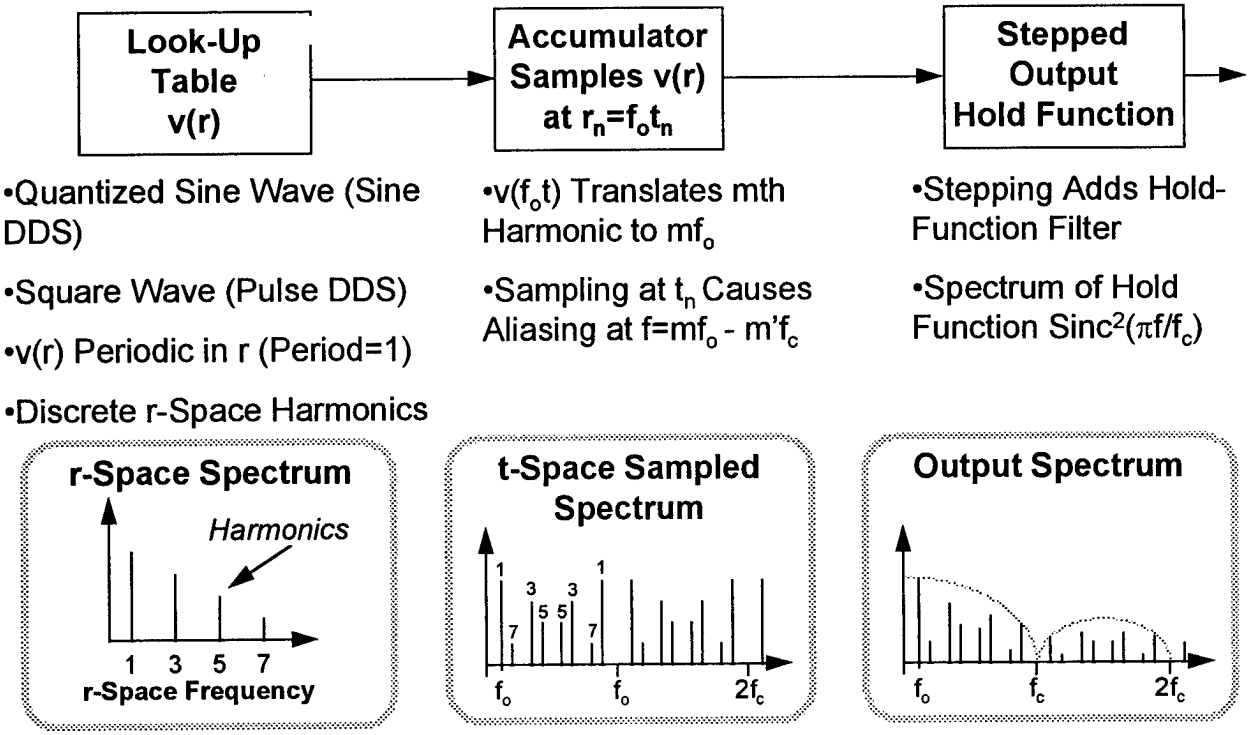


Figure 7. Model of uniformly sampled DDS and spur generation process.

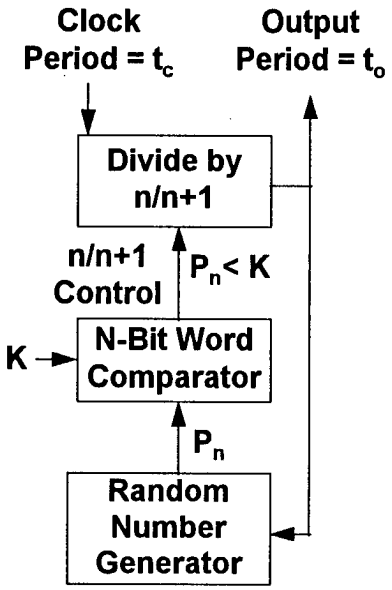


Figure 8. Spurless fractional divider block diagram

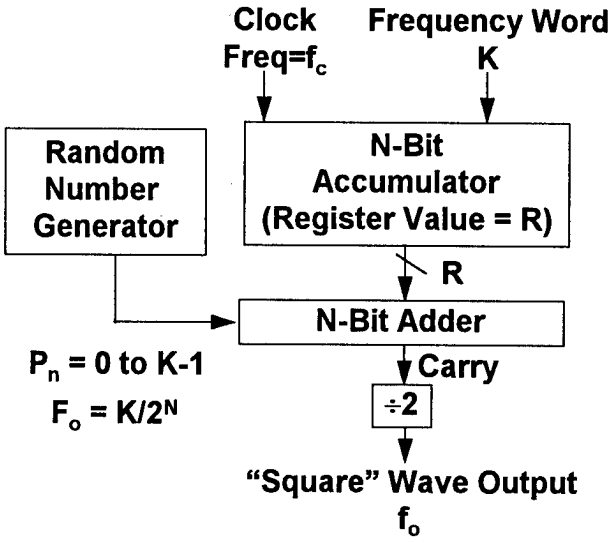


Figure 9. Wheatley Jitter injection DDS

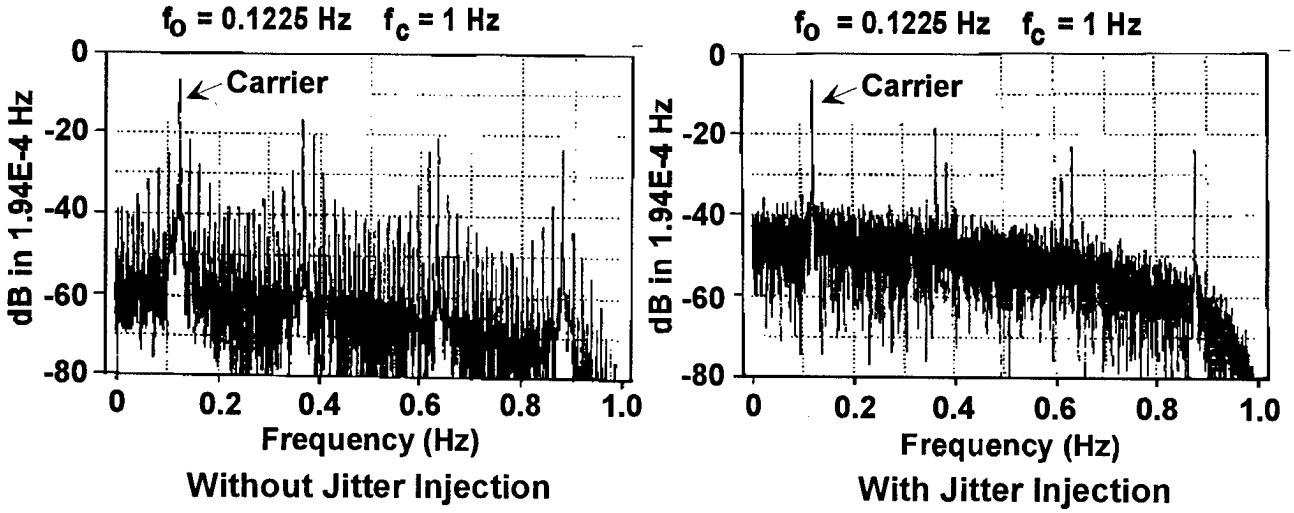


Figure 10. Wheatley jitter injection DDS frequency spectrum.

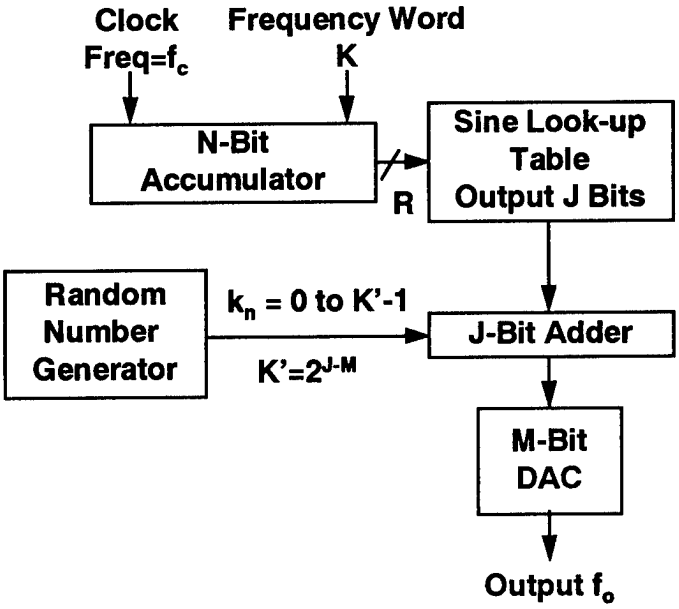


Figure 11. Randomized DAC DDS

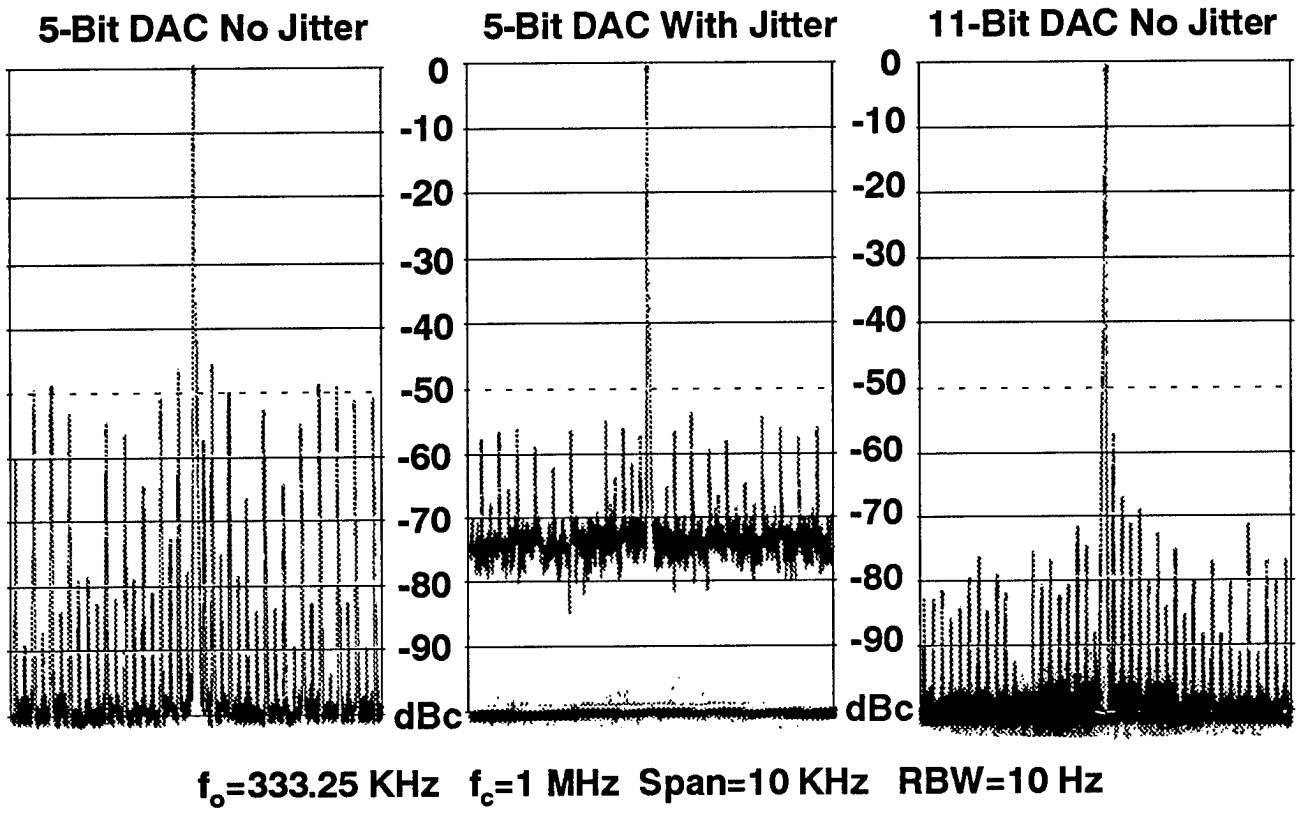


Figure 12. Comparison of Spectrums for sine output DDS and randomized DAC DDS.

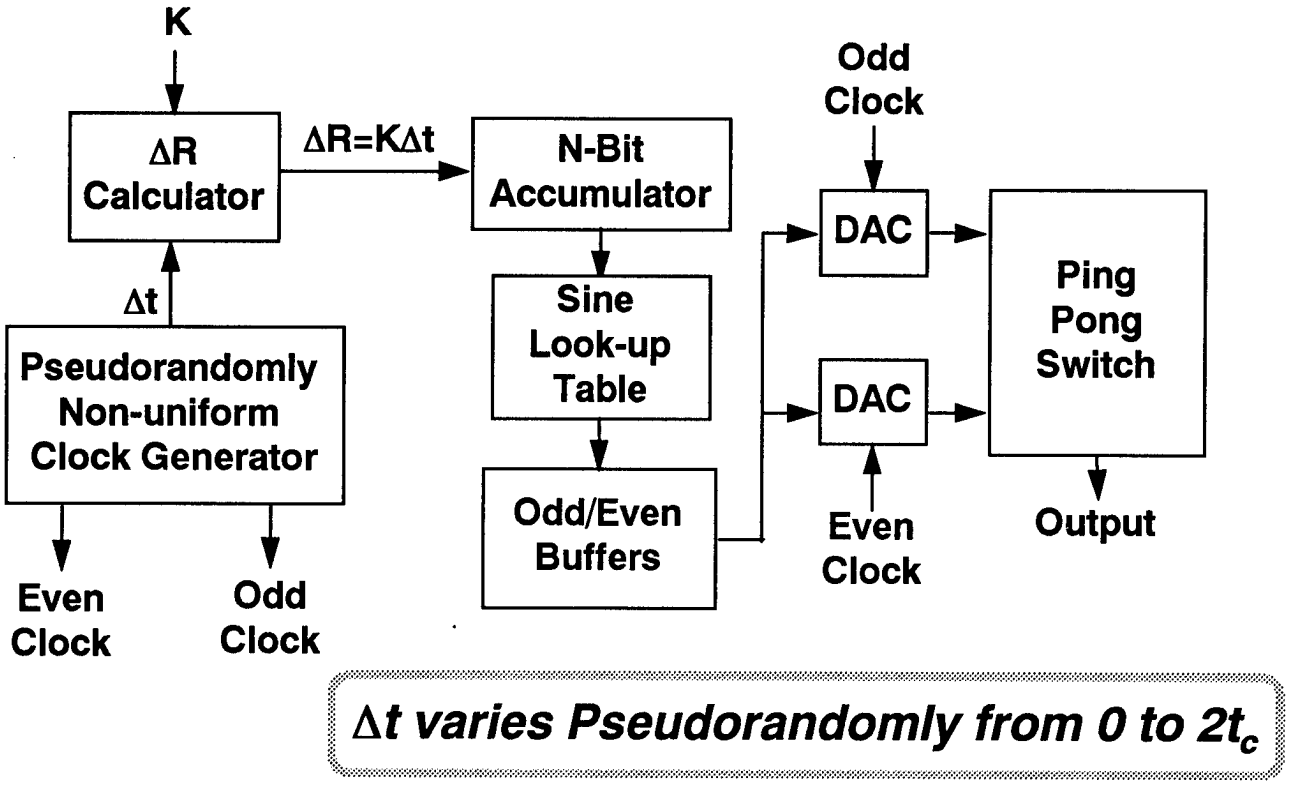


Figure 13. Nonuniform Clock DDS

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DISCRETE SPURIOUS SIGNALS AND BACKGROUND NOISE IN DIRECT DIGITAL FREQUENCY SYNTHESIZERS

Věnceslav F. Kroupa
The Institute of Radio Engineering and Electronics
Academy of Sciences of the Czech Republic
182 51 Praha, Czech Republic

Abstract: In the present paper we shall discuss some sources of output spurious signals in direct digital frequency synthesizers (**DDFS**).
First, we provide a closed form solution for spurious signals generated by the phase truncation. Then we pay attention to the accuracy of sine look-up tables and arrive at an estimated level of background spurious signals generated this way. Further, we investigate the behaviour of digital to analog convertors (**DAC**), their bit resolution, nonlinearity, and clock signal leakage.
Finally we will investigate the eventual phase noise level.

I. Introduction.

Perfection of digital integrated circuits in the last 20 years made it possible to introduce new techniques into frequency synthesis, namely, direct digital procedures. The difficulty with these simple systems is generation of single output frequency only and a rather large spurious signals accompanying the desired carrier [1,2].

The first problem was solved some 20 years ago after changing the rectangular output to a sine wave with the assistance of a phase accumulator and a sine look-up table [3]. Thus direct digital frequency synthesizers (DDFS) solve even to a large extent although the second, above mentioned difficulty, namely the reduction of undesired spurious signals. However, means for estimation of their level remained for a long time unsolved.

A first attempt about a rigorous mathematical solution was made by Nicholas et alii in 1987 [4] followed later with other papers [5,6]. There also exists an experimental analysis published by Garvey and Babitch 1990 [7].

Simultaneously the author worked on the theory of spurious signals in DDFS [8,9] and arrived at some practical results which may serve as leading ideas for designers. They will be discussed in following paragraphs.

II. Major sources of spurious signals in DDFS's.

The block diagram in Fig. 1 illustrates the set-up of DDFS. Its closer investigation may reveal major sources of spurious signals which are schematically shown in Fig. 2 and listed below:

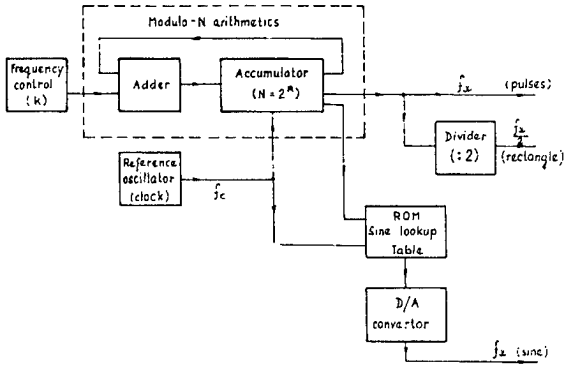


Fig. 1 Block diagram of a DDFS based on modulo-N aritmetics.

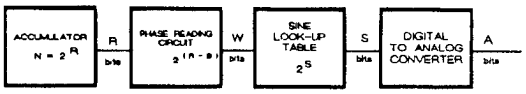


Fig. 2 Sources of spurious signals in DDFS due to the loss of information in different building bloks of DDFS.

A/ First we encounter truncation of the phase information stored in the accumulator. Generally we require the resolution of the output frequency f_x as small as possible. Since f_x is given by

$$f_x = \frac{k}{N} f_c \qquad (1)$$

where f_c is the input or clock frequency, k is the frequency control word and N the accumulator capacity generally in binary mode

$$N = 2^R \qquad (2)$$

For investigations of DDFS properties it is of advantage to change eq. (1) into a normalized form i.e.

0-7803-0905-7/93 $3.00 © 1993 IEEE

$$\frac{f_x}{f_c} = \xi_x = \frac{X}{Y} \qquad (3)$$

where both X and Y are mutually prime consequently

$$Y \le 2^R \qquad (4)$$

and the numerator X is always an odd number.

In practice the exponent R is often very large to achieve small output frequency steps Δf_x . The magnitudes of R in actual devices, we encounter, are e.g. *24, 32, 48* or even *64*.

There are two difficulties with such large N's, namely a long observation time on one hand and a limited number of phase bits, actually only W most significant bits (**MSB**), used for evaluation of sine values in look-up tables.

B/ Further we arrive at sine look-up tables (see Fig. 2) where the information is limited to S bits long words. We shall see later this particular limitation often generates the so called background spurious signals.

C/ Another source of errors is the digital to analogue convertor (**DAC**). In actual devices we use *8* bits in instances where higher clock frequencies, say over *100 MHz*, are used. Generally *10* or *12* bit converters are encountered. Only in some cases higher bit **DAC** are used [10].

D/ Finally the **DAC** characteristics are never ideal. As a consequence they are generators of intermodulation signals. The situation is worsened by the inevitable leakage of the clock frequency into the output circuits.

III.Truncation of the phase bits.

A/ Fourier analysis of the quantized output sine waves.

The output wave of **DDFS**'s is not an ideal sine but a staircase approximation as shown on an example in Fig. 3.

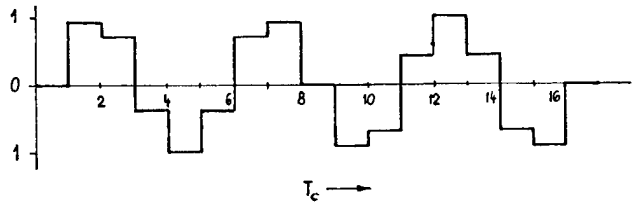


Fig. 3 Staircase approximation of the "sine" wave for the normalized frequency $\xi_x = 3/16$.

We can see that it repeats itself with the period YT_c. To compute the output spectrum we apply the Fourier analysis on each rectangle in the output staircase curve; see the quantized sine wave in Fig. 3. With the assistance of the complex form we get for the contribution to the ***n-th*** harmonic from the ***m-th*** rectangle

$$\begin{aligned} c_{n,m} &= \frac{1}{YT_c}\int_{mT_c}^{(m+1)T_c} \sin(2\pi m\frac{X}{Y})e^{-jn(\frac{2\pi}{YT_c})t}\,dt \\ &= \frac{1}{\pi n}\sin(\pi\frac{n}{Y})e^{-j\pi\frac{n}{Y}}\sin(2\pi m\frac{X}{Y})e^{-2j\pi\frac{nm}{Y}} \end{aligned} \qquad (5)$$

After performing summation over all m (from 0 to Y-1) we find out that all c_n are equal to zero for all $n \neq X$ and only for $n = X$ we arrive for onesided Fourier coefficient

$$c_X = \frac{\sin(\pi X/Y)}{\pi X/Y} \qquad (6)$$

B/ Spectral analysis of phase truncated sine waves.

As emphasized above the high frequency resolution of DDFS requires very large accumulator capacity (32, 48, or even 64 bits). It is evident that the would be complexity of the hard ware prevents us from referring to all this minute phase increments. Generally, we use only a small number of most significant bits (**MSB**), W, and disregard all the remaining least significant bits (**LSB**), B - c.f. Fig. 2. Thus from the original number stored in the accumulator [2]

$$x(m) = mX - rY \le Y \le 2^R \qquad (7)$$

we refer to

$$X'(m) = integer(\frac{mX - rY}{2^B}) \qquad (8)$$

Since the information passed to the "sine look-up table" must be of the same order as the original $X(m)$ we have to multiply $X'(m)$ by 2^B . However, the phase readings are still performed at each mT_c (m = 0,1,...,Y-1). As a consequence amplitudes of individual rectangles in the output sine wave change into

$$\begin{aligned} &\sin[2\pi\frac{2^B}{Y}integer(\frac{mX}{2^B})] = \\ &\sin\{2\pi\frac{2^B}{Y}[\frac{mX}{2^B} - s(m)]\} \end{aligned} \qquad (9)$$

where

$$s(m) = \frac{mX}{2^B} - integer(\frac{mX}{2^B}) \le 1 \qquad (10)$$

evidently

$$\frac{2^B}{Y}s(m) < 1 \qquad (11)$$

and relation (9) can be further simplified to

$$\sin\frac{2\pi mX}{Y} - 2\pi\frac{2^B}{Y}s(m)\cos\frac{2\pi mX}{Y} \qquad (12)$$

The situation with eq. (9) is such that we face a nearly ideal sine wave with amplitude 1 and a disturbing cosine wave with a rather small amplitude cf. eq. (12). We easily conclude that the largest spurious amplitude a_{sp} does not exceeds the amplitude of the disturbing cosine wave i.e.

$$a_{sp} \le 2\pi\frac{2^B}{Y} = 2\pi 2^{-W} \qquad (13)$$

which is in good agreement with earlier results [e.g. 4].

The exact solution of the output spectrum would require to introduce the relation (12) as amplitude coefficients into (5) and compute the output spectrum. But this is generally prohibitive because of the large $Y \le 2^R$. The other way is to estimate some major disturbing components. To this end we shall investigate properties of the phase modulation function $s(m)$. After reverting to the earlier results found by the author for the quasiperiodic rectangular waves [1,8,9] we can expand $s(m)$ into a periodic series with the assistance of modified continued fraction expansion

$$s(m) = \frac{mX}{2^B} - integer(\frac{mX}{2^B}) =$$

$$m(A_0 + \frac{a_1}{B_1} - \frac{a_1a_2}{B_1B_2} + \ldots + \qquad (14)$$

$$+ \frac{a_{n-1}a_n}{B_{n-1}B_n}) - integer(\frac{mX}{2^B})$$

Examples of phase truncation functions $s(m)$ are shown in Fig. 4.

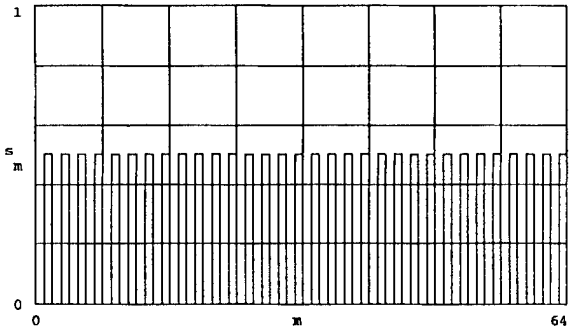


Fig. 4a) Plot of the phase truncation function s(m) for the normalized frequency ξ_x = 207/1024 for 1 bit LSB discarded; note the rectangular shape s(m) = m(103+1/2) - integer(207m/2).

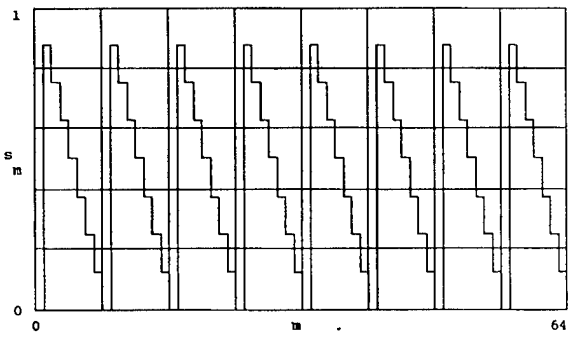


Fig. 4b) Three bits discarded - one sawtooth wave predominates - s(m) = m(26 - 1/8) - integer(207m/8).

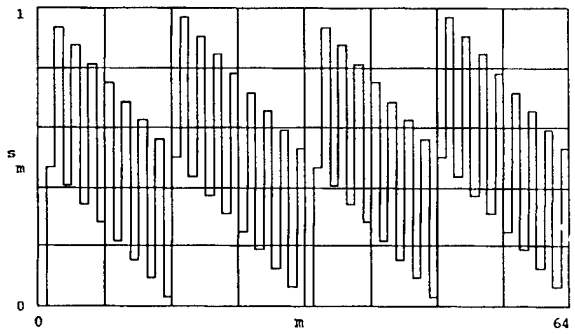


Fig. 4c) 5 bits discarded - superposition of rectangular and sawtooth waves - s(m) = m(7 - 1/2 - 1/32) - integer (207m/32).

Examination of the plots reveals that the first order approximation of $s(m)$ by A_1/B_1 is either a rectangular or a simple sawtooth wave with amplitude $1/2$ or 1 and the repetition period B_1T_c. The second order approximation adds another sawtooth wave with the amplitude $1/B_1$ and the repetition period B_2T_c .The same happens with every other approximation.

A numerical example for ξ_x = 207/1024 with 5 neglected LSB, i.e. B=5, (plot Fig. 4c) reveals

$$s(m) = m[6 + \frac{1}{2} - \frac{1}{2*17} - \frac{1}{17*32}] - integer(\frac{m207}{32}) \qquad (15)$$

Note that since X and 2^B are relatively prime integers (X is odd) the latest divisor B_n is equal to 2^B.

Series expansion (14) makes it possible to compute or estimate the levels and frequencies of spurious signals quite easy in comparison (to the author knowledge) with earlier approaches [4, 11]. In accordance with **Appendix** we shall start with finding the remainder R_b for $s(m=1)$

$$s(m=1) = \frac{X}{2^B} - integer(\frac{X}{2^B}) = R_b \qquad (16)$$

and than taking into account that ***s(m)*** is a periodic function we arrive at a sum of sine waves, i.e.

$$s(m) \doteq \frac{1}{\pi}\sum_{r=1,2,..}^{2^{(B-1)}} \frac{1}{r}\sin(2\pi r m R_b) \qquad (17)$$

Note that we can always held the value of the remainder in the range

$$0 \le R_b = \frac{P}{Q} \le \frac{1}{2} \qquad (Q = 2^B) \qquad (18)$$

which simplifies location of spurious signals.

After introducing the ***r-th*** harmonic from (17) into (12) we can expand the 2nd term into

$$\frac{2^B}{rY}\{\sin[2\pi m(\frac{rP}{Q} + \frac{X}{Y})] + \sin[2\pi m(\frac{rP}{Q} - \frac{X}{Y})]\} \qquad (19)$$

Evidently both terms of the above expression are staircase values of spurious sine waves around the carrier. Consequently we can compute their amplitudes with the assistance of (5) and in accordance with (6) we arrive at

$$c_{sr} = \frac{2^B}{Y}\frac{\sin[\pi(\frac{X}{Y} \pm \frac{rP}{Q})/Y]}{\pi(\frac{X}{Y} \pm \frac{rP}{Q})} \qquad (20)$$

After dividing the above relation by (6) we get for the ***r-th*** order spurious signal level

$$\frac{c_{sr}}{c_X} = \frac{2^B}{Y}\frac{X}{X/Y \pm rP/Q}\frac{\sin(\pi(X/Y \pm rP/Q)/Y)}{\sin(\pi\frac{X}{Y})} \doteq \frac{2^B}{rY} = \frac{2^{-R+B}}{r} = \frac{2^{-W}}{r} \qquad (21)$$

or in dB measure

$$\frac{c_{sr}}{c_X} \doteq -6W - 20\log(r) \qquad [dB] \qquad (22)$$

The above relation is plotted in Fig. 5 for ***r=1*** and reveals the worst case spurious signal level as function of the number ***W*** of the **MSB** used for phase reading from the accumulator of **DDFS**.

With the assistance of eq. (9) and (5) we have computed the spurious side bands for the case of the above mentioned numerical example. The result is shown in Fig. 6 and reveals that the largest spurious signals are about -30 dB as expected and extend down to about -50 dB - from eq. (22) we get -54 dB, which is a fairly good result since we have neglected attenuations due to the sampling effect, to the overlapping of alias, to the lower harmonics of basic sawtooth wave, etc.

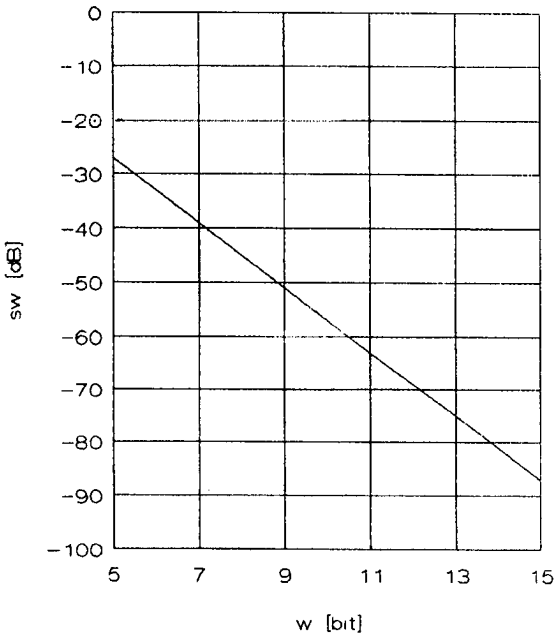


Fig. 5 Approximate spurious signal level s_w for W MSB retained.

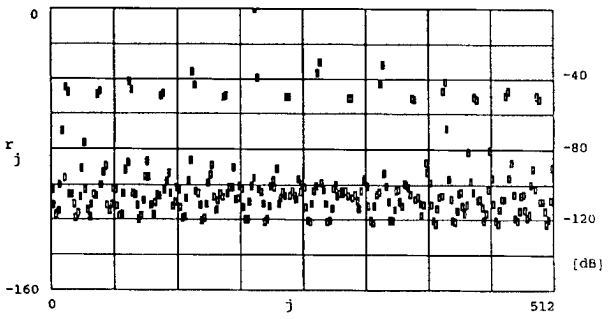


Fig. 6 An example of spurious side bands of $\xi_x = 207/2^{10}$ with 5 LSB neglected (S=12).

There are several remarks to be made:

First we would like to know the number of the spurious signals, caused by phase truncation, in the DDFS pass band which in normalized notation is ideally equal to ***Y/2***. From eq. (20) we can deduce the spurious line numbers (the carrier line number is X). However, we must also take into account that some ***rP/Q*** exceed the

pass band $Y/2$ but are returned as alias. As a consequence the spurious spectral line numbers must meet condition

$$n_r = X + (r\frac{PY}{2^B} + sY) < Y/2 \qquad (s = ...,-2,-1,0,1,2,...) \tag{23}$$

Investigation of the above inequality reveals different n_r only for $r < 2^B/2$ and a double one for $r=2^{B-1}$. After taking into account that there are two side bands around the carrier for each r (i.e. both positive and negative) their number is given by (leaving out the carrier)

$$2^B - 1 \tag{24}$$

Second: with the assistance of the above relation we easily arrive at actual spurious frequencies

$$f_{sr} = \frac{n_r}{Y} f_c = f_x + (r\frac{P}{2^B} + s)f_c \tag{25}$$

Third: investigation of Fig. 6 reveals a lot of "background" spurious signals which are much smaller than those due to the phase truncation. The problem will be discussed in the next section.

VI. Truncation of sine words in look-up tables.

Investigation of eq. (5) with ideal sine wave amplitudes, i.e.

$$\sin(2\pi m\frac{X}{Y}) \tag{26}$$

would reveal no spurious components in the output signal. However, the economy of actual **DDFS** requires that the information about the sine values in look-up tables is generally stored only as S-bits long words. The consequence is that besides those "high level" spurious signals caused by the phase truncation there is a large set of much smaller spurious signals, the level of which depends on the number of bits used in look-up table words. Consequently, truncation of stored sine values results in changing actual amplitudes into

$$\frac{1}{2^S} interger[2^S \sin(2\pi m\frac{X}{Y})] \tag{27}$$

or even into

$$\frac{1}{2^S} integer\{2^S \sin[2\pi \frac{2^B}{Y} integer(\frac{mX}{2^B})]\} \tag{28}$$

With the assistance of the FFT algorithm, i.e. after inserting the above relation into (5), we can compute all major spurious signals caused by the phase truncation and also the "background" spurious signals due to the finite word length in the sine look-up tables. Some examples are shown in Figs. 6,7,8.

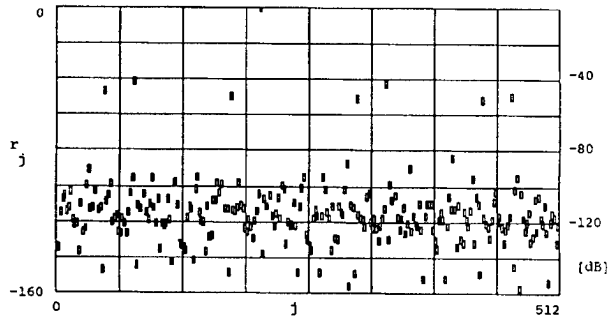


Fig. 7 An example of spurious side bands in DDFS with the normalized frequency $\xi_x = 207/2^{10}$ with 3 LSB neglected, i.e. B=3 and W=7; S=12.

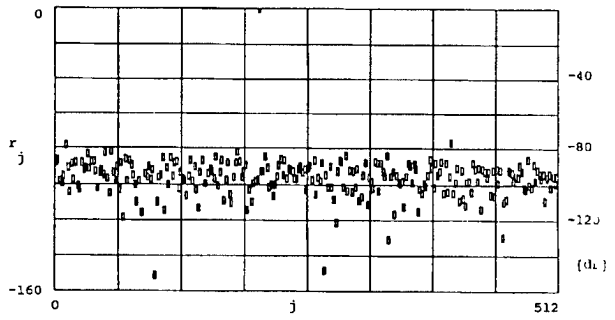


Fig. 8 Background spurious signal level for $\xi_x = 207/2^{10}$ with B=0, W=10 and S=10.

A/ Estimation of background spurious signal level.

At the first sight at these figures one has a feeling that amplitudes of these background spurious signals are randomly distributed. However, this cannot be true since we have evaluated a periodic system. To get more insight we shall compute and plot the error signal

$$e(m) = \sin(2\pi m\frac{X}{Y}) - \frac{1}{2^S} integer\{2^S \sin[2\pi \frac{2^B}{Y} inetger(\frac{mX}{Y})]\} \tag{29}$$

From the above relation we easily find out that

$$e(m)_{max} < |1/2^S| \tag{30}$$

and a plot in Fig. 9 confirms the above inequality. Its close investigation reveals an odd symmetry and some

other periodicities, i.e. the process ***e(m)*** is periodically stationary [12]. However, at the first sight we have a feeling of randomness and from this point of view we shall start farther investigations.

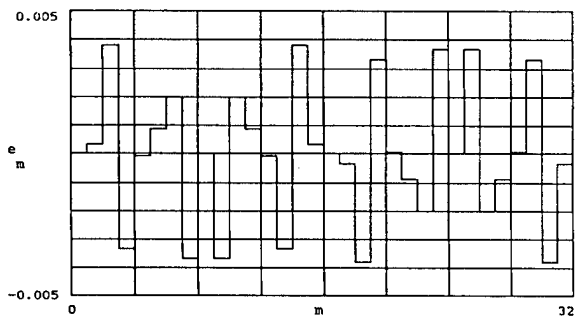


Fig. 9 The error signal e(m) due to finite length of words, S, in sine look-up tables for the normalized frequency ξ_x = 7/32, S=8, B=0.

For the amplitude distributions of ***e(m)'s*** s we shall use the simplest one, namely, a rectangular distribution between $e(m)_{max+}$ and $e(m)_{max-}$ - i.e.

$$p(e) = 1/(2*2^{-S}) \quad \textit{inside the interval}$$
$$p(e) = 0 \quad \textit{outside the interval} \qquad (31)$$

As a consequence we find for the variance

$$\sigma^2(e) \doteq (1/3)*2^{-2S} \qquad (32)$$

After computing actual variances in accordance with (29) and comparing the results with values arrived at with the assistance of (32) we have found quite small differences and concluded that we are justified to use relation (32) in our pursuit of the estimation of the "background" spurious signal levels. Investigation of Figs. 6, 7, and 8 reveals that there might exist a mean level $<S(n)>$ with a dispersion of about ±20 dB.

With these encouraging findings we can look for closed form solution of $<S(n)>$ for different ξ_x and length S of the used sine words.

First, we shall recall the relation between autocorrelation and the spectral density

$$R(0) = \int_0^\infty S(f)df \qquad (33)$$

where $S(f)$ is one-sided power spectral density (PSD).

Second, because of the odd symmetry of the error signal $e(m)$ its first moment is zero and consequently the autocorrelation is equal to the variance. Finally, by taking into account that the investigated PSD $S(f)$ is formed by spectral lines the number of which is ideally only

$$Y/4 - 1 \qquad (34)$$

since we must leave out the carrier and all spurious signals with even line numbers, which are negligibly small. Consequently relation (33) can be replaced with

$$\sigma^2(e) \approx (Y/4 - 1)*<S(n)> \qquad (35)$$

where $<S(n)>$ is the mean value of all spectral lines $S(f=n_s.f_c)$. After introducing (32) into the relation (35) we get for the mean value of the background spurious signals

$$<S(n)> = \frac{\sigma^2(e)}{Y/4-1} = \frac{1}{3}*2^{-2S}*\frac{4}{Y-4} \qquad (36)$$

For confirmation of the above theory we have computed actual background spurious and their means and found a good agreement with the above estimation, particularly, for large Y.

For a quick information of the readers we have plotted eq. (36) in Fig. 10 for two different Y (Y=32 and Y=1024) as function of the length of sine words, S, in look-up tables. Rectangles are actually computed means. Note a small dispersion of the computed means for Y=32 around the estimation characteristic and a very good agreement for Y=1024. Note that $r_s = <r_j>$.

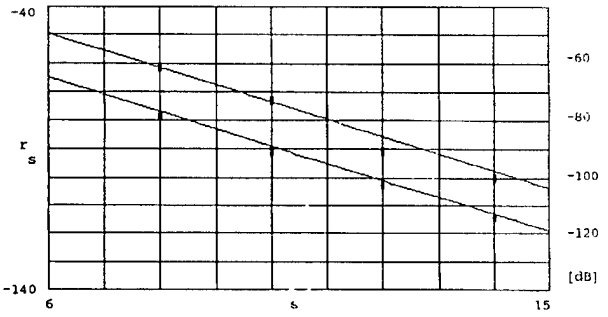


Fig. 10 Estimated means of background spurious signal level in accordance with the relation (36) for two different denominators "Y" (32 and 1024) of ξ_x = X/Y; rectangles are actually computed means.

The last but one question which remains is about the dispersion of actual values of background spurious signals around the mean. To this end we shall compute the mean of the cepstrum of $S(n)$ [13], i.e.

$$<r_j> = <10*\log[S(n_j)> \quad (j=1,3,...j\neq X) \qquad (37)$$

and the respective standard deviation which in accordance with our earlier investigations [14] is in the range of ***10 dB***. Further our experimental evidence taught us that there were only few "background" signals that exceeded the level given by

$$10*\log[<S(n)>] + 2\sigma \qquad (38)$$

This generally happens in the presence of phase truncation. A mere investigation of Figs. 6 and 7 reveals that spurious sidebands due to this origin are much larger and rather few. As a consequence these can be omitted as outlayers [13] and the above arrived results are valid also in these instances.

B/ Estimation of the spurious harmonics.

A closer investigation of many simulations of background spurious signals in DDFS reveals that there are few of them exceeding the level given by relation (38) - see Fig. 11. We have looked for their source and found it in higher harmonics of the output signal f_x.

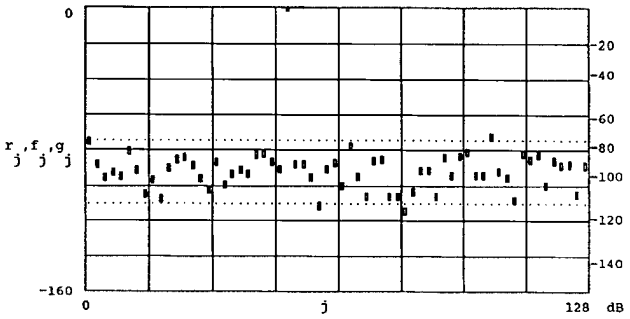


Fig. 11 FFT simulation of background spurious signals for S=10 and ξ_x = 51/256; points indicate 2σ dispersion lines around the mean (f_j and g_j).

Let us revert to eq.(29) and compute the mean value of $e(m)$ both for the positive and negative half period of the output sine wave. We get a rectangular wave with the amplitude $2^S/2$. Its Fourier series expansion reveals only odd harmonics with amplitudes

$$a_h = \frac{2*2^{-S}}{\pi} * \frac{1}{h} \quad (h = 3,5,\ldots) \qquad (39)$$

For the level of the 3rd spurious harmonic we get in dB measure

$$20*\log(\frac{c_{s,3x}}{c_x}) \approx -6S - 14 \quad [dB] \qquad (40)$$

Note that we have again neglected the effect of sampling on amplitudes. The above relation is plotted in Fig. 12 and it is in a good agreement with many computer simulations. Unfortunately for the 5th and higher harmonics the dispersion of the spurious level is much larger - we can only be sure that it generally does not exceed the value given by (40).

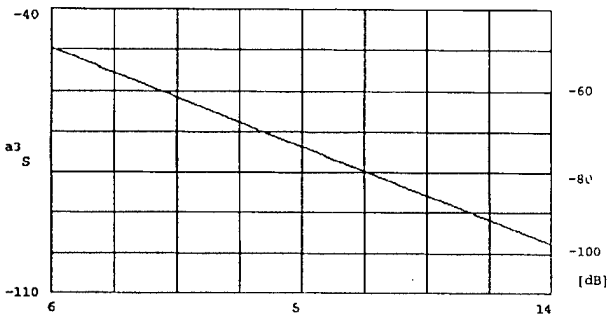


Fig. 12 The expected value of the 3rd harmonic of the output frequency ($3f_x$) as function of the number of S bits used.

V. Errors due to digital to analog converters.

There are three major sources causing degradation of the output spectral purity by **DAC**, namely, finite bit resolution, nonlinearity, and transient effects or glitches.

A/ Finite bit resolution.

From the point of view of present technology we are not able to build **DAC** with any arbitrary resolution on one hand. On the other hand the higher the resolution the more expensive **DAC** are. Consequently every designer has a difficult task to choose between the spectral purity and economy.

For discussion of the **DAC** bit resolution effect we must revert to the relation (27) or more often to (28) and take into account that from the S bit long sine word we use only A bits for conversion the digital information about the amplitude into its analog version. Note that in instances where a D bit **DAC** is used in 2's complement version

$$A = D - 1 \qquad (41)$$

As a consequence all conclusions about the output spectrum can be arrived at if the sine word length S in (27) or (28) is replaced by the bit resolution A (cf. paper by Garvey et al.[7]). The conclusion is that the extent of the sine look-up table and the bit resolution of **DAC** are closely related in properly designed **DDFS**.

B/ DAC nonlinearity.

Investigation of real output spectra of **DDFS** reveals also even harmonics of the output signal and odd

with (40). The reason is the nonlinearity of the **DAC** transfer characteristics which is difficult to predict since these differ between individual **DAC**.

C/ Leakage of the clock signal.

We can hardly prevent the leakage of the clock signal into **DAC**. Its nonlinearity and its behaviour as a switching modulator [cf.7] provides source of intermodulation signals, in the output pass band, of the type

$$f_{u,v} = f_x u + f_c v \quad (u,v = 0,\pm1,\pm2,...) \tag{42}$$

V. Output phase noise.

In previous paragraphs we have seen that the output signal of nearly each **DDFS** is accompanied by many spurious signals of different amplitudes. In the case of accumulator phase truncation their number is given by (24) and as a consequence the frequency difference between individual lines is (cf. Fig. 6,7)

$$\Delta f_{s,p} = \frac{f_c}{2^B} \tag{43}$$

Similarly in instances of sine word truncation the number of spurious spectral lines is larger, in accordance with (34), and the frequency difference smaller (cf. Fig. 8)

$$\Delta f_{s,s} = \frac{2}{Y} * f_c \tag{44}$$

Eventually, in the *1 Hz* frequency band we have

$$1/\Delta f_{s,s} = Y/2f_c \tag{45}$$

spectral lines; evidently for

$$Y > 2f_c \tag{46}$$

we get for the white mean background noise

$$S(f) \approx \frac{<S(r)>}{\Delta f_{s,s}} \doteq \frac{2}{3} \cdot \frac{2^{-2A}}{f_c} \tag{47}$$

or in *dB* measure approximately

$$-2 \ -6*A \ - \ 10*\log(f_c) \quad [dB] \tag{48}$$

In instances where Y is a small integer condition (46) is not met [cf. 4] and the background noise between individual spurious signals is that of the phase noise of the clock oscillator reduced by the division factor ξ_x or the phase noise of "dividers" and amplifiers in the respective **DAC** which is approximately [15, 16, 17]

$$S(f) \approx \frac{10^{-14.7}}{f} + 10^{-16.5} \tag{49}$$

VI. Conclusions.

In the present paper we have derived a simple closed form solution for frequencies and power level of spurious signals caused by truncation of the phase information to some **MSB**'s.

In the second part we have discussed the problem of the "background" spurious signals caused by the finite resolution of the sine look-up tables and **DAC**. We have found a strong dependence on the numerator Y of the normalized frequency. As a consequence the respective mean spurious level is different for different output frequencies of the same **DDFS**.

Finally we have paid attention to some problems connected with **DAC**.

References:

[1] V.F. Kroupa, "Spectra of Pulse Rate Frequency Synthesizers," Proc. IEEE, Vol.67, pp. 1680-82, December 1979.

[2] V.F. Kroupa, "Spectral purity of direct digital frequency synthesizers," Proc. 44th Annual Symposium on Frequency Control, pp. 498-510, May 1990.

[3] J. Tierney, C.M. Radar,and B. Gold,"A Digital Frequency Synthesizer," IEEE Trans. AU-19, pp. 43-48, March 1971.

[4] H.T. Nicholas and H. Samueli," An Analysis of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase-Accumulator Truncation," Proc. 41st AFCS, pp. 495-502, 1987.

[5] H.T. Nicholas, H. Samueli and B. Kim, "The Optimization of Direct Digital Frequency Synthesizer Performance in the Presence of Finite Word Length Effects". Proc. 42nd AFCS, pp. 357-363, 1988.

[6] H.T. Nicholas and H. Samueli, "A 150-MHz Direct Digital Frequency Synthesizer in 1.25-um CMOS with -90 dB Spurious Performance", IEEE J. of Solid State Circuits, vol. 26, No. 12, pp. 1959-1969, December 1991.

[7] J.F. Garvey and D. Babitch, "An exact spectral analysis of a number controlled oscillator based synthesizer", Proc. 44th AFCS, pp. 511-521, May 23-25, 1990.

[8] V. F. Kroupa, "Spurious signals in direct digital frequency synthesizers (DDFS) caused by accumulator truncation," not yet published.

[9] V.F. Kroupa, "Principles of direct digital frequency synthesizers," presented at Kleinheubacher Tagung October 1992.

[10] R.P. Giffard and L.S. Cutler, "A low frequency, high resolution digital synthesizer", Proc. of the IEEE 1992 Frequency Control Symposium, pp. 188-192, May 1992.
[11] S. Mehrgardt, "Noise Spectra of Digital Sine-Generators Using the Table-Lookup Method", IEEE Trans. ASSP-31, pp. 1037-39, August 1983.
[12] A. Papoulis, Probability, Random variables and Stochastic Processes. New York: McGraw-Hill, 1965.
[13] C.K. Yuen and D. Fraser, Digital Spectral Analysis. London: Pitman, 1979.
[14] V.F. Kroupa, unpublished results.
[15] V.F. Kroupa, "Noise properties of PLL systems", IEEE Tr. COM-30, pp. 2244-2252, October 1982.
[16] W.F. Egan, "Modeling phase noise in frequency dividers",IEEE Tr. UFFC-37, pp. 307-315, July 1990.
[17] J.A. Connelly and K.P. Taylor, "An Analysis Methodology to Identify dominant Noise Sources in D/A and A/D Converters", IEEE Trans. CS-38, pp. 1133-1144, October 1991.

Appendix.

Investigation of spurious signals generated by the phase truncation starts by computing the remainder R_b. We easily change eq. (16) into

$$R_b = \xi_x * 2^W - integer(\xi_x * 2^W) \qquad (50)$$

with the assistance of the modified continued fraction expansion - cf. eq. (14) where all a_1, a_2, etc. are equal either to +1 or to -1 [2]. The first order approximation of R_b is

$$R_b \approx (\pm)\frac{1}{B_1} \qquad (51)$$

and $s(m)$ in the time domain reveals a sawtooth wave with the amplitude 1 and the repetition period B_1T_c and its Fourier series expansion reveals

$$\frac{1}{\pi}\sum \frac{1}{r}\sin(\frac{2\pi r}{B_1T_c}t) \qquad (52)$$

The 2nd order approximation of R_b is

$$R_b \approx (\pm)\frac{A_2}{B_2} \qquad (53)$$

Again $s(m)$ in the time domain would reveal a sawtooth wave with amplitude 1, however, with the repetition period very close to that of the 1st order approximation, namely $(B_2/A_2)T_c$. With the assistance of the Fourier series expansion we get

$$\frac{1}{\pi}\sum \frac{1}{r}\sin(2\pi r\frac{A_2}{B_2}t) \qquad (54)$$

Proceeding further this way we finally arrive at the relation (17).

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A NEW TECHNIQUE OF FREQUENCY SYNTHESIS

Wei Zhou

Dept. of Measurement and Instrumentation, Xidian University
Xi'an, 710071, P.R.China

Abstract

This paper describes a new design method of frequency synthesizer. It is based on some precision frequency measurement techniques. The frequency synthesis can be accomplished by measuring synthesized frequency signal, comparing it with desired frequency by soft ware, and time (puls width) - voltage transformation. In this synthesizer there are not any frequency multipliers, mixers and any other frequency transformation circuits which are used in other frequency synthesizers. The hardware circuits are very simple. A lot of functions are completed by soft ware. This technique is suitable for different cases, especially for synthesis of complicated frequency signals. For different synthesized frequencies, the circuits are almost the same. Therefore it can be used widely.

1. Introduction

In conventional frequency synthesizers a lot of frequency transformation ciecuits (frequency multipliers, mixers and filters) are used. The hardware circuits of the instrument are complicated, whose cost is high, and it is difficult to integrate the circuit.

By using some new frequency measurement techniques, the different frequency signals can be measured with great accuracy by simple equipment. Therefore, a new technique of frequency synthesis based on the precision frequency measurement techniques can be developed. The basic principle of this method is that the frequency of a voltage control oscillator VCO or VCXO is first measured accurately, and the measured frequency is compared with the desired frequency by soft ware, then according to their difference frequency the control voltage is varied to obtain correct synthetic frequency. The measurement and control are made alternately, and the accuracy of the synthesizer depends on the accuracy of measurement and control.

2. Some new frequency measurement techniques

The frequency measurement techniques which are used to design the frequency synthesizers ought to be highly accurate and the frequency measurement equipment ought to be simple.

Some frequency measurement techniques based on the characteristics of the greatest common factor frequency and the phase coincidence detection method can satisfy the demands above mentioned. In the references [1], [2] and [3], the new techniques are described in detail. They are "The frequency measurement technique by broad- band phase detection" and "The dual broad- band phase detection method". These techniques are highly accurate in a very wide frequency range, and the equipment is very simple, almost all circuits are composed of digital integrated circuits. The time response can be decided according to the demands. The lower frequency measurement limitation of these techniques is about 10KHz, and the upper frequency measurement limitation is almost unlimited. The different equipment has different accuracy. The measurement accuracy can be $10^{-10}s/\tau$ to $10^{-12}s/\tau$, respectively, where τ is the gate time. Because the measurement is accomplished by a microcomputer, the computer can also be used to control the frequency of VCO to form frequency synthesizer and the intellectual characteristic of the computer can also influence the synthesizer.

3. A technique of frequency synthesis by directly measuring and controlling frequency

The frequency measuring methods of the new synthesis technique are different sometimes. "The frequency measurement technique by broad- band phase detection" and "The dual broad-band phase detection method" based on the characteristics of the greatest common factor frequency and the phase coincidence detection methods can be used for this purpose. These two techniques are of high accuracy and of wide frequency range. The former can be used more simply. It is suitable for frequency signal synthesis with some frequencies in certain frequency range. The latter is somewhat more complicated than the former, but it has better measurement continuity. Therefore, it can be used to design synthesizers with wide frequency range. One can choose the measurement method according to the synthesized frequency and accuracy.

The block diagram of the frequency synthesizer designed by the new method is shown in Figure 1.

0-7803-0905-7/93 $3.00 © 1993 IEEE

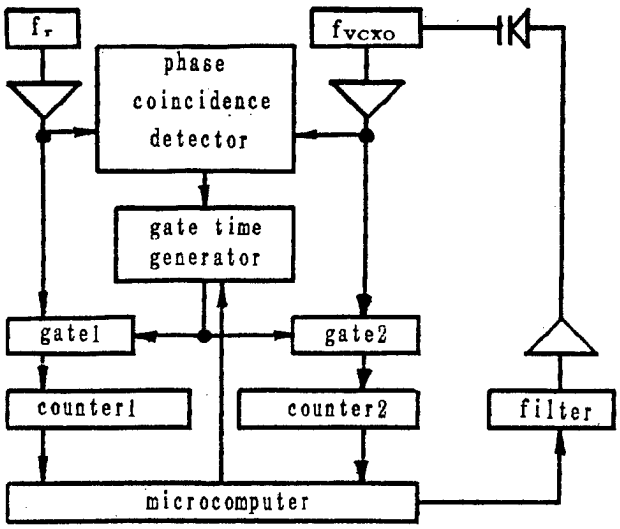


Figure 1.

The new technique differs from conventional ones (direct analog synthesis and indirect synthesis) in that the frequency measurement method with the high accuracy is used instead of the conventional frequency transformation and phase comparison methods to accomplish the frequency synthesis. This frequency synthesizer is composed mainly of a simple accurate digital frequency measuring device, one voltage control oscillator, a time-to-voltage converter and a microcomputer. The VCO or VCXO frequency is measured in terms of frequency measurement technique by broadband phase detection or dual broad- band phase detection. When the frequencies of the two compared signals are not very close or there is no a multiple relationship between them, the first method can show a high accuracy. Especially when the phase difference variation between two compared signals increases or decreases uniformly in the least common multiple period, or a period of difference frequency or a period of equivalent harmonic difference frequency, it can show a much higher accuracy. The accuracy of this method can be from $3\times10^{-10}s/\tau$ to higher than $1\times10^{-11}s/\tau$. After the measured frequency is calculated by the microcomputer, it is compared with the desired frequency which has been stored in the EPROM of the microcomputer or is set up by the keyboard. According to the difference of the two frequencies, the pulse width of the square-wave control signal generated by the microcomputer is varied. After being filtered and amplified, this signal is converted into cotrol voltage to control the frequency of the VCXO. Here the time (pulse width)-to-voltage convertion method is used to form the control voltage, because it is simpler, more inexpensive and sometimes accurate than the conventional D/A convertion method. (However, in some cases, the D/A converter can also be used. This method has a better time response than that of the former. The accuracy of ordinary D/A converters is not very high, they must be combined with other methods, then the control range and accuracy can be satisfied.) The time response of the former method is not very fast. The measuring gate time and the period of the square-wave signal can be chosen according to the characteristic of the VCXO and the performance of the frequency synthesizer. If the highly-stable crystal oscillator is used as the VCXO, the measuring time can be long and the period of the square-wave signal can be long too. In this case, the accuracy of measurement and control is higher, the short term frequency stability of the synthesizer depends on the controlled crystal oscillator while its long term performance depends on the reference frequency signal. If an ordinary crystal oscillator or other oscillators are used as the VCXO or VCO, the measuring time and the period of the square-wave signal ought to be short. In this case, the VCXO can be locked stably, the short and long term stabilities of the synthesizer both depend on the reference frequency signal and the measurement and control circuits.

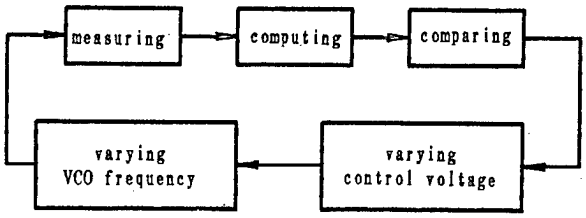


Figure 2.

The operation procedure of the synthesizer is shown in Figure 2. The operation period is from measurement, computation, comparison, control voltage variation, to VCXO frequency variation, then another period ..... . The periodic square wave control signal generated by the microcomputer is continuous. Its pulse width is varied by the result of the measurement. In the period of the operation procedure, computation and comparison take less than 1ms, the measurement takes longer time. The period of the square-wave control signal must be much less than the operation period, since the filter time constant is larger than the square-wave period. The operation period is the multiple of the period of the control signal. The time diagram of the operation procedure is shown in Figure 3.

When the synthesizer is switched on, the duty cycle (pulse width/period) of the square wave control signal is set as 1/2 or other more suitable values automatically. Then it is varied by the result of measurement. By means of different measuring methods the computation equations may be different. When the method in Fig. 1 is used, the computation equation is as follows:

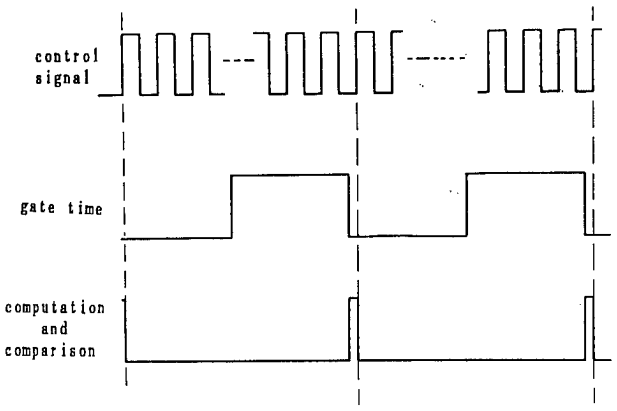


Figure 3.

$$f_{VCXO}=f_r \cdot N_{VCXO}/N_r \qquad (1)$$

where f_r is frequency of the reference signal, N_{VCXO}, the counted VCXO cycles, and N_r, the counted cycles of the reference signal.

When using the measuring result to vary the control voltage, we assume that there exists a linear relationship between the variations of the control voltage and the VCXO frequency. If in EPROM the desired VCXO frequency is set up as f'_{VCXO}, and the measured VCXO frequency is f_{VCXO}, the desired variation is $\Delta f=f'_{VCXO}-f_{VCXO}$. The relationship between it and the variation of the control voltage is:

$$\Delta f=K_1 \Delta V \qquad (2)$$

where K_1 is the ratio of the frequency variation of VCXO to the variation of control voltage. There is a linear relationship between the variations of control voltage and the pulse width of the square wave control signal

$$\Delta V=K_2 \Delta T \qquad (3)$$

where ΔT is the variation of the pulse width. It is from 0 to the period of the square wave signal. K_2 is the ratio of the variation of control voltage to the variation of the pulse width. ΔV is obtained by ΔT through filtering and amplfying. From equations (2) and (3),

$$\Delta T=\Delta f/K_1K_2 = J\Delta f \qquad (4)$$

where J is relevant to the voltage - frequency characteristic of VCXO and the characteristics of the filter and amplifier. It can be decided by experiment and stored in EPROM. For different quartz crystals, filters and amplifiers, J is obviously different. It is convenient to decide J experimantally. ΔT is added to the pulse width of the square wave signal. After several periods of measurement and control the frequency of VCXO can be locked at f'_{VCXO}. Then through periodic measurement and control the frequency value can be kept constant.

The same hardware circuit can also be used to lock another VCXO with other frequencies, while the controlled crystal and the soft ware are changed. Two frequency synthesizers have been built now. We employ this device to lock a highly-stable crystal (BA14 crystal resonator), and built a 5.000017MHz frequency synthesizer. With the help of the same device another ordinary crystal (JA15 crystal resonator) is locked, and a 5.0001 MHz frequency synthesizer is built. For the former, the measuring gate time is larger than 1 second. In this case, the measuring accuracy is higher than 2×10^{-12}, the control accuracy can also be higher, and the high short term stability of the crystal can be kept in the synthesizer. For the latter the measuring gate time is 20 ms, the measuring accuracy is higher than 5×10^{-10}. The clock frequency of the microcomputer is 12 MHz, Its machine period is equal to 12/12MHz = 1 μs. It is considered to be the discrimination of the pulse width variation of the square wave signal. If the period of the square wave control signal is different for different synthesizers, the control accuracy and controlled frequency range is also different. For the former highly-stable VCXO, the period of the square wave can be 100 ms, and the cotrol discrimination is 1 μs/100ms= 1×10^{-6}. The control accuracy depends on the control discrimination and the controlled frequency range. The measurement and control operation period of the 5.000017MHz synthesizer is less than 2 seconds. For the latter ordinary VCXO, the period of the square wave is 4 ms, and the control discrimination is 1 μs/4ms= 2.5×10^{-4}. Its operation period is 40ms. The frequency stabilities of two VCXOs before and after their being locked are shown in Figure 4.

The measurement is accomplished with the reference signal of the synthesizers as the standards.

By using this new technique, some signal generators can be changed into synthesizers. One can instal a varactor circuit in the oscillation circuit of a generator to control its frequency. The result of the measurement for the output frequency of the generator is used as preset coarse frequency of the synthesizer, and is stored in the RAM of the microcomputer. This frequency can be varied as the desired synthesis frequency through a keyboard. By measuring the frequency of the generator further, computing, and comparing it with the set frequency, and varying the control voltage, the frequency of the generator can be varied accurately. In the new synthesizer, the regulator of the generator is now used as the coarse regulator, and the keyboard, as the fine regulator. The two regulators are combined, a high accuracy and stable lock can thus be accomplished in a very wide frequency range. The

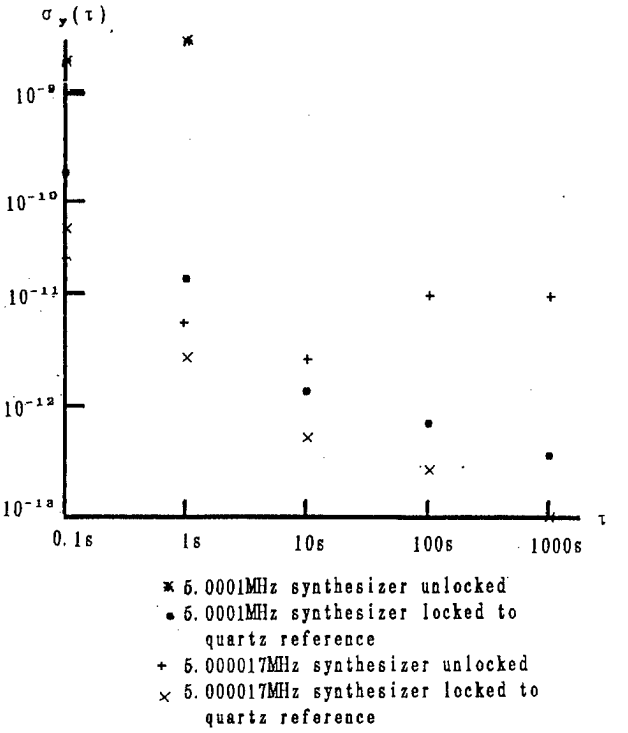


Figure 4

operation procedure of the device is shown in Fig. 5.

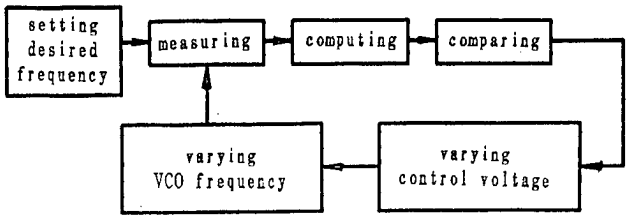


Figure 5.

In this synthesizer the frequency measuring device operates on the basis of the "dual broad-band phase detection method". In the frequency measuring device two different common oscillators are used alternatively, in a very wide frequency range the frequency can be measured accurately, and time response can be fast. A stable commercial generator can be rebuilt into a wide frequency range synthesizer by the use of this method and some frequency dividers. The frequency range is from 1Hz to over 100 MHz. After it has been rebuilt, the frequency drift of the generator (1×10^{-4}/half hour) can be eliminated, and the frequency stability is improved up to two orders (from 10^{-8}/second to 10^{-8}/second). The step can be 0.01 Hz to 10 Hz according to practical frequency range.

4. Conclusion

This new technique of frequency synthesis uses simple precision frequency measurement technique instead of the conventional frequency transformation and phase comparison methods. Its hardware circuit is simple and can be integrated, with its accuracy enhanced further and its frequency range very wide. For frequency synthesizers with different frequencies, the circuits are almost the same. Especially for the synthesis of complicated frequency signal, it has more advantage over other techniques. We can also use the soft ware instead of more hardwares. When combined with other techniques, it can find more applications. Besides synthesizers, many accurate signal generators can also be designed by means of this technique.

References:

1. Wei Zhou "The Greatest Common Factor Frequency and Its Application in the Accurate Measurement of Periodic Signals" 1992 IEEE Frequency Control Symposium, May, 1992

2. Wei Zhou, Zhougqiang Xuan "A Simple Precision Frequency Standard Comparator" 1993 IEEE International Frequency Control Symposium, June, 1993

3. Wei Zhou "A New Measurement Technique of Frequency and Periodic Signals - The Frequency Measuring Technique by Broad- band Phase Detection" will be published in Chinese Journal of Scientific Instrument.

4. Wei Zhou "A New Method of Time Difference Measurement: The Time Difference Method by Dual 'Phase Coincidence Points' Detection" 24th PTTI Applications and Planning Meeting, Dec. 1992

5. D.W.Allan, Herman Daams "Picosecond Time Difference Measurement System" 29th Annual Symposium on Frequency Control Symposium, pp. 404 - 411, 1975

6. Wei Zhou "A New Principle of Linear Phase Comparison - Irregular Phase Discrimination" 18th Annual PTTI Applications and Planning Meeting, Dec. 1986

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

The Composite DDS – A New Direct Digital Synthesizer Architecture

Lawrence J. Kushner
Lincoln Laboratory
Massachusetts Institute of Technology
Lexington, MA 02173-9108

Abstract–A new, low-power, high-speed, Direct Digital Synthesizer (DDS) architecture is presented, called the Composite DDS (CDDS). A low-speed, high-resolution DDS is combined with a high-speed, low-resolution phase accumulator and phase shifter via the serrodyne modulation technique. The low-speed circuitry provides the fine tuning while the high-speed circuitry provides coarse tuning. By minimizing the amount of circuitry required to clock at high-speeds, dc power is conserved.

The results from numeric simulations and a low-frequency proof-of-concept breadboard are presented. Progress on our 800 MHz CDDS development effort is described, followed by proposed enhancements to the CDDS architecture which promise improved performance.

Introduction

ALTHOUGH Direct Digital Synthesizers (DDSs) were invented decades ago [1], they did not come to play a dominant role in wideband frequency generation until recent years. Initially, DDSs were limited to producing narrow bands of closely spaced frequencies, due to the limitations of digital logic and Digital-to-Analog Converter (DAC) technologies. Larger bandwidths were generated by other means, such as phaselock loops, direct analog synthesis, or frequency multiplication, using the DDS to generate the fine resolution. For applications requiring fast tuning, these approaches resulted in expensive, complex, power-hungry, frequency generators.

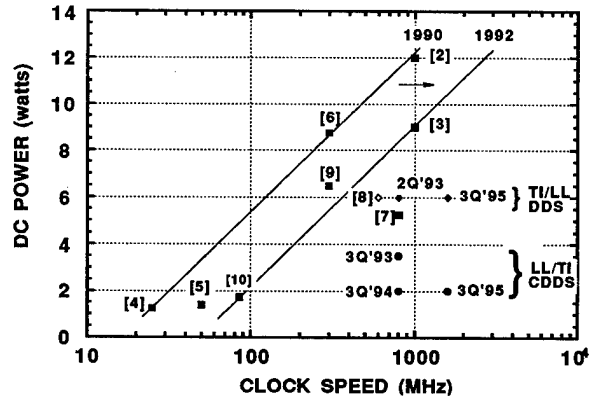


Figure 1 - DDS Power Consumption vs. Clock Speed. Data points based on references [4] - [10], and Lincoln / Texas Instruments projections (dates).

This work was sponsored by the US Army under Air Force Contract F19628-90-0002

Due to the rapid advance of digital technology, there is now an alternative approach. Direct Digital Synthesizers clocking at 1 GHz have recently become available [2]. Most, if not all, of the needed bandwidth can be generated by the DDS, resulting in reduced system complexity, size, and weight. While it is likely that DDS technology will continue to improve as digital technology advances, Figure 1 illustrates a key tradeoff that will remain: the faster the clock rate, the higher the dc power consumption.

While dc power may be unimportant in some systems, it is of utmost concern in many applications, such as portable, battery-powered communications equipment, and satellites. Additionally, even systems with power to burn may still have difficulty removing the heat generated by a high-power DDS.

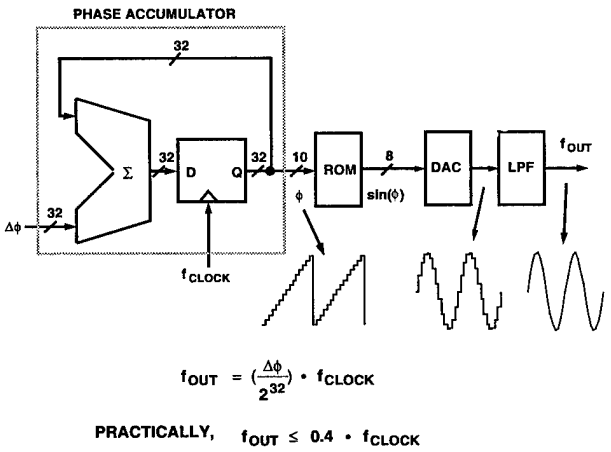


Figure 2 - Direct Digital Synthesizer Architecture. (Typical data-path widths shown)

The high-level architecture of a conventional DDS is shown in Figure 2. Since the frequency resolution of a DDS is equal to $f_{clock}/(2^N)$, where N is the number of bits in the phase accumulator, the greater the clock frequency, the more bits are needed to maintain a given resolution. 32-bit wide phase accumulators are commonly used in today's commercial DDSs. For a 1.6 GHz clock, 32 bits yields 0.37 Hz resolution.

The problem with this conventional approach is that on every clock cycle, the phase accumulator has to accumulate, the ROM has to produce a new value, and the DAC must convert another sample from digital to analog. All of this high-speed activity results in high dc power consumption.

0-7803-0905-7/93 $3.00 © 1993 IEEE

Composite Direct Digital Synthesizer (CDDS)

In contrast, the Composite DDS [11] uses a new architecture (Figure 3), allowing it to achieve high speed and low power consumption simultaneously. Instead of clocking the entire circuit at the maximum clock rate, the CDDS partitions the circuitry into a small, high-speed circuit that generates coarse frequency steps, and a larger, low-speed circuit to generate the fine frequency resolution. The fine-resolution frequency (f_1) is upconverted by the coarse-resolution frequency (f_2) via the serrodyne modulation technique (Figure 4).

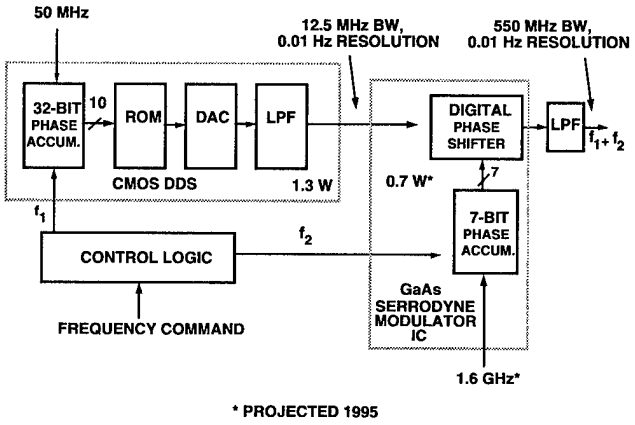


Figure 3 - Composite DDS Architecture (nominal clock frequencies and data-path widths shown for clarity).

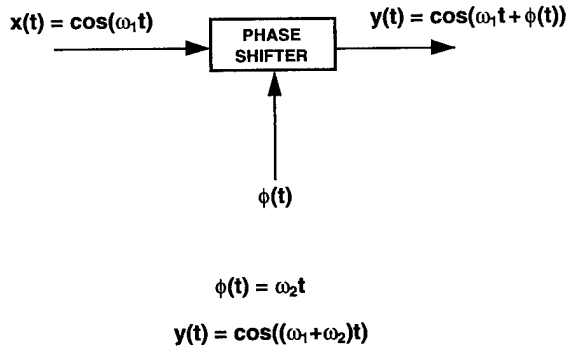


Figure 4 - Serrodyne Modulation Principle [12]-[14].

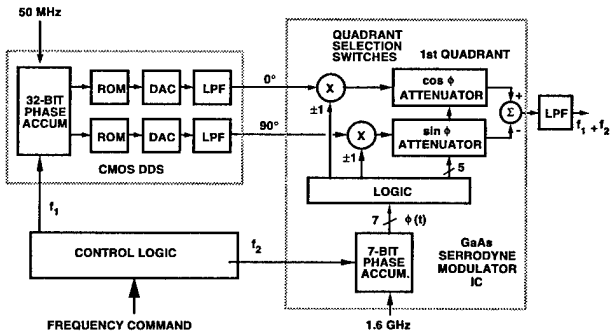
Since its inception in the late 1950's, the serrodyne principle has been used in radar applications to chirp fixed, microwave carriers at a relatively low rate [12]-[18]. In modern implementations of these serrodyne systems, a fixed microwave carrier is passed through microwave phase-shifters driven from a digital phase-accumulator, resulting in small frequency shifts and a narrow-band (< 1 octave) output.

In contrast, the CDDS uses a coarsely tuned, high-speed phase accumulator driving a fast-switching, low-frequency phase-shifter to translate a finely-tuned, low-frequency "carrier" by large amounts. The CDDS uses the serrodyne technique for broadband bandwidth expansion, resulting in a decade or more of output bandwidth. For the specific clock frequencies and bandwidths shown in Figure 3, the low-frequency DDS produces 12.5 MHz of bandwidth with 0.01 Hz of resolution, which is then upconverted to one of 44 different "bins" by the Serrodyne Modulator, resulting in 550 MHz of output bandwidth.

For GigaHertz operation with today's technology, the CDDS's high-speed circuitry could be implemented with GaAs MESFETs, GaAs HBTs, or high-speed Si bipolar transistors, while the low-speed circuit might be built in CMOS. Alternatively, a lower-frequency CDDS might be created entirely in CMOS, with a high frequency clock of ≈80 MHz and low-frequency clock of ≈2.5 MHz, again resulting in considerable power savings over a conventional DDS in which all of the logic clocks at the maximum clock frequency (80 MHz).

As digital logic technologies advance, so will these clock rates, but the underlying advantage of the CDDS will remain. By minimizing the amount of circuitry clocking at high speeds, dc power is conserved. With the addition of control circuitry (Figure 3), the CDDS will appear to the outside world as a standard DDS . By choosing the two CDDS clock frequencies to be related by a factor of 2, the amount of control logic needed can be minimized.

While phase-shifters could be built at these low frequencies using broadband all-pass networks [19], [20], the low-frequency poles and zeros associated with these networks would limit the phase-shifters' switching speeds. Instead, the Composite DDS (Figure 5) employs a vector modulator phase-shifter [21], made up of simple switches and attenuators (or amplifiers). These components are inherently broadband, and switching speed is limited only by energy storage in the switches' parasitic capacitances.



$\phi(t) = 2\pi f_2 t$

$\cos(2\pi f_1 t+\phi(t)) = \cos(2\pi f_1 t)\cdot\cos(\phi(t)) - \sin(2\pi f_1 t)\cdot\sin(\phi(t))$

Figure 5 - Composite DDS Architecture with 4-Quadrant Vector-Modulator Digital Phase-Shifter.

Additionally, this approach is much more amenable to integrated-circuit implementation than an allpass network-based design. The primary drawback to the vector modulator approach is that there are now two channels that must closely track each other for optimum performance. As discussed below, channel mismatch is the source of one of the close-in spurious signals in the Composite DDS's output spectrum.

Simulations and Frequency-scaled Proof-of-Concept Hardware

In order to study the spurious performance of a "real world" Composite DDS, a frequency-scaled, proof-of-concept CDDS was constructed (Figure 6). In this demonstration, the high-speed accumulator was clocked at 16 MHz while the low-speed circuitry was clocked at 500 kHz. Figure 7 shows the typical measured output spectrum from this proof-of-concept CDDS. Besides the desired 3.1 MHz output signal, a number of spurious components are also present. None of these would be present in a perfect CDDS implementation. Before attempting to build a higher frequency CDDS, it was essential to understand how all of these spurious frequency components were generated.

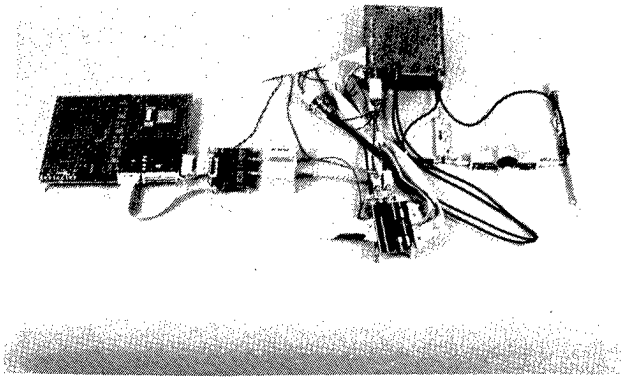
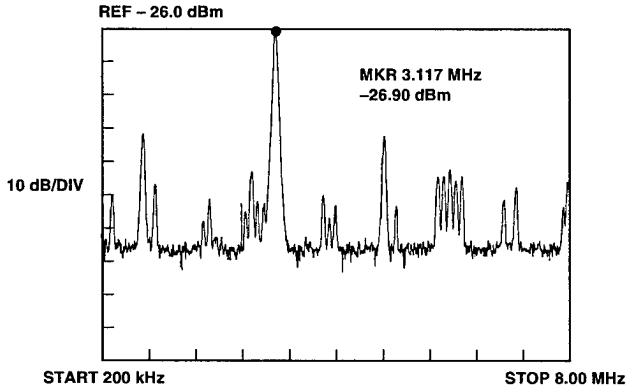


Figure 6 - Proof-Of-Concept Breadboard.



	f_{CLOCK}	f_{OUT}
7-BIT ACCUMULATOR	16 MHz	3 MHz
32-BIT DDS	0.5 MHz	0.1 MHz

Figure 7 - Proof-Of-Concept Breadboard Output Spectrum.

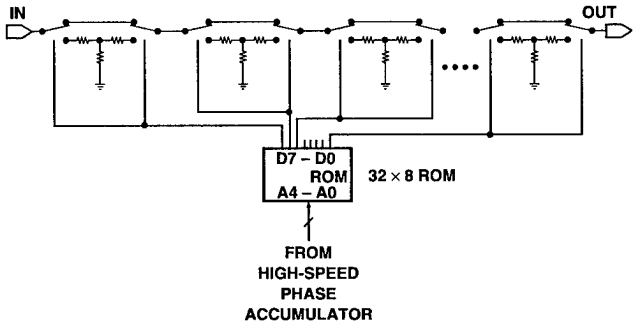


Figure 8 - Switched-Attenuator – ROM / Cascade Implementation.

In this breadboard CDDS, the vector modulator's switched-attenuator was implemented as a cascade of eight switched-attenuator sections controlled by a ROM lookup table (Figure 8). This approach has two fundamental problems: amplitude quantization (analogous to a DDS's DAC) and timing skew. Amplitude quantization leads to amplitude errors, even with perfect attenuators. For our breadboard CDDS these quantization effects were relatively small. We used a 7-bit high-speed phase accumulator, with the bottom 5 bits driving the switched-attenuator. We chose our eight attenuator bit weights to be approximately 16 dB, 8 dB, 4 dB, ... 0.125 dB, so with ideal attenuators, our worst case amplitude error would be 0.0625 dB.

The timing problem associated with this switched-attenuator implementation is even more insidious. Due to the finite time delay required for the signal to pass through each attenuator section, the switch control-signals must be skewed in time to line up with the signal as it flows through the attenuator chain. Even if the control-signal skews could be optimized at one temperature, maintaining the desired skew over temperature would present a difficult challenge. Additionally, there is typically some dispersion through the switched-attenuator sections due to their finite bandwidth, further complicating matters.

Figure 9 identifies most of the CDDS spurious sources. Besides the two sources described above, other sources include attenuator-section VSWR interaction, switch distortion products, switch glitches, quadrant-switch LO leakage, and channel imbalance. A numeric simulation of the CDDS was written (using MATLAB [22]) to further study the spurious sources. The simulation modeled phase and amplitude errors, harmonic distortion, intermodulation distortion, leakage, and imbalances. The simulation did not include the effects of switching glitches or timing errors.

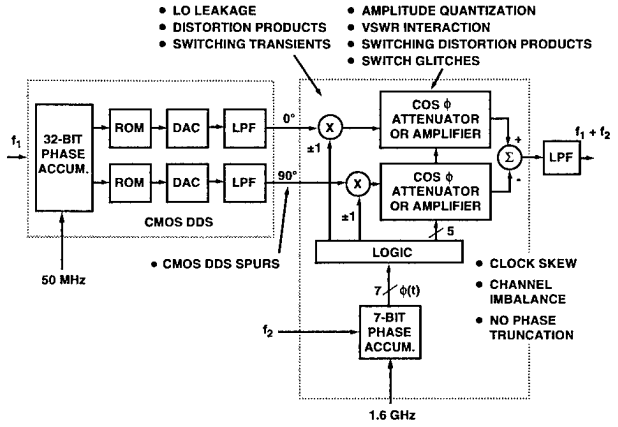
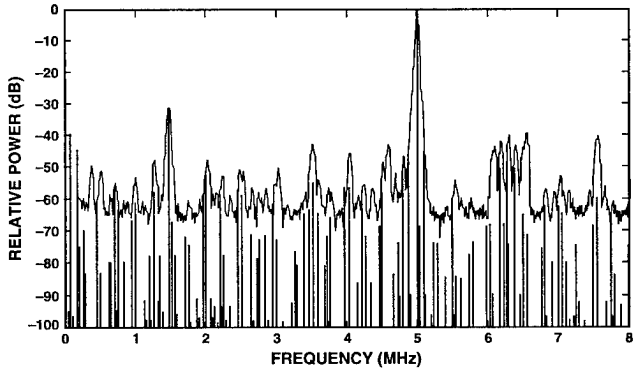
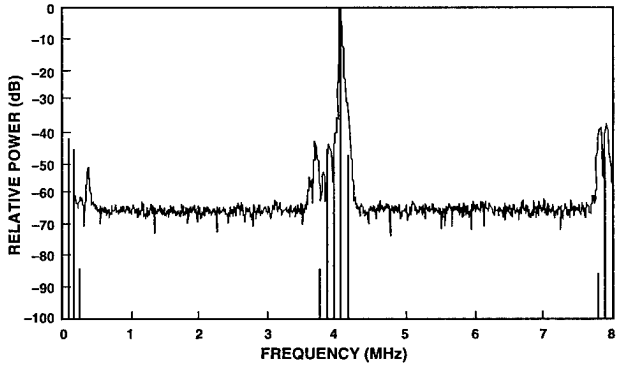


Figure 9- CDDS Spurious Sources.



	f_{CLOCK}	f_{OUT}
7-BIT ACCUMULATOR	16 MHz	4.875 MHz
32-BIT DDS	0.5 MHz	0.1 MHz

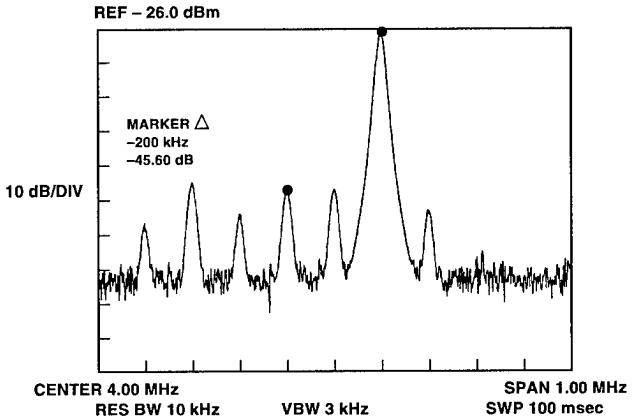
Figure 10 - Measured and Simulated Spectrum, f_{OUT} = 4.875 MHz.



	F_{CLOCK}	F_{OUT}
7-BIT ACCUMULATOR	16 MHz	4 MHz
32-BIT DDS	0.5 MHz	0.1 MHz

Figure 11 - Measured and Simulated Spectrum, f_{OUT} = 4.1 MHz.

Figures 10 and 11 show simulated spectrum superimposed on measured output spectrum, for a "dirty" and "clean" frequency, respectively. Since the simulation neglected glitch transients and timing skews, it was overly optimistic in its prediction of the amplitudes of some of the spurious signals. However, it did a good job of predicting their locations, allowing most of the sources of the spurious signals to be identified. The simulation also allowed parameters to be varied in order to study the output spectrum's sensitivity to them.



	f_{CLOCK}	f_{OUT}
7-BIT ACCUMULATOR	16 MHz	4 MHz
32-BIT DDS	0.5 MHz	0.1 MHz

Figure 12 - Measured Close-in Spectrum, f_{OUT} = 4.1 MHz.

Most of the spurious sources associated with the quadrant switches, along with channel imbalance errors, result in close-in spurious components. Figure 12 shows a zoomed-in plot of the spectrum of Figure 11. This 4.1 MHz output is generated from a 0.1 MHz fine-tune signal from the low-speed DDS, upconverted by a 4 MHz signal from the 7-bit high-speed accumulator. Notice the 4 MHz LO leakage, the 3.9 MHz lower sideband, and the 2-1 and 3-1 mixer products. Improving the mixer symmetry reduces the absolute level of the LO leakage, while increasing the rf signal level reduces the relative amplitude of this LO leakage. This rf signal level cannot be increased arbitrarily, however, since the larger the signal level, the larger the 2-1 and 3-1 spurs.

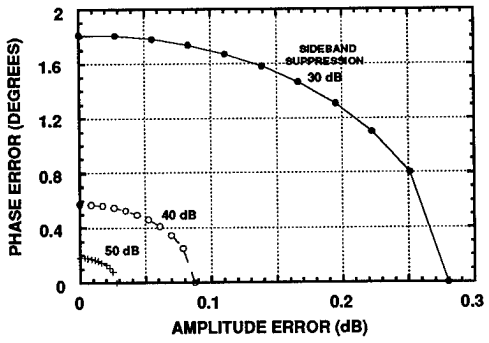


Figure 13 - Single-Sideband Mixer Phase/Amplitude Tracking Requirements.

The 3.9 MHz lower sideband level is determined totally by channel matching, just as in a single-sideband mixer. Figure 13 shows the channel tracking requirements that have to be maintained for a given level of sideband rejection. In order to achieve -40 dBc lower sideband levels, amplitude and phase tracking must be within ± 0.05 dB and ± 0.4°, respectively. While this level of matching would be difficult to achieve in discrete form, we hope to meet these goals by building the two channels on the same integrated circuit chip.

800 MHz Development Plans

Having learned from our low-frequency breadboard and simulations, we decided to replace the cascade switched-attenuator implementation of Figure 8 with the decoded-amplifier architecture shown in Figure 14. In this scheme, every phase state has its own fixed-gain feedback amplifier, eliminating any amplitude quantization effect. Since the signal now only passes through a single switch, the timing issue associated with the attenuators is eliminated. The VSWR interaction problem has been virtually eliminated, since there are now much fewer stages that the signal must pass through, and since the amplifiers provide input/output isolation. One last advantage of this switched feedback amplifier approach is that each switch sees the same drive impedance, which should improve transient response uniformity and push much of the switching noise up to the clock frequency.

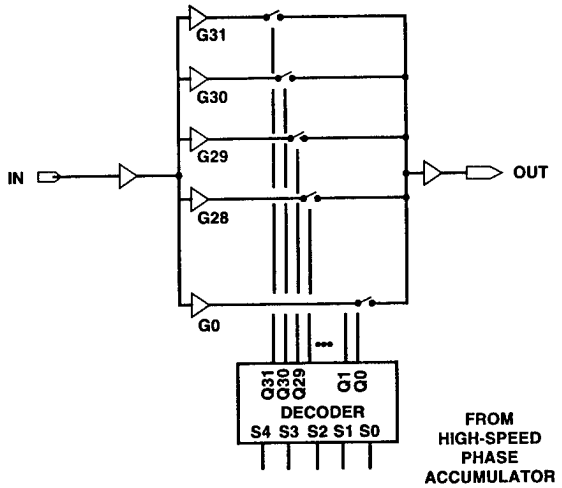


Figure 14 - Switched-Attenuator – Decoded Amplifier Implementation.

An 800 MHz CDDS "breadboard" has been designed and is currently being fabricated. This first high-speed CDDS consists of a commercially-available CMOS low-speed DDS and a custom GaAs single-chip Serrodyne Modulator IC (Figure 15). This GaAs chip is a collaboration between Lincoln Laboratory and the GaAs Signal Processing Technology Group of Texas Instruments, using TI's HI^{2}L HBT logic [23]. The architecture and analog design of this chip are Lincoln's responsibility, while TI is responsible for the digital design and I.C. fabrication. The first GaAs chips are expected to be delivered by July, 1993, and will be described in a later paper. The goal of this first breadboard is an 800 MHz CDDS, consuming <4W of dc power, and delivering -35 dBc worst-case spurious performance.

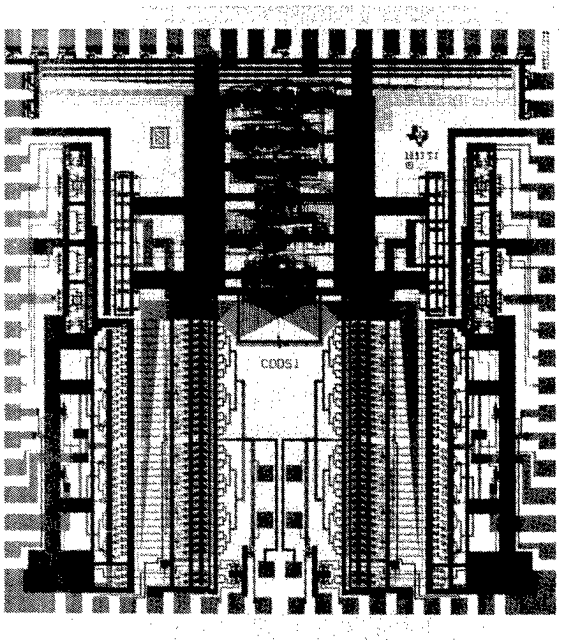


Figure 15 - 800 MHz GaAs HBT Serrodyne Modulator Integrated Circuit – CDDS1.

CDDS Architecture Enhancements:

Figure 16 shows a proposed improvement to the CDDS architecture. The order of the quadrant switches and the

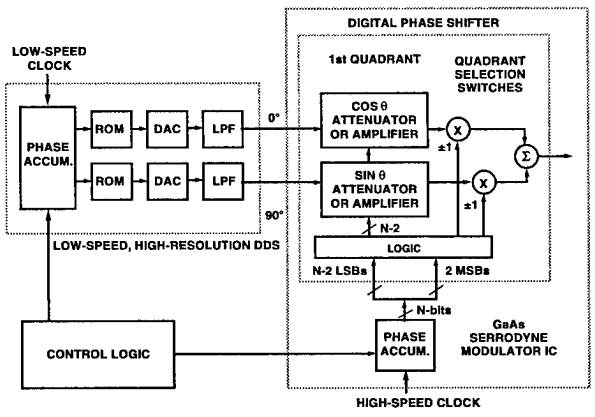


Figure 16 - CDDS – Improved Architecture.

first-quadrant switched-amplifiers has been reversed. While at first glance this seems like a trivial change, this architecture provides at least two improvements over our earlier approach: reduced clock-skew error, and reduced switched-amplifier bandwidth. While the clock skews inside the switched-attenuators themselves have been eliminated by going to the decoded-amplifier version of

Figure 14, there is still the timing of the quadrant-switches relative to the switched-amplifiers to contend with. Since the switched-amplifier sections have the N-way switch on their outputs, reversing the order as in Figure 16 puts these amplifier switches immediately preceding the quadrant switches, simplifying the timing requirements.

Since in this new implementation the switched-amplifiers only must pass the low frequencies of the low-speed DDS, the hardware implementation is made much easier. For example, in our 800 MHz CDDS, the amplifier bandwidth drops from 350 MHz to 20 MHz. This is quite significant for two reasons: it makes tight amplitude / phase tracking through the two channels easier to achieve, and it allows higher impedances to be used, resulting in reduced dc power consumption.

A second proposed improvement is shown in Figure 17. In addition to the order-reversal of Figure 16, the design has been made fully balanced. While the switched-amplifiers and quadrant-switches could be individually balanced, instead, the two circuits were merged, resulting in a balanced amplifier/switch combination. This merging reduces dc power consumption and complexity, eliminating the need for two broadband 180° signal splitters. Balancing the design has a number of advantages, including common-mode rejection (switching glitches, digital noise, and dc level shifts) and even-order spur. cancellation. The complexity of the balanced design may not be any higher than that of the unbalanced design, since the dc-level compensation circuitry that was present in the unbalanced design may be eliminated.

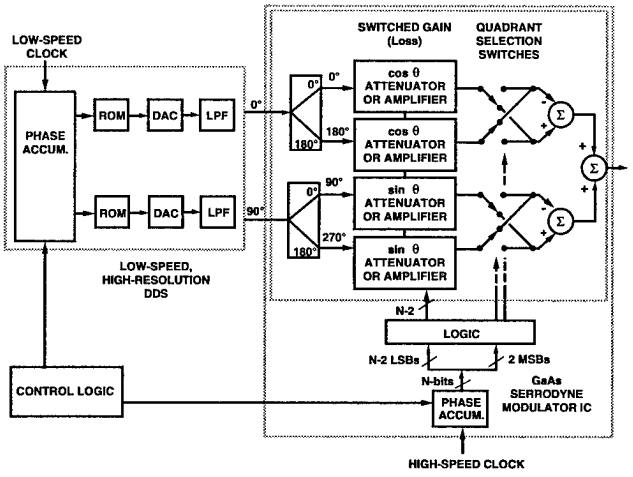


Figure 17 - Balanced CDDS.

After simulating these enhancements, a second pass GaAs Serrodyne Modulator I.C. design will be fabricated that will incorporate them along with additional improvements. By eliminating unneeded circuitry and by optimizing the design, the dc power of this second pass design is expected to drop to the 2 W goal, with the CMOS DDS consuming 1.3 watts, and the GaAs Serrodyne modulator IC consuming 0.7 watts. Advances in technology will continue to push the power down, and the speed up. We hope to achieve our 1.6 GHz, 2 W, CDDS goal by 1995.

Acknowledgments:

The author would like to thank Dave Snider, Ron Bauer, and the rest of Lincoln's management for their support and encouragement, and Marcus Ainsworth for building all of the CDDS breadboard hardware. The author would also like to acknowledge the contributions of Texas Instruments in making the GaAs Serrodyne Modulator chip: Van Andrews for the digital design along with the overall TI program management responsibilities, Bill White for consultation with the analog design, and Paul Garner and Dallas White for the layout of the chip.

References:

[1] J. Tierney, C. M. Rader, and B. Gold, "A Digital Frequency Synthesizer,", **IEEE Transactions on Audio Electroacoust.**, AU-19, 43-57, March 1971.

[2] Stanford Telecom, STEL-2373 data sheet, May 1991.

[3] Stanford Telecom STEL-2373, personal communication with P. Chizinski of STEL, Dec. 1992.

[4] Sciteq DDS-1 data sheet, 30 April 1991.

[5] Stanford Telecom STEL-1177 NCO data sheet, February, 1990, and Fujitsu 40978 DAC data sheet, Edition 3, September 1990.

[6] Stanford Telecom STEL-2172 NCO and Honeywell HDAC51400, as measured by Harry Wolfson & John Murphy of Lincoln Laboratory, October 1989.

[7] Plessey SP2002 data sheet (effective clock) May 1991.

[8] V. Andrews, M. Brown, W. White, P. Wang, R. Meyer, J. Vaal, "A Monolithic GaAs DDS for a Digital Radio Application," **1992 GOMAC Conference Digest**, pp. 431-434.

[9] Motorola 120DDSEVK data sheet, 1992.

[10] Analog Devices AD9955 data sheet, Rev.0, and AD9712B data sheet, Rev.0.

[11] patent pending.

[12] R. C. Cumming, "The Serrodyne Frequency Translator," **Proc. of the IRE**, Feb. 1957, pp. 175-186.

[13] —, "Method and Apparatus for Translating the Frequency of a Signal," U.S. Patent # 2,927,280, March 1960.

[14] —,, "Serrodyne Performance and Design," **The Microwave Journal**, Sept. 1965, pp. 84-87.

[15] G. Klein and L. Dubrowsky, "The DIGILATOR, a New Broadband Microwave Frequency Translator," **IEEE Transactions on Microwave Theory and Techniques**, vol. MTT-15, March 1967, pp. 172-179.

[16] S. Mitchell, J. Wachsman, G. Lizama, F. Ali, A. Adar, "Wideband Serrodyne Frequency Translator," **Applied Microwave**, Summer 1990, pp. 58-67.

[17] M. Topi, "An Eight-Phase Broadband Serrodyne Modulator," **1983 IEEE MTT-S Symposium Digest**, pp. 432-434.

[18] I. W. Smith and M. J. Schindler, "Serrodyne Modulator," U.S. Patent # 5,063,361, Nov. 5, 1991.

[19] R. B. Dome , "Wideband Phase Shift Networks," **Electronics**, December 1946, pp. 112-115.

[20] F. R. Shirley, "Shift phase independent of frequency," **Electronic Design** 18, 1 September, 1970, pp. 62-66.

[21] M. Tuckman, "I-Q Vector Modulator–The Ideal Control Component?," **Microwave System News and Communications Tech.**, May 1988, pp. 105-115.

[22] MATLAB, The MathWorks, Inc., Natick MA.

[23] H-T. Yuan, H-D Shih, J. Delaney, C. Fuller , "The Development of Heterojunction Integrated Injection Logic," **IEEE Transactions on Electron Devices**, vol. 36, no. 10, October 1989, pp. 2083-2092.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

INDUCED END-OF-LIFE ERRORS IN A FAST SETTLING PLL

Thomas J. Endres and Gar T. Calvetti

HUGHES AIRCRAFT COMPANY
Bldg. R10, M/S S320
P.O. Box 92919
Los Angeles, CA 90009

James B. Kirkpatrick

NIPPONDENSO TECHNICAL CENTER
LA Laboratories
2131 Palomar Airport Rd., Suite 200
Carlsbad, CA 92009

ABSTRACT

Undesired exponential transients, for example those caused by voltage-controlled oscillator (VCO) post tuning drift, loop filter component mismatch [17], timing errors, etc., may influence a fast settling phase locked loop (PLL) with satisfactory beginning-of-life (BOL) performance such that end-of-life (EOL) performance is not guaranteed when EOL errors are incurred by the system. An unintrusive method for inducing predicted EOL errors is presented. Corresponding effects are studied and verified via PSpice computer simulations and empirical data from a 5-GHz spaceborne PLL.

FAST SETTLING PLL INTRODUCTION

A PLL architecture can be employed for fast hopping frequency synthesis (see Figure 1). The assumption of linearity allows the loop to be treated as a classical control system and modeled via Laplace techniques. The following notation and dimensions will be used throughout this text:

MODULE	GAIN	UNITS
Phase Detector	K_d	volts/rad
VCO	K_v	rad/volt/sec
Divide-by-N	1/N	—
Loop Filter	F(s)	—

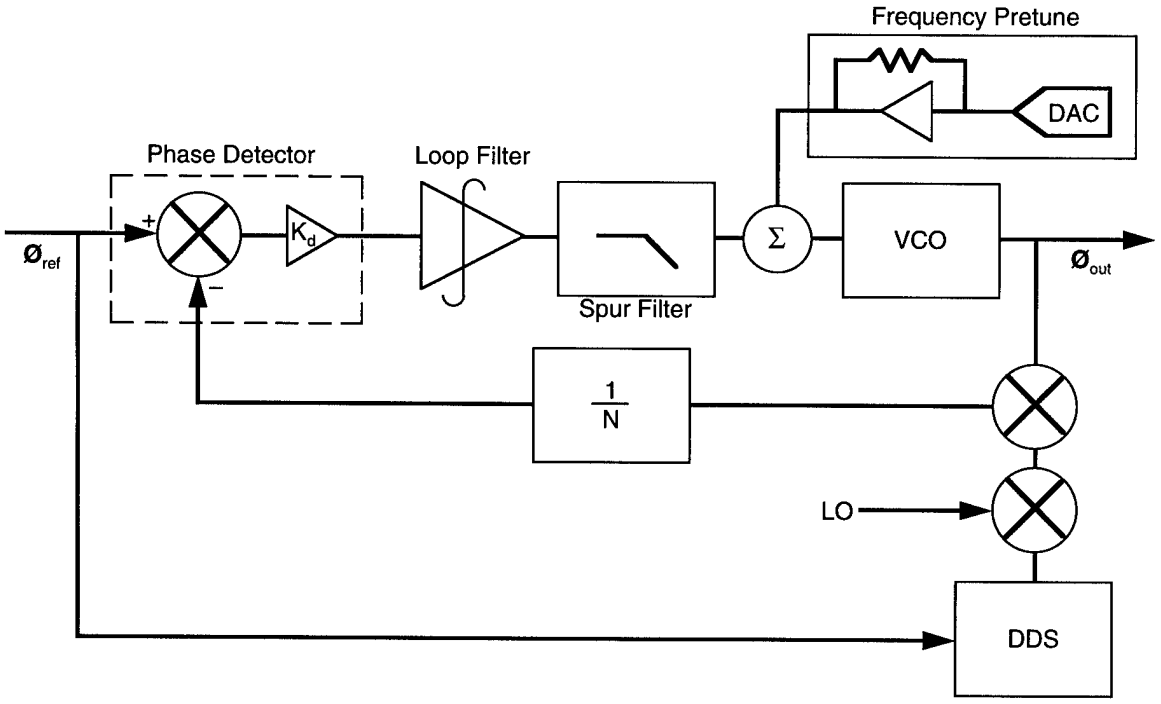


Figure 1. A PLL With DDS Injection As a Frequency Synthesizer

0-7803-0905-7/93 $3.00 © 1993 IEEE

When the loop is locked, the phase detector compares the divide-by-N output frequency with the reference frequency such that the two are equal. The phase detector's output is a voltage or current based on the phase difference between these two signals. This phase-sensitive signal is then integrated by the loop filter, which in turn slews the VCO to the proper frequency. In this manner the loop remains locked to the reference frequency.

Typically, a second order, type 2 loop (two integrators in the loop path) is chosen to permit a small static phase error and nonzero tuning voltage for the VCO [4], [6], [20]. The loop filter for this implementation is ideally modeled as [13]

$$F(s) = \frac{s\tau_2 + 1}{s\tau_1}$$

Second order system parameters can be defined

$$\omega_n = \sqrt{\frac{K_d K_v}{N\tau_1}}$$

$$\zeta = \frac{\omega_n \tau_2}{2}$$

and the closed loop transfer function is then

$$\frac{\varnothing_{out}(s)}{\varnothing_{in}(s)} = \frac{N\left(2\zeta\omega_n s + \omega_n{}^2\right)}{s^2 + 2\zeta\omega_n s + \omega_n{}^2}$$

The ideal time response due to a reference frequency step for an underdamped system is easily found [6].

$$\varnothing_{error}(t) = \frac{\Delta\omega}{\omega_n\sqrt{1-\zeta^2}} \sin\left(\sqrt{1-\zeta^2}\,\omega_n t\right) e^{-\zeta\omega_n t}$$

The phase error present at the VCO output is then N times that at the phase detector. This is analogous to an output frequency step of N times that at the reference [4].

The direct digital synthesizer (DDS) and upconversion chain in Figure 1 are used for fine frequency resolution and are discussed under LO Perturbation.

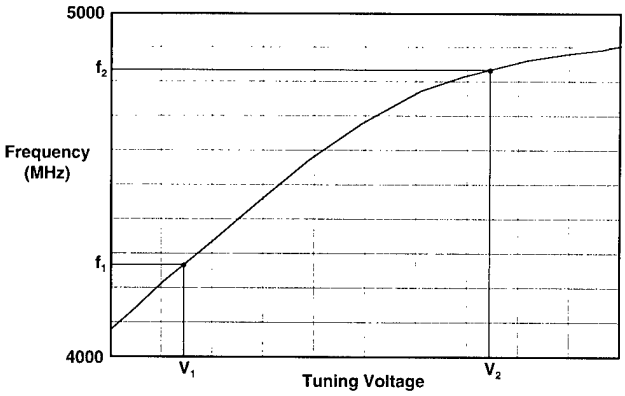


Figure 2. VCO Tuning Curve With Gain Compensation

LOOP STRESS ERROR

A fast settling PLL may employ a VCO with a nonlinear tuning curve (Figure 2).

Specifically, the gain, defined as the slope of the curve in Figure 2 or modulation sensitivity, may be chosen to compensate for the divide-by-N ratio; K_V/N is approximately constant. The loop dynamics thus do not wander as the loop is hopped and critical parameters such as phase noise are more predictable.

Typically, the loop requires some type of steering or help circuitry to settle quickly for a moderate frequency step. Figure 1 includes a pretune circuit that receives digital commands to output a stepped dc voltage. This voltage is superimposed with the loop filter output voltage to tune the VCO.

$$V_{TUNE} = V_{PT} - V_{LF}$$

The difference in the loop filter voltages for two tuning states is defined as loop stress error (LSE). For two points on the VCO tuning curve, f2 > f1, if $V_{LF2} > V_{LF1}$, the LSE is defined as positive; otherwise, LSE is negative.

$$LSE = V_{LF2} - V_{LF1}$$

When the loop is hopped, the pretune is commanded synchronously with the divide-by-N and coarsely tunes the VCO to a frequency close to $N \times f_{REF}$. The effective LSE necessary for the loop filter to track out is decreased,

aiding the loop in reaching final phase more quickly. This idea is demonstrated in Figure 3; an ideal loop (ζ = 0.9, ω_n = 290 krad/sec) is stepped at the output 500 MHz, as without a pretune, and 60 MHz (400 mV × 150 MHz/V), as with the aid of a pretune.

The ideal response is loosely dependent on the LSE.

DISTURBANCE MODEL

This dependence on LSE is significant when non-ideal inputs are considered. Specifically, many phenomena (loop filter component mismatch [17], VCO post tuning drift, power supply modulation, timing errors, crosstalk, etc.) manifest themselves as undesired exponential transients. Assuming a linear system and employing Laplace techniques, the exponential transient can be modeled as a disturbance into the loop and superposition employed (see Figure 4) [5], [11].

Suppose the VCO is stepped as from a divide-by-N change. The resultant phase error is the superposition of the ideal second order system's response and a component due to the disturbance input.

$$\varnothing_{out}(s) = \frac{\Delta\omega_n\left(2\zeta\omega_n s + {\omega_n}^2\right)}{s^2\left(s^2 + 2\zeta\omega_n s + {\omega_n}^2\right)} + \frac{AK_v s}{\left(s^2 + 2\zeta\omega_n s + {\omega_n}^2\right)\left(s + \frac{1}{\tau}\right)}$$

For small BOL LSEs, the amplitude of the undesired disturbance may be negligible so that its superposition with the ideal response is not noticeable. The system's behavior will then approximate ideal. However, as parameters in the loop age, the LSE seen by the loop filter may grow so that the disturbance amplitude, when multiplied by a high-gain VCO, may significantly contribute to the above superposition. Indeed, for $\tau >> 1/\zeta\omega_n$, the ideal second order system response is dominated by the disturbance response such that

$$\varnothing_{out}(t) \cong \frac{AK_v}{{\omega_n}^2\left(\tau_2 - \tau\right) - \frac{1}{\tau}}\, e^{\frac{-t}{\tau}}$$

This disturbance response (τ = 20 μsec ~ 5 × $1/\zeta\omega_n$, A = 10 mV) and ideal second order system response due to an output frequency step of 60 MHz are plotted in Figure 5.

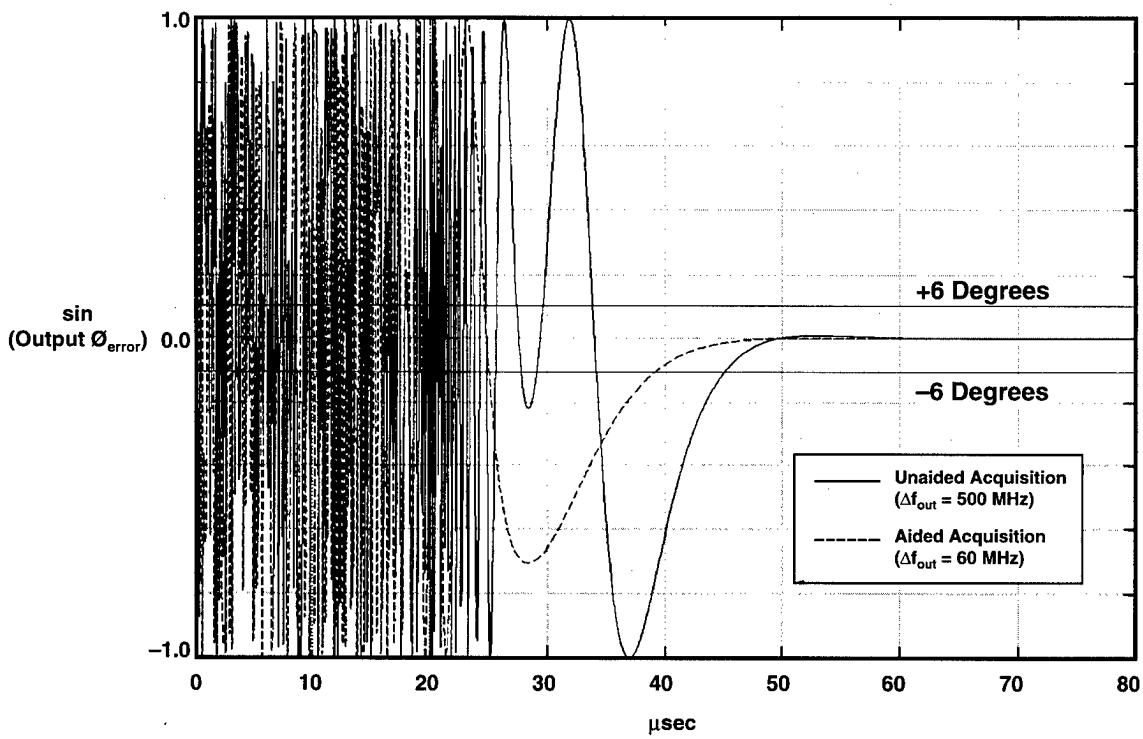


Figure 3. Ideal Phase Error Versus Tine, ζ = 0.9, ω_n = 290 krad/sec, (a) fstep = 500 MHz and (b) fstep = 60 MHz

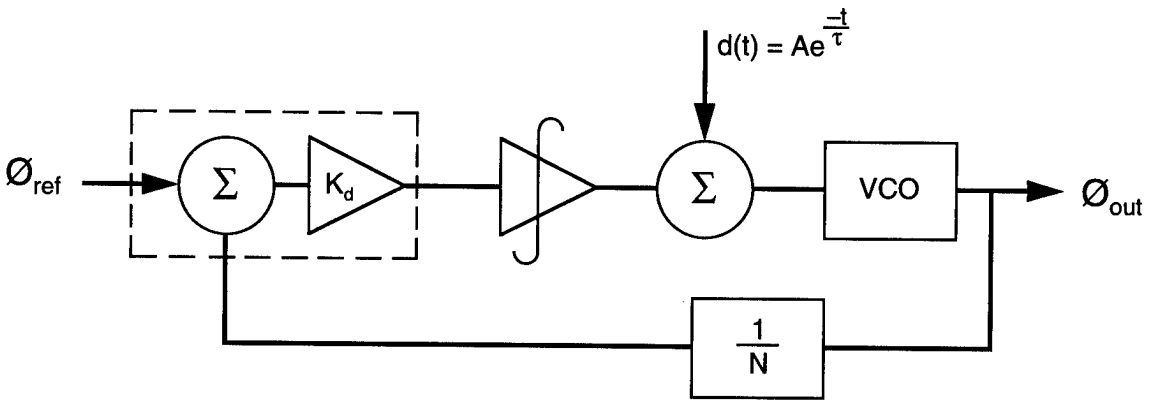


Figure 4. Undesired Exponential Transient Modeled As a System Disturbance

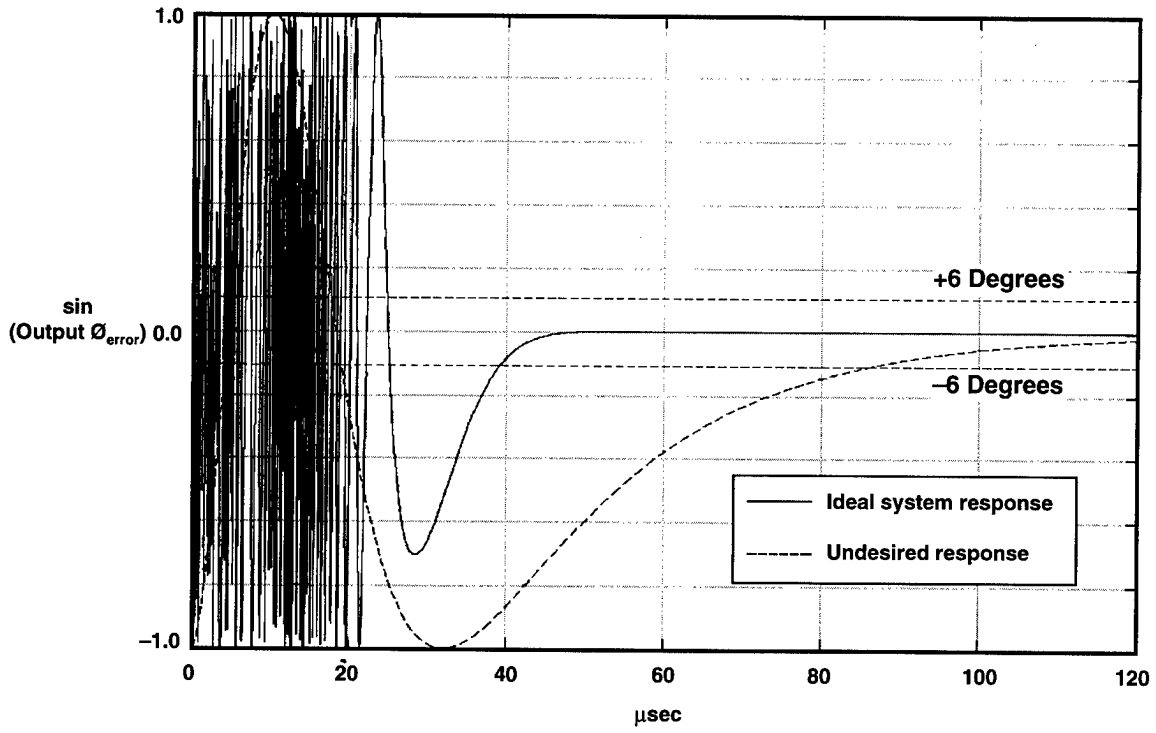


Figure 5. (a) Ideal Response to an Out Frequency Step of 60 MHz and (b) Undesired Response to Disturbance (τ = 20 μsec ~ 5 × 1/$\zeta\omega_n$, A = 10 mV)

As much as 100-percent degradation in 6-degree settling time is seen due to the exponential disturbance. Such a response may be typical at EOL, but nonexistent at BOL. Hence, inducing an expected EOL LSE may be desirable to predict EOL performance, revealing a potential catastrophe.

WCCA

A priori knowledge of expected EOL LSEs is required so that sensitivity to this parameter may be determined. Typically, worst case circuit analysis (WCCA) is required.

If a digital pretune circuit (Figure 6) is used for aided acquisition, the circuit can be shown to be a source of LSE. Variations between the pretune mapping of an ideal and an actual VCO tuning curve manifest themselves as LSE.

An upper and lower bound of pretune voltage for a given command is needed. Specifically, the stepped voltage from the pretune is translated from a current at the digital-to-analog converter (DAC) output through an op amp acting as a transimpedance amplifier. Variations in DAC output current may arise from power supply variations, DAC nonlinearities, and radiation effects. These components will act to produce an upper and lower bound of the DAC output current. It is reasonable to assume the processes are random and independent so that each parameter's contribution to DAC output current variation may be root-sum-squared (RSS'd), rather than linearly added [18], [19]. The expected current variation is then translated to a voltage error window through the op amp circuit. Resistor tolerance of the op amp's feedback resistor will act to broaden the voltage error window, as well as any other resistive division following the op amp.

Variations in the VCO tuning curve may translate directly to LSE contributions. Such variations may arise

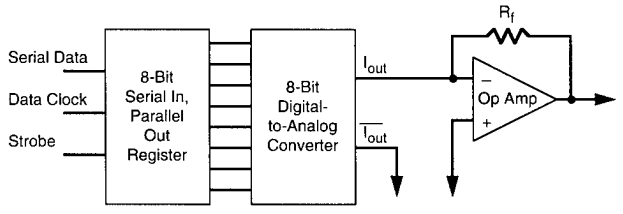


Figure 6. Digital Pretune Circuit

because of temperature changes, power supply variations, load pulling, aging, radiation, and other similar factors. Typically, tuning curve sensitivity to these parameters is empirically measured and changes in tuning voltages translate directly to LSEs. Note that a linear shift in frequency for a nonlinear VCO tuning curve will correspond to an LSE, whereas for a linear tuning curve, such a change incurs no LSE. Typically, temperature effects cause a linear shift and therefore may be the largest contributor to LSE caused by VCO tuning curve changes.

Pretune errors and VCO contributions are added, yielding an expected worst case upper and lower bound on LSE seen by the loop. Ideally, these extremes should be exercised by the system for transient sensitivity.

PSPICE LOOP MODEL

A simple PLL model is developed in PSpice (Figure 7).

The loop input is a stepped dc voltage to the VCO, representing an LSE and causing a frequency step out of the VCO.

The differential phase detector is modeled with ideal voltage-controlled voltage sources (VCVSs) with positive and negative gains. This model is reasonable because the loop does not slip cycles.

Similiarly, the VCO is modeled as an ideal integrator. The gain, K_V, is changed easily via choice of its feedback capacitor.

The divide-by-N is simply a resistor divider because the state variable is voltage.

The low-pass filter contains Rs and Cs and is used to attenuate the reference frequency and its harmonics.

The dual function (differential amplifier and integrator) loop filter is explicitly modeled with Rs and Cs and an ideal VCVS. The disturbance input is realized by offsetting the loop filter components by 2 percent. This offset has the effect of an exponential disturbance at the loop filter's output with $\tau \sim 5 \times 1/\zeta\omega_n$ [17]. Hence, the disturbance model presented earlier is valid.

Consider loop dynamics $\zeta = 0.9$, $\omega_n = 290$ krad/sec. Phase settling time to 6 degrees of final phase is simulated and plotted versus induced LSE (Figure 8) with (1) no

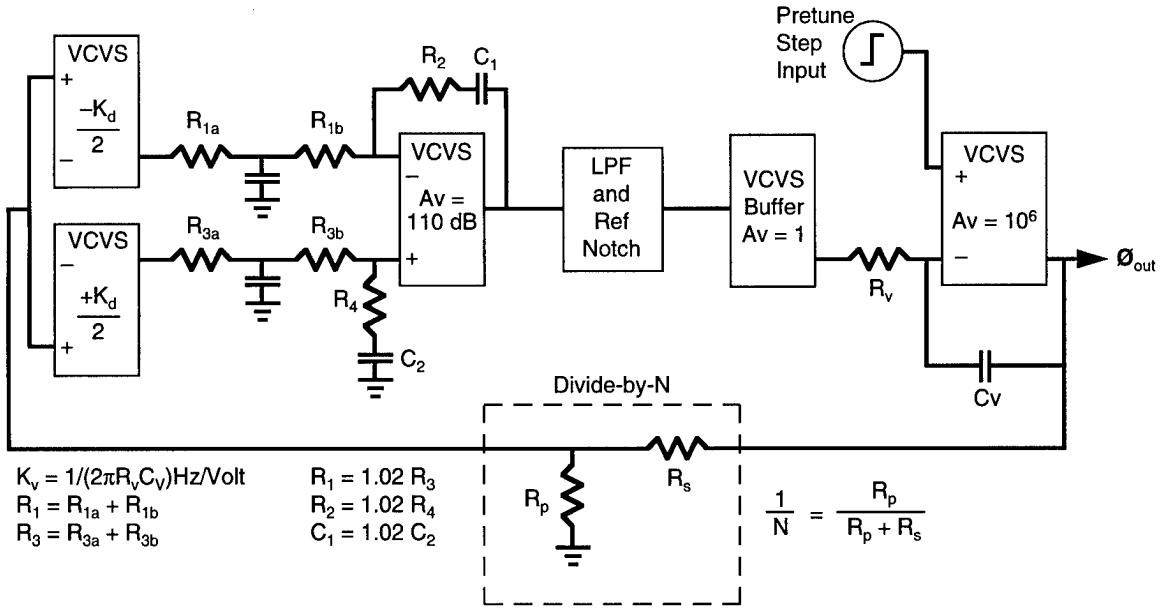


Figure 7. PSpice PLL Model

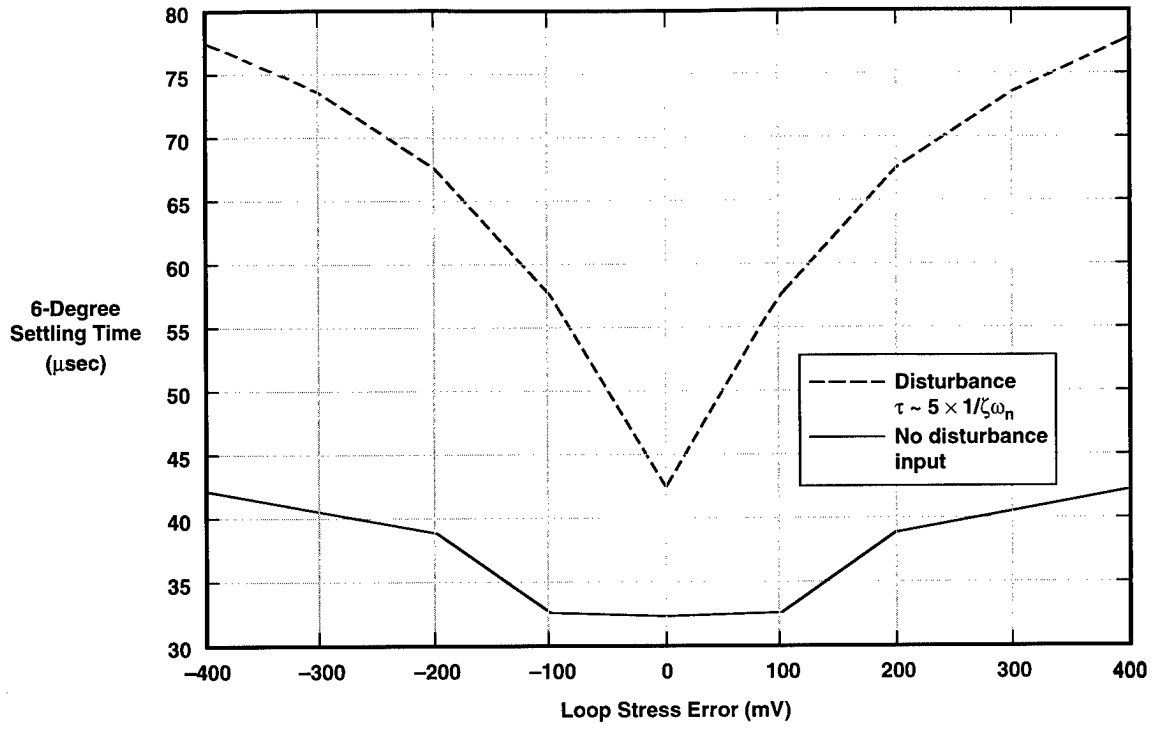


Figure 8. PSpice 6-Degree Settling Time Versus LSE

disturbance input or equal loop filter components and (2) disturbance $\tau \sim 5 \times 1/\zeta\omega_n$.

The plot shows settling time is sensitive to LSE when the undesired disturbance is present. Indeed, for moderate or BOL LSEs, the disturbance response approaches the ideal second order system response such that if only small, BOL LSEs are seen by the system, the undesired disturbance would have little effect and may go unnoticed. This becomes especially probable in a real system as it deviates from ideal. At EOL, when an undesired exponential transient is present and LSE increases as predicted, phase settling may degrade as much as 100 percent.

LO PERTURBATION

Most practical PLLs used as synthesizers require a fine step size in output frequency. One common approach to decrease the effective step size of the PLL while still maintaining the advantages of a PLL architecture is to inject a low frequency DDS into the feedback path of the loop via a mixer. (See Figure 1.) The injected DDS spectrum acts to fill in the natural coarse resolution, $N \times f_{REF}$, of the PLL at the expense of DDS spurious translated directly to the loop's output [16]. Note that the same effect could be realized by supplying the reference frequency via a DDS spectrum, but this scenario results in $20 \times \log(N)$ enhancement of DDS spurious at the loop's output and therefore is discouraged [15].

Employing Figure 1's architecture, a high frequency PLL typically requires the low frequency DDS to be upconverted before injection with the feedback path. Alternatively, a single sideband mixer can be used to inject the DDS frequency directly, but this approach suffers from higher spurious signals. A local oscillator (LO) is mixed with the DDS output, filtered, and mixed into the feedback path such that the frequency into the divide-by-N is the difference between the mixer's output frequency and the loop's output frequency. With proper LO selection, this scheme allows for high synthesized frequencies while the divide-by-N still sees moderately low frequency inputs. Because most high-speed, digital dividers are specified with an allowable window in input power and frequency much less than the desired synthesized frequencies, this architecture allows for fast divide-by-N change with high, wideband output.

A reasonable view of the wideband, high-frequency PLL is as a "black box" with certain input/output, the goal of which is to induce an expected EOL LSE. Forcing the pretune from nominal is an ideal solution. This offset could be realized by perturbing the pretune's digital commanding. However, commanding for the pretune may not be separate from other commands to the loop, such as divide-by-N ratio. Commanding variations may not realize the desired effect. Instead, consider the LO in the DDS chain. This LO can be perturbed from nominal with the loop still remaining in lock. For a static loop, any change in this LO translates to the same change at the loop's output; the frequency into the divide-by-N remains unchanged. This offset in output frequency forces the VCO from its proper frequency. The VCO tuning voltage shifts at the expense of an induced LSE.

It is desirable to predict the exact output frequency for a given induced LSE. However, variations in VCO tuning curves usually prevent such exactness. Moreover, if phase settling tests are automated, a valid output frequency is required or software must be modified. Ideally, it is desirable to switch the LO instantaneously back to its nominal value when the loop is hopped. The loop would then have the nominal output frequency and the loop filter sees the additional LSE. To avoid measuring the settling time of the LO, two LOs are used (Figure 9).

A fast (settling time much smaller than the loop) switch is used to select between the two LOs. The first LO is perturbed from nominal such that the desired LSE is induced. The second LO is at its nominal value. The perturbed LO is selected for the frequency the loop is hopped from, while the nominal LO corresponds to the frequency the loop is hopped to. The loop, therefore, is always hopping to the correct frequency so that no software or commanding change is necessary. The switch is controlled by a TTL signal that is clocked with the loop via the loop's hop clock. An LSE then is induced unintrusively into the loop.

Any filtering in the DDS upconversion chain will act as a limitation to this method by limiting the frequency offset able to be induced.

EMPIRICAL DATA

Final verification is done with empirical data from a spaceborne 5-GHz PLL. The actual loop contains a pretune circuit and reference spur filter per Figure 1. A DDS chain is upconverted with two external LOs for synthesized frequencies in the 5-GHz range. The second LO is chosen for perturbation to induce an EOL LSE. The FET switch has a settling time of ~200 nsec. The loop dynamics are similar to that of the PSpice model. A

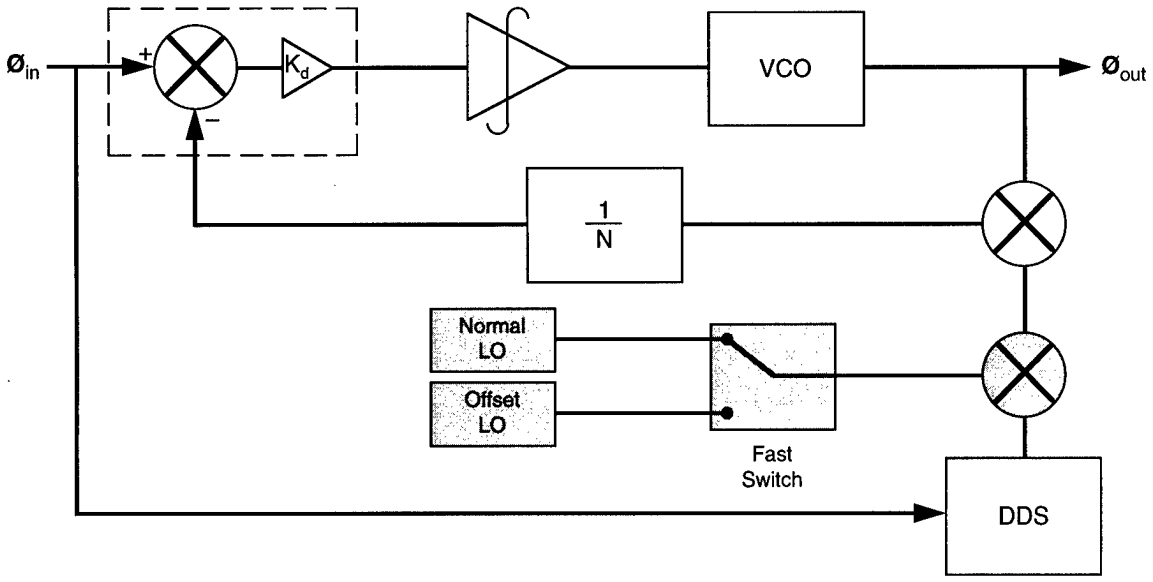


Figure 9. Two Switched LOs in the DDS Chain to Induce an LSE

transient is induced into the loop by mismatching the loop filter components [17], causing an exponential with time constant $\tau \sim 17$ μsec. Settling time to 6 degrees versus LSE is plotted in Figure 10 for this empirical loop and the aforementioned PSpice model.

The shift in LSE between the two curves is caused by a timing error in the real system between the pretune and divide-by-N. A small negative LSE acts to compensate for this time delay. The PSpice data are minimum for zero LSE because no such timing error exists. Both curves

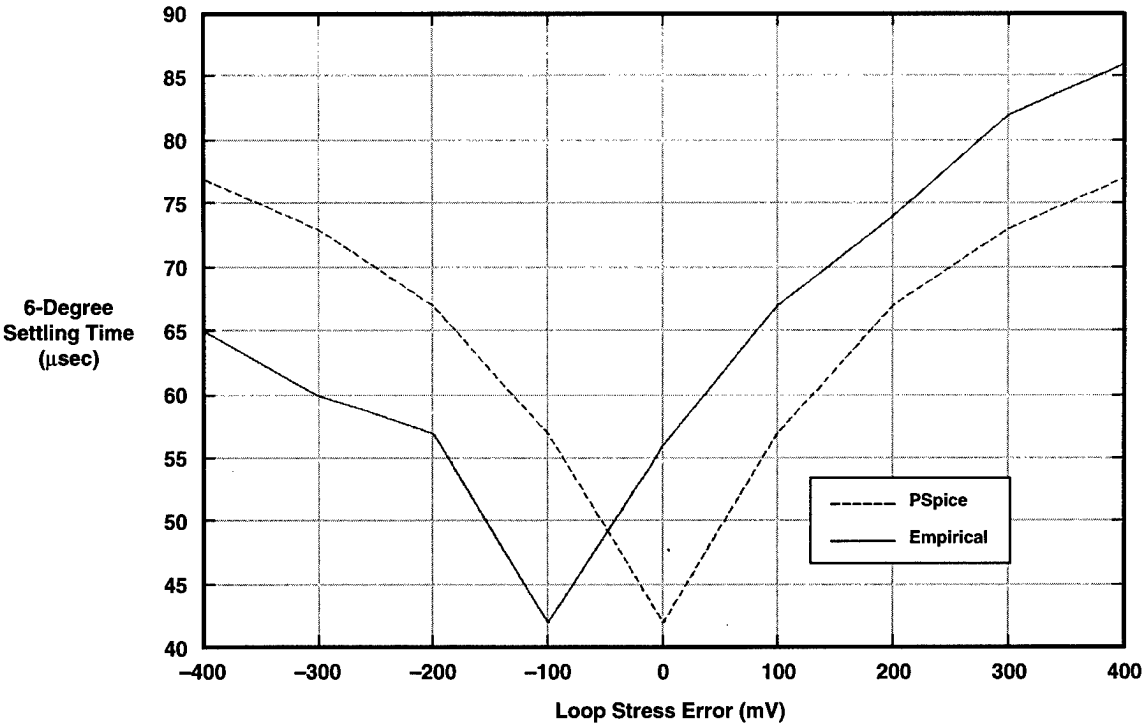


Figure 10. Six-Degree Settling Time Versus LSE for (a) Empirical Loop and (b) PSpice Model

suggest as much as 100-percent degradation in 6-degree settling times if the undesired transient is not accounted for with expected EOL LSE bounds. These data correlate well with the disturbance model presented earlier.

CONCLUSION

We have presented a mathematical description of the second order, type II PLL, which may deviate from ideal when undesired inputs, commonly arising in hardware, are incurred by the system. Satisfactory EOL performance is guaranteed when such transients are accounted for. We have presented a methodology for predicting expected EOL LSE excursions and demonstrated the system's dependence on this parameter. Together, a PSpice model and empirical PLL data verify acceptable degradation at EOL when an EOL transient is present.

ACKNOWLEDGMENTS

Parts of this work were performed under subcontract with TRW Incorporated, Electronic Systems Group, Space Communications Division. The authors wish to acknowledge the support of Mike Lenzo, Paul Sako, Scott Meyer, and everyone else who helped to further this investigation.

BIBLIOGRAPHY

[1] J. Blake, "Design of Wideband Frequency Synthesizers," *RF-Design,* May 1988.

[2] J.A. Crawford, "Understanding the specifics of sampling in synthesis," *Microwaves and RF,* August 1984.

[3] J.A. Crawford, "Extending sampling to Type II phase-locked loops," *Microwaves and RF,* September 1984.

[4] W.F. Egan, *Frequency Synthesis by Phase Lock,* New York: John Wiley and Sons, Inc., 1981.

[5] G.F. Franklin, J.D. Powell, A. Emami-Naeini, *Feedback Control of Dynamic Systems,* Reading, MA: Addison-Wesley, 1986.

[6] F.M. Gardner, *Phaselock Techniques,* 2nd edition, New York: John Wiley and Sons, Inc., 1979.

[8] R.W. Hall, "Synthesizer analysis clarifies PLL circuits," *Microwaves and RF,* May 1985.

[9] R.G. Irvine, *Operational Amplifier Characteristics and Design,* Englewood Cliffs, NJ: Prentice Hall, 1981.

[10] V. Manassewitsch, *Frequency Synthesizers,* New York: John Wiley and Sons, Inc., 1987.

[11] A.V. Oppenheim, A.S. Willsky, I.T. Young, *Signals and Systems,* Englewood Cliffs, NJ: Prentice Hall, 1983.

[12] C.W. Price, II, "Modeling Discrete-Time Phase Detection in the Phase-Locked Loop," *RF-Design,* May 1990.

[13] A.B. Przedpelski, "PLL Primer-Part IV, Description of Loop Integrator Circuit," *RF-Design,* November 1987.

[14] B. Sklar, *Digital Communications*, Englewood Cliffs, NJ: Prentice Hall, 1988.

[15] G.D. Parsons, "Frequency Synthesizer Fundamentals With Emphasis on Direct Digital Synthesis," MSEE Thesis, California State University, Northridge, December 1992.

[16] V. Reinhardt, K. Gould, K. McNab, and M. Bustamante, "A Short Survey of Frequency Synthesizer Techniques," Proceedings of the 40th Annual Frequency Control Symposium, May 1986.

[17] T.J. Endres and J.B. Kirkpatrick, "Sensitivity of Fast Settling PLLs to Differential Loop Filter Component Variations," Proceedings of the 1992 IEEE Frequency Control Symposium, May 1992.

[18] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd edition, New York: McGraw-Hill, 1991.

[19] W.H. Beyer, *CRC Standard Mathematical Tables,* 26th edition, Boca Rotan, FL: CRC Press, 1981.

[20] D.H. Wolaver, *Phase-Locked Loop Circuit Design,* Englewood Cliffs, NJ: Prentice Hall, 1991.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A DIGITAL EQUIVALENT OF AN ANALOG SPECTRUM ANALYZER

by
James A. Barnes
Austron, Inc.
May 5, 1993

I. Introduction

There are many means of estimating power spectral densities (PSD's). For this paper, we will consider the analog spectrum analyzer and the digital spectrum analyzer. The conventional analog spectrum analyzer incorporates a narrow band filter that is tuneable over some significant frequency range. The analyzer squares, the filter output, averages it, and scales it to provide an output in useful units (for example, Watts per Hertz).

Digital spectrum analyzers often compute an autocovariance function from the data, apply a "lag window," and then compute the Fast Transform (FFT) [Jenkins and Watt]. We replace the analog filters used in the first analyzer with digital filters. The digital filters are realized by using ARIMA models (more specifically AR(2) models) [Box and Jenkins]. An AR(2) model can be expressed in the form:

$$X_n = \phi_1 X_{n-1} + \phi_2 X_{n-2} + a_n \quad \text{for } n=1,2,..N \quad (1)$$

where X_n is the filter output, and a_n is the input. The filter is narrow band and approximately Lorentzian in shape if ϕ_1 and ϕ_2 satisfy the relations:

$$\phi_1^2 + 4\phi_2 < 0 \quad (2)$$

$$-2 < \phi_1 < 2 \quad \text{and} \quad -1 < \phi_2 < 0 \quad (3)$$

Unlike the conventional digital spectrum analyzer mentioned above, the digital filter method does not make use of the FFT. The digital filter method is slow compared to the conventional method. On the other hand, the digital filter method is not restricted to data lengths of 2^N and frequencies of integral multiples of 1/N. Indeed, the digital filter method can use any frequency in the range 0 to 1/2. Thus, it is capable of producing very smooth plots of the PSD. Like the conventional digital spectrum analyser the digital filter method re-uses all of the data for each spectral point. This is in contrast to the analog spectrum analyzers which are constantly being fed new data as the analysis proceeds.

Like the analog spectrum analyzer, one can cascade filters to realize sharp skirts on the resonance line width and, hence, it can separate near sidebands. In this paper, we will develop analyzers with a single-stage filter and a two-stage filter corresponding to skirts on the line shape with -6 and -12 dB/octave roll-off respectively. It is intuitively very satisfying to those of us who are used to analog spectrum analysers.

0-7803-0905-7/93 $3.00 © 1993 IEEE

As an added convenience, the digital filters can automatically reduce the size of the frequency increments (that is, reduce the slew rate) to follow steep features in the PSD. This materially improves the appearance of the resulting spectrum plots and saves time relative to the case of constant increments.

Filter Design -- Single Stage

As noted above, the filter is an AR(2) model [Box and Jenkins] in the form of Equation (1). The coefficient ϕ_1 essentially determines the resonant frequency of the filter and ϕ_2 determines the bandwidth. The PSD of the filter output, $S_x(f)$, where f is the Fourier frequency, is given by:

$$S_x(f) = \frac{S_a(f)}{|1 - \phi_1 e^{i\omega} - \phi_2 e^{i2\omega}|^2} \quad \text{where } \omega = 2\pi f \qquad (4)$$

and $S_a(f)$ is the PSD of the input, a_n -- the quantity we are trying to estimate. Throughout this paper, PSD's are taken to be one sided: that is, for any (real) z_n, Parsival's theorem becomes:

$$\sigma_z^2 = \int_0^{1/2} S_z(f)\, df \qquad (5)$$

The upper limit is the Nyquist frequency of one cycle per two data spacing intervals. Thus, the magnitude squared of the transfer function, $|K_1(f)|^2$, of the filter defined by Eq. (1) is [Box and Jenkins]:

$$|K_1(f)|^2 = \frac{1}{|1 - \phi_1 e^{i\omega} - \phi_2 e^{i2\omega}|^2} \qquad (6)$$

In order to normalize the transfer function we need to evaluate the integral of the magnitude squared of the transfer function:

$$Q_1 = \int_0^{1/2} \frac{df}{|1 - \phi_1 e^{i\omega} - \phi_2 e^{i2\omega}|^2} = \frac{1}{2} \frac{1-\phi_2}{1+\phi_2} \left[\frac{1}{(1-\phi_2)^2 - \phi_1^2}\right] \qquad (7)$$

(Dr. S. Jarvis [Jarvis] developed Eq. 7)

Resonant Frequency of Filter

We consider two possible definitions of the resonant frequency of the AR(2) filter:

(1) That frequency, $f_m=\omega_m/2\pi$, for which the magnitude squared of the transfer function, $|K_1(f)|^2$, is maximum.

$$\phi_1 = \frac{4\phi_2 \cos(\omega_m)}{\phi_2 - 1} \quad (8)$$

which can be solved for f_m and the maximum gain (squared) at f_m is:

$$|K_1(f_m)|^2 = \frac{4\phi_2}{(1+\phi_2)^2 (\phi_1^2+4\phi_2)} \quad (9)$$

(2) That frequency, $f_r=\omega_r/2\pi$, for which the filter rings upon removal of all inputs. (The turn-off transient is an exponentially decaying sine wave.)

$$\phi_1 = -2 \sqrt{-\phi_2} \cos(\omega_r) \quad (10)$$

and the gain (squared) at that frequency is:

$$|K_1(f_r)|^2 = \frac{1}{(1+\phi_2)^2 - \phi_1^2 \frac{(\sqrt{-\phi_2} - 1)^2}{\sqrt{-\phi_2}}} \quad (11)$$

There are various reasons for choosing one or the other of these two definitions but probably the best choice is the frequency at maximum gain. For the ringing frequency, the normalization coefficient becomes indeterminate at f = 0 and 1/2. Of course, one must use the ringing frequency to calculate the power in the turn-off transient.

Analyzer Bandwidth

The true (noise) bandwidth, BW, is the width of a rectangular window centered on the filter's resonant frequence, f_m. The filter input is white noise. The filter bandwidth, then, is the total output power of the filter ((that is, the integral of $|K_1(f)|^2$ as given in Eq. (7)) divided by the peak gain, $|K_1(f_m)|^2$, Eq. (9). That is:

$$BW = \frac{1}{8\phi_2} \frac{(1-\phi_2^2)(\phi_1^2 + 4\phi_2)}{(1-\phi_2)^2 - \phi_1^2} \quad (12)$$

Equations (8) and (12) determine the resonant frequency and the bandwidth of the filter as functions of ϕ_1 and ϕ_2. In principal, one can invert these relations to get ϕ_1 and ϕ_2 as functions of the frequency and bandwidth. This is the way one usually states the problem. Unfortunately, this is not a simple problem and we have chosen to make an approximation. If we choose to set ϕ_1=0, then Eq. (12) simplifies to just:

$$BW' = \frac{1}{2}\,\frac{1+\phi_2}{1-\phi_2} \qquad (13)$$

or, equivalently:

$$\phi_2 = \frac{2\ BW' - 1}{2\ BW' + 1} \qquad (14)$$

where BW' is the nominal bandwidth used to compute ϕ_2. Fortunately, Equations (13) and (14) are good approximations for most values of ϕ_1 (and, hence, most frequencies, f, as well). Some differences in bandwidth are observable near the spectral frequency limits of 0 and 1/2. For most uses, the distortions are not serious. If necessary, one can develop empirical corrections. For this paper, Eqs. (1), (8), and (14) are the defining relations. Note that Eq. (12) gives the exact bandwidth as a function of ϕ_1 and ϕ_2 and, therefore, the computed spectral densities (based on Eqs. (7) and (9)) are unbiased even near the frequency limits. What one looses by using Eq. (14) For the nominal bandwidth is a precise value of the number of degrees of freedom near the limits. Thus, the precise confidence intervals of the estimate are not reliable.

The filter response when tuned to the center of the spectrum (f = 0.25) is symmetric and nearly Lorentzian in shape, Fig. 1. As one tunes the filter away from center, the shape is no longer perfectly symmetric or Lorentzian although the changes are gradual (see Fig. 2). Equation 15 gives the general expression for the filter gain at any frequency f when tuned to f_m.

$$|K_1(f)|^2 = \frac{A_1^2}{1 - 4\phi_2\ A_1^2\ [\cos(\omega_m) - \cos(\omega)]^2} \qquad (15)$$

where A_1^2 is the maximum (power) gain at $f=f_m$ (see Eq. 9).

$$A_1^2 = |K_1(f_m)|^2 = \frac{4\phi_2}{(1+\phi_2)^2\ (\phi_1^2+4\phi_2)} \qquad (16)$$

TURN-OFF TRANSIENT

Given N data points, a_n, one can iterate the filter of Eq. (1) N-times and accumulate the squares of the filter outputs, X_n^2. The filter output rings with an exponential decay with a time constant of $\sqrt{-\phi_2}$. For n larger than N and a_n=0 there will be a non-zero output that represents input energy to the filter that is not included in the first N output values. There are two ways of correcting for this missing energy: (a) As just noted one can simply let the filter ring out and accumulate the sum of squares over N+M data points where M is several time constants of the exponential decay and a_n is zero for n>N: (b) Another approach is to note that the complete ringing of the filter is determined by the final two output values X_N and X_{N-1} (since this is an AR(2) process). One can then calculate analytically the ring-down "energy" and add it to the sum of squares for the first N output values, at least in principal.

We derived an equation for the turn-off transient for the second method discussed above for the single stage spectrum analyzer. For the two-stage analyzer we could not get an analytic formula for the ring-down energy so we used method 1 above. This approach is slower than the analytical method (see below). For the single stage analyzer, the remaining energy is given by:

$$E = \frac{X_N^2\,(1-\phi_2) + 2\,X_N\,X_{N-1}\,\phi_1\,\phi_2 + X_{N-1}^2\,\phi_2^2\,(1-\phi_2)}{(1+\phi_2)\,((1-\phi_2)^2 - \phi_1^2)} - X_N^2 \qquad (17)$$

The derivation of Eq. (17) is lengthy. We can, however, compute the sum of squares directly using method (a), above, and compare the results with Eq. (17), i.e., method (b). The results should be the same to within the precision of the computer (15 or 16 decimal places). That is, $S_N + E = S_{N+M}$ for M suitably large.

Spectral Density Computation:

The protocol to compute the Power Spectral Density, PSD, is:

A. Select ϕ_2 from Eq. (14) for the nominal bandwidth, BW'.

B. Select ϕ_1 from Eq. (8) for the desired frequency, $f = \omega/2\pi$.

C. Compute the sum of squares, S_N, of X_n for n=1 to N using Eq. (1).

D. compute the normalization coefficient, Q, from Eq. (7).

E. Compute the "ring-down" energy, E, from Eq. (17).

F. Compute the PSD using Eq. (19):

$$PSD = (S_N + E)/(QN) \quad \text{or} \quad = S_{N+M}/(QN) \qquad (19)$$

G. Increment frequency for next PSD value.

Program 1, below, is the listing of a program segment in BASIC that performs the seven statements in the above protocol for a single stage spectrum analyzer. A complete program, Prog. (2), (including the segment, Prog. 1) follows:

PROGRAM 1, Segment of Power Spectral Density Program (Single stage)

```
INPUT "ENTER BANDWIDTH = ";BW
INPUT "ENTER INCREMENT = ";FI
F=0#
PH2 = (BW+BW - 1#)/(BW+BW +1#)
FILTER:
  PH1 = 4# * PH2 *COS(TWOPI * F) / (PH2 - 1#)
  X1 = 0#: X2 = 0#: S = 0#
  FOR K = 1 TO N
     X = PH1*X1 + PH2*X2 + A(K)
     S = S + X*X
     X2 = X1: X1 = X
  NEXT K
  DX = (1# + PH2) * ((1# - PH2)^2 - PH1*PH1)
  Q = .5# * (1# - PH2) / DX
  A = 1# - PH2: B = 2# * PH1*PH2: C = PH2*PH2 * (1# - PH2)
  E = ((X1*X1 * A + X1*X2 * B + X2*X2 * C) / DX) - X1*X1
  PSD = (S + E) / (Q * N): IF F = 0# THEN PSET (0,PSD), 13
  LINE -(F, PSD), 13
F = F + FI: IF F < .50001# THEN GOTO FILTER
```

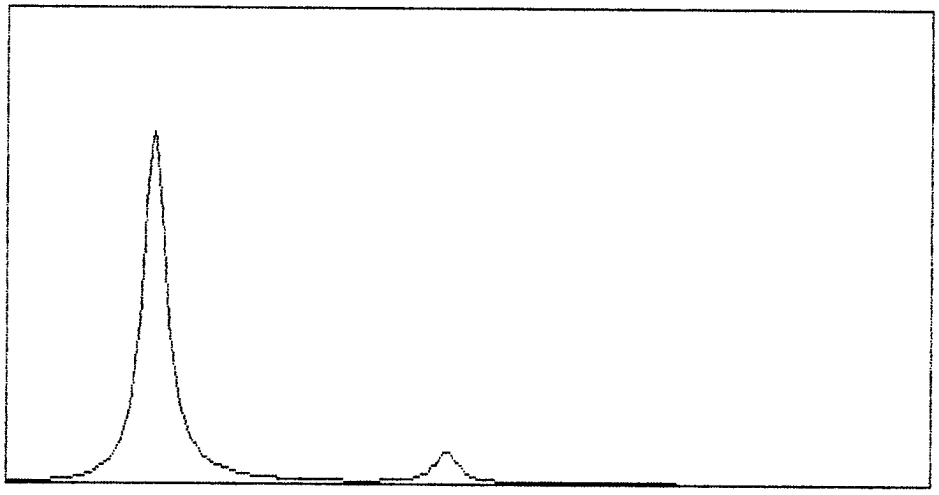
It is surprising that all of the mathematics developed above is accomplished in only 19 lines of computer code. Of course, to get a fully operable program many more lines of code are required.

SPECTRUM WITH TWO CASCADE, RESONANT STAGES OF FILTERING

There is little difference between the protocols for a single stage of filtering and two stages. There are differences in the equations, however. The basic filter relations are the cascade of two AR(2) filters with identical AR-coefficients, ϕ_1 and ϕ_2, as shown below:

$$Y_n = \phi_1 Y_{n-1} + \phi_2 Y_{n-2} + a_n$$
$$X_n = \phi_1 X_{n-1} + \phi_2 X_{n-2} + Y_n \qquad (18)$$

for n=1, 2, ... N. In order to assure against an inadvertant loss in significant digits, it is wise to keep the model as the cascade of two separate filters instead of an "equivalent" AR(4) model (which is formally possible). Equations (18) correspond to Eq. (1) for a single stage analyzer. The conditions on ϕ_1 and ϕ_2 are the same as given in Eqs. (2) and (3) for the single stage.



Press any key to continue

```
REM  PROGRAM NO. 2  SINGLE STAGE SPECTRUM ANALYSER
BW = .02: PI = 3.1415926535#: TWOPI = PI + PI: N = 200
DIM A(2100)
FI = .002: F = 0: CLS
PH2 = (BW + BW - 1) / (BW + BW + 1)
FOR I = 1 TO N: A(I) = SIN(I / 2) + .3 * COS(1.5 * I): NEXT I
'  A(N - 5) = N / 2!
'
SCREEN 2
WINDOW (-.05, -4)-(.5, 30)
LINE (0, 0)-(.5, 30), , B
'
PSET (0, 0)
FILTER:
   PH1 = 4 * PH2 * COS(TWOPI * F) / (PH2 - 1)
   X1 = 0: X2 = 0: S = 0
   FOR K = 1 TO N
      X = PH1 * X1 + PH2 * X2 + A(K)
      X2 = X1: X1 = X: S = S + X * X
   NEXT K
   DX = (1 + PH2) * ((1 - PH2) ^ 2 - PH1 * PH1)
   Q = (1 - PH2) / (DX + DX)
   A = 1 - PH2: B = 2 * PH1 * PH2: C = PH2 * PH2 * (1 - PH2)
   E = ((X1 * X1 * A + X1 * X2 * B + X2 * X2 * C) / DX) - X1 * X1
   PSD = (S + E) / (N * Q): LINE -(F, PSD)
F = F + FI: IF F < .5001 THEN GOTO FILTER
```

Since we are cascadeing two identical filters, the magnitude squared of the transfer function for the cascade is the square of that for the single stage. That is,

$$|K_2(f)|^2 = \frac{1}{\left[\, |1 - \phi_1 e^{i\omega} - \phi_2 e^{i2\omega}|^2 \,\right]^2} \qquad (19)$$

which corresponds to Eq. (6) for the single stage. The evaluation of the integral of Eq. (19):

$$Q_2 = \int_0^{1/2} |K_2(f)|^2 \, df \qquad (20)$$

is much more involved. The result (again due to Jarvis) can be written in the form:

$$A = \frac{1}{\sqrt{-\phi_2}} \qquad B = \frac{1 + \phi_2}{\sqrt{-\phi_2}}$$

$$C = \frac{\sqrt{(1 - \phi_2)^2 - \phi_1^2}}{\sqrt{-\phi_2}} \qquad D = \frac{\sqrt{-(\phi_1^2 + 4\phi_2)}}{-\phi_2}$$

$$\tan(\alpha) = \frac{\sqrt{-(\phi_1^2 + 4\phi_2)}}{\phi_1} \qquad \tan(\gamma) = \frac{\sqrt{-(\phi_1^2 + 4\phi_2)}\,(1 - \phi_2)}{\phi_1\,(1 + \phi_2)}$$

These six relations can be combined to give the normalization coefficient, Q, for the two-stage analyzer:

$$Q_2 = A \left(\frac{A}{\phi_2 BCD}\right)^2 \left[(3/A - 2/B)\cos(2\gamma) - (2/C)\cos(\alpha - 3\gamma) - (2/D)\sin(\alpha - 2\gamma)\right] \qquad (21)$$

These equations correspond to Eq. (7) for the single stage analyser. The computation of ϕ_1 is the same for the two stage analyser as for the single, Eq. (8) i.e.:

$$\phi_1 = \frac{4\phi_2 \cos(\omega_m)}{\phi_2 - 1} \qquad (22)$$

The peak power gain, $|K_2(f_m)|^2$, for the two stage filter is just the square of the single stage gain (9). That is,

$$|K_2(f_m)|^2 = \left[\frac{4\phi_2}{(1+\phi_2)^2(\phi_1^2+4\phi_2)}\right]^2 \qquad (23)$$

The exact expression for the bandwidth of the two stage filter corresponding to Eq. (12) for the single stage filter is:

$$BW = \frac{Q_2}{|K_2(f_m)|^2} \qquad (24)$$

The general expression for the (power) gain at any frequency, f, when the filter is tuned to f_m, for the two stage filter is just the square of that for the single stage filter, Eq. (15). That is:

$$|K_2(f)|^2 = \frac{A_2^2}{\left[1 - 4\phi_2 A_2 [\cos(\omega_m) - \cos(\omega)]^2\right]^2} \qquad (25)$$

where:

$$|K_2(f_m)|^2 = A_2^2 = \left[|K_1(f_m)|^2\right]^2 = [A_1^2]^2 \qquad (26)$$

and,

$$A_2^2 = \left[\frac{4\phi_2}{(1+\phi_2)^2 (\phi_1^2+4\phi_2)}\right]^2 \qquad (27)$$

Turn-off Transient

Unfortunately, we do not have the equivalent of Eq. (17) for the two stage filter. Instead, the two stage analyser accumulates the sum of the squares of the output transient and adds it to the sum of the squares of the first N output values.

Program 3, below, for the two stage filter is comparable to Prog. 1 for the single stage filter. Program 3 is a segment of Prog. 4 which is complete in the sense that it has all the necessary code to run and plot out results. Program 3 is the segment which performs this protocol noted above for prog. 1.

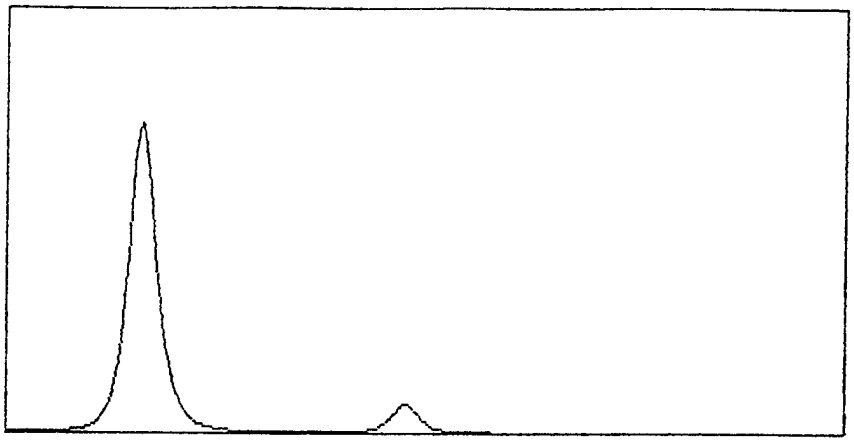
PROGRAM 3. Segment of Power Spectral Density Program (Two stage)

```
INPUT "ENTER BANDWIDTH = ";BW
ITAIL = INT(4/BW)
FOR J = N+1 TO N+ITAIL: A(J) = 0#: NEXT J
F=0#
INPUT "ENTER INCREMENT ";FI
PH2 = (4#*BW - 1#) / (4#*BW + 1#)
FILTER:
  PH1 = 4*PH2 * COS(TWOPI*F) / (PH2 - 1#)
  X1=0#: X2=0#: Y1=0#: Y2=0#: S=0#
  FOR I=1 TO N + ITAIL
     Y = PH1*Y1 + PH2*Y2 + A(I)
     X = PH1*X1 + PH2*X2 + Y
     Y2=Y1: Y1=Y: X2=X1: X1=X: S = S + X*X
  NEXT I
  A = 1#/SQR(-PH2): B=1#+PH2: C = A*SQR((1# - PH2)^2 - PH1*PH1)
  E = SQR(-(PH1*PH1 + 4#*PH2)): B = A*B: D = E/(-PH2)
  DK = A/(PH2*B): DK = A*DK*DK
  IF PH1=0# THEN AL=PI: GAM=AL: GOTO 450
  AL = ATN(E/PH1): GAM = ATN(E*(1# - PH2) / (PH1*(1# + PH2)))
  IF PH1<0# THEN AL = AL+PI: GAM = GAM-PI
450: T = (3#/A-2#/B) * COS(GAM+GAM) - (2#/C) * COS(AL-3#*GAM)
  Q = (DK*(T - (2#/D) * SIN(AL-GAM-GAM))) / (C*C*D*D)
  PSD = S/(Q * N): LINE -(F,PSD), 13
F = F+FI: IF F<.500001# THEN GOTO FILTER
```

While the two stage filter is a fair amount more complicated to design, the computer code is not much different -- 20 lines compared to 26 lines of code for Progs (1) and (3), respectively. It is a bit slower than the single stage operation.

CONCLUSION

We can devise useful power spectrum analyzers based on digital filters. Although the paper discusses only one and two stage filters, one could have any number. The algorithm does not involve the fast Fourier transform. While this spectrum analyzer is not as fast computationally as the more conventional ones, the great speed of present-day computers renders this aspect relatively unimportant. The computer program is quite short.



Press any key to continue

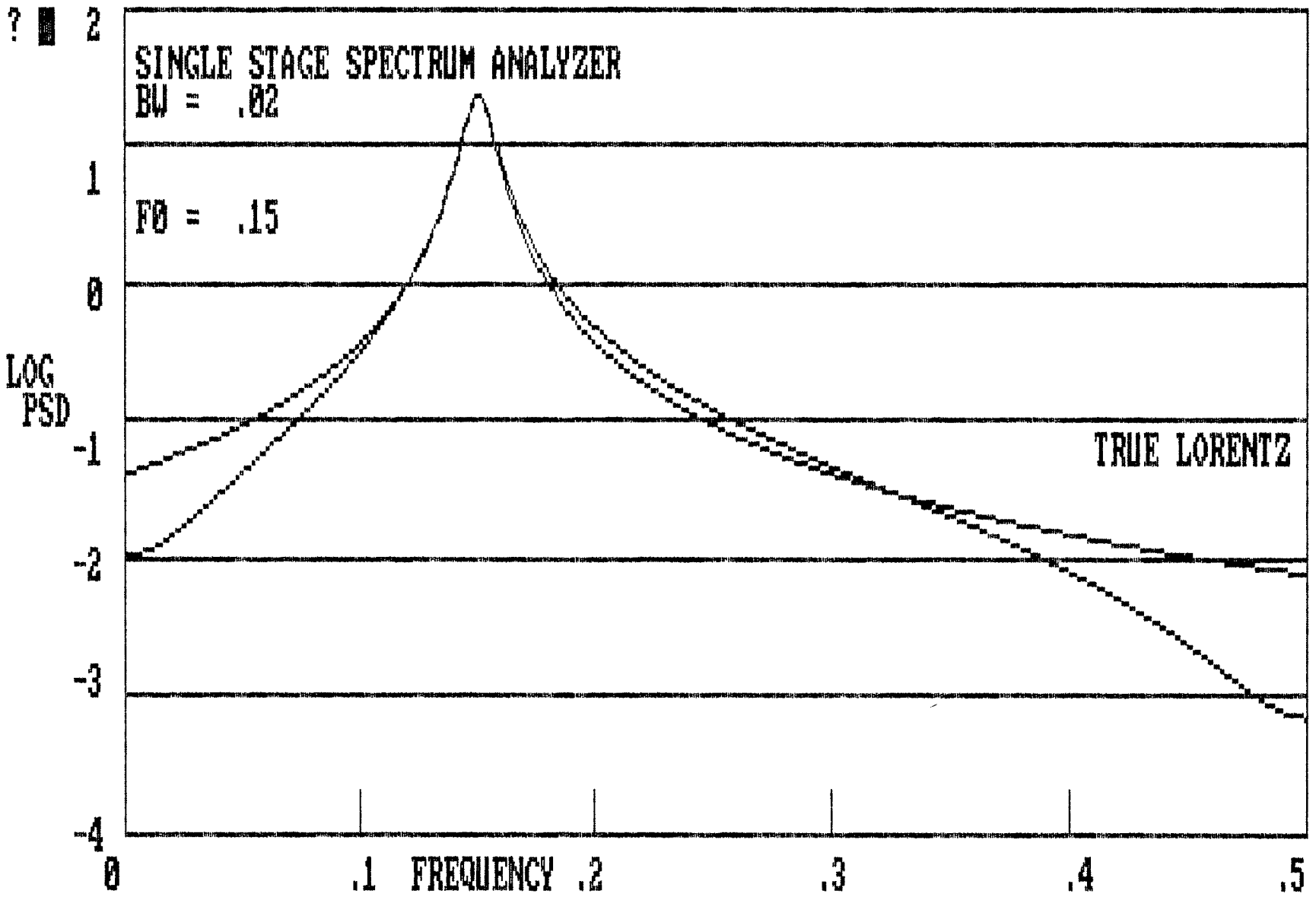
```
REM  PROGRAM NO. 4 TWO-STAGE SPECTRUM ANALYZER
BW = .02: PI = 3.141592: TWOPI = PI + PI: N = 200
DIM A(2100)
FI = .002: F = 0: CLS
PH2 = (4 * BW - 1) / (4 * BW + 1)
ITAIL = INT(3 / BW)
FOR I = 1 TO N: A(I) = SIN(I / 2!) + .3 * COS(1.5 * I): NEXT I
'  A(190) = N / 2
FOR I = N + 1 TO N + ITAIL: A(I) = 0: NEXT I
'
SCREEN 2
WINDOW (-.05, -4)-(.5, 30)
LINE (0, 0)-(.5, 30), , B: PSET (0, 0)
'
FILTER:
   PH1 = 4 * PH2 * COS(TWOPI * F) / (PH2 - 1)
   X1 = 0: X2 = 0: Y1 = 0: Y2 = 0: S = 0
   FOR I = 1 TO N + ITAIL
      Y = PH1 * Y1 + PH2 * Y2 + A(I)
      X = PH1 * X1 + PH2 * X2 + Y
      X2 = X1: X1 = X: Y2 = Y1: Y1 = Y: S = S + X * X
   NEXT I
   A = 1 / SQR(-PH2): B = 1 + PH2: C = A * SQR((1 - PH2) ^ 2 - PH1 * PH1)
   E = SQR(-(PH1 * PH1 + 4 * PH2)): B = A * B: D = E / (-PH2)
   DK = A / (PH2 * B): DK = A * DK * DK
   IF PH1 = 0 THEN AL = PI / 2: GAM = AL: GOTO 450
   AL = ATN(E / PH1): GAM = ATN(E * (1 - PH2) / (PH1 * (1 + PH2)))
   IF PH1 < 0 THEN AL = AL + PI: GAM = GAM - PI
450 : T = (3 / A - 2 / B) * COS(GAM + GAM) - (2 / C) * COS(AL - 3 * GAM)
   Q = (DK * (T - (2 / D) * SIN(AL - GAM - GAM))) / (C * C * D * D)
   PSD = S / (N * Q): LINE -(F, PSD)
F = F + FI: IF F < .50001 THEN GOTO FILTER
```

REFERENCES

Box, G.P.E. and G.M. Jenkins, (1970), TIME SERIES ANALYSIS, FORECASTING AND CONTROL, Holden-Day San Francisco, CA.

Jankins, G.M. and D.G. Watts, (1969), SPECTRAL ANALYSIS AND ITS APPLICATIONS, Holden-Day San Francsco, CA.

Jarvis, S. , (1978 ?) Private communication.


SINGLE STAGE SPECTRUM ANALYZER
BW = .02
F0 = .15
TRUE LORENTZ
LOG PSD
2
1
0
-1
-2
-3
-4
0
.1
FREQUENCY
.2
.3
.4
.5

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Measurement of the Phase Noise Characteristics of an Unlocked Communications Channel Identifier

W.K. Lam, K.W. Wan, J. Austin, E. Vilar (University of Portsmouth, U.K.)

and D.F. Bacon (Radiocommunications Agency, U.K.)

Abstract – **This paper reviews the technique of phase noise stability measurement using complex demodulation and applies the approach to an unlocked spread spectrum communications probe. This system has recently been developed to characterise a transhorizon communications channel at 11.6GHz under both tropospheric scatter and anomalous propagation conditions. The receiver has a bandwidth of 31.25MHz but has no carrier recovery and therefore no absolute phase reference. However, provided that the phase jitter and the frequency drift are small enough during the channel measurement period then it is possible to obtain an accurate channel transmittance estimate. The system has the capability of implementing *time domain signal averaging* "on the fly" thus permitting high speed channel identification under low SNR regimes. As the averaging period is increased (to improve the SNR), low phase jitter in the oscillators becomes imperative. The SNR improvement achieved is thus controlled by the maximum averaging period which is a direct reflection of the stability of the oscillators.**

I. Introduction

A *wide–band transhorizon* experiment began in 1990 involving collaboration between the University of Portsmouth (UP), CNET (Paris/Issy Centres) and the U.K. Radiocommunications Agency (RA).

The main objective of the experiment is to capture the wide–band channel frequency response of the transhorizon link between Cap d' Antifer (northern France) and Portsmouth (southern England) under the limitations of low signal–to–noise ratio (SNR) and finite channel stationarity.

The experimental system is based upon a suppressed carrier frequency unlocked system[1]. Because no absolute phase reference is recovered from the carrier, accurate channel identification can only be obtained if the oscillator phase jitter and the frequency drift are small within the data collection period. The medium to long term frequency stability of the system is maintained by using digitally synthesized crystal oscillators. These oscillators are phase locked to 10MHz frequency references obtained from LORAN–C receivers located at the two ends of the link.

As a result of the analysis of several years of narrow–band data, the median signal strength during tropospheric scatter conditions has been found to be –106.5dBm. This results in a signal–to–noise (SNR) ratio of –13.1dB within the 31.25MHz bandwidth of the receiver[1]. In order to recover the small signal and remove the embedded noise, a fast *time–domain signal averaging* system has been developed. The processing gain of the system is given by the relationship

$$\Delta SNR = 10\log(\mathcal{N}_a) \qquad (1.1)$$

where $\mathcal{N}_a$ is the number of averages.

However, as averaging is increased to improve the SNR, low phase jitter in the system becomes imperative. The phase instability causes averaging errors which results in signal distortion. Therefore, a compromise has to be made between the averaging period and the phase noise characteristics of the system.

The method of phase noise measurement is based on the *phase unwrapping technique* previously developed at Portsmouth[2]. With suitable configuration of the receiver system, the complex in–phase (I) and in–quadrature (Q) demodulator can be modified to measure the instantaneous "DC phase difference" (i.e. the phase noise) between the transmitter and the receiver. The complex I and Q demodulator turns the phase noise signals into a rotating phasor. As a result, the instantaneous principal phase can be obtained by a simple inverse tangent and the continuous phase subsequently reconstructed by phase unwrapping.

II. The Experimental System

Figures 2.1 and 2.2 show the simplified block diagrams of the transmitter and receiver. The system is designed to operate from baseband to baseband with the transmission carrier at X–band (11.6GHz). For the phase noise measurement, the X–band output is connected to the receiver for demodulation back to baseband. In addition, back–to–back measurement at L–band have also been carried out. A comparison between the measurements

0-7803-0905-7/93 $3.00 © 1993 IEEE

back–to–back at L–band and X–band are given later in the paper.

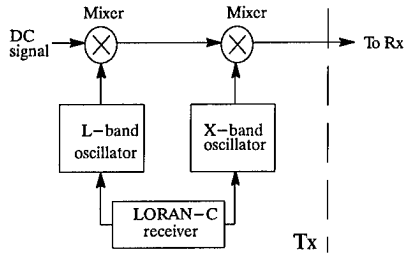


Figure 2.1 Transmitter

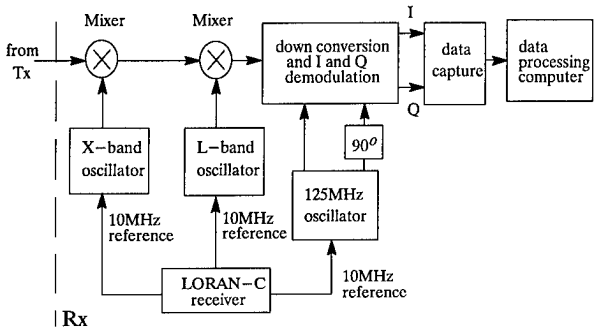


Figure 2.2 Receiver

Transmitter: The tailored spread spectrum channel probing signal is substituted by a DC signal at the input to the transmitter. This DC signal is then frequency up–converted to X–band. The 10MHz frequency reference from the LORAN–C receiver is used as a common phase reference for the L–band and X–band oscillators. The LORAN–C receiver is long–term phase locked to the 100kHz groundwave signal from the closest LORAN–C station in Lessay (northern France).

Receiver: In the receiver, the X–band "DC signal" is first down–converted to L–band and then further down–converted to 125MHz. The signal thereafter splits into two paths for I and Q demodulation. In a similar manner to the transmitter, all the three oscillators in the receiver are phase locked to the 10MHz frequency reference from a LORAN–C receiver.

III. The Theory

The basic operation of the frequency conversion processes in the system can be simplified to a single stage up/down conversion model as shown in figure 3.1.

The output of the oscillators in the transmitter and receiver can be represented by equations (3.1) and (3.2) respectively.

$$V_{\zeta_t}(t) = [V_c + \varepsilon_t(t)]\cos\left[2\pi(\nu_c + \nu_{\zeta_t})t + \phi_{\zeta_t}(t)\right] \quad (3.1)$$

$$V_{\zeta_r}(t) = [V_c + \varepsilon_r(t)]\cos\left[2\pi(\nu_c + \nu_{\zeta_r})t + \phi_{\zeta_r}(t)\right] \quad (3.2)$$

The subscript $_t$ and subscript $_r$ refer to the parameters in the transmitter and the receiver respectively. $\varepsilon_t(t)$ and $\varepsilon_r(t)$ are the random amplitude fluctuations around V_c, also known as amplitude modulated (AM) noise. ν_c is the nominal frequency of the carrier. ν_{ζ_t} and ν_{ζ_r} are the linear frequency drifts. $\phi_{\zeta_t}(t)$ and $\phi_{\zeta_r}(t)$ are the phase noise.

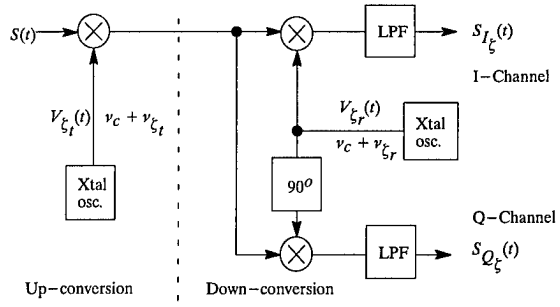


Figure 3.1 Simplified Block Diagram of the Up/Down Frequency Conversions and the I and Q Demodulator

Let the overall frequency and phase drifts be represented by ν_ζ and $\phi_\zeta(t)$ respectively. i.e.

$$\nu_\zeta = \nu_{\zeta_t} - \nu_{\zeta_r} \quad (3.3)$$

$$\phi_\zeta = \phi_{\zeta_t} - \phi_{\zeta_r} \quad (3.4)$$

Thus, for double sideband (DSB) up/down conversions, the baseband I and Q outputs can be shown to be

$$S_{I_\zeta}(t) = \frac{[V_c + \varepsilon_t(t)][V_c + \varepsilon_r(t)]}{2}\cos\left[2\pi\nu_\zeta t + \phi_\zeta(t)\right]S(t) \quad (3.5)$$

$$S_{Q_\zeta}(t) = \frac{[V_c + \varepsilon_t(t)][V_c + \varepsilon_r(t)]}{2}\sin\left[2\pi\nu_\zeta t + \phi_\zeta(t)\right]S(t) \quad (3.6)$$

For the application to phase noise measurement, the signal $S(t)$ is substituted by a DC signal of K volts. The sum of the rotating phasors $S_{I_\zeta}(t)$ and $S_{Q_\zeta}(t)$ gives the complex DC phase difference which is denoted by $M_\zeta(t)$

$$M_\zeta(t) = \sqrt{S_{I_\zeta}^{\,2}(t) + S_{Q_\zeta}^{\,2}(t)}\; e^{\left[j(2\pi\nu_\zeta t + \phi_\zeta(t)\right]} \quad (3.7)$$

$$|M_\zeta(t)| = \frac{K}{2}[V_c + \varepsilon_t(t)][V_c + \varepsilon_r(t)]$$

The AM noise in the system is related to the amplitude fluctuation of $|M_\zeta(t)|$ about its mean. Therefore by removing the mean amplitude $\overline{|M_\zeta(t)|}$, the overall AM noise $\varepsilon(t)$ can be obtained.

Let the sampled instantaneous *principal phase* of $M_\zeta(t)$ be represented by $\psi_\zeta(n)$ which can be obtained from the inverse tangent. i.e.

$$\psi_\zeta(n) = 2\pi\nu_\zeta n + \phi_\zeta(n) = \tan^{-1}\left(\frac{S_{Q_\zeta}(n)}{S_{I_\zeta}(n)}\right) \quad (3.8)$$

The *continuous phase* is then reconstructed by means of phase unwrapping in accordance with the recurrent relationship

$$\psi_\zeta(n) = \psi_\zeta(n-1) + Arg\left[\left(S_{I_\zeta}(n) + jS_{Q_\zeta}(n)\right)\left(S_{I_\zeta}(n-1) + jS_{Q_\zeta}(n-1)\right)^*\right] \quad n \in 1 \rightarrow N_l - 1 \quad (3.9)$$

where the initial phase is $\psi_\zeta(0) = Arg\left[S_{I_\zeta}(0) + jS_{Q_\zeta}(0)\right]$.

* represents the complex conjugate and N_l is the sequence length.

IV. Signal Analysis and Results

Figure 4.1 depicts the details of the signal analysis. Three different analyses have been carried out. They include:–

(i) Averaged phase noise spectral density $\overline{S_\psi(k)}$

(ii) Probability density of the Maximum Phase Excursion $\hat{\psi}_\zeta(\tau_a)$within a sampling interval τ_a

(iii) Probability density of the Differential Phase $\psi_\zeta'(n)$

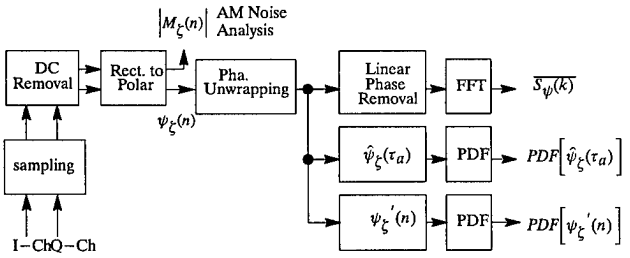
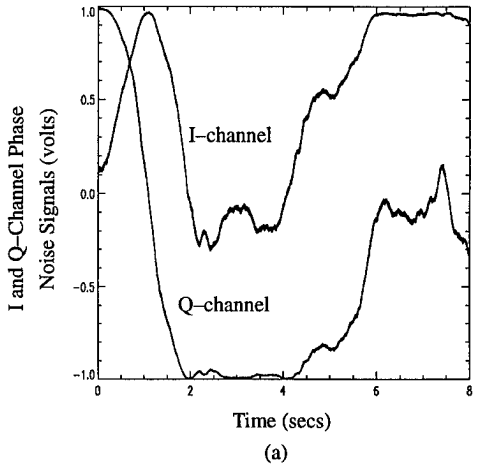
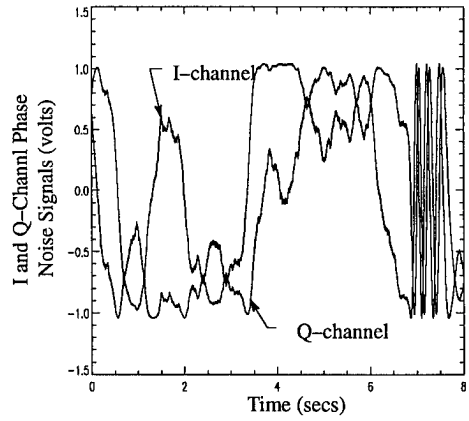


Figure 4.1 Block Diagram of the Signal Analysis

Figure 4.2 (a) and (b) are examples of I and Q phase noise signals measured back–to–back at L–band and X–band respectively. Figure 4.3 (a) and (b) are the corresponding XY–plots of the signals. Both of the L–band and X–band data use the same time interval τ_a of about 8seconds. Notice that only a partial circle is obtained in the L–band measurement; this implies, as expected, a slower phase variation at L–band. As mentioned above, the AM noise is represented as the amplitude fluctuations of $|M_\zeta(t)|$ about its mean and can be seen to be very small when compared to the magnitude of the signal (i.e. the radius of the circle).

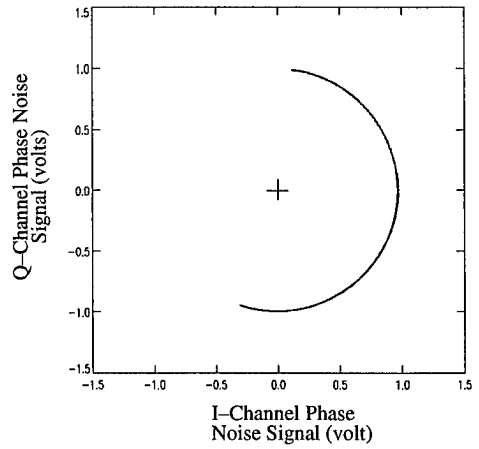


(a)



(b)

Figure 4.2 Examples of I and Q Phase Noise Signals at (a) L–band (b) X–band



(a)

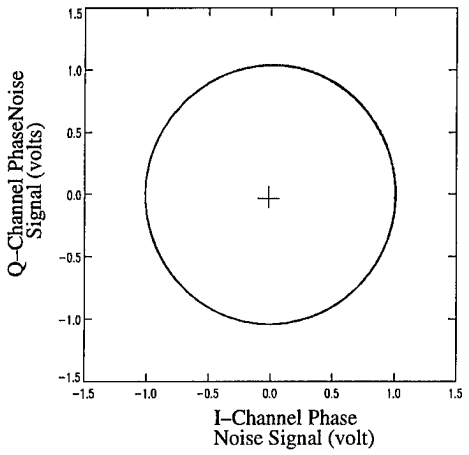


Figure 4.3 XY– plots of the I and Q Phase Noise Signals in figure 4.2 (a) L–band (b) X–band

Figure 4.4 shows examples of unwrapped phase signals at L–band and X–band with their linear phase components removed. The removal of the linear phase from the signal avoids the first order discontinuity at the boundary of the periodic extension which is assumed when using the FFT.

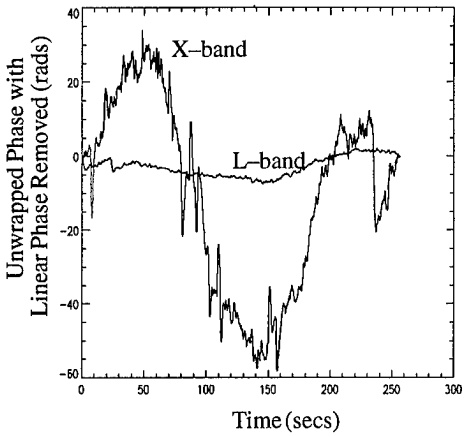


Figure 4.4 Unwrapped Phase at the L–band and X–band with Linear Phases Removed

A. Averaged Phase Noise Spectral Density $\overline{S_\psi(k)}$

A power spectral density (PSD) estimate, based on the FFT approach is known as a *periodogram*. By using this method, the two sided phase noise spectral density can be defined as[3]

$$S_\psi(f) = \lim_{N_l \to \infty} E\left\{ \frac{1}{(2N_l+1)\Delta t} \left| \Delta t \sum_{n=-N_l}^{N_l} \psi_\zeta(n) \exp(-j2\pi f n \Delta t) \right|^2 \right\} \quad (rad^2/Hz) \quad (4.1)$$

where E is the expectation of the spectral density function and Δt is the sampling interval.

However, spectral densities are theoretical concepts involving infinite duration processes as described in equation (4.1). In practice, only finite–duration processes are available, therefore, experimental knowledge of spectral densities suffers from many limitations when dealing with signals that are statistically non–stationary[4]. As described in [5], it is thus necessary to have some form of ensemble averaging or smoothing of the sample spectrum.

Low sidelobe windows can effectively reduce *leakage* but at the expense of broadening the spectrum and losing resolution of closely spaced frequency components. In order to mitigate this resolution problem, a long FFT together with a low leakage Blackman–Harris window[6] has been used for the spectral estimation. In addition, the high variance spectra (due to the long FFT) are necessarily further refined by *spectral averaging*.

In spite of the finite data length, it can be assumed that the spectra thus obtained are *approximately* the true phase noise spectral density.

The 4–term Blackman–Harris window used in the analysis is defined as

$$w(n) = a_0 - a_1 \cos\left(\frac{2\pi}{N_l} n\right) + a_2 \cos\left(\frac{2\pi}{N_l} 2n\right) - a_3 \cos\left(\frac{2\pi}{N_l} 3n\right)$$

where $n = 0, 1, 2, \ldots, N_l - 1$ (4.2)

The coefficients used in (4.2) are [6].

$a_0 = 0.35875, a_1 = 0.48829, a_2 = 0.14128$ and $a_3 = 0.01168$

Thus, the windowed phase sequence is given by

$$\psi_\zeta^w(n) = \psi_\zeta(n) w(n) \quad (4.3)$$

and the corresponding phase noise spectral density is

$$S_\psi(k) \approx \frac{\left|\Psi_\zeta^w(k)\right|^2}{f_s/N_l} \quad (rad^2/Hz) \quad (4.4)$$

where $\Psi_\zeta^w(k)$ represents the Fourier transform of the windowed phase sequence $\psi_\zeta^w(n)$, $f_s = 2048Hz$ is the sampling frequency and $N_l = 524288$ is the sequence length.

Thus the *averaged phase noise spectral density* is given by

$$\overline{S_\psi(k)} = \frac{N_l}{N_a f_s} \sum_{i=1}^{N_a} \left|\Psi_\zeta^w(k - iN_l)\right|^2 \quad (4.5)$$

where $N_a = 8$ is the number of averages. Typical measurements made at both L and X–band are shown in figure 4.5.

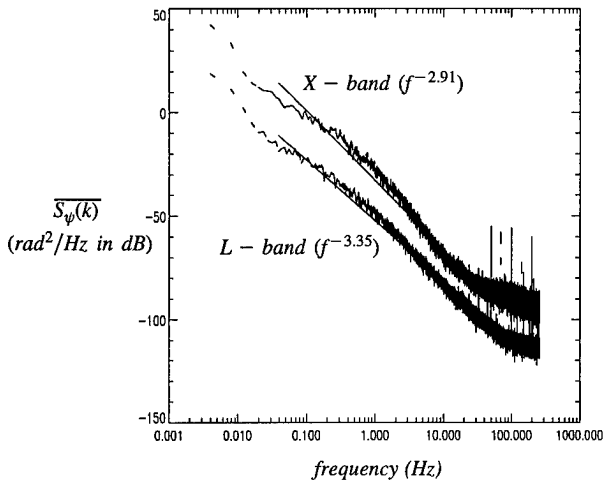


Figure 4.5 Averaged Phase Noise Spectral Densities at L–Band and X–Band

B. Maximum Phase Excursion $\hat{\psi}_\zeta(\tau_a)$

The processing error of the averaging algorithm is associated with the variation of $\psi_\zeta(t)$ within the averaging interval τ_a. The term $\hat{\psi}_\zeta(\tau_a)$ denotes the maximum phase excursion corresponding to the averaging interval τ_a. The relationship is depicted in figure 4.6.

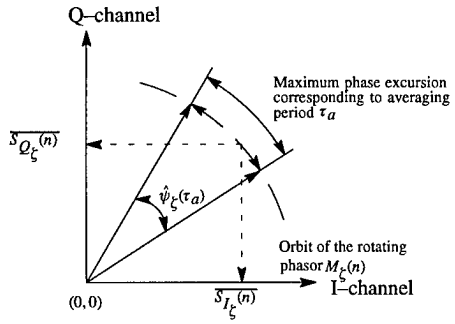


Figure 4.6 Relationship between the Maximum Phase Excursion $\hat{\psi}_\zeta(\tau_a)$ and the Averaging Interval τ_a

Provided that the $\hat{\psi}_\zeta(\tau_a)$ is *small* and the AM noise is negligible, it can be shown that the averaged I and Q outputs are given by

$$\overline{S_{I_\zeta}(n)} \approx \frac{V_c^2}{2} S_{I_\zeta}(n) \cos\psi_\zeta(n) \qquad (4.6)$$

$$\overline{S_{Q_\zeta}(n)} \approx \frac{V_c^2}{2} S_{Q_\zeta}(n) \sin\psi_\zeta(n) \qquad (4.7)$$

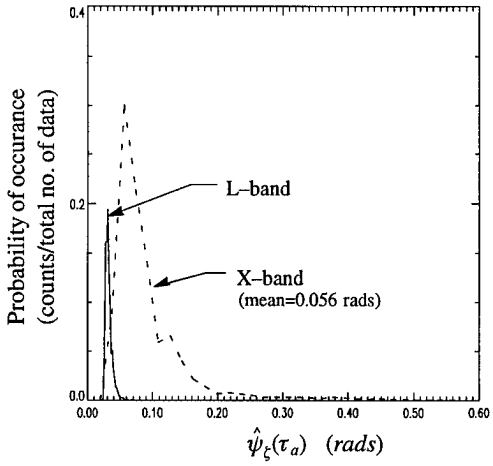


Figure 4.7 Examples of Probability Distribution of $\hat{\psi}_\zeta(\tau_a)$ with $\tau_a = 16ms$ at L–band and X–band

Figure 4.7 illustrates the probability distribution of $\hat{\psi}_\zeta(\tau_a)$ with the averaging interval $\tau_a = 16ms$ which is equivalent to $\mathcal{N}_a = 2048$. The mean and standard deviation of the X–band distribution are 0.056 rads and 0.038 rads respectively. These values are such that equations (4.6) and (4.7) can be considered valid. Therefore, during transfer function estimate, the system phase noise is such that *time domain signal averaging* of at least 2048 sequences becomes viable.

C. The Differential Phase $\psi_\zeta'(n)$

Since a finite phase sequence $\psi_\zeta(t)$ is used for spectral estimation, inconsistent estimates may result from different truncated sequences unless $\psi_\zeta(t)$ closely resembles a random process. In order to examine this problem, the probability distribution of the differential phase $\psi_\zeta'(n)$ at L–band and X–band have been computed.

The differential phase corresponding to the averaging interval τ_a is defined as

$$\psi_\zeta'(n) = \frac{\Delta_{\tau_a}\psi_\zeta(n)}{\tau_a} \qquad n \in integer \qquad (4.8)$$

Figure 4.8 shows the experimental results obtained during back–to–back at L–band and X–band tests. Gaussian distributions with the same mean and standard deviation are superimposed for comparison purposes.

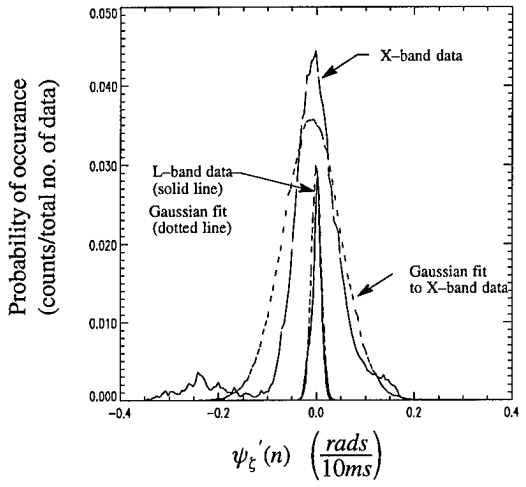


Figure 4.8 Probability Distribution of the Differential Phase $\psi_\zeta'(n)$ with $\tau_a = 16ms$ at L–band and X–band

V. Conclusions

A. *Phase Noise Measurement*

The use of complex I and Q demodulation and phase unwrapping applied to phase noise measurement has been presented. Although the particular equipment described in this paper is not suitable for the measurement of the phase noise characteristics of an individual oscillator, it provides a useful configuration for measuring the *overall* phase noise characteristics of an I and Q demodulation system.

The averaged phase noise spectral density $\overline{S_\psi(k)}$ shows the expected f^{-3} flicker frequency noise in the range $\sim 0.1Hz$ to $\sim 50Hz$. However, the method still has some difficulties in characterizing the low frequency spectrum in the milli–hertz range. Better estimates could be obtained if longer phase sequences were to be captured for the analysis.

A strong 50Hz component together with its harmonics were found in the spectrum. This is induced from the 50Hz AC source of the local power supplies and is extremely difficult to eliminate in a complete system. Better screening of the power supplies or the use of low noise battery sources might be the solution to this problem.

B *Transhorizon Experiment*

Both the maximum phase excursion $\hat{\psi}_\zeta(\tau_a)$ and the differential phase $\psi_\zeta'(n)$ relate to the processing error of the averaging process. A compromise between the averaging period and the phase noise characteristics of the system has to be made in order to ensure **minimum averaging error** whilst maintaining **optimum processing gain.**

During the development of the transhorizon experimental system, it was found that the oscillator phase noise was limiting the acceptable averaging period. As a result all oscillator crystals were replaced giving a 20dB improvement in phase noise. The system now is capable of carrying out a high number of averages (e.g. $\mathcal{N}_a = 2048$) before distortion occurs. SNR improvements can now be achieved to cater for weak signals such as those prevalent under tropospheric scatter propagation conditions.

Acknowledgment

The author would like to thank the U.K. Radiocommunications Agency for the sponsorship of the experiment.

References

[1] E. Vilar, J. Austin, K.W. Wan, W.K. Lam, X. Shen, and M. Juy, "A wide–band Transhorizon Experiment at 11.6GHz," *The 8th International Conference on Antennas and Propagation (ICAP), Edinburgh (U.K.),* Mar., 1993, pp. 1.441–1.445.

[2] K.W. Wan, J. Austin and E. Vilar, "A Novel Approach to the Simultaneous Measurement of Phase and Amplitude Noise in Oscillators," *Proceedings of the 44th Annual Symposium on frequency Control, USA,* May 90, pp 140–144.

[3] S. Lawerence Marple, Jr., "Digital Spectral Analysis with Applications," *Prentice Hall Signal Processing Series*, 1987.

[4] J. Rutman, "Characterization of Phase and Frequency Instabilities in Precision Frequency Sources: Fifteen Years of Progress," *Proceedings of the IEEE, vol. 66, No.9,* Sept.,1978, pp. 1048–1075.

[5] S.M. Kay and S. L. Marple, " Spectrum Analysis – A Modern Perspective", *Proceedings of the IEEE, vol., 69, No. 11,* Nov., 1981, pp. 1380–1419.

[6] F.J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform", *Proceedings of the IEEE, vol., 66, No. 1,* Jan., 1978, pp. 51–83.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

REDUCING ERRORS, COMPLEXITY, AND MEASUREMENT TIME OF PM NOISE MEASUREMENTS

F. L. Walls
Time and Frequency, National Institute of Standards and Technology
325 Broadway, Boulder, CO 80303

Abstract

This paper shows that a new measurement technique based on the two-oscillator technique and the addition of a noise source in series with the reference oscillator can significantly reduce calibration time for accurate PM measurements in oscillators and other components as compared to the traditional two-oscillator technique. This technique also significantly reduces the measurement time and improves the accuracy of 3-cornered-hat measurements. Measurement complexity is greatly reduced. The noise source is used to generate a known level of PM noise (PMCAL) for calibrating the product of mixer sensitivity and amplifier gain with Fourier frequency. This can be used to correct for PLL effects when PMCAL is larger than the residual phase noise in the oscillator under test. PMCAL is typically constant to ± 0.1 dB for Fourier frequencies from 0 to 5% of the carrier (maximum width typically less than 500 MHz). When the PMCAL is off, the noise added to the reference signal is typically less than -150 dBc/Hz at 1 Hz and -190 dBc/Hz at 10 kHz for carrier frequencies of 5 to 100 MHz. A similar system also works in the microwave range.

INTRODUCTION

Many applications require repetitive phase modulation (PM) noise measurements at a few standard frequencies. One of the primary factors limiting the accuracy of the traditional two-oscillator technique is the measurement of the mixer sensitivity and the calibration of amplifier gains versus frequency. The phase-locked-loop (PLL) and the mixer-amplifier interaction can also lead to errors. Changing the oscillator's output power, impedance, or length of the cable changes k_d. As a result the k_d requires a new determination. Correction for amplifier or mixer gain with Fourier frequency usually requires a complex computerized measurement system. Overall accuracy is dependent on operator skill for manual systems and ± 2-4 dB in commercial computerized systems [1].

Contribution of U.S. Government, not subject to copyright.

This paper shows that a new measurement technique based on the two-oscillator technique and the addition of a noise source in series with the reference oscillator can significantly reduce calibration time for accurate PM noise measurements in oscillators and other components as compared to the traditional two-oscillator technique [2,3]. This technique greatly reduces the measurement complexity as well. A Gaussian noise source is used to generate a known level of PM noise (PMCAL) for calibrating the product of mixer sensitivity and amplifier gain with Fourier frequency. This can be used to correct for PLL effects when PMCAL is larger than the residual phase noise in the oscillator under test. PMCAL is typically constant to ± 0.1 dB for Fourier frequencies from 0 to 5% of the carrier (maximum width typically less than 500 MHz). When the PMCAL is off, the noise added to the reference signal is typically less than -150 dBc/Hz at 1 Hz and -190 dBc/Hz at 10 kHz. Insertion losses are typically less than 1 dB.

When the noise in available reference sources and measurements is too high to determine the PM noise in a source, two references equipped with calibrated noise sources can be used in a cross-correlation configuration to reduce the noise contribution of the reference and the noise floor of the measurement system by approximately 15-20 dB. Similar cross-correlation techniques can be applied to the measurement of PM noise in other devices [1,4].

CALIBRATION OF PHASE NOISE IN PRODUCTION OSCILLATORS

A. Traditional Approaches

To show the improvements obtained from this new approach, it is necessary to review the traditional

U.S. Government work not protected by U.S. copyright

approach. Figure 1 shows the block diagram of a traditional measurement configuration for measuring PM noise in oscillators when the noise of the reference oscillator and the measurement system can be neglected [1]. The calibration sequence is:

1. Measure the mixer sensitivity k_d, typically by allowing the oscillators to beat and measuring the period and slope of the waveform at the zero crossings.

2. Phase-lock the oscillators together. If measurements close to the carrier are required, determine the action of the PLL on the phase variations.

3. Measure the power spectral density (PSD) of the noise voltage $V_n(f)$. Table 1 gives the 95% confidence intervals versus the number of averages.

4. Compute $\mathcal{L}(f)$ or $S_\phi(f)$ from $\mathcal{L}(f) = \frac{1}{2}S\phi(f) = \frac{1}{2}\text{PSD } V_n(f)/(k_dG(f))^2$.

Step 4 actually measures

$$S_\phi(f) = S_\phi(f)_{DUT} + S_\phi(f)_{REF} + S_\phi(f)_{MS} + \beta\, S_a(f)_{DUT} + \beta\, S_a(f)_{REF}, \quad (1)$$

where $S_\phi(f)_{DUT}$ is the PM noise of the device under test (DUT), $S_\phi(f)$ is the PM noise of the reference, $S_\phi(f)_{MS}$ is the noise floor of the measurement system (in this configuration), β is the AM to PM conversion factor of the mixer, $S_a(f)$ is the AM noise of the DUT and $S_a(f)$ is the AM noise of the Reference [1]. The complete error model for the measurement is given in Table 2. The various error parameters are in general dependent on f. Overall accuracy is dependent on operator skill for manual systems and approximately ± 2-4 dB in commercial computerized systems. An oscilloscope or other recording device is usually required to determine k_d.

The measurement of $k_d(f)$ typically takes many minutes and can be one of the major factors limiting the accuracy. Changing the oscillator output level, the driving impedance, or length of the cable changes k_d. As a result a new determination of k_d is measured. Correction for amplifier or mixer gain with Fourier frequency usually requires a complex (often computerized) measurement system.

B. Phase Noise Standard Approach

Figure 2 shows an alternate approach that reduces the calibration time and usually improves the accuracy. A calibrated noise source centered about the carrier frequency has been added to the reference signal using a directional coupler. When the calibrated noise source is off the residual PM noise on the reference signal due to the directional coupler and other components is negligible. The phase noise added when PMCAL is on is typically flat to ± 0.1 dB for Fourier frequencies from dc to about 1/4 the bandwidth of the bandpass filter following the noise source. The calibration of the added phase noise is similar to that developed for the NIST phase noise standards [3] and covered by US patent [7]. This calibration only needs to be done occasionally. We have found ours to be stable to ± 0.3 dB for several years. The calibration procedure now becomes:

1. Phase lock the oscillator under test to the reference oscillator.

2. Turn on the calibrated PM noise. Measure PSD of $V_n(f)_{on} = \text{PMCAL}(k_dG(f))^2$. If the mixer sensitivity and amplifier gain are constant with Fourier frequency, PSD $V_n(f)_{on}$ is a constant above the PLL bandwidth. If measurements need to be taken close to the carrier one can measure $\text{PSDV}_n(f)_{on}$ over the same range to account for effect of the PLL.

3. Turn the calibrated PM noise off and measure PSD of $V_n(f)_{off}=\text{PSD } \delta\phi(f)(k_dG(f))^2$.

4. Compute $S_\phi(f)$ from $S_\phi(f) = \text{PMCAL}(\text{PSDV}(f)_{off}/\text{PSDV}(f)_{on})$.

The error model for this approach is shown in Table 3. The accuracy is approximately ± 0.2 dB ± 0.1-0.25 dB ± $10\log(1 + 1.9/N^{1/2})$, where the accuracy of the calibration of PMCAL is typically ± 0.2, the uncertainty of the measurement of $\text{PSDV}_n(f)_{on}$ is ±0.1-0.25 dB, and N is the number of measurements of $V_n(f)$. Changes in amplifier or mixer gain with Fourier frequency, even errors in spectrum analyzer voltage reference and noise bandwidth are calibrated by measuring the ratio of PSD $V_n(f)_{on}$ to PMCAL on the spectrum analyzer. The resulting accuracy of approximately 1 dB exceeds the accuracy of most traditional approaches. The skill required to make the measurements is lower than that for most of the computer controlled systems and much lower than that required for the manual systems. Less

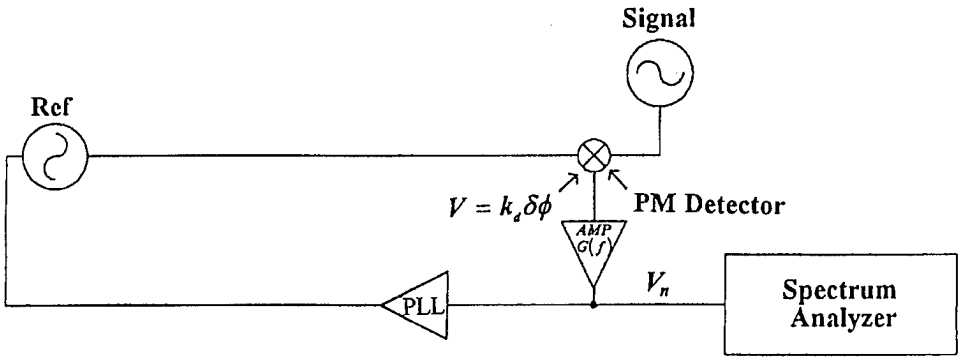


Figure 1. Block diagram of traditional two oscillator measurement system

TABLE 1. STATISTICAL UNCERTAINTY OF SPECTRAL DENSITY MEASUREMENTS

$S_m(f) = S(f)\ [1 \pm k/N^{1/2}]$ for FFT measurements, $S_m(f) = S(f)\ [1 \pm k\ (VIDEO_{BW}/NRES_{BW})^{1/2}]$ for swept measurements, k = 1 yields a confidence interval with 68% probability, k = 1.9 yields a confidence interval with 95% probability when the number of samples averaged N is greater than approximately 30, $VIDEO_{BW}$ is the video bandwidth, RES_{BW} is the resolution bandwidth and should be less than f/10. To avoid leakage biases f should be greater than the FFT span/75 for Hanning windows and the FFT span/23 for flat top windows [5].

Number of Samples	k ≃ 1 (approx. 68%) $S_m = S\ [1\pm]$ S = S_m +dB	k ≃ 1.9 (approx. 95%) $S_m = S\ [1\pm]$ S = S_m +dB
4	0.54 -2 , +3.3	2.5 -3 , +6
6	0.42 -1.5 , +2.3	1.4 -2.5 , +5
10	0.32 -1.2 , +1.7	0.61 -2.1 , +4
30	0.18 -0.72, +.86	0.35 -1.3 , +1.8
100	0.1 -0.41, +0.46	0.19 -0.76, +0.92
200	0.058 -0.24, +0.25	0.14 -0.46, +0.51
1000	0.032 -0.13, +0.13	0.06 -0.26, +0.28
3000	0.018 -0.08, +0.08	0.035 -0.15, +0.15
10000	0.01 -0.04, +0.04	0.019 -0.08, +0.08

TABLE 2. PM ERROR MODEL FOR TRADITIONAL "2 OSCILLATOR" MEASUREMENTS

5 MHz CARRIER, f = 10 Hz to 100 kHz

	TYPICAL (dB)	TIME
1. DETERMINATION OF k_d	0.1	300 s
2. DETERMINATION OF G(f) VERSUS f	0.5-2	--
3. PLL EFFECTS AT LOW f	?	0 - 600 s
4. LINEARITY OF SPECTRUM ANALYZERS	0.1	
5. CONTRIBUTION OF AM	0.05	
6. 95% STAT. CONFIDENCE FOR N = 390	0.4	210 - 780 s
7. ACCURACY OF PSD FUNCTION	0.1	
8. HARMONIC DISTORTION EFFECTS	?	
9. SYSTEM NOISE FLOOR CONTRIBUTION	?	
10. UNFOLDING 3-CORNERED-HAT	?	
TOTAL	1-4 dB	10-30 MIN

TABLE 3. PM ERROR MODEL FOR 2 OSCILLATOR MEASUREMENTS WITH PMCAL

5 MHz, f = 10 Hz to 100 kHz

	TYPICAL (dB)	TIME
1. DETERMINATION OF k_d	0.45	10 s
2. DETERMINATION OF G(f) VERSUS f	INCLUDED	--
3. PLL EFFECTS AT LOW f	?	0 - 230 s
4. LINEARITY OF SPECTRUM ANALYZERS	0.1	
5. CONTRIBUTION OF AM	0.05	
6. 95% STAT. CONFIDENCE FOR N = 390	0.4	210 s
7. ACCURACY OF PSD FUNCTIONS	INCLUDED	
8. HARMONIC DISTORTION	?	
9. SYSTEM NOISE FLOOR CONTRIBUTION	?	
10. UNFOLDING 3-CORNERED-HAT	?	
TOTAL	1 dB	4 - 8 MIN

equipment is also required since one does not need the counter or recording device used to measure k_d in the traditional approach. The measurement of $PSDV_n(f)_{on}$ is typically very fast as well because the noise is approximately white [5,6]. Taking 1000 measurements at 10 kHz using a 100 kHz scan takes an FFT analyzer less than 10 s. This corresponds to a 95% confidence interval of ± 0.25 dB. This improves to ±0.1 dB by averaging 7 adjacent points [5,6].

C. Cross Correllation Phase Noise Standard Approach

The configuration shown in Figure 3 can be used to improve the noise floor of figures 1 and 2 approximately 20 dB [1,4]. This is useful in cases where the phase noise of the oscillator under test is close to or lower than the reference oscillator or measurement system. Two channels are used to measure the PM noise between the oscillator under test and two independent references, each of which are equipped with the PMCAL technology. Both independent oscillators are phased locked to the oscillator under test. A two channel FFT is used to measure the $PSDV_{n1}$ calibration from channel 1 and the $PSDV_{n2}$ from channel 2. The sensitivity for the cross-spectrum (CS) is the mean of the measured values for channel 1 and 2. The two-channel FFT analyzer measures channel 1, channel 2, and the cross-spectrum simultaneously. The measurement sequence is

1. Phase lock both reference oscillators to the oscillator under test.

2. Turn on the calibrated PM noise in both reference oscillators (PMCAL1 and 2). Measure PSD of $V_n(f)_{on} = PMCAL(k_dG(f))^2$ for both channels at the highest Fourier frequency of interest. If measurements need to be taken close to the carrier one can measure $PSDV_n(f)_{on}$ over the same range to account for effect of the PLL.

3. Turn the calibrated PM noise "off" and measure the $PSDV_n(f)_{off}$ for each channel and the cross spectrum $PSD(V_{n1} \times V_{n2})_{off}$.

4. Compute $S_\phi(f)$ from:

$$S_\phi(f) = PSD(V_{n1} \times V_{n2})_{off}/[k_dG(f)_1(k_dG(f)_2)]$$

$$= PSD(V_{n1} \times V_{n2})/[(PSDV_{n1on}/PMCAL_1)^{1/2} (PSDV_{n2on}/PMCAL_2)^{1/2}].$$

Step 4 actually measures

$$S_\phi(f) = S_\phi(f)_{DUT} + (S_\phi(f)_{REF} + S_\phi(f)_{MS1} + S_\phi(f)_{MS2})/N^{1/2} + \beta S_a(f)_{DUT} + S_a(f)_{REF}. \quad (2)$$

The contributions of the reference and the two measurement systems to the noise floor are reduced by $N^{1/2}$. The accuracy of the measurement is approximately ± 0.2 dB ± 0.1-0.25 dB ± 10 log(1 + {(PSD V_{n1} +PSD V_{n2})/PSD CS}/$N^{1/2}$), where the accuracy of the calibration of PMCAL is typically ± 0.2, the uncertainty of the measurement of PMCAL for each channel is of order ± 0.1-0.25 dB.

This accuracy far exceeds the accuracy of traditional three-cornered-hat techniques. First, errors in the calibration of each channel are linear in the resulting estimate of the oscillator phase noise, whereas in a traditional three-cornered-hat the estimate of the oscillator phase noise is derived by subtracting large numbers to obtain a small value. As a result the estimate of the phase noise of the oscillator converges much better (no negative PSDs). Second, the amount of time required for a given accuracy is greatly reduced because only one set of $V_n(f)_{off}$ data is required and the calibration of k_{d1} and k_{d2} is fast. With this approach it is possible to measure an oscillator that is 10 dB better than either reference to a precision of approximately ± 3 dB with 1500 samples. If the oscillator under test is 17 dB better than the references then 10 000 samples yields a precision of approximately ± 3 dB. Thirdly, the noise floor of the measurement system is reduced by approximately $N^{1/2}$ as compared to the traditional three-cornered-hat technique.

MEASUREMENT OF PASSIVE COMPONENTS

A. Traditional Approach

Figure 4 shows the block diagram of a traditional setup for measuring the phase noise added by a passive component. The calibration sequence is

1. Adjust the phase shift ϕ so that the dc output voltage of the mixer is nominally 0.

2. Measure the mixer sensitivity k_d, typically by terminating divider B output and substituting another oscillator of equal source impedance and power. The oscillators are allowed to beat and the period and slope of the waveform at the zero crossings measured to determine the slope in volts per radian.

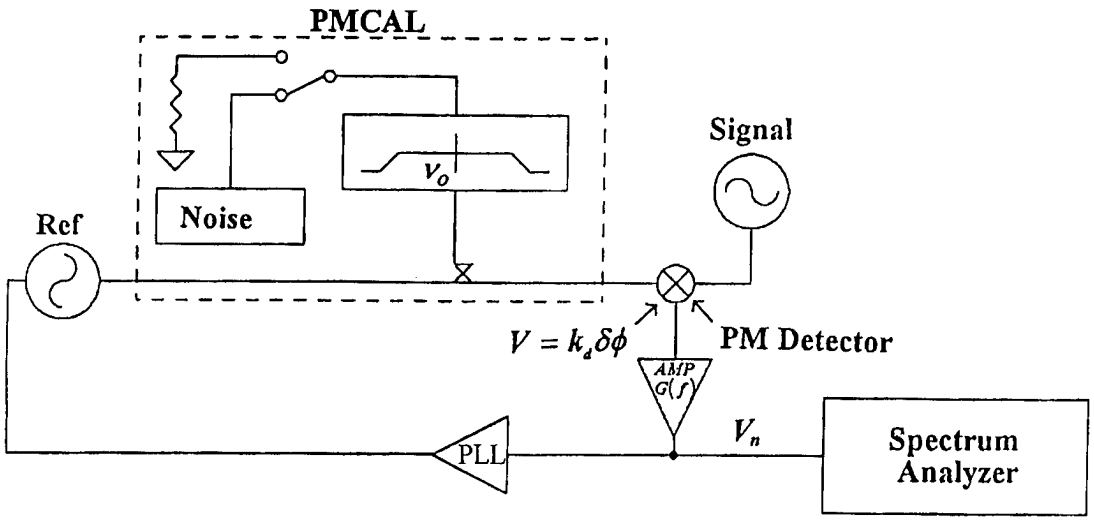


Figure 2. Block diagram of new measurement system using PMCAL to calibrate mixer sensitivity and amplifier gain with Fourier frequency.

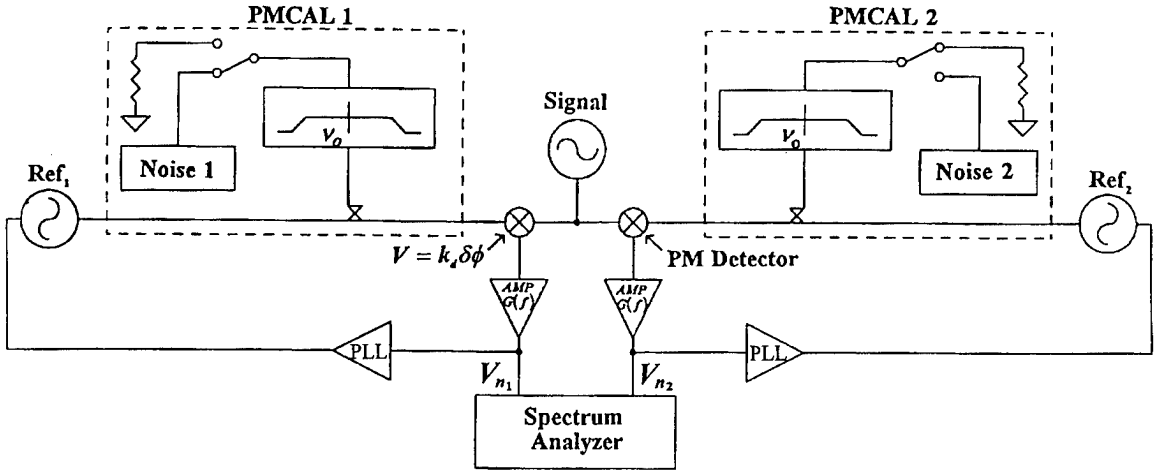


Figure 3. Block diagram of new system for 3-cornered-hat measurements using dual PMCAL noise sources to calibrate mixer sensitivity and amplifier gain with Fourier frequency.

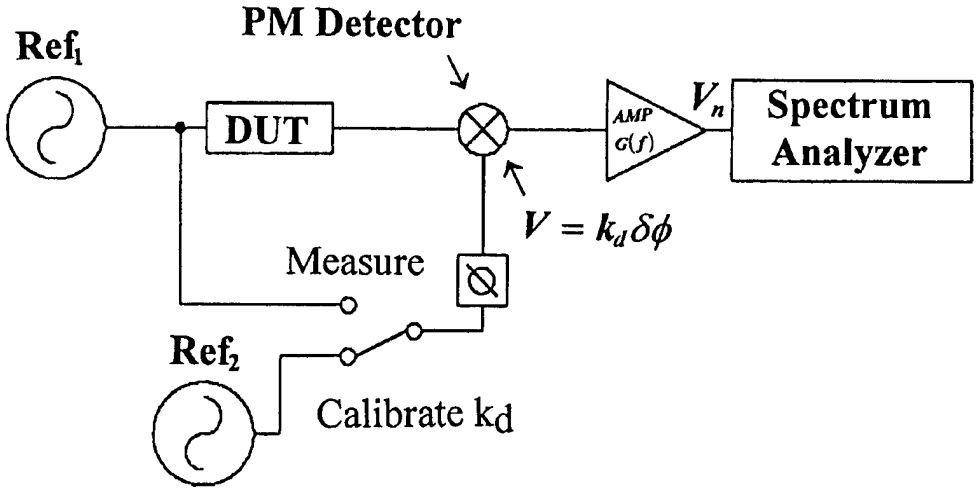


Figure 4. Block diagram of traditional two oscillator measurement system for measuring the phase noise added by a device under test (DUT).

3. Reattach Channel B.

4. Measure the power spectral density (PSD) of the noise voltage V_n.

5. Compute $S_\phi(f) = \text{PSD } V_n/(k_d G(f))^2$.

Step 4 actually measures

$$S_\phi(f) = S_\phi(f)_{DUT} + S_\phi(f)_{MS} + \delta S_\phi(f)_{REF} + \beta S_a(f)_{DUT} + \beta S_a(f)_{REF}, \quad (3)$$

where δ takes into account the decorrelation of the PM noise in the reference [1]. The measurement of k_d takes several minutes, and the accuracy depends on the extent that the output level, the driving impedance, or length of the cable used with the substitution oscillator matches the actual drive from divider output B.

B. Phase Noise Standard Approach

Figure 5 shows an alternate approach for this meausrement that reduces the calibration time and usually improves the accuracy. This approach uses the PMCAL technology described above. The added phase noise is typically flat to $\pm$ 0.1 dB for Fourier frequencies from dc to about 1/4 the bandwidth of the bandpass filter following the noise source. The calibration procedure now becomes:

1. Turn on the calibrated PM noise. Measure PSD of $V_n(f)_{on} = PMCAL(k_d G(f))^2$ at the highest Fourier frequency of interest. For 1000 measurements at 10 kHz this typically takes an FFT analyzer less than 10 s. This corresponds to an uncertainty in the measurement of $\pm$ 0.25 dB. By averaging 7 adjacent points this improves to $\pm$ 0.1 dB. The uncertainty of PMCAL is typically $\pm$ 0.25 dB. There are no PLL effects for this measurement.

2. Turn the calibrated PM noise off and remeasure PSD $V_n(f)_{off}$.

3. Compute $S_\phi(f) = PMCAL(PSDV_n(f)_{off}/PSDV_n(f)_{on})$.

The accuracy of this approach is approximately $\pm$ 0.2 dB $\pm$ 0.1-0.25 dB $\pm$ 10 log(1 + $1/N^{1/2}$), where the accuracy of the calibration of PMCAL is typically $\pm$ 0.2, 0.1 is the uncertainty of the measurement of $PSDV_n(f)_{on}$, and N is the number of averages of V_n. The resulting accuracy summarized in Table 4 of approximately $\pm$ 1 dB far exceeds the accuracy of most traditional approaches and it is obtained in a much shorter time.

If the PM noise of the passive component under test is close to or better than the PM noise of the measurement system, cross correlation techniques similar to those described above for oscillators can be used to improve the noise floor by approximately 15-20 dB [1,4].

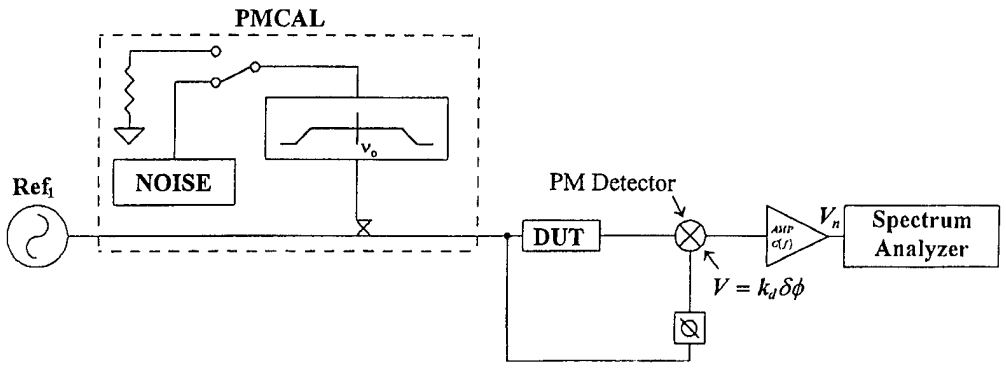


Figure 5. Block diagram of new measurement system for measuring the phase noise added by a device under test (DUT) using PMCAL to calibrate mixer sensitivity and amplifier gain with Fourier frequency.

TABLE 4. PM ERROR MODEL FOR PASSIVE COMPONENTS WITH PMCAL

5 MHz, f = 10 Hz to 100 kHz

	TYPICAL (dB)	TIME
1. DETERMINATION OF k_d	0.45	10 s
2. DETERMINATION OF G(f) VERSUS f	INCLUDED	--
3. PLL EFFECTS AT LOW f	NONE	
4. LINEARITY OF SPECTRUM ANALYZERS	0.1	
5. CONTRIBUTION OF AM	0.05	
6. 95% STAT. CONFIDENCE FOR N = 390	0.4	210 s
7. ACCURACY OF PSD FUNCTION	INCLUDED	
8. HARMONIC DISTORTION	?	
9. SYSTEM NOISE FLOOR CONTRIBUTION	?	
10. UNFOLDING 3-CORNERED-HAT	?	
TOTAL	1 dB	4 MIN

CONCLUSION

Common to all the three new measurement systems discussed is the use of a calibrated noise source to add broadband Gaussian noise to the reference oscillator. By using the existing house standard(s) for the reference(s), we obtain considerable savings over a system that requires new internal references. Significant reduction in the complexity of the measurement process is also obtained. No longer is it necessary to use an elaborate computer program to correct for changes in amplifier gain or PLL effects with Fourier frequency. As a result, measurements can be made much faster than before and will have improved accuracy. Also, both of the oscillator measurements schemes (Figures 2 and 3) can easily be adapted to large scale testing by using a matrix switch to connect the oscillators under test to the measurement system.

ACKNOWLEDGEMENTS

The author thanks F. Ascarrunz, L. M. Nelson, and P. Pond for help in construction and to D. A. Howe for helpful comments. This work was supported in part by the Calibration Coordination group of JCTG/GMT-JLC and the Space and Terrestrial Communications Evaluation Division, Tactical Support Branch, Ft. Monmouth, NJ.

REFERENCES

1. F. L. Walls, A. J. D. Clements, C. M. Felton, M. A. Lombardi, and M. D. Vanek, "Extending the Range and Accuracy of Phase Noise Measurements," Proc. of 42nd Frequency Control Symposium, 1988, pp. 432-441; also "Characterization of Clocks and Oscillators," NIST Tech Note 1337, March 1990.
2. F. L. Walls, A. J. D. Clements, C. M. Felton, and T. D. Martin, Precision Phase Noise Metrology," Proc. NCSL, Albuquerque, NM, August, 1991, pp. 257-275.
3. F. L. Walls, "Secondary Standard for PM and AM Noise at 5, 10, and 100 MHz," Proc. of 1992 IEEE Frequency Control Symposium, Hershey, PA, May 278-29, 1992, pp. 290-299.
4. F. L. Walls, "Cross-correlation Phase Noise Measurements," Proc. of 46th Annual Frequency Control Symposium, 1992, pp. 257-261.
5. W. F. Walls, D. B. Percival, and W. R. Irelan, "Biases and Variances of Several FFT Spectral Estimators as a Function of Noise Type and Number of Samples," Proc. of 43rd Annual Frequency Control Symposium, 1989, pp. 336-341.
6. D. B. Percival and A. T. Walden, "Spectral Analysis for physical Application," Cambridge Univ. Press, 1993; B. N. Taylor and C. E. Kuyatt, NIST Technical Note TN1297, 1993.
7. F. L. Walls, "Frequency Calibration Standard Using a Wide Band Phase Modulator," United States Patent - No. 5,101,506, 1992.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

RELATIONSHIP OF AM TO PM NOISE IN SELECTED RF OSCILLATORS ♦

L. M. Nelson, C. W. Nelson, and F. L. Walls
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, CO 80303

Abstract

We have studied the amplitude modulation (AM) and phase modulation (PM) noise in a number of 5 MHz and 100 MHz oscillators to provide a basis for developing models of the origin of AM noise. To adequately characterize the AM noise in high performance quartz oscillators, we found it necessary to use two-channel cross-correlation AM detection. In the quartz oscillators studied, the power spectral density (PSD) of the f^{-1} and f^{0} regions of AM noise is closely related to that of the PM noise. The major difference between different oscillators of the same design depends on the flicker performance of the resonator. We therefore propose that the f^{-1} and f^{0} regions of AM and PM noise arise from the same physical processes probably originating in the sustaining amplifier.

Introduction

To our knowledge there has been little quantitative work analyzing the relationship between AM and PM noise in oscillators. Most of the literature assumes that the AM noise is much smaller than the PM noise. Leeson, Parker, and others have described models of PM noise in oscillators [1-3]. We are not aware, however, of any theory for the spectral dependence of the AM noise close to the carrier. We have therefore undertaken a detailed study of PM and AM noise in several quartz oscillators at 5 and 100 MHz to provide a basis for developing models of the origin of AM noise.

To adequately characterize the AM noise in high performance quartz oscillators, we found it necessary to use two-channel cross-correlation AM detection. The signal is split by a reactive power splitter into two signals which are each AM detected. The resulting baseband signals are then amplified and the power spectral density (PSD) of the cross spectrum between the two channels is computed. This technique provides a noise floor of order -190 dBc/Hz for AM noise at signals of approximately +13 dBm.

An improved three-corner-hat technique using cross correlation makes PM noise measurements more quickly and accurately than the traditional three-corner-hat measurement technique [4].

In many of the quartz oscillators studied, the PSD of the f^{-1} and f^{0} regions of AM noise is closely related to that of the PM noise over a substantial range in Fourier frequency. The major difference between different oscillators of the same design depends on the flicker performance of the resonator. AM and PM noise generally differ only in the region where the PM noise varies as f^{-3} (flicker FM noise). For example, we have found that the AM noise of 5 MHz oscillators is quite similar to the PM noise for Fourier frequencies above approximately 3 Hz. Likewise, the AM and PM noise of the 100 MHz quartz oscillators studied were quite similar for Fourier frequencies above approximately 300 Hz.

The relationship between the AM and PM noise can be exploited in many areas of analysis to assist in the design of new oscillators. It is tedious to determine the PM performance of a new oscillator design, especially one that exhibits performance superior to existing references. At the very minimum three oscillators of similar performance are required. (The requirements on the performance of the other oscillators can be relaxed considerably if the modified three-corner-hat approach is used.) On the other hand the AM performance of a single superior oscillator can be easily assessed using the two-channel AM detection techniques presented in this paper. Therefore, it can be beneficial to use the readily obtained AM performance to evaluate the noise more quickly.

Description of AM Detection Circuits

Figure 1 shows the single-sideband noise floor of a variety of available AM detectors at 5 MHz. All exhibit noise floors or resolution far above the phase noise of available oscillators at Fourier frequencies above 1 kHz and hence are unsuitable for measurements of the AM noise in these oscillators. Parker has used double-balanced mixers as phase detectors with good results at

♦ Work of U.S. government. Not subject to U.S. copyright.

U.S. Government work not protected by U.S. copyright

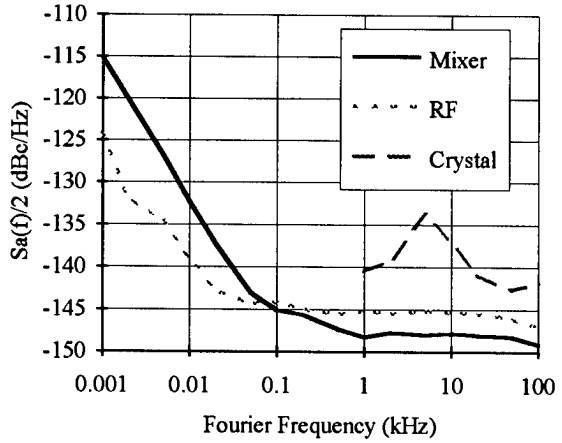


Figure 1. AM noise floor of various detectors. The noise of a mixer is included for comparison. The mixer is a WJ M9HC, the RF Detector is a WBE model A33, and the Crystal Detector is a HP 8472B.

drive powers of +23 dBm [5]. Typically the oscillators with the best performance phase noise at 5 and 100 MHz have outputs of +13 dBm (see Figure 2). This leaves us with a severe measurement problem due to the output power level being lower in these oscillators. The problem has been overcome by combining the approach of Parker with a two-channel cross-correlation technique [6-8]. Figure 3 shows the full AM measurement system schematic that was used for the cross-correlation AM noise detection.

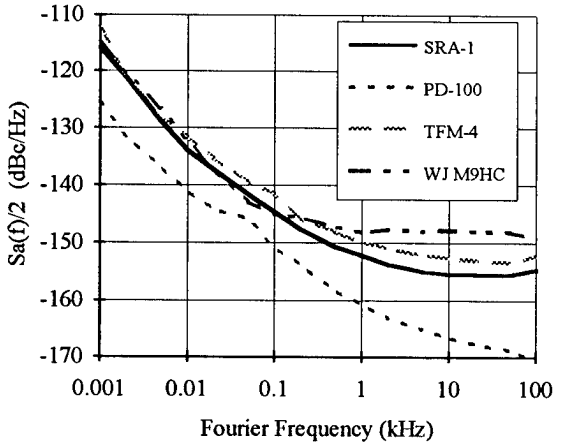


Figure 2. AM noise floors of two mixer types [9]. The PD-100 is a single-balanced mixer, the SRA-1 is a double-balanced mixer, the TFM-4 is a double-balanced mixer, and the WJ M9HC is a double-balanced mixer.

Measurement Techniques

At the output of the spectrum analyzer we obtain $S_a(f)/2$ for AM noise, or $S_\phi(f)/2$ for PM noise measurements. These are the single-sideband spectral densities of our signal. The practical definition for the power spectral density of amplitude fluctuations is given by

$$S_a(f) = \frac{(\Delta\varepsilon(f))^2}{V_o^2}\frac{1}{BW} \qquad (1)$$

Where $\Delta\varepsilon(f)$ is the change in amplitude measured at Fourier separation f from the carrier, Vo is the average carrier voltage and BW is the noise bandwidth of the spectrum analyzer. For the measurement system shown in Figure 3 Equation (1) became

$$S_a(f) = \frac{PSD(V_n)}{(k_a G(f))^2} \qquad (2)$$

where PSD denotes power spectral density, Vn is the noise voltage out, k_a is the sensitivity factor of the detector for converting $\Delta\varepsilon/V_0$ to a voltage, and G(f) is the amplifier gain.

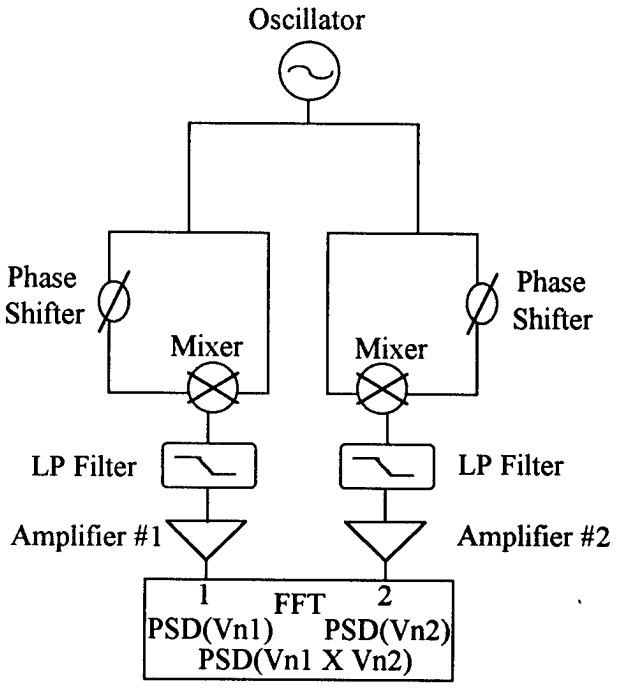


Figure 3. AM cross-correlation measurement system.

At the output of the spectrum analyzer we obtain the power spectral density (PSD) of each of the two channels. This provides us with the noise from the source and noise in each respective channel (see Equations (3) and (4)). By using cross-correlation we can obtain the source noise as Equation (5). The uncorrelated noise between the channels averages away as one over the square root of the number of averages.

$$PSD(V_{n1}) = \alpha(AM_{Source} + System_1) \quad (3)$$

$$PSD(V_{n2}) = \alpha(AM_{Source} + System_2) \quad (4)$$

$$PSD(V_{n1}XV_{n2}) = \alpha[AM_{Source}(+\frac{System_1 + System_2}{N^{1/2}})] \quad (5)$$

$$PSD(V_{n1}) - PSD(V_{n1}XV_{n2}) = \alpha(System_1) \quad (6)$$

$$PSD(V_{n2}) - PSD(V_{n1}XV_{n2}) = \alpha(System_2) \quad (7)$$

where the proportionality constant $\alpha = (k_a G(f))^2$. Subtracting the source noise from Equations (3) and (4), respectively, we are able to obtain the noise in each channel (see Equations (6) and (7)).

AM and PM Noise from Selected Oscillators

AM and PM noise measurements on various oscillators are described below. The models for their relationships arise in the following section.

Figure 4 shows the AM and PM noise for a 5 MHz oscillator that uses an AT-cut resonator driven at low power. The AM noise was measured using the techniques described in the previous section. Calibration of the PM noise was done using a three-corner-hat technique outlined in [8]. The PM noise in the 0.3 to 3 Hz range shows flicker frequency (f^{-3}) as expected with white PM noise at large Fourier frequencies. The AM noise shows f^{-2} (white frequency) behavior as does the PM noise until 30 Hz and continues to follow the white PM noise to higher Fourier frequencies. The models for the AM and PM noise are both similar for the f^{-2} and the f^0 components of the noise.

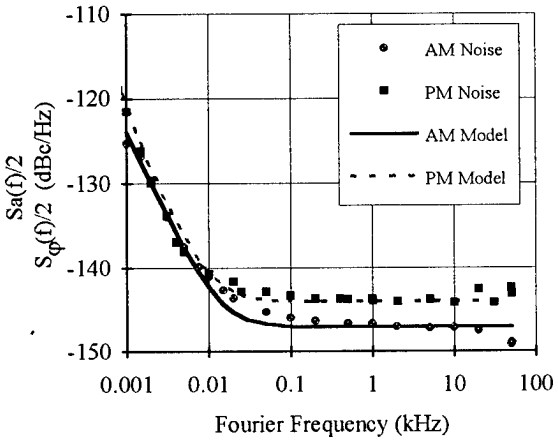


Figure 4. AM and PM noise with models for AT-cut 5 MHz oscillator.

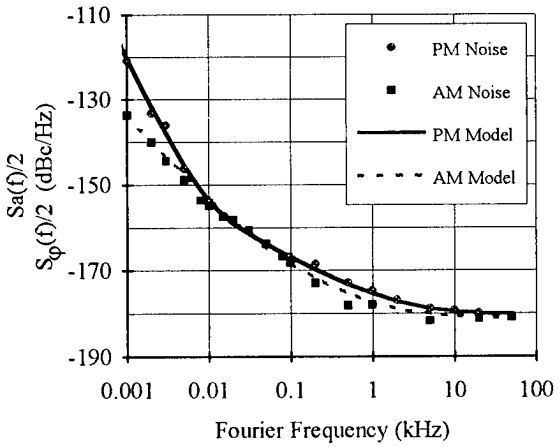


Figure 5. AM and PM noise with models for SC-cut 5 MHz oscillator.

Figure 5 shows AM and PM noise of a 5 MHz high performance oscillator that uses an SC-cut resonator driven at relatively high power. The AM and PM noise were measured with the techniques described above. The PM noise exhibits flicker FM, flicker PM, and white PM noise. The AM noise tracks the PM noise from approximately 10 Hz to at least 50 kHz. The models for the AM and PM noise are both similar for f^{-1} and f^0 components of the noise.

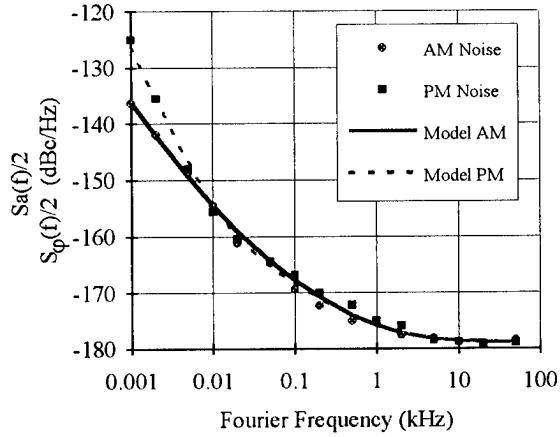


Figure 6. AM and PM noise with models for SC-cut 5 MHz oscillator.

Figure 6 shows AM and PM noise of another 5 MHz high performance oscillator of the same SC-cut design as Figure 5. The PM noise was measured using a cross-correlation technique described by W. Walls [9]. Again the PM noise exhibits flicker FM, flicker PM, and white PM noise. The AM noise close to the carrier has a white frequency component but follows the PM noise in both the flicker PM and white PM noise characteristics at higher frequencies.

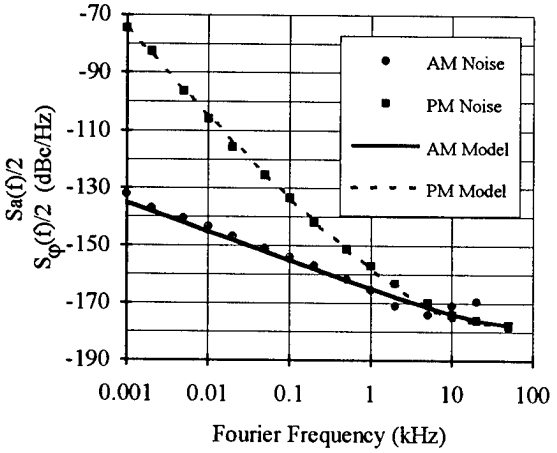


Figure 7. AM and PM noise with models for SC-cut 100 MHz oscillator.

Figure 7 shows AM and PM noise for a high performance 100 MHz oscillator. The AM noise was measured with the technique previously described. A three-corner-hat technique was used to obtain the PM noise. The PM noise fits the classic model described by Parker with flicker FM, white frequency, and white PM noise [10]. The AM noise falls as f^{-1} and meets up with the PM noise at higher frequencies.

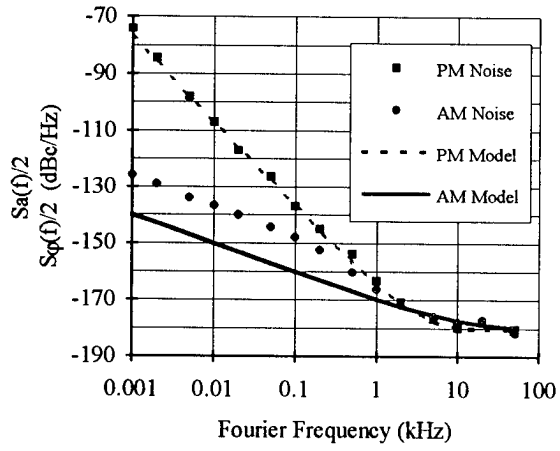


Figure 8. AM and PM noise with models for SC-cut 100 MHz oscillator.

Figure 8 shows AM and PM noise for another high performance 100 MHz oscillator. The AM noise was measured with the same technique as described above. The PM noise was measured using the modified three-corner-hat technique outlined in [4]. The PM noise follows Parker's model with flicker FM, white frequency, and white PM noise [10]. The AM noise falls as f^{-1} and follows the white PM noise of the PM noise at higher frequencies from the carrier.

Discussion

The AT-cut 5 MHz oscillator fit the following models with minimal variations. The f^{-2} noise only differs by 1 dB and the f^{0} noise differs by 3 dB.

$$\text{PM Model} \quad \frac{10^{-12.5}}{f^4}+\frac{10^{-13.9}}{f^3}+\frac{10^{-12.3}}{f^2}+10^{-14.4} \qquad (8)$$

$$\text{AM Model} \quad \frac{10^{-12.4}}{f^2}+10^{-14.7} \qquad (9)$$

The SC-cut 5 MHz oscillator in Figure 5 follows the following models

$$\text{PM Model} \quad \frac{10^{-12.1}}{f^4}+\frac{10^{-12.8}}{f^3}+\frac{10^{-14.7}}{f}+10^{-18.02} \qquad (10)$$

$$\text{AM Model} \quad \frac{10^{-13.4}}{f^2}+\frac{10^{-15}}{f}+10^{-18.1} \qquad (11)$$

The f^{-1} noise differs by 3 dB and the f^{0} only by eight tenths of a dB.

The 5 MHz oscillator in Figure 6 shows a similar AM and PM model relationship as the previous SC-cut 5 MHz oscillator with only slight variations in the exponents of the model.

$$\text{PM Model} \quad \frac{10^{-12.65}}{f^3}+\frac{10^{-14.87}}{f}+10^{-17.9} \qquad (12)$$

$$\text{AM Model} \quad \frac{10^{-13.65}}{f^2}+\frac{10^{-14.87}}{f}+10^{-17.9} \qquad (13)$$

The f^{-1} and f^{0} noise are identical in this case.

The 100 MHz oscillator in Figure 7 resembles the models from Figure 8 quite closely except for a slight difference in the f^{-2} exponent of the PM noise model. The AM and

PM f^0 models differ by approximately 3 dB.

PM Model $$\frac{10^{-7.46}}{f^3}+\frac{10^{-9.9}}{f^2}+10^{-17.77} \quad (14)$$

AM Model $$\frac{10^{-13.5}}{f}+10^{-18} \quad (15)$$

The 100 MHz oscillator in Figure 8 matches the following models with slight discrepancies between the model and the close in data. In this case the f^0 noise characteristics are identical.

PM Model $$\frac{10^{-7.7}}{f^3}+\frac{10^{-10.8}}{f^2}+10^{-18.06} \quad (16)$$

AM Model $$\frac{10^{-14}}{f}+10^{-18.06} \quad (17)$$

We do not understand the differences between the models for our 100 MHz AM and PM noise measurements. However, the model characteristics of Figures 7 and 8, whose oscillators were of similar cut and brand, have similar AM and PM models for their f^0 component. The f^0 model differs at most by 3 dB. The f^{-2} component; however, differs by about 9 dB.

Conclusion

The model relationships between the various 5 MHz oscillators present strong evidence to suggest that there is a connection between the AM and PM noise for a specific oscillator. The f^{-1} and f^0 noise processes in both the AM and PM measurements seem to come from the same source. The 100 MHz oscillator measurement relationships are inconclusive in this respect, although they maintain similar model relationships between the two SC-cut oscillators tested. We speculate that the difference in the model and the data in Figure 8 could be due to compression in the output amplifier of our system. The f^0 component of the 100 MHz oscillators; however, seems to have some AM and PM noise correspondence and could give some information about the thermal noise characteristics in our system.

Acknowledgments

The authors thank Huascar D. Ascarrunz, Eva S. Ferre, and Franklin G. Ascarrunz for their help with the measurements. They would also like to thank the Calibration Coordination group of JCTG/GMT-JLC and the Space and Terrestrial Communications Evaluation Division, Tactical Support Branch, Ft. Monmouth, New Jersey, for their partial financial support.

References

[1] D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," Proc. of the IEEE, vol. 54, pp. 329-330, February 1966.

[2] T. E. Parker and G. K. Montress, "Spectral Purity of Acoustic Resonator Oscillators," Proc. 46th Ann. Frequency Control Symp., 1992, pp. 340-348.

[3] G. S. Curtis, "The Relationship between Resonator and Oscillator Noise, and Resonator Noise Measurement Techniques," Proc. 41st Ann. Frequency Control Symp., 1987, pp. 420-428.

[4] F. L. Walls, "Reducing Errors, Complexity, and Measurement Time for PM Noise Measurements," To be published in the Proceedings of the 47th Annual International Frequency Control Symposium, 1993.

[5] G. K. Montress, T. E. Parker, and M. J. Loboda, "Residual Phase Noise Measurements of VHF, UHF and Microwave Components," Proc. 43th Ann. Frequency Control Symp., 1989, pp. 349-359.

[6] A. L. Lance, W. D. Seal, and F. Labaar. Infrared and Millimeter Waves. San Diego: Academic Press, Vol. II, 1984, ch. 7, pp. 284-286.

[7] F. L. Walls, A. J. D. Clements, C. M. Felton, and T. D. Martin, "Precision Phase Noise Metrology," Proc. of National Conference of Standards Laboratories (NCSL), 1991, pp. 257-275; also found in TUTORIAL 3A.

[8] W. F. Walls, "Cross-correlation Phase Noise Measurements," Proc. 46th Ann. Frequency Control Symp., 1992, pp. 257-261.

[9] Commercial equipment is identified in this paper in order to specify adequately the experimental design. This identification does not imply recommendation or endorsement by NIST, nor does it imply that the equipment is the best available for the purpose.

[10] T. E. Parker, "Characteristics and Sources of Phase Noise in stable oscillators," Proc. 41st Ann. Frequency Control Symp., 1987, pp. 99-110.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

INVESTIGATIONS OF AM AND PM NOISE IN X-BAND DEVICES*

F. G. Ascarrunz, E. S. Ferre and F. L. Walls
National Institute of Standards and Technology
325 Broadway
Boulder, CO 80303

Abstract

In this paper we report on measurements of phase modulation (PM) and amplitude modulation (AM) noise in a variety of amplifiers, dielectric resonator oscillator (DRO) sources and mixers at 10.6 GHz. There is little information on AM noise and only limited information on PM noise in microwave devices. Two channel measurement systems and cross-correlation analysis were used in these noise measurements. In amplifiers there are at least two random noise processes that generate equal levels of PM and AM noise when the amplifier is in the linear operating range. The noise processes correspond to a flicker noise process and a white noise process. In the DRO sources we show that these noise mechanisms are present in the active components and determine the AM and PM characteristics of the oscillator. We suggest a noise model for oscillators incorporating Leeson's PM noise model and a model for AM noise in which the active components have 2 common random noise processes. We report AM noise levels of 140 dB below the PM levels at 1 Hz and typical noise floors of -169 dBc/Hz at 10 MHz. We have no evidence for AM to PM conversion in these sources. We report on noise residuals in mixers from three different companies. Typical 1 Hz intercepts vary from -110 to -127 dBc/Hz. Noise floors are typically ~ -173 dBc/Hz. Cross-correlation techniques can be used to lower this value to below -190 dBc/Hz.

I. Introduction

In this paper we report on the measurement of amplitude modulation (AM) and phase modulation (PM) noise in a variety of X-band double balanced mixers, amplifiers and DRO oscillators. There is little information of AM and PM noise in these devices, and this information is critical for the design of many microwave systems and measurements. These measurements are meant to be a beginning for characterizing microwave components and developing noise models. The measurement techniques outlined here permit us to measure the residual noise in virtually any present X-band device. Clearly there are thousands of models available on the market, some of which probably have lower noises than the models tested. Since only 1 or 2 samples were tested, we cannot be certain that these results are typical of these models nor should the inclusion of a model in this study imply that it is the best for this use. The models chosen for this study were those that happened to be available in the laboratory.

To measure the noise in these devices we generally used a two channel measurement system with cross-correlation or, more accurately cross-spectrum analysis to reduce the noise contribution of the measurement process [1-3]. On the basis of these data we then suggest general noise models for these devices that include both AM and PM noise. Specific samples often deviate from the models, but they still give us a better insight into the total noise process within the device. The noise model for X-band double balanced mixers is similar to that of rf mixers [4,5]. The noise models for amplifiers are, to our knowledge, the first to show that the PM and AM noise added by an amplifier originates from two common sources. Specifically, we show that for most of the amplifiers tested the 1/f noise, which has always been assumed to be only PM noise [6,7], also produces AM noise of approximately equal amplitude. Our models for oscillator noise are among the first to include both PM and AM noise. As in the amplifier case, we find that there is often a 1/f component of AM and PM noise of nearly equal amplitude. Close to the carrier the $1/f^3$ PM noise dominates as is generally accepted [2,4,8]. There has been speculation that PM to AM conversion with the oscillator would cause the AM noise close to the carrier to vary as $1/f^3$. We have not been able to observe this conversion in the oscillators tested. We present a model that suggests a relationship between the PM noise and AM noise in oscillators. The PM noise roughly follows Leeson's Model.

* Work of U.S. government. Not subject to U.S. copyright.

U.S. Government work not protected by U.S. copyright

II. PM and AM Noise in X-Band Amplifiers

This section discusses PM and AM noise measurements and characterization of commercial GaAs FET amplifiers designed for 10-11 GHz frequency range. All measurements were made with the amplifiers operating in the linear region.

A. PM Noise:

The two channel measurement system shown in Fig. 1 was used to measure the PM noise added by various amplifiers [1,3]. In this configuration the oscillator signal is split into two channels, and the amplifier under test is placed in one of the channels. Each channel is split again and fed into two phase noise detectors, composed of a phase shifter and a double-balanced mixer. The phase shifters are adjusted so that there is a 90^{o} phase difference between the input signals (the signals are in quadrature). At this point the output of each mixer is proportional to the phase fluctuations between the input signals to the mixer. The signal is then low pass filtered, amplified, and fed into a two-channel cross-correlation FFT signal analyzer. Calibration of mixer sensitivity and amplifier gain can be performed by adding a known level of wideband Gaussian noise and measuring the power spectral density in the FFT [9].

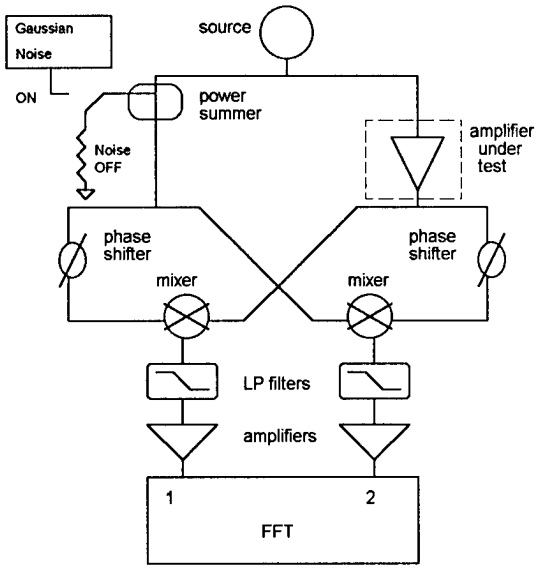


Figure 1. PM Noise Measurement System for Amplifiers

The calibrated gain of each channel ($k_dG(f)$) is obtained from the equation

$$PSD(V_n)_{noise\ ON} = PMCAL(k_dG(f))^2, \quad (1)$$

where PMCAL is the calibrated PM noise added to the system, k_d is the mixer sensitivity to phase fluctuations, and G(f) is the gain of the output amplifier. Three different spectral density functions are obtained from the FFT signal analyzer: the power spectral density of channel 1 ($PSD(V_{n1})$), the power spectral density of channel 2 ($PSD(V_{n2})$ and the power spectral density of the cross-spectrum of the two channels $PSD(V_{n1} X V_{n2})$ (where X denotes cross-spectrum). $PSD(V_{n1})$ includes the PM noise of the amplifier under test in addition to the noise added by the mixer and amplifier in channel 1. Similarly, $PSD(V_{n2})$ includes the PM noise of the amplifier under test in addition to the noise added by the mixer and amplifier in channel 2. $PSD(V_{n1} X V_{n2})$ includes only the coherent noise between the two channels, that is, the PM noise of the amplifier. The uncorrelated noise added by each channel is averaged away as $N^{-1/2}$, where N is the number of averages taken by the FFT. Eqs. (2), (3), and (4) summarize these relations:

$$PSD(V_{n1}) = (k_dG(f))^2 [PM_{amp} + System_1] \quad (2)$$

$$PSD(V_{n2}) = (k_dG(f))^2 [PM_{amp} + System_2] \quad (3)$$

$$PSD(V_{n1}XV_{n2}) = (k_dG(f))^2 PM_{amp}. \quad (4)$$

The noise floor of the system, given by $PSD(V_{n1} X V_{n2})$ when the amplifier under test is taken out of the system, is shown in Fig. 2. The figure also illustrates the PM noise of two X-band amplifiers. The PM noise of these amplifiers is at least 15 dB higher than the noise floor of the system throughout most of the frequency range (from 10 Hz to 10 MHz). The PM noise of various amplifiers is illustrated in Figs. 3 and 4. Amplifier 1, shown in Fig. 3, has a 1/f (flicker) PM noise component from approximately 1 Hz to 1 MHz from the carrier. Around 10 MHz from the carrier, the PM noise is white at -173 dBc/Hz. This white noise originates from the thermal PM noise of an amplifier [4,5] given by the expression

$$S_{\phi,thermal}(f) = \frac{2kTFG(f)}{P_o}, \quad (5)$$

where k is Boltzmann's constant, T the temperature in degrees kelvin, F the noise figure of the amplifier, G(f) the gain of the amplifier, and P_o the output power of the amplifier. For this amplifier the noise figure is ~3

dB, the gain 16 dB, and the output power 18 dBm; therefore,

$$S_{f,thermal} = -173.9 \text{ dBc / Hz} + 3 \text{ dB} + 16 \text{ dB} - 18 \text{ dB}$$

$$S_{f,thermal} = -172.9 \text{ dBc / Hz}.$$

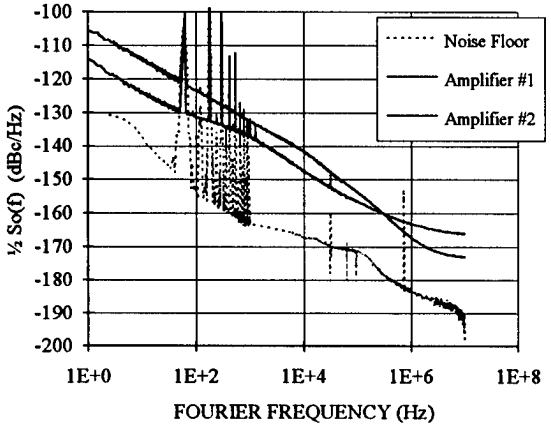


Figure 2. Noise floor of PM Measurement System.

This calculated value is almost equal to the thermal noise obtained from the PM curve in Fig. 3. Hence, amplifier 1 follows very closely Parker's PM noise model for amplifiers [4,5], in which the spectral density of phase fluctuations, $S_\phi(f)$, is given by

$$S_\phi(f) = \alpha_E \frac{1}{f} + \frac{2kTFG(f)}{P_o}. \qquad (6)$$

The first term is flicker PM noise, characterized by the flicker noise coefficient α_E, and the second term is one half the thermal (or Johnson) noise of the amplifier. α_E depends on both the gain element and the circuit configuration [6]. Fig. 4 shows the PM noise of amplifier 2. Although excess noise is observed at certain frequencies, throughout most of the frequency range the PM noise exhibits a 1/f dependence. At 10 MHz the PM noise reaches the white noise level. It is likely that the intrinsic noise of this amplifier (excluding the excess noise) follows the model presented by Eq. (6). The excess noise above the intrinsic flicker component may be caused by noise from the bias supplies in the amplifier.

B. AM Noise

The AM noise added by the amplifiers was measured using the two-channel measurement system schematically shown in Fig. 5 [1-3]. In this configuration, a carrier signal (split into two) is fed into two similar channels for AM noise detection. Each channel contains an AM noise detector, a low pass filter and an amplifier. The AM noise detector

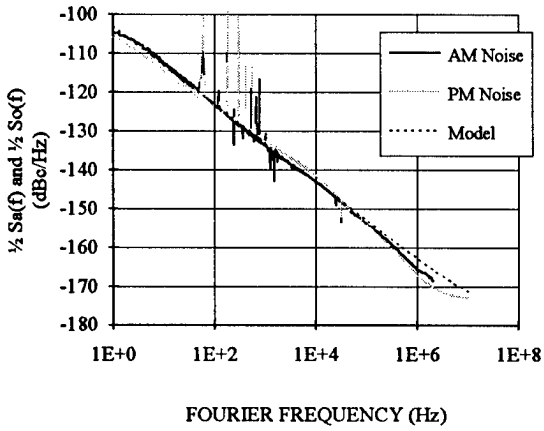


Figure 3. AM and PM Noise in Amplifier #1.

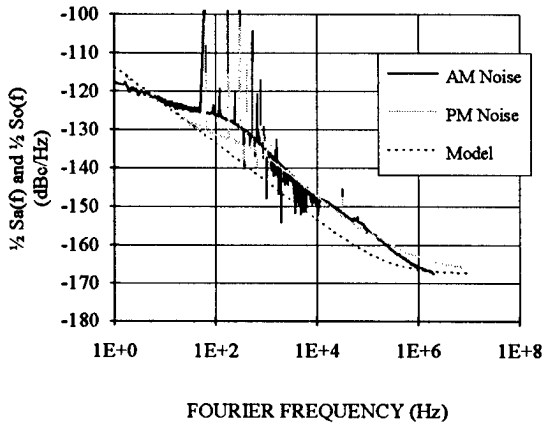


Figure 4. AM and PM Noise in Amplifier #2.

consists of a power splitter, a phase shifter and a double-balanced mixer. The phase shifter is set so that the input signals to the mixer are in phase ($\Delta\phi=2\pi n$). At this point the output voltage of the mixer is proportional to the amplitude fluctuations of the input signals. The PM noise suppression obtained in this system was very high, exceeding 140 dB at 1 Hz. The AM system includes the option of adding Gaussian noise to one of the channels, as discussed earlier in the PM noise section. Since added noise contributes equally to both AM and PM, the calibration method for the AM measurement system is similar to the calibration method for the PM measurement system discussed previously. The amplifier under test is added to one of the channels, ahead of the AM detector. The

PSD of this channel, measured by the FFT, includes AM noise from the source, AM noise added by the amplifier under test, and noise from the AM detector

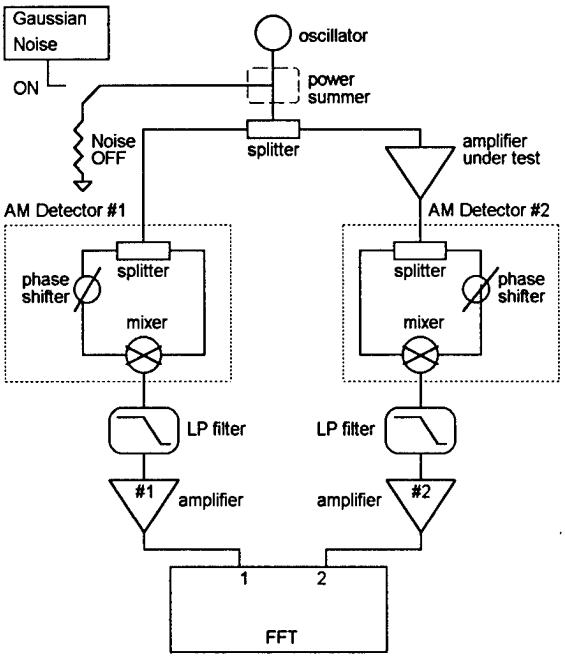


Figure 5. AM Noise Measurement System for Amplifiers.

and IF amplifier (Eq. (8)). The PSD of the other channel (when the Gaussian noise is OFF) includes AM noise from the source, and noise added by the AM detector and IF amplifier (Eq. (7)). The cross-spectrum of the two channels consists of the noise coherent or common to the two channels, that is, the source noise. The uncorrelated noise added by detectors and IF amplifiers is averaged away as $N^{-1/2}$, where N is the number of averages in the FFT signal analyzer. $PSD(V_{n1})$ - $PSD(V_{n2})$, the difference between spectral densities of the two channels, includes AM noise from the amplifier plus noise from channel 1 minus noise from channel 2 (Eq. (9)). Since the noise added by the channels is subtracted, $PSD(V_{n1})$ - $PSD(V_{n2})$ is a good measure of the AM noise added by the amplifier under test. Eqs. (7-10) summarize these relations.

$$PSD(V_{n1}) = (k_d G(f))^2 [AM_{Source} + System_1], \quad (7)$$

$$PSD(V_{n2}) = (k_d G(f))^2 [AM_{Source} + AM_{Amp} + System_2], \quad (8)$$

$$PSD(V_{n1}) - PSD(V_{n2}) = (k_d G(f))^2 [AM_{Amp} + System_1 - System_2], \quad (9)$$

$$PSD(V_{n1} \times V_{n2}) = (k_d G(f))^2 AM_{Source}. \quad (10)$$

The noise floor of the measurement system is obtained from $PSD(V_{n1})$ -$PSD(V_{n2})$ when the amplifier under test is not in the system. Results from this measurement are illustrated in Fig. 6. Also included in the figure are AM noise results of two sample amplifiers. The noise floor of the system is at least 15 dB lower than the AM noise of the sample amplifiers from 10 Hz to 10 MHz.

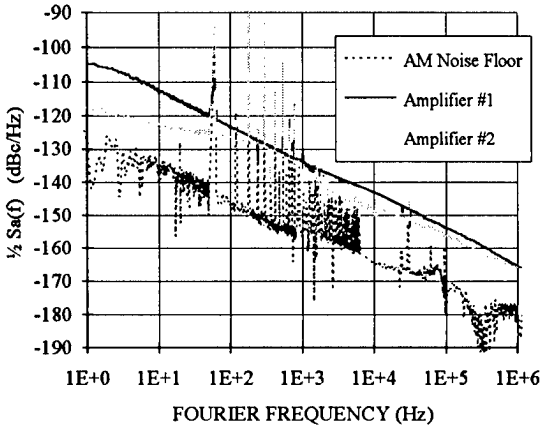


Figure 6. Noise Floor of AM Measurement System.

AM noise for amplifiers 1 and 2 is compared with their PM noise in Figs. 3 and 4. In both amplifiers the AM noise is comparable in value to the PM noise. In amplifier 1 the AM noise shows the same 1/f and white noise components as the PM noise. In amplifier 2, the AM noise closely follows the PM noise at most frequencies and shows the excess noise observed in the PM noise. These results suggest that the PM and AM noise of an amplifier originate from two common sources added to the signal, a 1/f modulation of the gain element and thermal noise. The 1/f component scales proportionally with the signal and the thermal component is independent of the signal level. Thus, the AM noise of a linear amplifier is the same as the PM noise given in Eq. (6); that is,

$$S_a(f) = S_\phi(f) = \alpha_E \frac{1}{f} + \frac{2kTFG(f)}{P_o}. \quad (7)$$

III. PM and AM noise in DRO Oscillators

In this section we apply the model for PM and AM noise presented in the previous section to the case where the amplifier is used as a feedback element in an oscillator. We compare the model to measurement data on commercial DRO sources operating at 10.6 GHz.

A. AM and PM Model for DRO Oscillators:

The noise characteristics of dielectric resonator oscillators (DRO) are necessarily dependent on the noise characteristics of the components that make up the oscillator. As a minimum an oscillator consists of a passive resonator and an active feedback element or amplifier. Some oscillators also contain a .buffer or power amplifying stage. In the following discussion we will consider the case of an oscillator without the buffer or amplifying stage.

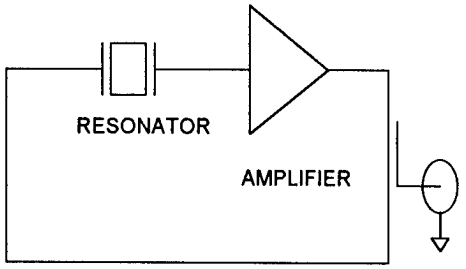


Figure 7. Oscillator Model

The open loop noise characteristics of resonators are not well understood, and a discussion of the subject is beyond the scope of this paper. However, measurements indicate that both the frequency and phase fluctuations across a resonator follow a 1/f power law. The frequency and phase fluctuations are given by Eq. (8) where α_R characterizes the resonator noise and Q_L is the resonator loaded quality factor:

$$S'_{\phi}(f) = (4Q_L)^2 S_y(f) = \left[\frac{\alpha_R}{f}\right]. \qquad (8)$$

The open loop noise characteristics of an amplifier as described in Section II B, suggest that two random processes produce equal amounts of AM and PM noise when the device is in a linear operating range. Since an amplifier is an integral part of an oscillator we expect these processes to be generating AM and PM noise in the oscillator. The open loop expression for the AM and PM noise in the amplifier is given by Eq. (7).

When the loop is closed, the conditions for oscillation are that the phase around the loop be an integer of 2π and that the gain be equal to the losses around the loop (G=1). Leeson derived the equation for the phase noise of the oscillator when the restriction on the phase is imposed [8]. The closed loop PM noise is given by

$$S_{\phi}(f) = \left(\frac{\upsilon_o}{2Q_L}\right)^2 \frac{1}{f^2}\left[\frac{\alpha_{A1}}{f} + \frac{2kTF_1G_1}{P_1}\right] + \left[\frac{\alpha_{A1}}{f} + \frac{2kTF_1G_1}{P_1}\right] + \left(\frac{\upsilon_0}{2Q_L}\right)^2 \frac{1}{f^2}\left[\frac{\alpha_R}{f}\right], \qquad (9)$$

where ν_0 is the carrier frequency, Q_L the loaded quality factor of the resonator and f is the offset frequency from the carrier. P is the power output of the amplifier and G is the gain of the amplifier. Inside the bandwidth of the resonator, the frequency of oscillation must vary so that the phase changes around the loop are an integral multiple of 2π [8]. Therefore the flicker phase noise of the amplifier and resonator determine the flicker frequency noise of the oscillator. Likewise the thermal noise floor of the amplifier determines the white frequency noise of the oscillator and its thermal noise floor as well.

The AM noise of the oscillator is generated by the amplifier. There is no evidence for the generation of AM inside the oscillating loop or an AM component generated by the resonator. Therefore the AM noise in the oscillator is given by Eq. (10) if the amplifier is operating linearly. In most DROs this is not the case. The open loop gain of the amplifier is usually much greater than one. This excess gain is designed into the oscillator to ensure oscillation over temperature variations or imperfections in the gain elements. The steady state gain of one is achieved by limiting in the amplifier after a particular power level is reached or by an active form of gain control. When limiting occurrs we can approximate the AM noise at the output by Eq. (11).

$$S_A(f) = \left[\frac{\alpha_E}{f} + \frac{2kTF_1G_1}{P_1}\right] \qquad (10)$$

$$S_A(f) = \frac{1}{C}\left[\frac{\alpha_E}{f} + \frac{2kTF_1G_1}{P_1}\right], \qquad (11)$$

where C is the gain compression in the amplifier. Some investigators have reported a degradation of the PM noise performance in amplifiers that are operating non linearly [4]. In this case Eq. (9) does not hold. We also expect a degradation of the PM noise in oscillators with feedback amplifiers operating non-linearly. Since both the AM and PM noise components of the amplifier come from the same source, the AM and PM noise of the oscillator is characterized by α_E, the flicker noise parameter and the thermal or Johnson noise. The noise in the dielectric resonator is probably smaller than the noise contributed by the amplifier in the DRO [10]. We can conclude that the AM and PM noise of the oscillator is solely determined by the noise processes in the amplifier and the loaded Q of the resonator. Resonators with high Q and high loop power will have low flicker frequency noise and low thermal noise floors.

B. Noise Measurements on DRO Oscillators:

We used two different methods to measure the PM noise in the DRO sources. For Fourier frequencies of 1 Hz to 100 kHz, we used the cavity discriminator method (for a detailed explanation see [11]). For Fourier frequencies of 10 kHz to 100 MHz we used the two oscillator measurement method. For the two oscillator measurements we used a very low noise reference at 10.6 GHz and the NIST automated phase noise measurement system [11]. The noise floor of the system is -172 dBc/Hz at 20 MHz for an input power of +16 dBm. The AM noise measurements were done using the two channel cross-correlation system shown in Fig. 5 without the amplifier. Dual balanced mixers were used as AM detectors by setting the phase of the input signals to zero degrees [4]. The noise floor of the AM measurement system was -190 dBc/Hz at 10 MHz for an input power of +16 dBm. We measured the AM and PM noise in 5 DRO's that were readily available in the laboratory. The operating frequency of these oscillators is 10.6 GHz and the output power varies between +13 dBm and +24 dBm. For modeling, we used a loaded Q of 1000 as specified by the manufacturer for one brand of DROs. The AM and PM noise measured for DRO 1 is illustrated in Fig. (8). We can indirectly estimate α_E for the amplifier and the thermal noise parameter of the oscillator from the AM or PM noise measurement. The dashed lines indicate the AM and PM noise predicted by the model if we use the AM noise data to estimate α_E and the thermal noise floor. The measured noise data from oscillator 1 does not match our model very well. Notice that the PM noise floor is about 6 dB higher than the AM noise floor. It is likely that the amplifier is operating under 6 dB of compression. If we assume 6 dB compression and recalculate our model we obtain a very good fit (Fig. 9). If we use the PM noise data we can get an estimate of

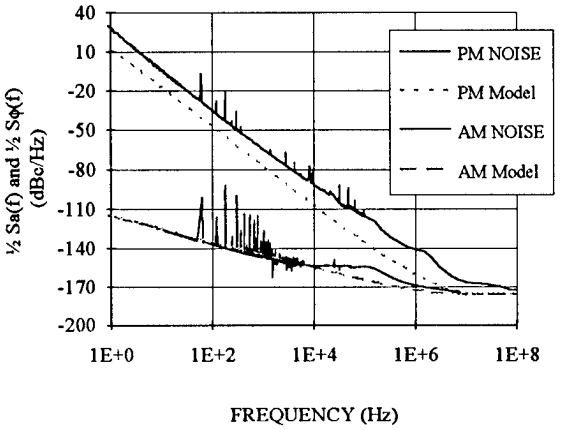


Figure 8. AM and PM Noise in DRO 1.

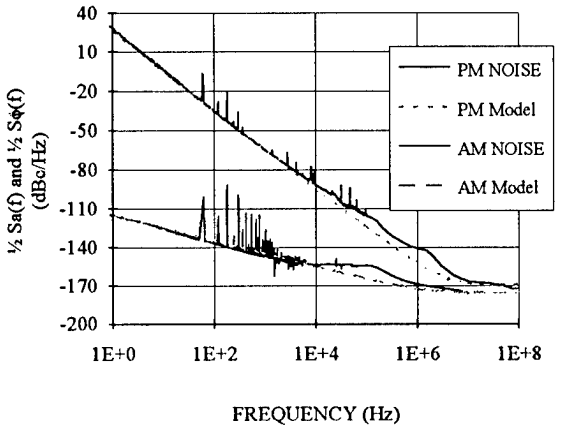


Figure 9. AM and PM Noise in DRO 1.

α_E and the thermal noise. However we can predict the AM noise performance of the oscillator only for the linear case unless the magnitude of compression occurring in the amplifier is known. There is an additional noise source that the model does not account for. Both the AM and PM noise of the DRO 1 (Fig. 9) deviate from the model at Fourier frequencies

of 10 kHz to 1 MHz. We believe that this noise is caused by excess noise in the FET bias circuitry. It contributes to both the AM and PM noise in the source.

In DRO 2, Fig. (10) we see the same noise characteristics seen in the previous source. The extraneous noise affects both the AM and PM noise, and the amplifier is self limiting or compressing. The difference between the AM thermal noise floor and the PM noise floor indicate at least 8 dB of compression in the amplifier. Adjusting the model for this amount of compression results in a good fit of the model to the data Fig. (10).

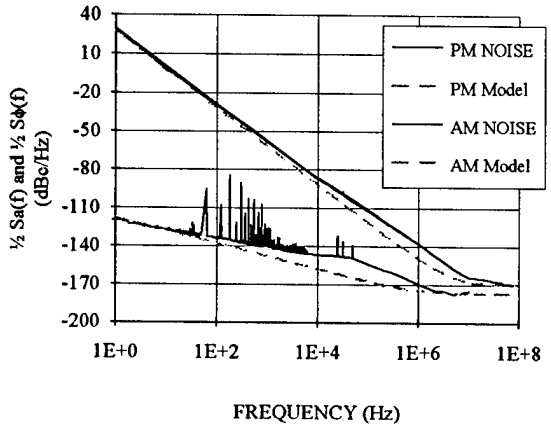


Figure 10. AM and PM Noise in DRO 2.

The same processes that generate AM noise in oscillators also determine the characteristics of the PM noise. We did not observe any PM to AM conversion in any of the oscillators even though the PM noise at 1 Hz from the carrier was over 140 dB higher than the AM noise. To reach a better understanding of these noise processes, we need to start characterizing both the AM and PM noise in these devices. In the past, the AM noise has been considered negligible; while this is true in oscillators at frequencies close to the carrier, the AM noise characteristics are still an important piece of the puzzle. Other noise processes can degrade the AM and PM noise characteristics in oscillators and amplifiers. Minimizing these noise processes can lead to better oscillators.

IV. Residual Noise in Mixers

In this section we discuss the residual noise in mixers. Mixers are a basic building block for many systems. The characterization of noise residuals in mixers is important, especially in measurement system applications where the residual noise in mixers can be the noise floor of the system.

Both PM and AM noise residual measurements were done on two pairs of mixers. In general the residual noises in mixers have two components. There is a 1/f component at low Fourier frequencies and a thermal noise floor. The levels of these noise processes are dependent on a variety of environmental factors, the drive powers, vswr, and the isolation in the mixer [12]. We were not able to draw any conclusive model but we can show some of the limiting factors of the noise performance. The PM residual measurements were done using a DRO at 10.6 GHz with known AM and PM characteristics. The signal was split in a reactive power splitter and one signal was phase shifted to obtain a 90^0 phase difference between the two mixer inputs. The input powers on the LO and RF ports were + 13 dBm. The same hardware setup was used in the AM residual measurements except that the phase between the mixer input ports was set at a multiple of 2 π. The powers at the mixer inputs were +10 dBm. Dual channel measurements were done on two pairs of mixers to try to separate the mixer residuals from other system noise. A comparison of the PM residuals in four different mixers is shown in Fig. 11. The 1/f component of the residual varies as much as 8 dB in these four different mixers. We think that part of this 1/f noise is due to AM to PM conversion occurring in the mixer. The AM noise levels of our source follow a 1/f power law with a 1 Hz intercept at -115 dBc/Hz. The noise floor of our IF amplifier is about -173 dBc/Hz. As shown in the Figs. 11,12 the noise floor of the mixers could possibly be lower than -173 dBc/Hz, but we cannot measure it with the present scheme. A two channel measurement system with real time cross-spectrum analysis capability is useful in separating

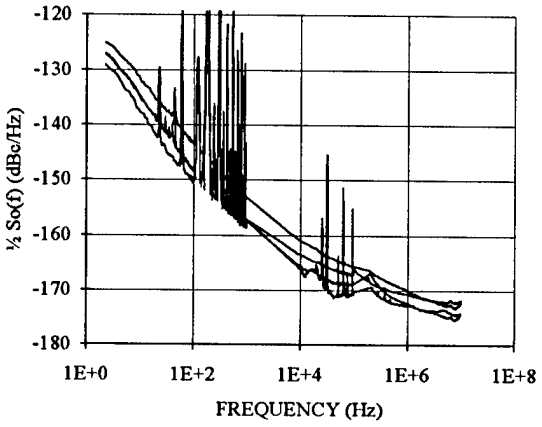


Figure 11. PM Noise Residuals in Various Mixers

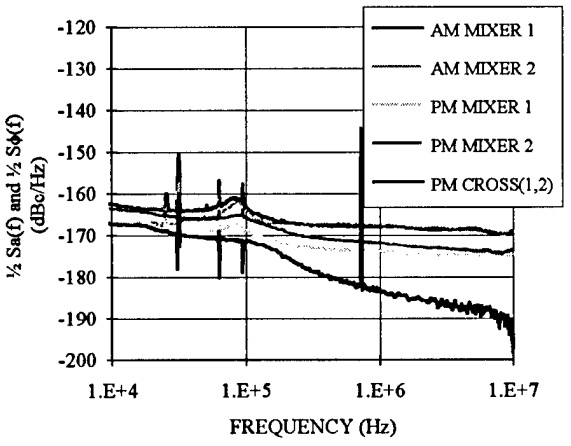


Figure 12. AM and PM Residual Noise in Mixer 12-18B

system noise from the residual noise measurement. The power spectral density of the cross-spectrum will contain only the coherent noise terms. The uncorrelated noise such as the noise generated in the mixers and the IF amplifiers will average down as $N^{-1/2}$, where N is the number of averages. Fig. 12 shows the PSD of the two mixers and the PSD of the cross-spectrum. The number of averages was 10 000, so the expected uncorrelated noise suppression is 20 dB. The cross-spectrum measurement indicates that for most of the frequency range the residual noise is not coherent in the two channels and probably is due to noise contributions by the mixers and the IF amplifiers. The AM residuals in the mixer can be obtained from two channel measurements. The power spectral density of channel 1 (Eq. (7)) minus the power spectral density of the cross-spectrum (Eq. (10)) will result in the residual noise in the mixer and the noise in the amplifier Fig. 12. There are limitations to how well you can determine residual noise from mixers. The limitations for PM single channel measurements are the noise floor of the IF amplifier following the mixer and the level of the AM noise in the source. Typically we have found the suppression of AM in mixers to be about 15-22 dB. The AM levels in the source do not average away as $N^{-1/2}$ in two channel cross-spectrum measurements since it is coherent between the two channels. The AM residual noise in a mixer cannot be measured without using a two channel measurement system because there is no way to sort out the noise contributions of the IF amplifier or the AM noise of the source. For both system calibration and actual AM or PM measurements it may be critical to measure both the AM and PM noise in the sources to be able to accurately interpret the measurement data.

V. Conclusion

There is a thermal noise source and a flicker noise source that generate equal levels of AM and PM noise in a linear amplifier. The thermal noise has been well characterized but the flicker noise source is still not understood. In summary, the PSD of the output voltage of an amplifier can be expressed by the following equation, where $\delta(\nu_0)$ is the carrier.

$$S_V(f) = \frac{V_o^2}{2}\left[\delta(\nu_o) + \frac{2\alpha_E}{f} + \frac{4kTFG(f)}{P_0}\right] \quad (8)$$

The thermal noise contribution of the amplifier, $4kTFG(f)/P_0$, contributes equally to both AM and PM noise. The wideband AM and PM noise in amplifiers are typically dependent on the signal level because they originate from the thermal noise and therefore scale as $1/V_0^2$. This noise model for amplifiers can be used to interpret the noise in oscillators. We propose a noise model for oscillators where the AM and PM noise due to the electronics is generated by the same noise processes, a thermal noise process and a flicker noise process. The PM noise in oscillators is also dependent on the noise and loaded Q of the resonator. The AM noise does not seem to be affected by the characteristics of the resonator and we did not observe any PM to AM conversion.

Acknowledgment

The authors thank Huascar D. Ascarrunz and Craig W. Nelson for their help in the measurements. They also thank the Calibration Coordination group of JCTG/GMT-JLC and the Space and Terrestrial Communications Evaluation Division, Tactical Support Branch, Ft. Monmouth, New Jersey, for their partial financial support of this project.

References

[1] F. L. Walls and S. R. Stein, "Accurate Measurements of Spectral Density of Phase Noise in Devices," Proc. 30th Ann. Frequency Control Symp., pp. 335-343,1977.

[2] A. L. Lance, W. D. Seal, and F. Labaar, "Phase Noise and AM Noise Measurements in the Frequency

Domain," Infrared and Millimeter Waves, Vol. 11, Chapter 7, pp. 284-286, 1984.

[3] W. F. Walls, "Cross-Correlation Phase Noise Measurements," Proc. 46th Ann. Frequency Control Symp., 1992, pp. 257-261.

[4] G. K. Montress, T. E. Parker and M. J. Loboda, "Residual Phase Noise Measurements of VHF, UHF, and Microwave Components," Proc. 43rd Ann. Frequency Control Symp., 1989.

[5] T. E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators," Proc. 41st Ann. Frequency Control Symp., 1987.

[6] S. G. Andresen and J. K. Nesheim, "Phase Noise of Various Frequency Doublers," IEEE Trans. Instrument Measurements, Vol. IM-22, pp. 185-188, June 1973.

[7] D. Halford, A. E. Wainwright and J. A. Barnes, "Flicker Noise of Phase in RF Amplifiers and Frequency Multipliers: Characterization, Cause and Cure," Proc. 22nd Ann. Frequency Control Symp., pp. 340-341, 1968.

[8] D. B. Leeson, , "A Simple Model of Feedback Oscillator Noise Spectrum," Proceedings of the IEEE, Vol.54, No. 2, pp.329-330, February 1966.

[9] F. L. Walls, "Secondary Standard for PM and AM Noise at 5, 10, and 100 MHz," Proc. 46th Ann. Frequency Control Symp., 1992.

[10] M. J. Loboda, T. E. Parker, and G. K. Montress, "Frequency Stability of L-Band, Two-Port Dielectric Resonator Oscillators," IEEE Trans. on Microwave Theory and Techniques, Vol. MTI-35, No. 12, Dec. 1987.

[11] F. L. Walls, A. J. D. Clements, C. M. Felton, M. A. Lombardi and M. D. Vanek, "Extending the Range and Accuracy of Phase Noise Measurements," Proc. 42nd Ann. Frequency Control Symp., 1988.

[12] L. M. Nelson and F. L. Walls, "Environmental Effects in Mixers and Frequency Distribution Systems," Proc. 46th Ann. Frequency Control Symp., 1992.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CROSS CORRELATION ANALYSIS IMPROVES TIME DOMAIN MEASUREMENTS

A. Lepek* and F. L. Walls
Time and Frequency Division
National Institute of Standards & Technology
Boulder, CO 80303

Abstract

We introduce cross correlation-based variances ($\#\sigma^2$) to estimate the Allan variance, the modified Allan variance, and the time variance in the presence of the measurement system noise. These variances substantially lower the short-term measurement noise floor; however, they require significantly more data. $\#\sigma^2$ is also used as a tool to analyze and improve the measurement system noise. $\#\sigma^2$ bears a precise relation to the 3-cornered-hat σ. We have reduced the short-term noise floor for $\sigma_y(1$ s) by a factor of 100 relative to our initial hardware and software configuration.

Introduction

The resolution or noise floor in time-domain measurement systems is limited by added noise in the measurement equipment. In this paper we introduce several approaches to analyzing time-domain measurement data which use cross correlation to reveal data obscured by the measurement noise. Looking for estimation of the Allan variance in such an environment we use $\#\sigma_y(\tau)$ (cross-sigma), $\#\mathrm{mod}\sigma_y(\tau)$ (cross- mod sigma), and $\#\sigma_x(\tau)$ which have close correspondence to the traditional variances $\sigma_y(\tau)$, $\mathrm{mod}\sigma_y(\tau)$, and $\sigma_x(\tau)$ [1-3].

The principle is to measure the time differences between two clocks by means of two independent simultaneous measurements and then to cross-correlate and average them to reduce the contributions of the measurement noise. This approach yields substantial improvements in the short-term noise floor of the system. However, it requires significantly more data. In this paper we show how this analysis is applied to the NIST Extended Dual Mixer Time Difference Measurement System [4,5] (EDM system) resulting in a substantial improvement in the estimates of $\sigma_y(\tau)$, $\mathrm{mod}\sigma_y(\tau)$, and $\sigma_x(\tau)$. Following similar guidelines, other measurement configurations such as the common view GPS time transfer may benefit from this method as well.

We also address measures of correlation, extensions to 3-cornered-hat measurements, the use of multiple clocks, and estimation of uncertainties. By studying the nature of these parameters as a function of τ we were able to identify noise components that corresponded to physical processes within the measurement system. Once identified, we reduced their effect on the noise floor, resulting in a better measurement system.

Definitions and Notations

The Allan variance is defined as

$$\sigma^2_y(\tau) = \langle\{X(t)-2X(t+\tau)+X(t+2\tau)\}^2\rangle/(2\tau^2) = \langle\delta^2\rangle = \langle\delta\cdot\delta\rangle \quad (1)$$

where $\langle\ \rangle$ denotes ensemble average or sample average, X(t) is a time series of time differences between two clocks spaces τ seconds apart, and δ is the normalized second difference. If τ is an integral number of τ_0, that is, $n\tau_0$ an overlapping average may be used by shifting the $\delta\cdot\delta$ terms by τ_0.

If we have two series a and b of time differences measured at the same instants, we obtain two sigmas which are $\sigma^2_y(t)_a = \langle\delta_a\cdot\delta_a\rangle$ and $\sigma^2_y(t)_b = \langle\delta_b\cdot\delta_b\rangle$. We would like to think of these time series as originated from the same two clocks i and j but from two measurement systems A and B. For two such time series we define $\#\sigma_y(\tau)$ (read "cross-sigma") as

$$\#\sigma^2_y(\tau)_{ab} = \langle\delta_a\cdot\delta_b\rangle, \quad (2)$$

where ab indicates the origin of the input series a and b of $\#\sigma_y(\tau)$, and

Contribution of the U.S. Government, not subject to copyright.
**Israel National Physical Laboratory, Hebrew University Campus, Jerusalem, Israel 91904*

U.S. Government work not protected by U.S. copyright

$$\#\sigma_y(\tau)_{ab} = \text{sign}(<\delta_a \cdot \delta_b>) \left| <\delta_a \cdot \delta_b> \right|^{1/2}, \quad (3)$$

so the sign of $<\delta_a \cdot \delta_b>$, which may be negative, is preserved. A change in the sign (as a function of τ) in $<\delta_a \cdot \delta_b>$ indicates a change in the causes for the correlation between the δ_a and δ_b time series. $\#\sigma^2(\tau)$ bears a strong relation to the traditional covariance. A similar sigma was investigated by Groslambert et al. [6].

Similarly, the modified Allan variance is defined as

$$\text{mod}\sigma^2_y(\tau) = <\{\underline{X(t)-2X(t+\tau)+X(t+2\tau)}\}^2>/(2\tau^2) = <\mu \cdot \mu>, \quad (4)$$

where the underline denotes an average over the interval τ, $<\ >$ denotes as average over the whole measurement time series, and μ is a normalized and averaged second difference. For the two time series, a and b, we define $\#\text{mod}\sigma_y(\tau)$ (read cross-mod-sigma) by

$$\#\text{mod}\sigma^2_y(\tau)_{ab} = <\mu_a \cdot \mu_b>, \text{ with} \quad (5)$$

$$\#\text{mod}\sigma_y(\tau)_{ab} = \text{sign}(<\mu_a \cdot \mu_b>) \left| <\mu_a \cdot \mu_b> \right|^{1/2}, \quad (6)$$

following same reasoning as for $\#\sigma$.

In the same fashion, we extend the definition of $\sigma_x(\tau)$ commonly called T-VAR or time variance to $\#\sigma_x(\tau)$, (read cross-T-VAR). That is, $\#\sigma^2_x(\tau)_{ab} = <\epsilon_a \cdot \epsilon_b>$, where [3]

$$\epsilon = \tau/3^{1/2}. \quad (7)$$

To simplify notation and to cover all types of σ and $\#\sigma$, we use d_a for δ_a, μ_a and ϵ_a. Then, for example, we have

$$\#\sigma^2_y(\tau)_{ab} = <d_a \cdot d_b>$$

and (8)

$$\text{mod}\sigma^2_y(\tau)_a = <d_a \cdot d_a> = <d^2_a>$$

So far, $\#\sigma$ was obtained by crossing d_a and d_b, where each such second difference was obtained within one time series. It is possible to cross elements differently as will be clear from the following definition. We denote this type of $\#\sigma$ by $\#_a\sigma^2_y(\tau)$ (a denotes alternate for alternate crossing).

$$\#_a\sigma^2_y(\tau) = \quad (9)$$

$$<\{X_a(t+2\tau)-X_b(t+\tau)-X_a(t+\tau)+X_b(t)\}\{X_b(t+2\tau)-X_a(t+\tau)-X_b(t+\tau)+X_a(t)\}>/2\tau^2$$

$$= <\{X_a(t+2\tau)-X_a(t+\tau)-X_b(t+\tau)+X_b(t)\}\{X_b(t+2\tau)-X_b(t+\tau)-X_a(t+\tau)+X_a(t)\}>/2\tau^2 . \quad (10)$$

Here we crossed X between the two series in the first line and y (the fractional frequency $[X(t+\tau)-X(t)]/\tau$) in the second line. The equality shows that it does not matter whether we cross X or y; both methods result in the same $\#_a\sigma$. Experimentally we find that in most cases $\#\sigma$ is lower than $\#_a\sigma$ and the differences are small; therefore we will not discuss these alternatives any further.

Measurement Setup

The measurement system is the Extended Dual Mixer Time Difference Measuring System [4,5] used by NIST and some other laboratories for many years. This system is called the EDM system. Figure 1 is a block diagram of the system. Channel 1 is used as the system master. An offset frequency reference, generated from channel 1, is mixed with all other input channels of same nominal frequency. (For these tests the input frequency was 5 MHz and the beat frequency was 10 Hz). The time differences between channel 1 and all other channels are obtained virtually simultaneously by means of time interval counters, one per channel. The counters are triggered by channel 1, which provides the start pulse. The time differences between any two clocks (channels) are then computed from the comparisons with channel 1.

To do the $\#\sigma$ analysis, we split each clock into two channels. This is done with a power splitter or a distribution amplifier. Figure 2 is a block diagram of the information flow in such a setup. The measurement blocks in this figure include both the hardware and the software functions. Thus for each two clocks we may obtain four time series. To be specific we choose two time series for a $\#\sigma$ analysis. We denote the channels due to clock i as c_{ia} and c_{ib}, and the channels due to clock j as c_{ja} and c_{jb}. Further, let us denote the time series (of second differences) d_a as originating from comparing c_{ia} with c_{ja} and d_b from comparing c_{ib} and c_{jb}. We may calculate two additional time series d_c and d_d originating from the other two channel combinations.

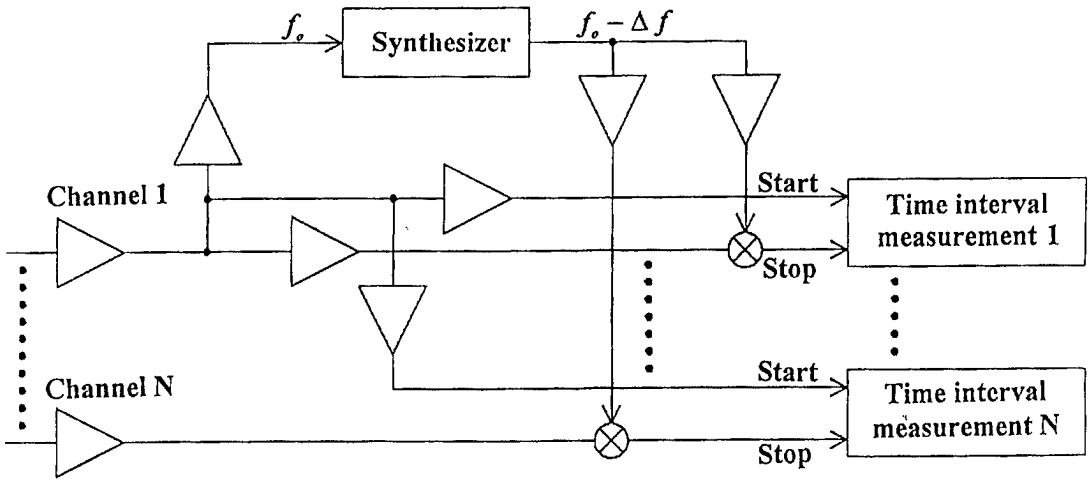


Figure 1. Block diagram of the extended dual mixer time difference measurement system.

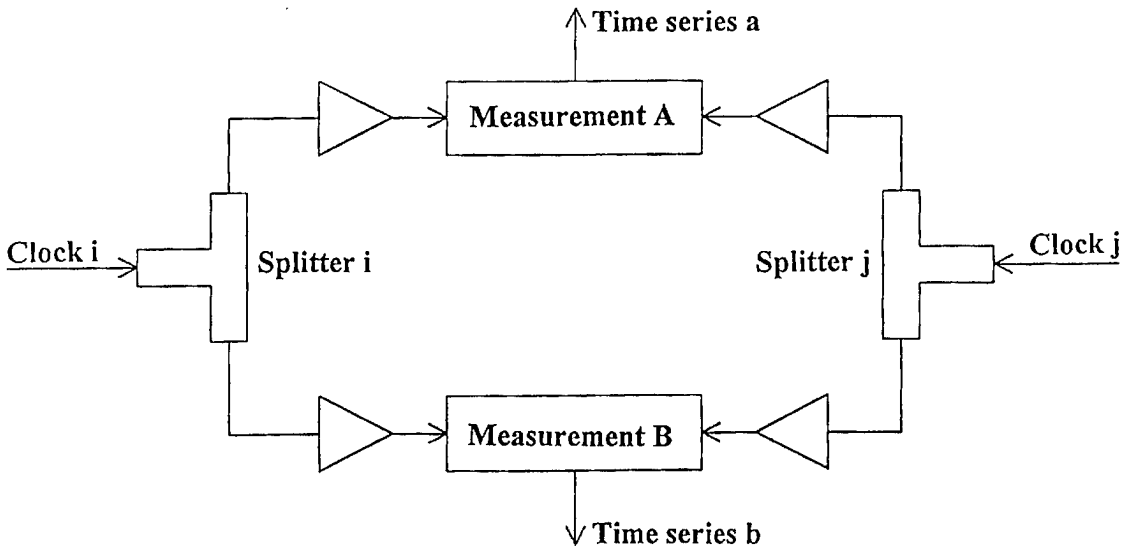


Figure 2. Information flow in the #σ setup.

If we use all four channel combinations, we may use an extended $\#\sigma$ denoted as $\#_4\sigma$ defined as

$$\#_4\sigma^2(\tau) = < (d_a \cdot d_b \cdot d_c \cdot d_d)^{1/2} >. \qquad (11)$$

No change in the setup is required to use this equation since all four channels are already present once each clock is split. However, experimentally $\#_4\sigma$ has an advantage over $\#\sigma$ only in limited cases. The reason is that it is averaging the best and the worst channels, whereas when using $\#\sigma$ we can choose the best performing channels (having the least-correlated noise).

This setup may be extended to include more than two channels per clock. If the number n of channels per clock is even, we may extend the definition of $\#\sigma$ to

$$\#_n\sigma_y^2(\tau) = < (d_a \cdot d_b \cdot d_c \cdot d_d \ldots)^{2/n} > \qquad (12)$$

A still different extension is to compare one clock to two different reference clocks. As will be shown, the $\#\sigma^2$ of this combination is exactly equivalent to the 3-cornered-hat Allan variance. Figure 3 shows a setup for 3-cornered-hat where clock i is split into two channels, each of which is compared to two other clocks through two combinations.

General Noise Relationships

We may break down d_a and d_b so that $d_a = Q + A$ and $d_b = Q + B$, where Q is the time series due to the clocks only (assuming ideal measurement) and A or B are due to the added noise in the measuring systems. The task is to estimate the Allan variance $<Q^2>$ in the presence of the added noise.

We note the following relations valid for any τ:

$$\sigma^2_a + \sigma^2_b = <(Q+A)^2 + (Q+B)^2> = 2<Q^2> + 2<Q(A+B)> + <A^2> + <B^2>, \qquad (13)$$

$$2\#\sigma^2_{ab} = 2<(Q+A)(Q+B)> = 2<Q^2> + 2<Q(A+B)> + 2<AB>. \qquad (14)$$

Therefore:

(1) If the measurement noise is low enough compared to the clocks that $2<Q^2> \gg <A^2> + <B^2>$ (then $2Q^2 \gg 2<AB>$) and

$$(\sigma^2_a + \sigma^2_b)/2 \approx \#\sigma^2_{ab}, \qquad (15)$$

then $\#\sigma^2_{ab}$ is a good estimate of the average Allan variances.

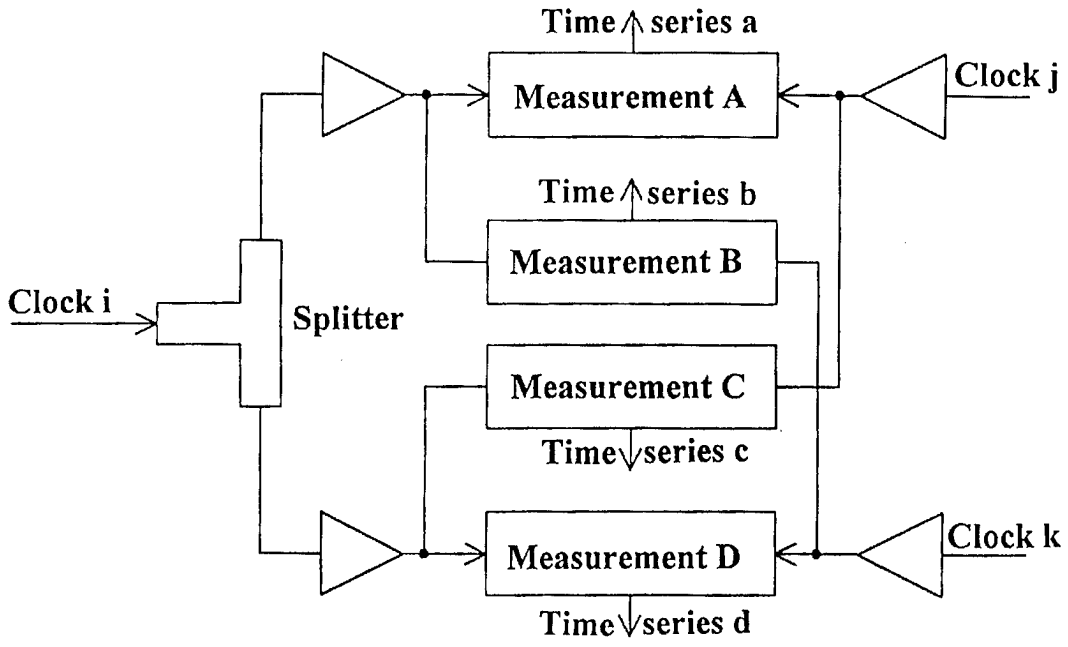


Fig. 3 Information flow in a $\#\sigma$ setup for 3-cornered-hat.

(2) Since $<A^2>+<B^2> \geq 2<AB>$ is always true, $\#\sigma^2_{ab}$ is never higher than $(\sigma^2_a+\sigma^2_b)/2$, $\#\sigma^2_{ab}$ is a better estimate of the Allan variance than the average $(\sigma^2_a+\sigma^2_b)/2$ and practically it is better than the best of these two sigmas. Following the same reasoning, the difference

$$D^2(\tau) = (\sigma^2_a+\sigma^2_b)/2 - |\#\sigma^2_{ab}| = <A^2+B^2>/2 - |<AB>| \quad (16)$$

is positive, is independent of Q, and characterizes the measurement system.

(3) To estimate the added noise by the system we connect the same clock to all inputs, then Q=0. Therefore Eqs. (8) and (9) become

$$\sigma^2_a+\sigma^2_b = <A^2> + <B^2>, \quad (17)$$
$$2\cdot\#\sigma^2_{ab} = 2<AB>. \quad (18)$$

Therefore the lower is the correlation $<AB>$ between A and B, the higher is the ratio

$$(\sigma^2_a+\sigma^2_b) / |2\cdot\#\sigma^2_{ab}| = <A^2 + B^2> / 2|<AB>|. \quad (19)$$

This ratio is a measure of the improvement in the noise floor due to the use of $\#\sigma$ instead of σ.

(4) We define the function $R(\tau)$ as

$$R(\tau) = \#\sigma^2_{ab}/(\sigma_a\cdot\sigma_b). \quad (20)$$

When measuring the noise floor using the above method Q=0, and we obtain

$$R(\tau) = <A\cdot B>/\{<A^2>\cdot<B^2>\}^{1/2}. \quad (21)$$

Equation (20) shows that $R(\tau)$ is between 0 and 1. The definition of $R(\tau)$ is similar to the definition of the Pearson correlation coefficient r [7]. In statistics r^2 is a measure of the correlated portion of $\sigma^2(x1)$ of the variable x1 due to the linear regression of a second variable x2. Also, $(1-r^2)$ is a measure of the uncorrelated portion. Similarly we use $R^2(\tau)$ as a measure of the correlation between different channels in the measurement system. We use R^2 as a measure of the improvement due to the use of $\#\sigma^2$ and as a figure of merit reflecting the amount of correlated noise in the system.

(5) Since $\#\sigma^2_{ab}=<d_a\cdot d_b>$, $\sigma^2_a=<d_a\cdot d_a>$ and $\sigma^2_b=<d_b\cdot d_b>$, the uncertainty in $\#\sigma$ (due to its spread) is expected to be of the order of the uncertainties in σ_a and σ_b. When using overlapping averages, a simple way to deal with uncertainties, especially when the noise is not white is to divide the time series interval (total number of points) into N equal intervals (we use four) and for each of them compute $\#\sigma_{ab}$, σ_α, and σ_b. Then the standard deviations of $\#\sigma_{ab}$, σ_a and σ_b are computed from these n values. They are then divided by $\sqrt{n}$ to obtain an estimate of the uncertainties $\#Er_{ab}$, Er_a and Er_b. This method avoids any assumptions and the need to know the type of noise when estimating the uncertainty of σ_y. The method used to compute the partial $\#\sigma_{ab}$, σ_a and σ_b (overlapping intervals) is the same as for the complete series. An additional advantage of dividing into n segments is it reveals how quickly $\#\sigma$ converges to a limit with the number of points. The uncertainty in $\#\sigma$ as a function of n can be used to detect this convergence.

(6) Application to 3-cornered-hat measurements. Comparing clock i to two other different clocks j and k results in two measured time series, a=ij and b=ik. A third time series c=jk may be computed from these two. We use here the ij notation to indicate the origin of the time series. As before,

$$\sigma^2_{ij} = <d_{ij}\cdot d_{ij}> \quad (22)$$

the 3 corner hat σ_i is

$$\sigma^2_i = < d_{ij}\cdot d_{ij} + d_{ik}\cdot d_{ik} - d_{jk}\cdot d_{jk} >/2, \quad (23)$$

where d_{jk} was calculated from the other series from

$$d_{jk} = d_{ik} - d_{ij}, \quad (24)$$

as is done in the EDM system. Substituting d_{jk} yields

$$\sigma^2_i = < d_{ij}\cdot d_{ij} + d_{ik}\cdot d_{ik} - (d_{ik} - d_{ij})\cdot(d_{ik} - d_{ij}) >/2$$
$$= < d_{ij}\cdot d_{ik} + d_{ik}\cdot d_{ij} >/2 = <d_{ij}\cdot d_{ik}> = \#\sigma^2_{ij_ik} \quad (25)$$

where ij_ik indicate that the crossing is between ij and ik. This means that to get the 3-cornered-hat-sigma we do not need to compute all three sigmas; we can use two time series and compute the cross-sigma.

Note that

$$\sigma^2_j = \#\sigma^2_{ji_jk} = -\#\sigma^2_{ij_jk}, \quad (26)$$

because

$$\langle d_{ji} \cdot d_{jk} \rangle = - \langle d_{ij} \cdot d_{jk} \rangle .$$

Usually, the 3-cornered-hat equation of σ_i is based on the assumption that the three sigmas are independent. This is not the case in the EDM system since the jk time series is calculated from the other two. This relationship occurs because all clock comparisons are derived by their comparison to clock 1.

Going one step further, we can derive a 3-cornered-hat for the $\#\sigma$. Let us assume that each one of the three clocks i, j, and k is split into two channels. We can choose now two time series, for example d_{i1_j1} and d_{i2_j2}, and compute $\#\sigma$ for them:

$$\#\sigma^2_{(i1_j1 \cdot i2_j2)} = \langle d_{i1_j1} \cdot d_{i2_j2} \rangle . \quad (27)$$

Following Eq. (23) the 3-cornered-hat $\#\sigma^2_i$ for clock i is defined as

$$\#\sigma^2_i = \langle d_{i1_j1} \cdot d_{i2_j2} + d_{i1_k1} \cdot d_{i2_k2} - d_{j1_k1} \cdot d_{j2_k2} \rangle /2. \quad (28)$$

As before we obtain the jk series from the two other by

$$d_{j1_k1} = d_{i1_k1} - d_{i1_j2}, \quad (29)$$
$$d_{j2_k2} = d_{i2_k2} - d_{i2_j2}. \quad (30)$$

Substituting d_{j1_k1} and d_{j2_k2} yields

$$\begin{aligned} \#\sigma^2_i &= \langle d_{i1_j1} \cdot d_{i2_j2} + d_{i1_k1} \cdot d_{i2_k2} - (d_{i1_k1} - d_{i1_j2}) \cdot (d_{i2_k2} - d_{i2_j2}) \rangle /2, \\ &= \langle d_{i1_j1} \cdot d_{i2_k2} + d_{i1_k1} \cdot d_{i2_j2} \rangle /2, \\ &= (\#\sigma^2_{i1_j1 \cdot i2_k2} + \#\sigma^2_{i2_j2 \cdot i1_k1})/2. \end{aligned} \quad (31)$$

$\#\sigma^2_i$ is the average of the two $\#\sigma$'s and is symmetrical in 1 and 2.

Our previous discussions show that similar equations are obtained for $\#\mathrm{mod}\sigma$ and $\#\sigma_x$.

Results

Analysis of the EDM system noise reveals several noise sources. These are counter quantization noise, instabilities in synthesizer mixer noise, digital level changes, distribution amplifiers, environmental effects (temperature and vibration), and spurious 60 Hz modulation. Some of these sources can be seen directly in the time series and some, such as digital levels, in statistical histograms of the time series. By analyzing $\#\sigma$, which is sensitive to correlated (or common) noise, we can trace the pollution of these types of noises in the system. It was impossible to use σ as a guide because of it's much smaller sensitivity. Figure 4 shows an initial $\#\sigma$ plot of the system noise floor. The input frequency was 5 MHz and the offset frequency is 10 Hz for Figures 4-8. In this figure we used the setup of Figure 2, where a and b result from two chosen time series and all inputs originate from the same hydrogen maser. The relatively small differences between the $\#\sigma$ and σ show that the correlated noise is high and probably accounts for most of the measurement noise. Using 10000 points for τ_0, the uncertainties in this and the following graphs are at the level of few percent. An effort was made to cancel the correlated noise to as low a level as possible. This included changes in the offset synthesizer to lower the PLL time constants, replacement of noisy components and especially the mixers, isolation and filtering of the power supplies, and the use of power splitters instead of distribution amplifiers. We have found also that dc isolation of the incoming signals reduces measurement noise. Figure 5 shows the present state of the system noise floor obtained by injecting the same clock to all channels. Figure 6 shows the $\#\mathrm{mod}\sigma$ analysis of the data of Figure 5. Note the sign change in $\#\sigma$ and $\#\mathrm{mod}\sigma$ at about 8 s in Figures 5 and 6; this indicates a change in the causes of the correlations. $|\#\sigma|$ may be very low at the regions of sign change due to the balance of these causes, resulting in effective uncorrelated noise. This change is supported by Figure 7 where $R^2(\tau)$ has two or three peaks in the correlated noise. The noise floor peaking around τ=30 s is probably due to the synthesizer's instability, as can be verified by changing components in the synthesizer. The improvement in noise floor is nearly a factor of 100 over the initial performance and seen in Figures 4-6.

Figure 8 shows the 3-cornered-hat $\#\sigma$ obtained for an active hydrogen maser splitting it (as clock i in Figure 3) and crossing each of the two channels with two other hydrogen masers. We note the influence of the synthesizer's instability peaking around τ=30 s. Comparing with Figure 4 shows that this measurement was impossible before the system improvements.

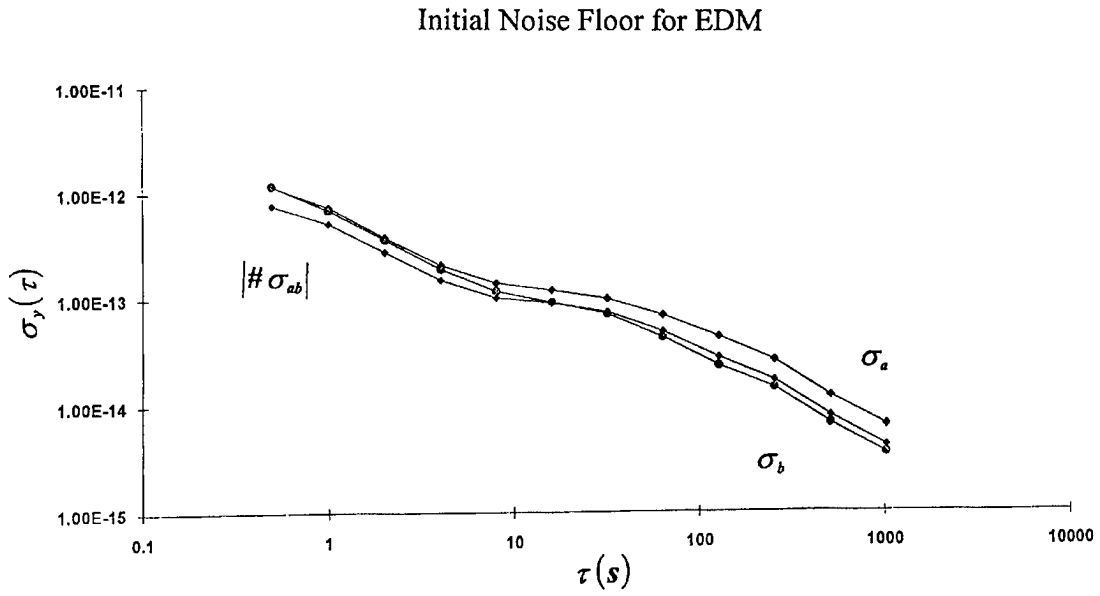


Figure 4. Initial noise floor for the NIST Extended Dual Mixer System (EDM) versus measurement time τ.

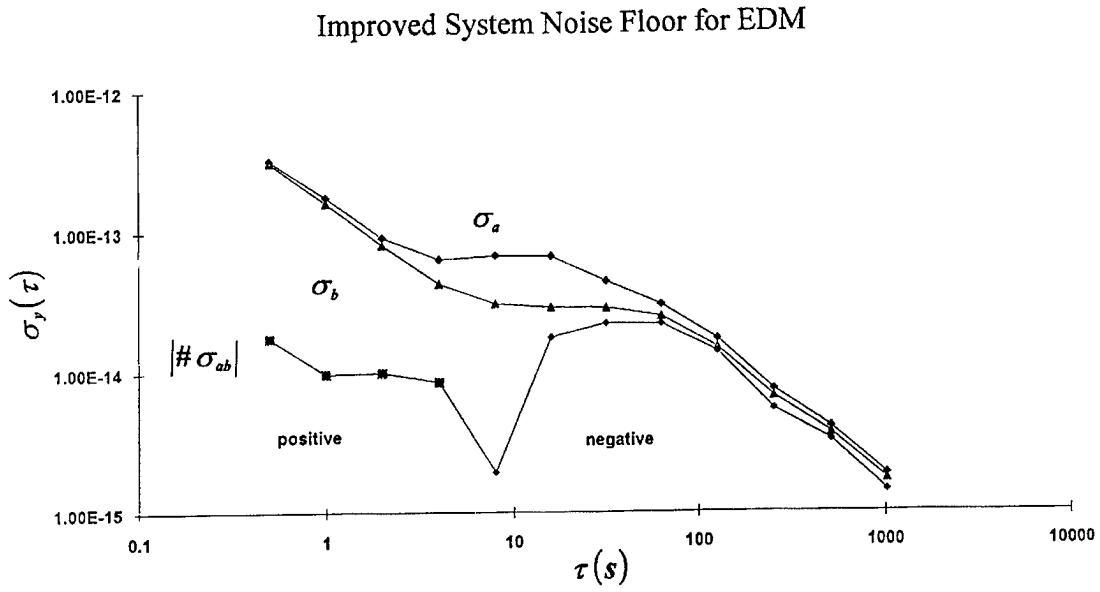


Figure 5. System noise floor for $\#\sigma_y(\tau)$ and $\sigma_y(\tau)$ after improving the measurement hardware.

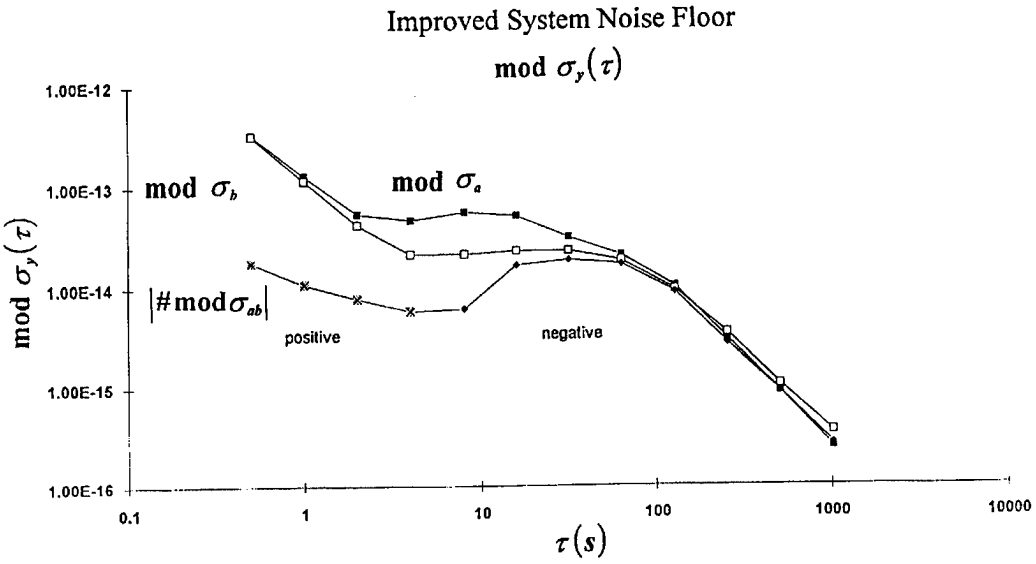


Figure 6. System noise floor for #mod$\sigma_y(\tau)$ and mod$\sigma_y(\tau)$ after improving the measurement hardware.

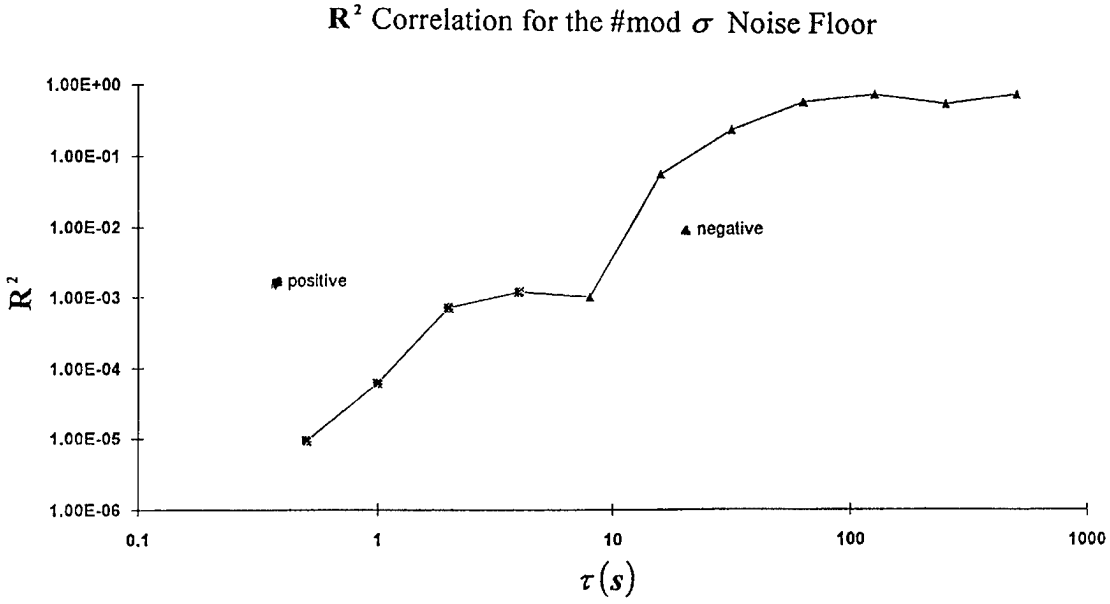


Figure 7. R^2 for #modσ versus measurement time τ. The increase in R^2 at longer τ indicated increased correlated noise. The dips in R^2 coincide with the change in the sign of #modσ indicating a change in the correlation mechanism.

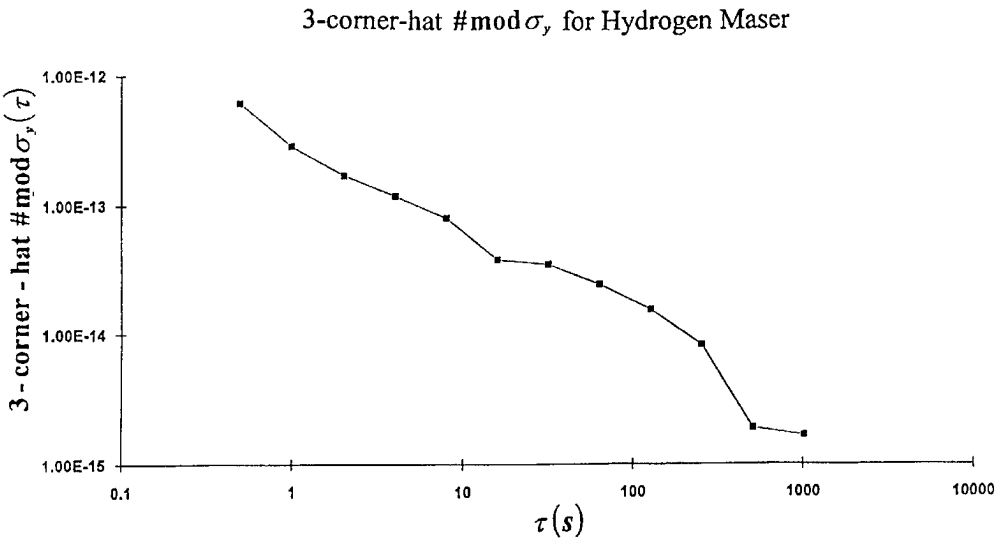


Figure 8. Estimates of #mod$\sigma_y(\tau)$ at 5 MHz reference frequencies for an active hydrogen maser versus measurement time τ.

Conclusions

We have shown the usefulness of using $\#\sigma$ and R as tools to lower the noise floor resulting so far in an overall improvement of about 100. These improvements enable low noise measurements that were not possible earlier at 5 MHz. We have shown as well that the $\#\sigma$ is a good estimate of the Allan and modified Allan variances and also some precise relationships concerning the 3-cornered-hat. Additional work should be done to further eliminate the common noise injected from the various sources to lower $\#\sigma$. Once the tools are established they can be used to improve the present system to its limits and to design a new and better system. Other systems having separate measurement channels, such as the common view GPS, may benefit from these methods as well.

Acknowledgments

We gratefully acknowledge helpful discussion with David W. Allan, David A. Howe, and Marc A. Weiss.

References

[1] James A. Barnes, et al., "Characterization of Frequency Stability, IEEE Transactions on Instrumentation and Measurement," IM-20, 105-120, 1971.

[2] David W. Allan and James A. Barnes, "A Modified "Allan Variance" with Increased Oscillator Characterization Ability," Proc. 35th Annual Frequency Control Symposium, 470-475, 1981.

[3] D. W. Allan, M. A. Weiss and J. L. Jespersen, "A frequency-domain view of time domain characterization of clocks and time and frequency distribution systems," Proc. Annual Symposium on Frequency Control, 667-678, 1991.

[4] S. Stein, D. Glaze, J. Levine, J. Gray, D. Hilliard, D. Howe, "Performance of an Automated High Accuracy Phase Measurement System," Proc. 36th Annual Frequency Control Symposium, 314-320, 1982.

[5] D. J. Glaze and S. R. Stein, "Picosecond time difference measurements utilizing CA-MAC based ANSI/IEEE -488 data acquisition hardware," NBS Tech Note 1056.

[6] J. Groslambert, D. Fest, M. Olivier, and J. J. Gagnepain, "Characterization of Frequency Fluctuations by Crosscorrelations and by Using Three or More Oscillators," Proc. 35th Annual Frequency Control Symposium, 458-463, 1981.

[7] J. F. Ratcliffe, Elements of Mathematical Statistics, Oxford University Press, 173-176, 1962.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

FLICKER NOISE PROCESS ANALYSIS

WEI Guo

Shaanxi Astronomical Observatory, Academia Sinica
P. O. Box 18 Lintong, Shaanxi 710600, P. R. China

Abstract

Since flicker noise is a non-stationary process whose spectrum is divergent at low frequency ($f \to 0$), its autocorrelation function cannot be derived from Wiener-Khintchine relation. By means of the method so-called *Order-Reducing and Integrating*, the analytic expression of autocorrelation function of flicker FM noise is obtained in the paper. According to the result, the general principle of characterization of frequency stability is discussed. It is pointed out that the essence of the characterization is transferring a non-stationary process into stationary and averaging it in time. Finally, Allan variance of flicker FM noise is deduced from the autocorrelation function.

Introduction

Flicker noise is one kind of processes whose spectrum fulfil f^{-1}. It is often observed in precise frequency sources,

$$S_y(f) = h_{-1} f^{-1} \qquad 0 < f \le f_h \qquad (1)$$

As a frequency parameter, above equation could be regarded as the property of flicker noise in frequency domain. However, because the knowledge of properties in frequency domain are not sufficient enough while a process being analyzed, the properties in time domain are often required. This is not a difficulty for a stationary process, since the gap between frequency domain and time domain could be bridged by Wiener-Khintchine relation which is a couple of Fourier transform between spectrum and autocorrelation function,

$$S(f) = \int_{-\infty}^{\infty} R(\tau) \exp(-2\pi f \tau)\, d\tau \qquad (2)$$

$$R(\tau) = \int_{-\infty}^{\infty} S(f) \exp(2\pi f \tau)\, df \qquad (3)$$

So, if the spectrum is known, the autocorrelation function could be deduced with Eq.(3).

But Wiener-Khintchine relation is not always valid, except that some conditions are satisfied. From the point of view of statistics, the process should be stationary; and as a condition of Fourier transform, $S_y(f)$ in Eq.(3) should be absolutely integrable. Obviously, $S_y(f)$ presented in Eq.(1) will not meet this condition, which will be divergent at $f=0$ even though the integral strip is limited by high cut-off frequency f_h. This is the difficulty to analyze the autocorrelation function of flicker noise process.

A method so called *Order-Reducing and Integrating* was presented in Ref.[2] to overcome this difficulty. First, $y(t)$ is differentiated to deduce its differential $z(t)$; $z(t)$ is a transitive variable which is a stationary process. Then, $S_z(f)$,

0-7803-0905-7/93 $3.00 © 1993 IEEE

the spectrum of $z(t)$, is derived from $R_z(t_1,t_2)$, the autocorrelation function of $z(t)$, based upon Wiener-Khintchine relation. Finally, $R_z(t_1,t_2)$ is integrated to deduce $R_y(t_1,t_2)$, the autocorrelation function of $y(t)$. This method is introduced briefly in the paper and then the analytic expression of $R_y(t_1,t_2)$ is presented. From the point of view of autocorrelation function, it is discussed about the general principle of characterization of frequency stability, and Allan variance is deduced.

$z(t)$ and its Autocorrelation Function

The transitive variable, $z(t)$, is defined as

$$y(t)=\int_0^t z(t')\,dt' \tag{4}$$

$$z(t)=\frac{dy(t)}{dt} \tag{5}$$

It could be proved that the spectrum of $z(t)$ is[3]

$$\begin{aligned} S_z(f) &= (2\pi)^2 f^2 S_y(f) \\ &= (2\pi)^2 h_{-1} f \qquad 0\le f\le f_h \end{aligned} \tag{6}$$

Obviously, $S_z(f)$ is absolutely integrable in the field of definition. $z(t)$ is obtained by taking Eq.(6) into Eq.(3),

$$\begin{aligned} R_z(\tau) &= \frac{1}{2}\int_{-f_h}^{f_h} S_z(f)\exp(2\pi f\tau)\,df \\ &= h_{-1}\left(\frac{\cos 2\pi f_h\tau}{\tau^2}+\frac{2\pi f_h \sin 2\pi f_h\tau}{\tau}-\frac{1}{\tau^2}\right) \end{aligned} \tag{7}$$

Autocorrelation Function of $y(t)$

Because $y(t)$ is a non-stationary process, its autocorrelation function is not only related to the time interval τ but also related to time evolution t. The general expression of autocorrelation function is,

$$\begin{aligned} R_y(t_1,t_2) &= Ey(t_1)y(t_2) \\ &= E\int_0^{t_1} z(t')\,dt' \int_0^{t_2} z(t'')\,dt'' \\ &= \int_0^{t_2} dt'' \int_0^{t_1} Ez(t')z(t'')\,dt' \end{aligned} \tag{8}$$

where $Ez(t'')z(t')$ is the general expression of autocorrelation of $z(t)$,

$$\begin{aligned} Ez(t')z(t'') &= R_z(t''-t') \\ &= h_{-1}\left[\frac{\cos 2\pi f_h(t''-t')}{(t''-t')^2} \right. \\ &\left. +\frac{2\pi f_h \sin 2\pi f_h(t''-t')}{t''-t'}-\frac{1}{(t''-t')^2}\right] \end{aligned} \tag{9}$$

Take Eq.(9) into Eq.(8), then

$$\begin{aligned} &Ey(t_1)y(t_2) \\ &= 2h_{-1}\{C(\pi f_h t_2)+C(\pi f_h t_1) \\ &\quad -C[\pi f_h(t_2-t_1)]\} \end{aligned} \tag{10}$$

where the function $C(u)$ is defined as

$$C(u)=\int_0^u \frac{\sin^2 x}{x}\,dx \tag{11}$$

When $u\gg 1$, $C(u)$ is approximately written as

$$\begin{aligned} C(u) &= \int_0^u \frac{\sin^2 x}{x}\,dx \\ &\approx \frac{4}{3}+\frac{1}{2}\ln 2u \qquad u\gg 1 \end{aligned} \tag{12}$$

Take Eq.(12) into Eq.(10), then

$$Ey(t_1)y(t_2)=h_{-1}\left(\frac{3}{2}+\ln\frac{2\pi f_h t_2 t_1}{|t_2-t_1|}\right) \tag{13}$$

Therefore, the autocorrelation function of flicker FM noise is derived. It could be seen from Eq.(13) that the autocorrelation function is not only related to time interval (t_2-t_1) but also related to time evolution of t_1 and t_2. When $t_1=t_2$, $Ey^2(t)$ is the variance of $y(t)$, assuming $Ey(t)=0$,

$$E\bar{y}_{k+1}^2 = h_{-1}\left(1+\ln\frac{2\pi f_h}{\tau}\right) + \frac{2h_{-1}}{\tau}\left[(t_k+2\tau)\ln(t_k+2\tau) - (t_k+\tau)\ln(t_k+\tau)\right] \quad (25)$$

$$E\bar{y}_{k+1}\bar{y}_k = h_{-1}\left(1+\ln\frac{2\pi f_h}{\tau}-2\ln 2\right) + \frac{h_{-1}}{\tau}\left[(t_k+2\tau)\ln(t_k+2\tau) - t_k\ln t_k\right] \quad (26)$$

Take Eq.(24)-(26) into Eq.(22), we obtain

$$R_\nabla(0) = h_{-1}4\cdot\ln 2 \quad (27)$$

Similarly, $R_\nabla(1)$ and $R_\nabla(j)(j>1)$ could be derived as

$$R_\nabla(1) = h_{-1}\left(\frac{9}{2}\ln 3 - 8\ln 2 - 2\ln\tau\right) \quad (28)$$

$$R_\nabla(j) = \frac{1}{2}h_{-1}\left[(j-2)^2\ln(j-2) - 4(j-1)\ln(j-1) + 6j^2\ln j - 4(j+1)^2\ln(j+1) + (j+2)^2\ln(j+2)\right] \quad (29)$$

Since $R_\nabla(j)=E\nabla\bar{y}_k\nabla\bar{y}_{k+j}$ is merely related to time interval j but not related to time evolution k, $\nabla\bar{y}_k$ is generally stationary. Also, it is easy to see from Eq.(29) that,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}R_\nabla(j) = 0 \quad (30)$$

So $\nabla\bar{y}_k$ is not only stationary but also ergodic.

It is an important nature of stationary and ergodic processes that the time average and the ensemble average are identical, that is,

$$\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M}(\nabla\bar{y}_k)^2 = E(\nabla\bar{y}_k)^2 \quad (31)$$

According to this nature, by Eq.(20), (22) and (27), then

$$\tau_y^2(\tau) = \frac{1}{2}\lim_{M\to\infty}\sum_{k=1}^{M}(\nabla\bar{y}_k)^2 = \frac{1}{2}E(\nabla\bar{y}_k)^2 = \frac{1}{2}R_\nabla(0) = h_{-1}\cdot 2\ln 2 \quad (32)$$

Therefore, we obtained the Allan variance of flicker FM noise from the autocorrelation function in the time domain which is exactly same as the result from frequency domain of Eq.(19).

As a comparison, we will discuss the standard variance of flicker FM noise. The statistic quantity of standard variance is average frequency itself, that is

$$\tau^2(N) = \frac{1}{N-1}\sum_{k=1}^{N}\bar{y}_k^2 \quad (33)$$

It is concluded by Eq.(24) that $\bar{y}_k$ is non-stationary since $E\bar{y}^2{}_k$ is related to time k. $\sigma^2(N)$ will not be convergent with increasing N. So, for any non-stationary processes like flicker FM noise, the standard variances are always divergent and could not be used to characterize them. In the method of Allan variance, the non-stationary process is transferred into a stationary and ergodic process which is averaged in time, and then, the uncertainty is measured.

Conclusion

Since flicker noise is a kind of non-stationary processes, it is difficult to derive its autocorrelation function by conventional methods. By the method of *Order-Reducing and Integrating*, the difficulty that Wiener-Khintchine relation is not valid for non-stationary processes is skirted, and the autocorrelation function of flicker FM noise is obtained. According to Eq.(13) and (14), the

$$
\begin{aligned}
Ey^2(t) &= 4h_{-1}C(\pi f_h t) \\
&\approx 4h_{-1}\left(\frac{3}{4}+\frac{1}{2}\ln 2\pi f_h t\right) \\
&= h_{-1}(3+2\ln 2\pi f_h t)
\end{aligned} \tag{14}
$$

The variance of $y(t)$ is also the function of time t; therefore, $y(t)$ is a typical non-stationary process.

Allan Variance Analysis

1) Allan Variance from Frequency Domain

The definition of Allan variance is

$$
\langle \sigma_y^2(N,T,\tau)\rangle = \left\langle \frac{1}{N-1}\sum_{n=1}^{N}\left(\bar{y}_n - \frac{1}{N}\sum_{k=1}^{N}\bar{y}_k\right)^2\right\rangle \tag{15}
$$

where <> denotes infinite time average, and $\bar{y}_k$ is average frequency in the interval τ,

$$
\bar{y}_k = \frac{1}{\tau}\int_{t_k}^{t_k+\tau} y(t)\,dt \tag{16}
$$

The relationship between Allan variance and the spectrum $S_y(f)$ is presented by a special filter $H(f)$,

$$
\langle \sigma_y^2(N,T,\tau)\rangle = \frac{N}{N-1}\int_0^{\infty} S_y(f)\,|H(f)|^2 df \tag{17}
$$

where

$$
|H(f)|^2 = \frac{\sin^2\pi f\tau}{(\pi f\tau)^2}\left(1-\frac{\sin^2\pi fTN}{N^2\sin^2\pi fT}\right) \tag{18}
$$

When $T=\tau$ and $N=2$, Eq.(15) becomes so-called two samples variance. For flicker FM noise, it could be deduced that[1],

$$
\sigma_y^2(\tau) = h_{-1}\cdot 2\ln 2 \tag{19}
$$

The essence of Allan variance is that the non-stationary processes are transferred into stationary by a filter and then are averaged in time. However, this procedure does not appear so clearly in above derivation. In the following discussion, we will obtain a clearer result.

2) Allan Variance from Autocorrelation Function

Two samples variance is the variance of the difference of average frequency, $\nabla\bar{y}_k=(\bar{y}_{k+1}-\bar{y}_k)$,

$$
\begin{aligned}
\sigma_y^2(\tau) &= \frac{1}{2}\langle(\bar{y}_{k+1}-\bar{y}_k)^2\rangle \\
&= \frac{1}{2}\lim_{M\to\infty}\frac{1}{M}\sum_{k=1}^{M}(\nabla\bar{y}_k)^2
\end{aligned} \tag{20}
$$

It is determined by the ergode of process $\nabla\bar{y}_k$ whether above time average has any practical meaning. The autocorrelation function $Ey(t_1)y(t_2)$ makes it possible to analyze the ergode of process $\nabla\bar{y}_k$.

Let $R_\nabla(j)$ denote the autocorrelation function of $\nabla\bar{y}_k$,

$$
R_\nabla(j) = E\nabla\bar{y}_k\nabla\bar{y}_{k+j} \tag{21}
$$

when $j=0$,

$$
\begin{aligned}
R_\nabla(0) &= E(\nabla\bar{y}_k)^2 \\
&= E(\bar{y}_{k+1}-\bar{y}_k)^2 \\
&= E\bar{y}_{k+1}^2 - 2E\bar{y}_{k+1}\bar{y}_k + E\bar{y}_k^2
\end{aligned} \tag{22}
$$

According to Eq.(16)

$$
\begin{aligned}
E\bar{y}_k^2 &= E\frac{1}{\tau}\int_{t_k}^{t_k+\tau} y(t)\,dt\cdot\frac{1}{\tau}\int_{t_k}^{t_k+\tau} y(t')\,dt' \\
&= \frac{1}{\tau^2}\int_{t_k}^{t_k+\tau}\int Ey(t)\,y(t')\,dt'dt
\end{aligned} \tag{23}
$$

where $Ey(t)y(t')$ is just the autocorrelation function of $y(t)$. Take Eq.(13) into Eq.(23), then

$$
\begin{aligned}
E\bar{y}_k^2 &= \frac{h_{-1}}{\tau^2}\int_{t_k}^{t_k+\tau}\int\left(\frac{3}{2}+\ln\frac{2\pi f_h t t'}{|t-t'|}\right)dtdt' \\
&= h_{-1}\left(1+\ln\frac{2\pi f_h}{\tau}\right) \\
&\quad + \frac{2h_{-1}}{\tau}\left[(t_k+\tau)\ln(t_k+\tau) - t_k\ln t_k\right]
\end{aligned} \tag{24}
$$

autocorrelation function and the standard variance are all related to time, so it becomes clear that ensemble average and time average could not be exchanged and it is impossible to measure these two quantities by the method of time average. It is answered why a standard variance could not be used to characterize flicker noise. By the autocorrelation function presented in this paper, the ergode of average frequency difference, $(\bar{y}_{k+1}-\bar{y}_k)$, was discussed and Allan variance was derived, which is exactly same as the result from frequency domain. This result not only proved the correctness of autocorrelation function but also, more important, uncovered the principle of Allan variance.

The method presented in this paper is suitable for all non-stationary processes in precise frequency sources, although only flicker noise was discussed. The results of other noise processes are presented in Ref.[6].

In time domain analysis, the role of autocorrelation function is irreplaceable by the spectrum. Besides of the characterization of noises discussed the paper, the autocorrelation function could be used in all analyses related to second moment operation, e.g., atomic time scale algorithm, parameter estimation and error analysis.

Reference

[1] J. A. Barnes, et al., "Characterization of Frequency Stability", IEEE Trans. on Instru. Measu., Vol.20, No.2, 1971

[2] WEI Guo, "An Expression of Flicker Autocorrelation", Publications of Shaanxi Astronomical Observatory, Vol.11, No.1-2, 1988

[3] P. Kartaschoff, Frequency and Time, Aca. Press Inc., London 1978

[4] D. W. Allan, "Statistics of Atomic Frequency Standards", Proc. IEEE, Vol.54, Feb. 1966

[5] WEI Guo, Thesis of Ph. D, 1991

[6] WEI Guo, "Characterization of Frequency Stability: Analysis in Time Domain", Science in China, Vol.36, No.3, 1993

[7] WEI Guo, "A Study of Atomic Time Scale Stability", Proc. 46th Annual Frequency Control Symposium, 1992

[8] WEI Guo, "The Limit of Frequency Estimation", Proc. 24th Annual PTTI Applications and Planning Meeting, 1992

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SIMULTANEOUS MEASUREMENT OF DRIFTS AND NOISE COEFFICIENTS OF OSCILLATORS : APPLICATION TO THE ANALYSIS OF THE TIME STABILITY OF THE MILLISECOND PULSARS

F. Vernotte*, E. Lantz**, F. Meyer*, F. Naraghi*

*Observatoire de Besançon, 41 bis avenue de l'Observatoire, BP 1615, 25010 BESANÇON Cedex - FRANCE
**Laboratoire d'Optique P.M.Duffieux, UA CNRS 214, Faculté des Sciences et Techniques, 25030 BESANÇON Cedex

Abstract - **The results presented in this paper prove that the least square method is unable to separate a linear drift and a low frequency noise. Moreover, the use of a single variance cannot distinguish a linear frequency drift and a f^{-2} frequency noise, if it is calculated in the frequency domain, or a linear frequency drift and a f^{-3} frequency noise, if it is calculated in the time domain. Since f^{-3} frequency noise may be encountered in pulsar timing data, we cannot use this last solution. Hence, the multi-variance method using variances calculated in both time and frequency domains seems to be the best solution in this case. Moreover, the great separability of a multi-variance system using these two types of variances increases the interest of this method.**

1. INTRODUCTION

Oscillators are affected by both deterministic (phase and frequency drifts) and random (phase and frequency noises) effects. Generally, the measurements of these alterations are performed sequentially : firstly, the drifts are estimated by the least square method, and secondly, the noise coefficients are assessed from the residual data (after removing the estimated drifts) by variance measurements. However, this method introduces a bias in the case of low frequency noises (flicker frequency noise and especially random walk frequency noise) since part of the random long term fluctuations can be assigned to drifts.

In order to avoid this effect, it is necessary to simultaneously estimate the drifts and the noise coefficients. This can be performed by variance measurements. It will be shown in this paper that the Allan variance may be considered as a tool to directly measure drifts and noise coefficients. However, the accuracy and the dynamics of the variance measurements decrease greatly if there are several different noise types and drifts in the signal.

The measurement accuracy and the dynamics can be increased by applying the multi-variance method [1, 2] which uses several variances calculated for several integration time values τ over the signal to study. In order to obtain the greatest possible dynamics, the different variances must be chosen according to their different sensitivities to the noise coefficients and drifts. Moreover, it has been shown in a previous paper [2] that the multi-variance method lends itself to an estimation of the uncertainty of the noise coefficient measurements (and, in this case, of the drift measurement), taking into account the dispersion of the variance results.

In classical oscillators, the frequency noises range between f^{-2} to f^{+2} in the frequency domain. A linear frequency drift yields the same slope as a f^{-3} frequency noise with the Allan variance calculated in the time domain. However, for the pulsar signals (after all corrections) the idea was emitted [3] that the primordial gravitational radiation could have provided a background of gravitational waves. Since these waves modify the metric of the space time, the timing residuals of the pulsar would contain a noise from this origin. This noise behaves in f^{-3} [4] and several methods were used [5, 6] in an effort to extract it from the pulsar signal. The multi-variance method provides us a different approach which could allow us to study this very low frequency noise without extracting the drifts by the least squares.

This method can be used also for classical time stability measurement. The results obtained with the classical sequential method will be compared in this paper with the results obtained with this simultaneous multi-variance method for different types of signals.

2. DRIFT MEASUREMENT

2.1. Classical method

Usually, drifts are determined by the least square method and variances are applied to the residual data to measure the level of each noise (i. e. the h_α coefficients [7]). However, from the point of view of signal processing, it is always better to determine simultaneously the different signal parameters (deterministic and random). Moreover, it is well known that the least square method is optimum only for data affected by gaussian uncorrelated noise, i. e. white noise. In the case of low frequency noises (f^{-1} noise, f^{-2} noise,

0-7803-0905-7/93 $3.00 © 1993 IEEE

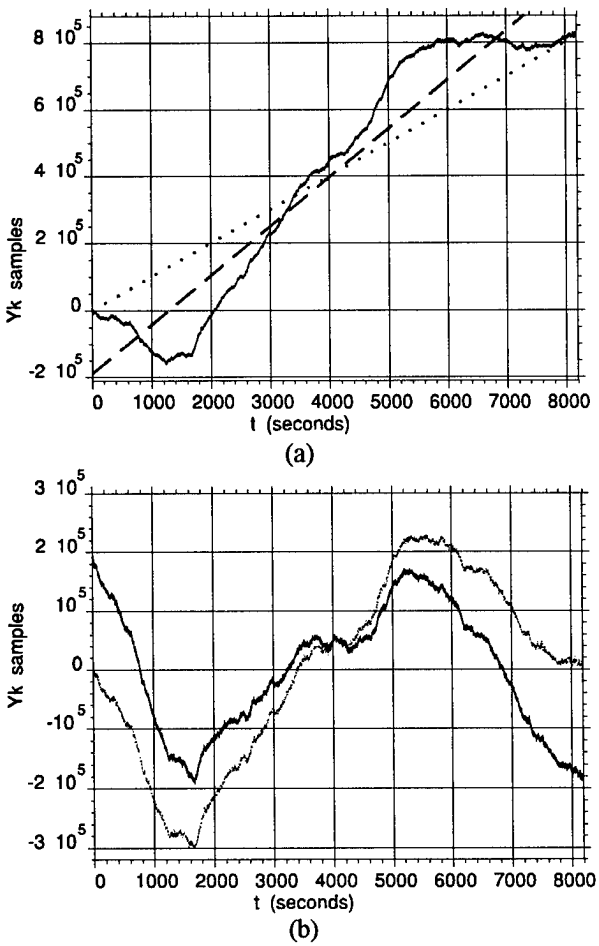


Figure 1: sequence of 8192 samples of low frequency noise with a linear drift. (a) The dotted line is the real drift, the disconnected line is the estimation by the least squares. (b) The dark curve is the residual data after removing the estimated drift, the light curve is the real low frequency noise.

...), which yield long term fluctuations, the least square method may give erroneous measurements. Figure 1.a shows a sequence of simulated data composed of a linear drift and low frequency noises. The dotted line is the real drift entered for the simulation and the disconnected line is the drift measured by the least square method. Here, the drift is overestimated by about 46%. The residual data are presented in figure 1.b (dark curve) with the real noise entered for the simulation (light curve). Thus, in this case, whereas the drift is overestimated, the noise is underestimated.

2.2. Drifts and variances

It is possible to calculate the response of a variance for a linear drift (linear frequency drift). This could be a solution to determine simultaneously drifts and noise coefficients. However, we must distinguish variances calculated in the time domain (from their measurement sequence) and variances calculated in the frequency domain (from their transfer function) since responses for each are different. We give below the example of the Allan variance.

2.2.1. Time domain

The Allan variance is defined in the time domain by its measurement sequence [8] as :

$$\sigma_y^2(\tau) = \frac{1}{2}\left\langle \left(\overline{y_{k+1}} - \overline{y_k} \right)^2 \right\rangle \quad (1)$$

For a pure frequency drift we have :

$$y(t) = C_1 t \quad (2)$$

and :

$$\overline{y_k} = C_1 t_k + C_1 \frac{\tau}{2} \quad (3)$$

$$\overline{y_{k+1}} = C_1 t_k + C_1 \frac{3\tau}{2} \quad (4)$$

then :

$$\sigma_y^2(\tau) = \frac{1}{2} C_1^2 \tau^2 \quad (5)$$

Thus, the Allan variance gives a τ^2 slope for a linear frequency drift. This response is different from all other noise responses of this variance (from f^2 to f^{+2}, see Table I). Unfortunately, in the case of pulsars, f^{-3} frequency noise (f^5 phase noise) could be present in the timing data. The response of the Allan variance for f^{-3} frequency noise is also a τ^2 slope (see Table I). Therefore, the Allan variance calculated in the time domain cannot be chosen for measuring both f^{-3} frequency noise level and linear frequency drift coefficient.

2.2.2. Frequency domain

A linear drift can be considered as the derivative of the sum of :

- a rectangular function whose width w is the width of the sampling window ;
- two Dirac peaks at each side of the rectangular function.

As shown in figure 2, the Fourier Transform of a linear frequency drift is :

$$D(f) = C_1 \left[\frac{\sin(\pi w f)}{2i\pi^2 w f^2} + \frac{\cos(\pi w f)}{2i\pi f} \right] \quad (6)$$

With a discrete Fourier Transform, we obtain $D(f_k)$ for frequencies $f = k/w$, where k is an integer. Hence :

$$D(f_k) = C_1 \frac{(-1)^k}{2i\pi f_k} \quad (7)$$

The spectral density is :

$$S_y(f_k) = C_1^2 \frac{1}{4\pi^2 f_k^2} \quad (8)$$

Allan variance	Linear drift	f^{-3} noise	f^{-2} noise	f^{-1} noise	f^{0} noise	f^{+1} noise	f^{+2} noise
Time domain	$\frac{1}{2} C_1^2 \tau^2$	$\pi^2 h_{-3} \tau^2 (-.2-2\ln\pi\tau f_b)$	$\frac{2\pi^2 h_{-2} \tau}{3}$	$2 \ln 2\, h_{-1}$	$\frac{h_0}{2\tau}$	$\frac{h_1[1.04+3\ln 2\pi\tau f_h]}{4\pi^2\tau^2}$	$\frac{3 h_2 f_h}{4\pi^2\tau^2}$
Frequency domain	$\frac{C_1^2 \tau}{6}$	$\pi^2 h_{-3} \tau^2 (-.2-2\ln\pi\tau f_b)$	$\frac{2\pi^2 h_{-2} \tau}{3}$	$2 \ln 2\, h_{-1}$	$\frac{h_0}{2\tau}$	$\frac{h_1[1.04+3\ln 2\pi\tau f_h]}{4\pi^2\tau^2}$	$\frac{3 h_2 f_h}{4\pi^2\tau^2}$

Table I : responses of the Allan variance calculated in the time domain (above) and in the frequency domain (below) for a linear frequency drift and for 6 types of noise (from f^{-3} to f^{+2} frequency noises)[7, 9].

This spectral density cannot be distinguished from that of f^2 frequency noise. The Allan variance has the same response as for this noise (see Table I) :

$$\sigma_y^2(\tau) = \frac{C_1^2 \tau}{6} \qquad (9)$$

In conclusion of this paragraph, using the Allan variance in only one domain does not allow us the separation of the drift from f^3 frequency noise (time domain) or f^2 frequency noise (frequency domain)

2.3. Using the multi-variance method

The same variance calculated in the time domain or in the frequency domain gives different responses for a linear drift. Hence, using these two variances over the same sequence is very efficient to determine simultaneously the drift coefficient and the noise coefficients. The use of several variances on the same sequence is the aim of the multi-variance method [1, 2]. Moreover, this method gives an estimation of the uncertainty for the determination of the noise coefficients and of the drift coefficient. In the following, the results presented were obtained over 8192 data sequences, with an 8 variances system :

- Allan variance (time domain)

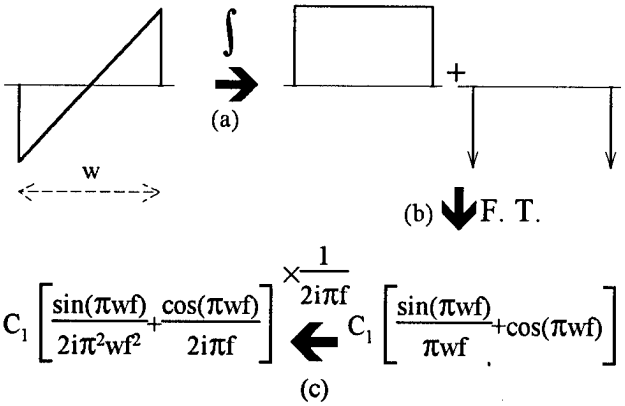


Figure 2 : the Fourier Transform of a drift is obtained by (a) integrating the drift, (b) performing the Fourier Transform of the result of this integration, (c) multiplying the result by $1/2i\pi f$.

- Allan variance (frequency domain)
- Picinbono variance (time domain) [10]
- Picinbono variance (frequency domain)
- Modified Allan variance (time domain) [11]
- filtered Picinbono variance (frequency domain) [9]
- high-pass variance [7]
- band-pass variance [7]

3. RESULTS AND DISCUSSION

Figure 3 shows the results obtained with the multi-variance system described above. The sequences were composed of low frequency noises and of an important linear frequency drift (10 times greater than that of figure 1). In this case, after extracting the linear drift from the sequences, the multi-variance is almost unable to measure the noise coefficients (about 50% of mean error), especially for very low frequency noise coefficients (h_{-3} and h_{-2}). On the other hand, the simultaneous determination of noise and drift coefficients permits a greater accuracy (about 20% of mean error for h_{-3} and 10% for h_{-1}). The most spectacular improvement is the determination of the linear frequency drift. The accuracy of the measurement of this coefficient is 40 times better (2% of mean error by the least squares, but 0.05% by simultaneous determination) !

Since the multi-variance method provides an assessment of the precision of measurement of the coefficients [1, 2], we observed a satisfactory correlation between measurement errors (difference of entered and measured coefficient) and measurement uncertainties (estimated from the dispersion of the variance measurements). Moreover, the multi-variance method quantifies the separability of the system of variances for two noises (or drifts) by supplying the angles between each couple of eigenvectors (each eigenvector corresponds to a noise or a drift)[1, 2]. In the case of this multi-variance system, the drift coefficient is well distinguished from both f^3 and f^2 noises (87° and 86°). This result shows the advantage of using variances calculated in the time domain as well as in the frequency domain.

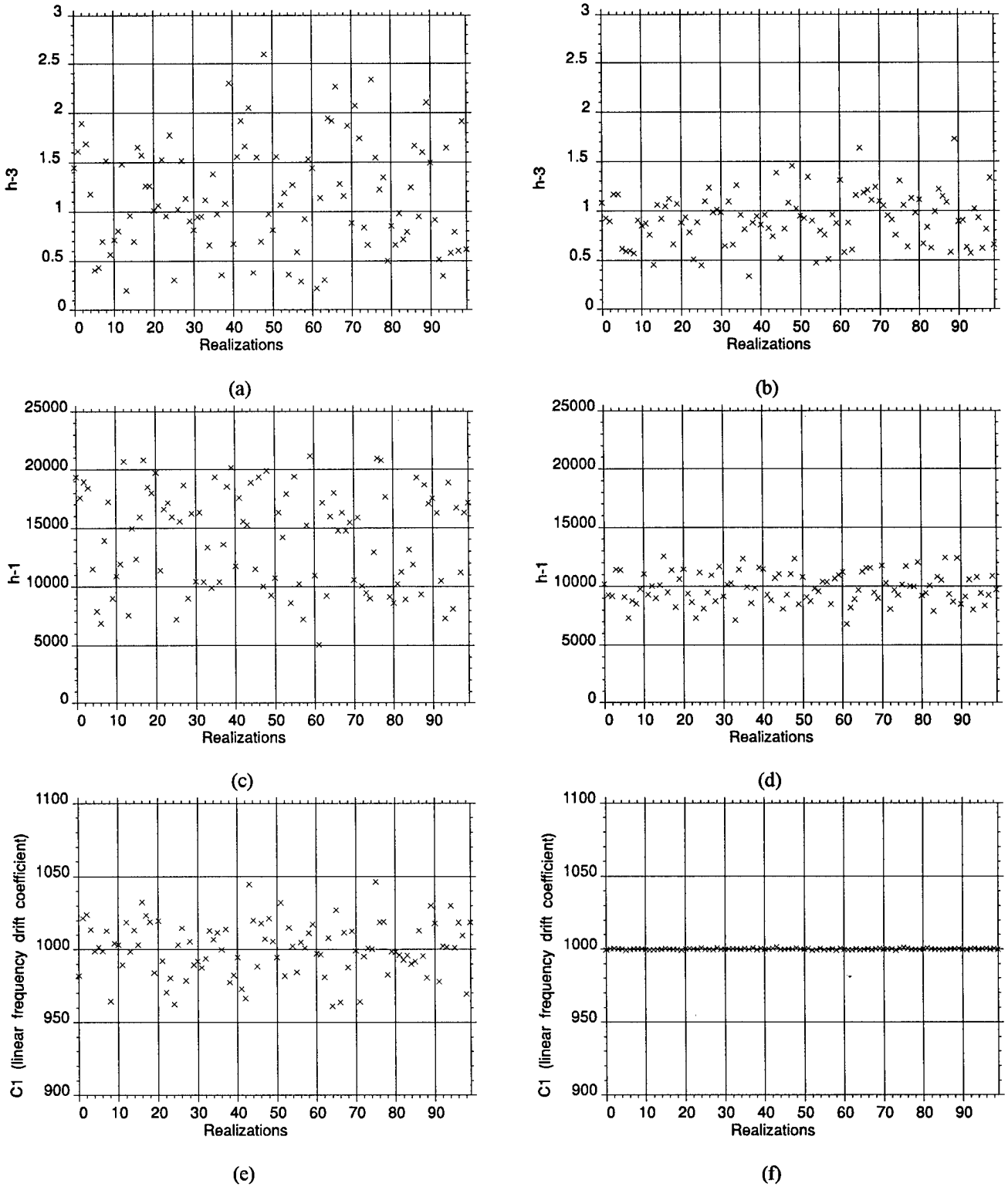


Figure 3 : results obtained over 100 realizations of sequences composed of f^{-3}, f^{-2}, f^{-1}, f^{0} frequency noises and a linear frequency drift (h_{-3}=1, h_{-2}=100, h_{-1}=10000, h_0=100000, C_1=1000). The multi-variance method was used (a), (c), (e) after extracting the drift by the least square method, (b), (d), (f) by simultaneous determination of the drift and of the noise coefficients.

We used also this multi-variance system for a sequence composed of a linear frequency drift and of white frequency noise. Since the least square method is optimum for white gaussian noise, we may expect that in this case the extraction of the drift by the least square method is more accurate than the simultaneous measurement of drift and noise coefficients. Indeed, we observed a mean error of about 0.05% for the determination of the drift coefficient with the least squares and 0.5% of mean error with the simultaneous determination. However, we noticed a slightly better accuracy for the determination of the white noise coefficient with the simultaneous determination (0.5% of mean error and 1% after extracting the drift).

4. CONCLUSION

These results prove that the least square method is indeed unable to separate a linear drift and a low frequency noise. Moreover, the use of a single variance cannot distinguish between a linear frequency drift and a f^{-2} frequency noise (if it is calculated in the frequency domain) or a linear frequency drift and a f^{3} frequency noise (if it is calculated in the time domain). Since f^{3} frequency noise may be encountered in timing data of pulsars, we cannot use only this last solution. Hence, the multi-variance method using variances calculated in both time and frequency domains seems to be the best solution in this case. Moreover, the great separability of a multi-variance system using these two types of variances increases the interest of this method.

Finally, an application of this method will be the analysis of the time stability of millisecond pulsars. It has been shown that these pulsars present a great long-term stability (durations greater than one year), which could be better than the long-term stability of the TAI. Consequently, the millisecond pulsars could be used to generate a new time scale. However, before studying the time stability of the millisecond pulsars, the raw data of pulsar timing must be corrected from several effects (earth motion, interstellar medium dispersion, relativistic effects, etc.). Thus, this new simultaneous multi-variance method will allow us to process directly the raw pulsar timing data and could avoid calculation bias.

5. REFERENCES

[1] F. Vernotte, E. Lantz, J. Groslambert and J. J. Gagnepain, "A new multi-variance method for the oscillator noise analysis", in Proceedings of the 46th annual Frequency Control Symposium, Hershey, USA, May 1992.

[2] F. Vernotte, E. Lantz, J. Groslambert and J. J. Gagnepain, "Oscillator noise analysis : multivariance measurement", IEEE Trans. Instrum. Meas., vol. 42, No. 2, 1993.

[3] M. V. Sazhin, "Opportunities for detecting ultralong gravitational waves", Soviet Astron., vol. 22, pp. 36-38, 1978.

[4] R. Blandford, R. Narayan and R. W. Romani, "Arrival-time analysis for a millisecond pulsars", J. Astrophys. Astr., vol. 5, pp. 369-388, 1984.

[5] J. E. Deeter, "Techniques for the estimation of red power spectra. II Evaluation of alternative methods", the Astrophys. J., vol. 281, pp. 482-491, 1984.

[6] D. R. Stinebring, M. F. Ryba, J. H. Taylor and R. W. Romani, "Cosmic gravitational wave background : limits from millisecond pulsar timing", Phys. Rev. Lett., vol. 65, pp. 285-288, 1990.

[7] J. Rutman, "Characterization of phase and frequency instabilities in precision frequency sources : fifteen years of progress", Proceedings of the IEEE, vol. 66, no. 9, pp. 1048-1075, September 1978.

[8] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leeson, T. E. McCunigal, J. A. Mullen, W. L. Smith, R. L. Sydnor, R. Vessot and G. M. R. Winkler, "Characterization of frequency stability", IEEE Trans. Instrum. Meas., vol. IM-20, pp. 105-120, 1971.

[9] F. Vernotte, "Stabilité temporelle des oscillateurs : nouvelles variances, leurs propriétés, leurs applications", Ph. D. Thesis, Université de Franche-Comté, No. 199, Besançon, France, February 1991.

[10] E. Boileau and B. Picinbono, "Statistical study of phase fluctuations and oscillator stability", IEEE Trans. Instrum. Meas., vol. IM-25, pp. 66-75, 1976.

[11] D. W. Allan and J. A. Barnes, "A modified "Allan variance" with increased oscillator characterization ability", in Proceedings of the 35th annual Frequency Control Symposium, Ft. Monmouth, USA, pp. 470-475, May 1981.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A NEW APPROACH TO CLOCK MODELING AND KALMAN FILTER TIME AND FREQUENCY PREDICTION

Wei Su
Geo Centers, Inc., 615 Hope Road, Eatontown, NJ 07724

Raymond L. Filler
U.S. Army Research Laboratory, AMSRL-EP-ME, Fort Monmouth, NJ 07703-5601

Abstract

An extension of a previously reported Kalman filter prediction algorithm is presented. In this new algorithm, the time variable is added to the clock model so both the time and frequency can be predicted. The long-term limitation of the previous prediction algorithm is removed by calculating the aging time constant using nonlinear fitting with the gradient search technique. Finally the noise covariances are discussed.

Introduction

The application of the Kalman algorithm[1] to precision clocks was extended recently by Stein and Filler[2]. Time, frequency, and aging rate, denoted by x(t), y(t), and $a(t)$ respectively, were described by the vector equation

$$\begin{bmatrix} \dot{x}(t) \\ \dot{y}(t) \\ \dot{a}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \\ a(t) \end{bmatrix} + \begin{bmatrix} \zeta(t) \\ \xi(t) \\ \phi(t) \end{bmatrix} \quad (1)$$

where $\zeta(t)$, $\xi(t)$, and $\phi(t)$ were assumed to be white noises. The frequency prediction at the discrete-time t_{k+1} was given by

$$\bar{y}_{k+1} = \hat{a}_k t_{k+1} + \hat{b}_k. \quad (2)$$

The aging rate $\hat{a}_k$ and the parameter $\hat{b}_k$ in Eq. (2) were estimated by the Kalman filter algorithm from measurements taken up to and including the one at time t_k. The deficiency of this model is that the aging rate $\hat{a}_k$ was assumed to be constant after the most recent measurement.

Su and Filler[3] extended the model to include the non-linear frequency vs. time function[4,5]

$$y(t) = \alpha \cdot \log(Bt+1) + c \quad (3)$$

A two dimensional discrete-time model, including just the frequency and the parameter α_k, was obtained from Eq. (3) resulting in

$$\begin{bmatrix} y_{k+1} \\ \alpha_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & \log\frac{Bt_{k+1}+1}{Bt_k+1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_k \\ \alpha_k \end{bmatrix} + \begin{bmatrix} 0 \\ \omega_k \end{bmatrix} \quad (4)$$

which was simplified in the long-term limit (where $|Bt_k| \gg 1$) to

$$\begin{bmatrix} y_{k+1} \\ \alpha_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & \log\frac{t_{k+1}}{t_k} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_k \\ \alpha_k \end{bmatrix} + \begin{bmatrix} 0 \\ \omega_k \end{bmatrix}; \quad (5)$$

$$|Bt_k| >> 1.$$

Thus knowledge of the aging time constant B is not needed if the long time assumption $|Bt_k| \gg 1$ is true.

The deficiencies of this model are (1) the time variable x(t) is not included, (2) when t_k is small, the prediction is poor, and (3) the covariance matrices Q_k and R_k have to be determined before the algorithm can be implemented. To overcome these problems, a new algorithm was formulated.

Clock Model Formulation

If we assume α and c in Eq. (3) are constants we have

$$\frac{dy(t)}{dt} = \frac{\alpha \cdot B}{Bt+1} \quad (6)$$

which gives the vector equation

$$\begin{bmatrix} \dot{x}(t) \\ \dot{y}(t) \\ \dot{\alpha}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & \frac{B}{Bt+1} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \\ \alpha(t) \end{bmatrix} \quad (7)$$

or simply

$$\dot{\chi}(t) = A(t) \cdot \chi(t) \quad (8)$$

0-7803-0905-7/93 $3.00 © 1993 IEEE

where

$$\chi^T(t) = [\, x(t) \quad y(t) \quad \alpha(t) \,] \tag{9}$$

and

$$A(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & \dfrac{B}{Bt+1} \\ 0 & 0 & 0 \end{bmatrix}. \tag{10}$$

Eq. (7) is the continuous-time model obtained from Eq. (3). However the unknown parameters α and c have not yet been determined. To estimate these unknowns, we use a discrete-time version of Eq. (7). The discrete-time vector χ_{k+1} is obtained by

$$\chi_{k+1} = \chi(t_{k+1}) = \Phi(t_{k+1},t_k)\cdot\chi(t_k) \tag{11}$$

where the state transition matrix $\Phi(t_{k+1},t_k)$ is calculated by a method known as the Peano-Baker method of integration[6],

$$\Phi(t_{k+1},t_k) = I + \int_{t_k}^{t_{k+1}} A(\lambda)d\lambda + \int_{t_k}^{t_{k+1}}[A(\lambda_1)\int_{t_k}^{\lambda_1} A(\lambda)d\lambda]d\lambda_1 + \tag{12}$$
$$+ \int_{t_k}^{t_{k+1}}\{A(\lambda_2)\int_{t_k}^{\lambda_2}[A(\lambda_1)\int_{t_k}^{\lambda_1} A(\lambda)d\lambda]d\lambda_1\}d\lambda_2 + \cdots\cdots.$$

Inserting Eq. (10) into Eq. (12) we have

$$\Phi_{k+1,k} = \Phi(t_{k+1},t_k)$$
$$= \begin{bmatrix} 1 & t_{k+1}-t_k & \dfrac{Bt_{k+1}+1}{B}\cdot\log\dfrac{Bt_{k+1}+1}{Bt_k+1} - t_{k+1}+t_k \\ 0 & 1 & \log\dfrac{Bt_{k+1}+1}{Bt_k+1} \\ 0 & 0 & 1 \end{bmatrix}. \tag{13}$$

Finally, Eqs. (10)-(13) yield the discrete-time clock model

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \\ \alpha_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & t_{k+1}-t_k & \dfrac{Bt_{k+1}+1}{B}\cdot\log\dfrac{Bt_{k+1}+1}{Bt_k+1} - t_{k+1}+t_k \\ 0 & 1 & \log\dfrac{Bt_{k+1}+1}{Bt_k+1} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ \alpha_k \end{bmatrix} \tag{14}$$

Time and Frequency Estimation

The Kalman algorithm is a recursive computation for obtaining the optimum estimate of the noisy state χ_k based on the measurement

$$z_k = H_k\chi_k + \mu_k \tag{15}$$

where μ_k is the measurement noise which is assumed to be white, and where H_k is the measurement matrix which is determined based upon the application. For instance, time measurements of the clock are described by

$$H_k = [1\ 0\ 0] \tag{16}$$

and frequency measurements of the clock are described by

$$H_k = [0\ 1\ 0]. \tag{17}$$

The estimate of χ_k, denoted by $\hat{\chi}_k$, is obtained from[1]

$$\hat{\chi}_k = (I - P_k^- H_k^T(H_kP_k^-H_k^T+R_k)^{-1}H_k)\Phi_{k,k-1}\hat{\chi}_{k-1} \tag{18}$$
$$+P_k^-H_k^T(H_kP_k^-H_k^T+R_k)^{-1}z_k$$

where

$$P_k^- = \Phi_{k,k-1}(P_{k-1}^- - P_{k-1}^-H_{k-1}^T(H_{k-1}P_{k-1}^-H_{k-1}^T+R_{k-1})^{-1}H_{k-1}P_{k-1}^-)\Phi_{k,k-1}^T + Q_{k-1} \tag{19}$$

The matrices Q_k and R_k in Eqs. (18) and (19) are known as the process noise covariance matrix and the measurement noise covariance matrix, respectively. An exact knowledge of these matrices are required in the Kalman algorithm.

Determining the Noise Covariance

Mehra[7] proposed an algorithm for estimating both Q_k and R_k using innovation sequences. This method does not apply to this clock model, Eq. (7), because Mehra's algorithm assumed a time-invariant system. Stein[8] determined Q_k and R_k by calculating the variance of innovations in terms of the true values of the parameters and the approximate gain and actual covariance. Here, we use a different approach starting with the variance of the Kalman filter innovation sequence[7]

$$E[(z_k - H_k\Phi_{k,k-1}\hat{\chi}_{k-1})(z_k - H_k\Phi_{k,k-1}\hat{\chi}_{k-1})^T] = H_kP_k^-H_k^T + R_k. \tag{20}$$

If we assume that the unknown parameters α and c are constant, Q_k will be the null matrix. Inserting Eq. (19) into Eq. (20), and rearranging yields

$$R_k = E[(z_k - H_k\Phi_{k,k-1}\hat{\chi}_{k-1})(z_k - H_k\Phi_{k,k-1}\hat{\chi}_{k-1})^T] \tag{21}$$
$$-H_k\Phi_{k,k-1}(P_{k-1}^- - P_{k-1}^-H_{k-1}^T(H_{k-1}P_{k-1}^-H_{k-1}^T+R_{k-1})^{-1}H_{k-1}P_{k-1}^-)\Phi_{k,k-1}^TH_k^T$$

We can estimate the expectation value of the innovation sequence by using the time average as follows[7]

$$E[(z_k - H_k\Phi_{k,k-1}\hat{\chi}_{k-1})(z_k - H_k\Phi_{k,k-1}\hat{\chi}_{k-1})^T]$$
$$\doteq \frac{1}{N+1}\sum_{l=k-N}^{k}(z_l - H_l\Phi_{l,l-1}\hat{\chi}_{l-1})(z_l - H_l\Phi_{l,l-1}\hat{\chi}_{l-1})^T \tag{22}$$

where N<k is an integer. Thus, R_k is estimated by

$$\hat{R}_k = \frac{1}{N+1}\sum_{i=k-N}^{k}(z_i - H_i\Phi_{i,i-1}\hat{x}_{i-1})(z_i - H_i\Phi_{i,i-1}\hat{x}_{i-1})^T \qquad (23)$$

$$-H_k\Phi_{k,k-1}(P_{k-1}^- - P_{k-1}^- H_{k-1}^T(H_{k-1}P_{k-1}^- H_{k-1}^T + \hat{R}_{k-1})^{-1}H_{k-1}P_{k-1}^-)\Phi_{k,k-1}^T H_k^T$$

Time and Frequency Prediction

If we assume that at time $t>t_k$, no reference frequency or time is available, the time and frequency at $t_p>t_k$, denoted by x_p and y_p, are predicted, based on the Kalman estimates obtained at t_k, by

$$\hat{x}_p^- = \hat{x}_k + (t_p - t_k)\cdot\hat{y}_k + \left(\frac{B\cdot t_p+1}{B}\cdot\log\frac{Bt_p+1}{Bt_k+1} - t_p + t_k\right)\cdot\hat{a}_k \qquad (24)$$

$$\hat{y}_p^- = \hat{y}_k + \hat{a}_k\cdot\log\frac{Bt_p+1}{Bt_k+1}$$

or

$$\hat{x}_p^- = \hat{x}_k + (t_p - t_k)\cdot\hat{y}_k + \left(t_p\cdot\log\frac{t_p}{t_k} - t_p + t_k\right)\cdot\hat{a}_k \qquad (25)$$

$$\hat{y}_p^- = \hat{y}_k + \hat{a}_k\cdot\left(\log\frac{t_p}{t_k}\right); \qquad |Bt_k| >> 1$$

Determining the parameter B

When $|Bt_k|\approx 1$ or smaller, the model in Eq.(4) and (24) should be used. Thus, the parameter B must be determined. To do this, we look back for N steps and fit the frequency measurements $z_k, z_{k-1}, \cdots, z_{k-N}$ (N<k) to the logarithmic function

$$y_j = \alpha\cdot\log(Bt_j+1) + c \qquad (26)$$

where α and c can be estimated either by the Kalman algorithm or by using a least squares fit. A search is conducted for the best parameter B which minimizes the fitting error

$$J = \sum_{j=k-N}^{k}(z_j - y_j)^2 \qquad (27)$$

We use a gradient search algorithm to update the trial parameter B by

$$B_{i+1} = B_i + g\cdot\frac{\partial J}{\partial B}\Big|_{B=B_i} \qquad (28)$$

until the best B is found, where g is a number which controls the step size of B_i.

We note that it is not necessary to determine B at each sampling time since the parameter B does not vary significantly with time. Thus the gradient search can be run periodically, say, once a week. The calculation will take little time since the search is only in one dimension.

Example

Figure 1 illustrates the frequency measurements of a quartz crystal oscillator for 100 days.

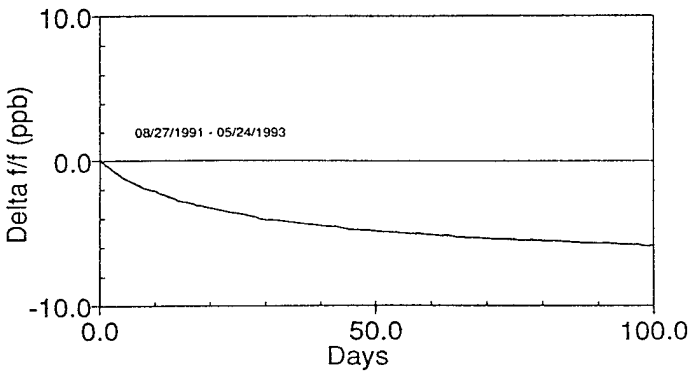


Figure 1 Frequency Measurements for 100 Days

When the time t is sufficiently large, the frequency model in Eq. (3) can be simplified to

$$y(t) = \alpha\cdot\log(t) + \beta \qquad (29)$$

where α and β can be estimated by the Kalman algorithm[3] and the frequency can be predicted without knowing the parameter B (See example 1 in [3]). However, if t is not large enough, these predictions will be quite poor. Figure 2 shows the short-term predictions using the long-time approximation, Eq. (29). The □'s are the frequency measurements taken at the discrete-times t_k (k=0,1,2,···), and the ▵'s are the frequency predictions y_{k+1} based on all of the measurements made up until time t_k. That is each ▵ is the prediction which would be generated just prior to the next measurement. The predictions in figure 2 do not agree with the measurements very well because y(t) in Eq. (29) tends to minus infinity as t approaches zero. To avoid this problem, we repeated the analysis using the logarithmic function Eq. (3) including the parameter B. The B is calculated with Eqs. (27) and (28), with N=20 and g=0.001. A much better result is obtained as shown in Figure 3.

Conclusions

The time and frequency of a quartz oscillator can be predicted by using the Kalman filter algorithm with a new clock model. The long-term limit applicable to the previous algorithm is no longer needed. The measurement noise covariance can be estimated if the process noise covariance is the null matrix.

References

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Trans. of the ASME-J. of Basic Engr., pp.35-45, March 1960.

[2] S. R. Stein and R. L. Filler, "Kalman Filter Analysis for Real Time Applications of Clocks and Oscillators," Proc. 42nd Annual Symposium of Frequency Control, pp.447-452, June 1988.

[3] W. Su and R. L. Filler, "Application of Kalman Filtering Techniques to the Precision Clock with Non-Constant Aging," Proc. 1992 IEEE Frequency Control Symposium, pp.231-237, May 1992.

[4] R. L. Filler and J. R. Vig, "Long Term Aging of Oscillators," Proc. 1992 IEEE Frequency Control Symposium, pp.470-484, May 1992.

[5] J. R. Vig, "Introduction to Quartz Frequency Standards," Proc. 23rd Annual Precise Time and Time Interval(PTTI) Applications and Panning Meeting, December 1991.

[6] L. A. Pipes, *Matrix Methods for Engineering*, Prentici-Hall Inc., Englewood Cliffs, N. J., pp.90-92, 1963.

[7] R. K. Mehra, "On the Identification of Variances and Adaptive Kalman Filtering," IEEE Trans. on Automatic Control, Vol. AC-15, No.2, pp.175-184, April 1970.

[8] S. R. Stein, "Kalman Filter Analysis of Precision Clock with Real-Time Parameter Estimation," Proc. 43rd Annual Symposium of Frequency Control, pp.232-236, June 1989.

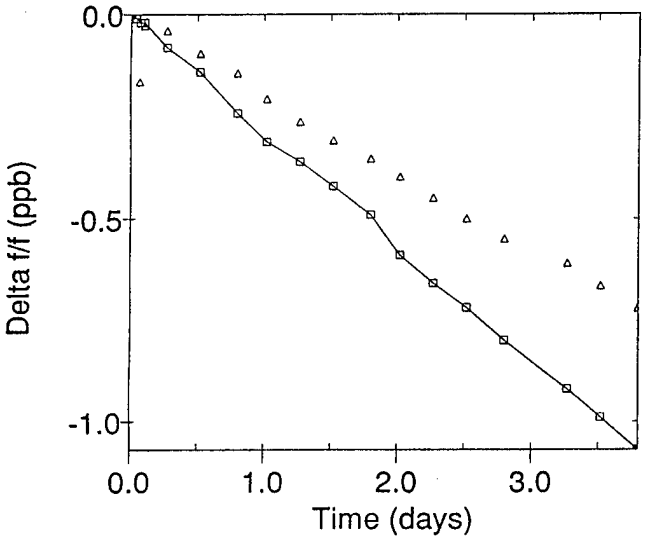


Figure 2 Short-term Predictions using the Long-term Assumption

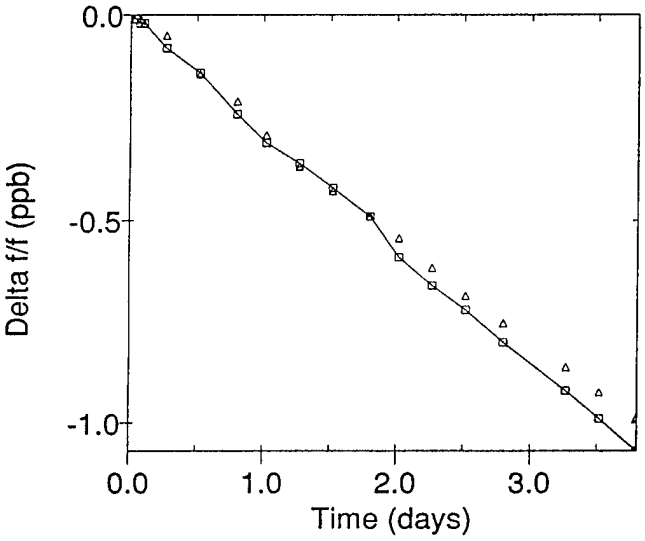


Figure 3 Short-Term Predictions after Calculating the B Parameter

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A SIMPLE PRECISION FREQUENCY STANDARD COMPARATOR

Wei Zhou, Zhongqiang Xuan

Dept. of Measurement and Instrumentation, Xidian University
Xi'an, 710071, P.R.China

Abstract

Conventional precision frequency standard comparators can only be used to compare the frequency signals with certain frequencies. This paper describes a frequency standard comparator based on the characteristics of the greatest common factor frequency and the phase coincidence detection technique. By using a stable and high frequency common oscillator, the comparator can not only compare frequency standards in a very high accuracy, but also measure frequency precisely in a wide frequency range. In the comparator almost all circuits are composed of digital integrated circuits. Because of its flexible uses, low cost and potential accuracy, it will be used widely.

1. Introduction

Using the characteristics of the greatest common factor frequency and the method of phase coincidence detection the much higher measuring accuracy than that of ordinary measuring instruments can be obtained in a very wide frequency range. The frequency and time measurement instruments designed with this technique usually show more than 1000 times the accuracy over that of ordinary instruments.

The greatest common factor frequency between two frequency signals is similar to the mathematical greatest common factor between two numbers. According to the characteristics of the greatest common factor frequency, the quantization phase variation discrimination between the two frequency signals f_1 and f_2 is as follows:

$$\Delta T = f_{maxc}/f_1 \cdot f_2 \qquad (1)$$

where f_{maxc} is the greatest common factor frequency between f_1 and f_2. The period of f_{maxc} is the least common multiple period T_{minc}. The characteristic of the phase difference variation between two signals depends on the frequency relationship between them. It is very complicated. If the quantization phase differences are rearranged by size sequence in a T_{minc} period, the variation of the adjacent phase differences must be ΔT.

In the measurement over a wide frequency range the phase coincidence detection precision of the circuit must be lower than ΔT, in order to be adjusted to the wide measurement range and fast response. The detection precision usually is from dozen times to more than one hundred times that of ΔT. Because ΔT is varied by different measured frequencies, one must consider the worst cases.

In comparison of some usual standard frequency signals, exact lengths of measuring gate time and higher accuracy are desirable. The frequencies of the standard signals often are 10 MHz, 5 MHz, 2.5 MHz, 1 MHz or 100 KHz. They are the same or proportional in frequency. From equation (1), it is unadvisable to make measurement with the standard frequency signals directly, because either the measuring accuracy is too low, or the measuring period is too long.

2. Dual broad-band phase detection method

By using a common oscillator with good short term frequency stability, the problems above mentioned can be solved. The demand for the common oscillator is very flexibile. Because the frequency of the common oscillator can be chosen randomly, in a very wide frequency range the frequency relationship between the common oscillator and two compared signals is also random. It is favorable for the measurement.

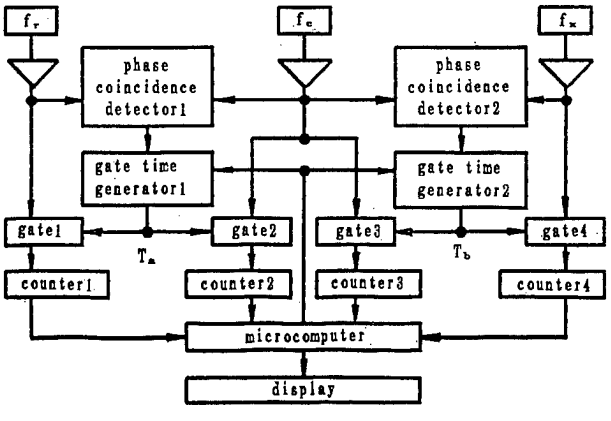


Fig. 1

A block diagram of the instrument designed by the new method is shown in Figure 1. The 'phase coincidences' between the common oscillator signal f_c and a reference frequency signal f_r, f_c and a measured

0-7803-0905-7/93 $3.00 © 1993 IEEE

frequency signal f_x are detected respectively. At the same time a reference gate time signal is obtained by either computer software or frequency divider from the standard frequency. Its length is 10ms, 0.1s, 1s or 10s, etc. Two measuring gate time signals T_a and T_b depend on the reference gate time signal and the 'phase coincidences'. The measuring gate time T_a is close to the reference gate time in length, but its starting and stoping moments correspond strictly to two 'phase coincidences' between f_r and f_c signals. The measuring gate time T_b is also close to the reference gate time in length, its starting and stoping moments correspond strictly to two 'phase coincidences' between f_x and f_c signals. Therefore, the gate time T_a is synchronized by the multiple periods of both f_r and f_c, and the gate time T_b is synchronized by the multiple periods of both f_x and f_c. The synchronization is exact, or very close to being exact. In gate time T_a the cycle numbers of f_r and f_c are counted respectively, and N_r and N_{c1} are obtained. In gate time T_b the cycle numbers of f_x and f_c are counted respectively, and N_x and N_{c2} are obtained. These 4 counted numbers represent four groups of complete or nearly complete cycle numbers respectively.

$$f_x=f_c \cdot N_x/N_{c2}$$

$$f_c=f_r \cdot N_{c1}/N_r$$

then

$$f_x=f_r\frac{N_x \cdot N_{c1}}{N_r \cdot N_{c2}} \qquad (2)$$

From the frequency of f_r and the four counted numbers the measured frequency f_x can be computed. According to equation (2), it is unimportant that the frequency of common oscillator is known or not.

In Fig. 1 the counted numbers of the four counters are fed into a microcomputer seperately. The microcomputer computes the numbers with equation (2) and drives a display device to display the measured result.

In the measurement there are the two greatest common factor frequencies, that are f_{maxc1} between f_c and f_r and f_{maxc2} between f_c and f_x. Hence the two quantization phase variation discriminations ΔT_1 and ΔT_2.

$$\Delta T_1=f_{maxc1}/f_c \cdot f_r \qquad (3)$$

$$\Delta T_2=f_{maxc2}/f_c \cdot f_x \qquad (4)$$

The two least common multiple periods are respectively:

$$T_{minc1}=1/f_{maxc1} \qquad T_{minc2}=1/f_{maxc2}$$

The frequency f_c of the common oscillator can be chosen very flexibly, to obtain the suitable f_{maxc1}, f_{maxc2} and less ΔT_1 and ΔT_2. The measuring accuracy is determined by ΔT_1 and ΔT_2, and is lower than anyone of them. For the high accuracy, better time response and wide measuring frequency range, generally, the common oscillator ought to have following characteristics: 1. It is of high frequency. 2. Its frequency or multiple, fractional frequency has some frequency difference compared with usual frequency standard signals. 3. It must have good short term frequency stability.

In the measurement, f_r is fixed and f_x is changeable. Therefore, ΔT_1 is also fixed and ΔT_2 is also changeable. In the T_{minc1} period the distribution of 'phase coincidences' between f_c and f_r is regular, if the frequency of the common oscillator is close to the frequency of the reference signal or its multiple or fractional frequency, and there will be some frequency difference with it. But in T_{minc2} period the distribution of 'phase coincidences' between f_c and f_x is complicated, sometimes it is regular and uniform, however, in most cases it is irregular. Because there are different frequency relationships and distributions of 'phase coincidences' between f_c and f_x, the accuracy will be different for different measured frequecy signals.

3. A precision frequency standard comparator based on the dual broad-band phase detection method

With the help of the dual broad-band phase detection method and a suitable common oscillator, a precision frequency standard comparator can be built. It can be used to compare the usual frequency standards with a very high accuracy, and can also be used to measure any frequency signals in a very wide frequency range with a high accuracy. The block diagram of the comparator is shown in Figure 2.

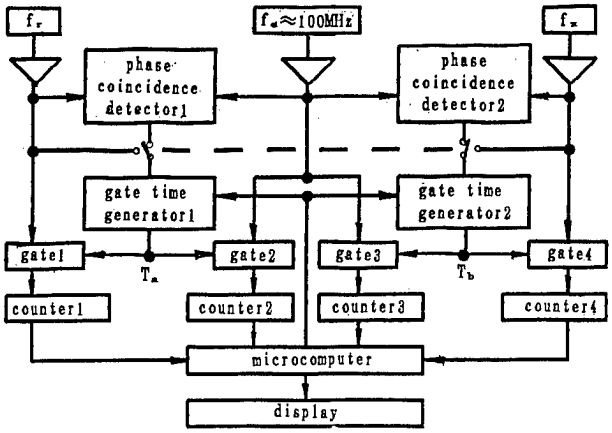


Figure 2

In this comparator the common oscillator is an unlocked quartz crystal oscillator which has good

short term stability. Its nominal frequency is 100 MHz, but its practical frequency has about 200Hz frequency difference with the nominal frequency. The short term stability of the common oscillator is better than 5×10^{-10}/ms, exerting no influence upon the accuracy of the comparator. In Figure 2, if the gates T_a and T_b are opened and closed exactly at the same time, the frequency instability of the common oscillator cannot influence the measurement accuracy. However, they cannot be synchronized exactly. The time difference between the starting or stoping time of the two gate times is about from zero to 5 millisecond. It is the frequency fluctuation in the time interval of two not synchronized gates which is much less than the most gate time that influences measuring accuracy. If the phase fluctuations of the common oscillator are small during the interval as compared with the phase fluctuations between f_r and f_x over a full gate time, the noise of the common oscillator is insignificant in the measurement noise error budget, which means the noise of the common oscillator can be worse than that of either f_r or f_x and still not contribute significantly.

Obviously, the main part (100 MHz) of frequency of the common oscillator is the multiple frequencies of the usual frequency standard (for example, 10 MHz, 5 MHz, 2.5 MHz, 1 MHz, 100 KHz, etc.). Therefore, there is a very regular distribution of 'phase coincidences' between f_o and f_r, f_o and f_x. In a period of the equivalent harmonic difference frequency between f_o and compared signal f_r or f_x, the variation of the phase difference is from large to small or from small to large regularly and uniformly. In the period of the equivalent harmonic difference frequency, the increment of the variation of the phase difference is:

$$\Delta T' = f_{hm}/f_o \cdot f \qquad (5)$$

where f_{hm} is the equivalent harmonic difference frequency between f_o and f_r or f_x, f is the frequency of f_r or f_x. To different frequency standard signals (for example, 10 MHZ, 5 MHz, 1 MHz), $\Delta T'$ can be about 0.2 ps, 0.4 ps and 2 ps. Obviously, $\Delta T'$ is not equal to ΔT. Generally, the former is much larger than the latter. In special case, we can use the latter to get higher accuracy.

The case mentioned above is shown in the waveform diagram Fig. 3.

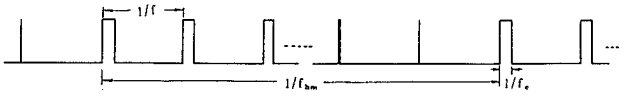


Figure 3

For this very regular variation of the phase difference, it is the stability of the phase coincidence detection circuit, rather than its precision, that influences the accuracy of the comparator. Generally, using digital integrated circuit one only can get merely several tenths nanosecond precision of the phase detection. However, the stability of some circuits can be much higher than that. Using this characteristic, one can get higher accuracy. In the comparator the main errors are the triggering error of the circuits and the instability of the circuits.

Using the same instrument other frequency signals can also be measured. The frequency range of the measured signals is from 10 KHz to over 100 MHz (the upper frequency limitation only depends on the the circuit speed). Because in this case, the distribution of the 'phase coincidences' and the variation of phase differnce between f_o and f_x are irregular, the measuring precision depends on the precision of the phase coincidence detection circuits. To accomplish the measurement in a very wide frequency range, the precision of the phase detection cannot be too high. Therefore, the precision of frequency measurement in a wide frequency range is lower than that of the frequency standard comparison. In Figure 2 the reference gate time can be 10 ms, 100 ms, 1s and 10s respectively. When this comparator is used to compare frequency standard, the measuring precision of frequency stability is higher than 1×10^{-10}/10ms, 1×10^{-11}/100ms, 1×10^{-12}/1s and 1×10^{-13}/10s at 5MHz and 10MHz. The measuring result of frequency stability can be displayed by Allan variance, and the sample interval is about 5 ms. When the Allan variance of the measured signal is computed, in equation

$$\sigma y(\tau) = \frac{1}{f_o}\sqrt{\frac{\sum_{i=1}^{N}(f_{i+1}-f_i)^2}{2N}}$$

the first result of the measurement of the measured signal is used as f_o. It has enough accuracy. By using this method, the comparator can be used to measure the frequency stability of any frequency signal in the measured frequency range.

When the display is in frequency, the display figures are 14 figures/10s, 13 figures/1s, 12 figures/100ms and 11 figures/10ms. When this comparator is used to measure frequency in a very wide frequency range, the measuring precision is higher than 3×10^{-8}/10ms, 3×10^{-9}/100ms, 3×10^{-10}/1s and 3×10^{-11}/10s over 1MHz, and higher than 6×10^{-8}/10ms, 6×10^{-9}/100ms, 6×10^{-10}/1s and 6×10^{-11}/10s under 1 MHz. According to the principle of the comparator, in the comparison of frequency standards its gate time is accurate, but in the measurement of any frequency signal its gate time is not equal to its nominal value strictly. This characteristic coincides with the most application demands.

This comparator can also be used to compare frequency standard with other frequencies (for example

4.608 MHz, 10.7 MHz etc.) with a very high precision, if other common oscillators with suitable frequencies are used.

In this comparator the phase coincidence detection circuits can also not be used. The measuring gate T_a is composed of f_o and f_r, T_b, of f_o and f_x directly as shown in Figure 2. In this case the measuring precision is lower, but wider frequency measuring range can be obtained. The lower limitation frequency is lower than 1Hz. The gate T_a is synchronized by f_r, and the gate T_b is synchronized by f_x. The equation for the computation is the same as equation (2). As the result of the measurement there is $\pm 1 \sim 2$ count error of the common oscillator and in most cases it is less than ± 1 cycle of f_r and f_x.

4. Conclusion

This new frequency standard comparator can be used to compare frequency standard with a very high precision, measure frequency signals in a wide frequency range with a high precision, and measure frequency in a wider frequency range. In this instrument almost all circuits are digital integrated circuits, it being very simple. This comparator can also be used to compare frequency standard with other frequencies precisely, if the frequency of the common oscillator is changed. This comparator serves as only an application example of some new measuring principles and techniques. They are the conception of the greatest common factor frequency and phase coincidence detection techniques (the frequency measurement technique by broad- band phase detection and the dual broad-band phase detection method). Because of their high accuracy, wide measuring range, simple structure and low cost, these new techniques can be used more widely.

References:

1. Wei Zhou "The Greatest Common Factor Frequency and Its Application in the Accurate Measurement of Periodic Signals" 1992 IEEE Frequency Control Symposium, May, 1992

2. Wei Zhou "A New Principle of Linear Phase Comparison - Irregular Phase Discrimination" 18th Annual PTTI Applications and Planning Meeting, Nov. 1986

3. Wei Zhou "A New Measurement Technique of Frequency and Periodic Signals - The Frequency Measuring Technique by Broad- band Phase Detection" will be published in Chinese Journal of Scientific Instrument

4. D.W. Allan, Herman Daams "Picosecond Time Difference Measurement System" 29th Annual Symposium on Frequency Control Symposium, pp. 404 - 411, 1975

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PECULARITIES OF TECHNOLOGY, PHYSICAL PROPERTIES AND APPLICATIONS OF NEW PIEZOELECTRIC MATERIAL LANGASITE($La_3Ga_5SiO_{14}$)

A.N.Gotalskaya, D.I.Drezin, V.V.Bezdelkin,
V.N.Stassevich

Scientific Research Institute "Fonon",Moscow,Russia

Abstract

The technology of synthesis and of growing langasite crystal are described. The features of crystallization process are discussed. Main defect types of langasite crystals are defined. The results of crystallografic investigation of langasite crystalls connected with correction of Bragg-angles and angular parameters of atomic planes in this material are given. The fundamental psysical properties of langasite and the application of this crystals in resonators and filters are discussed.

Introduction

For designing VCXOs and TCXOs and filters with a wide pass band while maintaining high stability and small insertion attenuation of these devices, respectively the necessity arose in discovering new piezoelectric crystals having intermediate properties between those of quartz and lithiuim tantalate.They are berlinite and langasite.Growing berlinite crystals and its analogues, however, is connected with considerable technological difficulties and with the twinning problem [1].Technological process of growing langasite crystals is simpler. It is grown up by Czochralski method in air atmosphere at the temperature of 1470°C. Langasite has no phase transitions up to the melting point and it does not need monodomenization with compared to lithium niobate and lithium tantalate[2,3].

Langasite crystal growing process

Langasite ($La_3Ga_5SiO_{14}$) belongs to the group of compositions with the structure of calcium-gallium-germanium garnet (CGGG) $Ca_3Ga_2Ge_4O_{14}$ and crystallizes in the same acentric spatial group as quartz.Langasite crystal density is of 5.754 g/cm^3, its Mohs hardness is 6.6, the linear expansion coefficients are α_{11} =5,1x10^{-6}C^{-1} and α_{33} =3,61x10^{-6} C^{-1}.The crystal is a dielectric [2-5].

The paper describes developments in mass production technology of synthesis and growing up langasite crystals with the diagonal diameter of 62 mm and the weigth up to 2,0 kg. The crystals are grown along the Z-axis by using the equipment called "Crystal" with the diameter weight control of the crystal being grown.The design chosen with the use of a Pt-screen enabled to grow run crystals with the cylindrical part up to 150 mm, the yield obtained was 60 %.

The charge was prepared from corresponding oxides by using solid-state phase reaction via an intermediate composition perovskite ($LaGaO_3$). Annealed weighted powders are mixed and synthetisized at the temperature of 1300°C by using quartz crucibles. The purity of starting reactives and the completness of chemical reaction are of prime importance, because the deviation from the stoichiometry within 0,3 - 2,0 % causes impurity phase accumulation in the melt and the deterioration of crystals being grown. Based on the results obtained by differential thermal gravimetric analysis (DTG) and X-ray phase analysis it was found that $La_3Ga_5SiO_{14}$ phase appears at the temperature range above 1000°C. The samples were synthesized at the temperatures of 1000°C, 1200°C, 1300°C 1400°C by using DTG-method at the derivatograph Q-1000.

The presence of small endoeffects at temperatures of 420°C and 580°C at samples annealed at 1000°C and 1200°C envidences of the presence of non-reactive phase La_2O_3 (see figure 1).

0-7803-0905-7/93 $3.00 © 1993 IEEE

The samples annealed at the temperature above 1300 °C did not show these peaks, which confirms the chemical reaction completness. The final is stage melting charge into Pt- crucible.

After melting, the melt is held for a certain time period for homogenization process of the system. Langasite single crystals are grown on the seeds cut along the Z-axis. The seeds were oriented with the accuracy not less that 10".The seeding moment was determined visually and by an abrupt weight increase by 0,5 g.

The column of 10 mm was grown with the rate of 10 mm per hour is then pulled. At this stage dislocation exit occurs. These dislocations can originate both from a seed crystal and can result from thermal shock during the process of seed insertion. By means of thyristor power decrease, a smooth broadening crystal begins and further growth with a constant diameter takes place.

At the end of the crystal growth run the crystal is smoothly taken off the melt and is slowly cooled down by means of a stepwise voltage change.

The features of langasite crystallization process are: dissociation presence and evaparation of one of the melt components (gallium oxide), crystallization front (CF) inversion and facial growth. Gallium oxide evaporation causes distortion of stoichiometric composition of the melt and deterioration of crystal quality at the end of the run.Gallium oxide dissociation contributes to platinum oxidation, its transfer to the melt and entering into crystal as inclusions. The most important condition for growing perfect crystals is obtaining of a flat crystallization front.A number of experiments has been made for detailed investigation of this problem. Inversion of the crystallization front was determined by the changes in incremental weight of a crystal being grown. CF-inversion was observed in the process of conical part increase of the crystal.At the initial growth stage the crystallization front has a convex shape directed to the melt , because the convection currents from the hot crucible walls are directed to the surface of the melt towards the cooler centre.This causes excessive defect formation in the conical part of the crystal. Having attained the critical diameter,the growing crystal suppresses, due to its rotation the naturial convection and changes the direction of convection currents to an opposite one. Crystallization front becomes flat as a result. In order to obtain a flat interface it is recommended to select such a rotation rate that the crystallization front inversion takes place at the end of the conical part of a boule. In the case of high rotation rates an unwanted turbulization of the melt currents occurs in the region of the crystallization front. Thus, these exits a sufficiently narrow region of optimum rotation rate values, beyond which the crystal quality deteriorates. For langasite crystals a facial growth is characteristic. Crystals grown along the Z-axis have well developed six prism faces $\{10\bar{1}0\}$ at its latteral surface. Sometimes they can be cut by the prism faces $\{11\bar{2}0\}$. From the point of view of the CF-stability the Z-orientation at which the cross section is near to a correct hexagon,should be considered to be the optimum one.We have grown up langasite crystals along the Y-axis with the diameter of 40 mm.While growing crystals along this direction, however, the CF-stability is lost at a certain stage of growth, because of asymetric crystal shape. The Y-orientated crystals grow with strongly distorted prism face with lower growth rate and have considerable residual stresses.

For quality evaluation of "as grown" crystals it is necessary to consider the real structure and main defect types which can be inherent to langasite crystals.

The sectorial inhomogeneity occurs when the crystallization front constitutes a complicated surface,consisting of regions of normal and facial growth. X-ray investigations of crystals with sectorial structure showed that facial growth regions have greater elementary cell parameter a = 0.007 Å. For this reason residual stresses can appear in the crystal. The analysis of double-refraction in different points of samples cut along the crystal length, showed that in crystals with a wide central region of normal growth, only axial stresses of tensile in the centre and those of compression at the periphery are acting. The voltage sign change occurs at the interface of normal and facial growth regions. In the crystals with a narrower central normal growth region non -axial stresses in the growth layers are superimposed on the axial ones. The indicatrix axis deviates from the crystal axis at the angle 5°-20°.

Considerable influence on the homogeneity of the growing crystal exert short-term temperature fluctuations in the vicinity of the crystallization front,which cause considerable changes in the solid phase composition.As a result of this the regions occur, parallel to the growth boundary, which differ in color, refraction index and in other characteristics,i.e. so-called striatation along the crystal length.

Langasite crystals grown with low rotation rates have a growth column - a macrodefect along the crystal,consisting of gaseous inclusions and microcracks. This defect appears because a standing

still region exists in the centre of the CF under dominance of natural convection in the melt. For elimination of this defect it is usually necessary to elevate the rotation rate. When natural convection predominates in the melt, voids (knags) can be formed at the conical part of the growing crystal. This is due to high growth rate in the axial direction and to the convex CF- shape.

For langasite crystals two types of inclusions are characteristic: Pt-inclusions and crystal inclusions of other phases, forming at the end of the growth run when stoichiometry deviations are present. Langasite melt distoys partially the Pt-crucible at high temperatures. This is caused by gallium oxide dissociation as a result of which the atomic oxygen is formed.

For development of new compositions with the CGGG-structure and new technology of crystal growing, the compositions $Ca_3Ga_2Ge_4O_{14}$, $Sr_3Ga_2Ge_4O_{14}$, $Ba_3Ga_2Ge_4O_{14}$ were synthetisized. These works are at the initial stage yet, but prelimilary results obtaind are rather promising. As far as piezoelectric moduli and ultrasonic attenuation are concerned, they surpass α-quartz and langasite. They have no phase transitions up to the melting temperature ($1350^{\circ}C$-$1470^{\circ}C$).

By correlating the behaviour of piezoelectric moduli d_{ij} with the relationship of two-valent cations and polyheders surrounding them, it is possible to observe, that the smaller is the difference $(M-O)_m^{VIII}$ and $(M+O)_m^{VIII}$ and the denser the two-valent cations are located in oxigen polyheders, the greater are the values piezoelectric moduli of the compositions. The greatest values of piezoelectric moduli in this series has $Ba_3Ga_2Ge_4O_{14}$ and the smallest ones has $Ca_3Ga_2Ge_4O_{14}$. In the series of compositions with CGGG - structure langasite takes an intermediate position with regard to its piezoelectric properties. With the implantation of lighter cations in CGGG-structure the stability and the quality of the material will increase. With the implantation of heavier cations with a greater radius (Sr,Ba), the piezoelectric moduli, the electromechanical coupling coefficient will increase and the pass band will be broader. At present we investigate the influence of dopants on piezoelectric properties of langasite. It was found that Al-dopants influence favourably the piezoelectric properties of the material.

Crystallografic and X-ray investigations of langasite crtstals

For initial sample orientation etch figures have been investigated on the surface of the basis plane cut 00.1[4]. The optimum etch conditions were: $HCl:HNO_3$=1:1 solution, 2 h, at the room temperature, or 15 min at the temperature of $120^{\circ}C$. The etch patterns had triangular form. Analogous to quartz, the crystallographic +X-direction goes out of the acute angle of the etch figure 2. For initial X-ray orientation of langasite crystals it is reasonable to use two atomic planes of X-fan with the diffraction indices 01.3 and $0\bar{1}.3$. Their relative intensities differ in more than an order of magnitude: 4.0 % and 0.3 %. X-ray control of "as grown" crystal was made at the basis plane 00.1 by means of piezoelectric goniometer or diffractometer in accordance with Figure 2. Analogously to quartz, we supposed that the positive end of +Y-axis goes out from under the minor rhombohedral face directed upwards (the index is 01.3 for langasite). For obtaining reliable information on elementary cell parameter definitions a large quantity of mass-produced samples has been measured. Their orientation was yxl/β and xyl/β. They have been processed by micropowder. The relative reflection intensity of atomic plane cuts was determined from the ratio in percent $I_{hkl}/I_{00.2}$, where $I_{00.2}$ is the peak intensity of corresponding atomic plane cut. The analysis of a great number of extremum values of the relative reflection intensity of langasite atomic planes gave the possibility to estimate standard values error. For the medium-grade crystals it is equal to 10-15 %.

The differences in stoichiometric composition, in the number of inclusions and in technological process of growth have significant influence on elementary cell parameter values a, c and Bragg's angles. The final specification of the elementary cell parameters of langasite crystals was obtained on the basis of experimental angle data ϑ°_{hkl} and Θ_{hkl} for atomic planes with diffraction indices (table 1). For calculation of the mean value of wavelegth CuK_{α} X-ray radiation was used $\lambda_{K_{\alpha 1,2}}$=1.5418 Å. The finally obtained parameter values were a=8.1783 Å and c=5.1014 Å. Based on them the final final calculation of angle parameters of atomic planes, giving reflections with the intensity of $\geqslant$ 0.5 % ($I_{00.2}$) was

made. The arithmetic mean of the deviation of experimental data from the calculated ones was 0'51". This condition confirms the correctness of elementary cell parameter finally values found for 85 atomic planes. The most interesting results have been obtained while X-ray investigations of langasite crystal lattice homogeneity by means of a reference three-crystal X-ray spectrometer, with the absolute angle measurement error of $\leqslant$0.95". The measured data show that the peak intensity value $I_{00.2}$ changes insignificantly along the sample length, practically at the level of a measurement error. More sensitive integral parameter of a crystal quality is a half width of a peak. It characterizes a disorientation of single crystal sections. For comparison, the natural disorientation is 2-10" for quartz, 30-60' for lithium niobate, 60-90' for lithium tantalate. This value is 12-17" for langasite.

Application of langasite crystals in devices

Langasite crystals grown at Scientific Reseach Institute "Fonon" are succesfully used in piezoelectric devices. As it has been reported at previous Frequency Control Symposia langasite is widely used in resonators and filters for covering the pass band of 0.4-0.8 % over a wide frequency range.Typical characteristic of the first samples of HF langasite resonators and filters are given in tables 2 and 3.

Investigation of experimental samples showed, that frequency - temperature characteristic (FTC) of langasite resonators and filters have the form of a quadratic parabola with the extremum in the region of the room temperature for yxl/0-3°-cuts,where nearly maximum value of electromechanical coupling factor ~18 % is observed,and the frequancy coefficient has a nearly minimum value.Their change in the vicinity of zero cut has a gently sloping character. This evidences of the small compared to quartz crystal units dependence of FTC inflection point, of electromechanical coupling coefficient and frequency constant (N) on the angle of cut α.This simplifies the manufacturing technology of langasite products.The steepness coefficient of FTC branches is $-7\text{x}10^{-8}\ ^{o}C^{-2}$. Thus as far as thermal stability of frequency is concerned, langasite resonators and filters are twice more stable than those of lithium tantalate.

The mean resonance spacing value for langasite resonators is 0.8 %,i.e.it is 3-4 times more compared to the AT-cut quartz crystal units. The frequency coefficient value is 1380 kHz.mm, which is 1,2 times smaller compared to AT-cut quartz crystal units.It is important to note, that parameters of langasite resonators and filters can be improved by employing optimum design and technology.

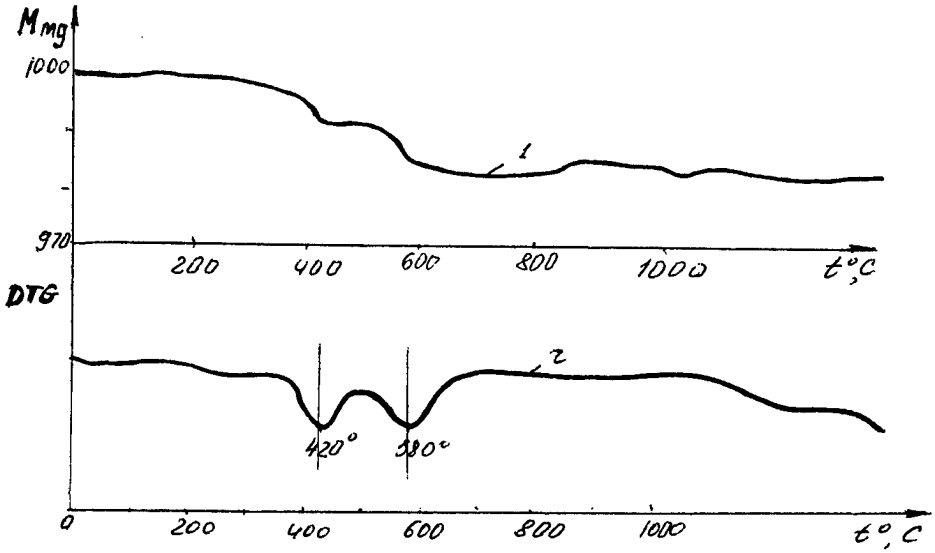


Figure 1. Results DTG-measurements of charge sampless glued at 1200°C;the peaks evidence of La_2O_3-presense

Table 1. Angle parameters of langasite atomic planes for mixed $K_{1,2}$ X-ray radiation of Cu...

Atomic plane indices $hk.\ell$	Trace angle $xC_{hk.\ell}$	Initial setting of angle sample $ZZ_{hk.\ell}$	Inclination angle of atomic plane $\varphi_{hk\ell}$	Wolf-Bragg angle $\theta_{hk\ell}$	Detector angle $2\theta_{hk.\ell}$	Relative intensity +	Relative intensity −
1	2	3	4	5	6	7	8
01.0	0°	0°	-7°06'00"	6°13'10"	12°26'20"	0,7	
02.0	—"—	—"—	-0°44'00"	12°35'10"	25°10'20"	16	
03.0	—"—	—"—	5°45'00"	19°04'10"	38°08'20"	23	
04.0	—"—	—"—	12°29'30"	25°48'40"	51°37'20"	8	
05.0	—"—	—"—	19°40'00"	32°59'10"	65°58'20"	20	
07.0	—"—	—"—	36°19'20"	49°38'30"	99°17'00"	8	
08.0	—"—	—"—	47°15'40"	60°34'50"	121°09'40"		
07.1	—"—	11°13'10"	37°40'00"	50°59'10"	101°58'20"	6	
06.1	—"—	13°01'40"	28°46'00"	42°05'10"	84°10'20"	0,6	
05.1	—"—	15°31'10"	21°04'00"	34°23'10"	68°46'20"	6	
04.1	—"—	19°08'30"	14°09'00"	27°28'10"	54°56'20"	0,6	1
03.1	—"—	24°05'00"	7°46'00"	21°05'10"	42°10'20"	4	
05.2	—"—	29°02'40"	25°10'00"	38°29'10"	76°58'20"	8	
02.1	—"—	34°46'00"	2°02'00"	15°21'10"	30°42'20"	45	38
04.2	—"—	—"—	18°41'00"	32°00'10"	64°00'20"	6	
03.2	—"—	42°47'10"	13°05'30"	26°24'40"	52°49'20"	13	10
04.3	—"—	46°09'30"	25°37'00"	38°56'10"	77°52'20"	0,6	
02.2	—"—	54°14'10"	8°32'30"	21°51'40"	43°43'20"	5	10
04.4	—"—	—"—	34°50'00"	48°09'10"	96°18'20"	1,6	
05.5	—"—	—"—	55°18'00"	68°37'10"	137°14'20"	1,9	3
03.4	—"—	61°37'20"	30°04'00"	43°23'10"	86°46'20"	3	
02.3	—"—	64°21'00"	16°52'00"	30°11'10"	60°22'20"	16	13
03.5	—"—	66°37'40"	42°05'00"	55°24'10"	110°48'20"	0,6	0,3
01.2	—"—	70°11'40"	5°25'20"	18°44'30"	37°29'00"	1	3
02.4	—"—	—"—	26°37'00"	39°56'10"	79°52'20"	0,6	
03.6	—"—	—"—	61°11'20"	74°30'30"	149°01'00"	2	
02.5	—"—	73°55'40"	38°31'00"	51°50'10"	103°40'20"	4	3
01.3	—"—	76°30'00"	14°29'30"	27°48'40"	55°37'20"	4	0,3
02.6	—"—	—"—	55°30'00"	68°49'10"	137°38'20"	3	1,6
01.4	—"—	79°47'30"	24°33'00"	37°52'10"	75°44'20"	0,3	
01.5	—"—	81°48'10"	36°25'00"	49°44'10"	99°28'20"	1	0,2
01.6	—"—	83°09'20"	52°37'00"	65°56'10"	131°52'20"	0,6	1,5
00.1	—"—	0°	-4°37'30"	8°41'40"	17°23'20"	20	
00.2	—"—	—"—	4°15'50"	17°35'00"	35°10'00"	100	
00.3	—"—	—"—	13°37'30"	26°56'40"	53°53'20"	11	
00.4	—"—	—"—	23°51'20"	37°10'30"	74°21'00"	45	
00.5	—"—	—"—	35°44'30"	49°03'40"	98°07'20"	4	
00.6	—"—	—"—	51°43'10"	65°02'20"	130°04'40"	24	
14.0	10°54'	0°	16°38'30"	29°55'40"	59°51'20"	16	
14.1	—"—	18°51'20"	18°06'00"	31°25'10"	62°50'20"	9	10
14.2	—"—	31°12'50"	22°21'20"	35°40'30"	71°21'00"	6	5
14.3	—"—	42°16'00"	29°03'40"	42°22'50"	84°45'40"	4,8	4,5
14.4	—"—	50°28'20"	38°16'40"	51°35'50"	103°11'40"	3,1	4,5
14.5	10°54'	56°34'10"	51°33'10"	64°52'20"	129°44'40"	3,2	2,3

Table 1(continued)

Atomic plane indices	Trace angle	Initial setting of angle sample	Inclination angle of atomic plane	Wolf-Bragge angle	Detector angle	Relative intensity	
hk.e	$x C_{hk.e}$	$ZZ_{hk.e}$	$\varphi_{hk.e}$	$\theta_{hk.e}$	$2\theta_{hk.e}$	+	−
1	2	3	4	5	6	7	8
13.0	13°54'	0°	9°47'40"	23°06'50"	46°13'40"	26	
13.1	—"—	21°03'40"	11°33'00"	24°52'10"	49°44'20"	20	24
13.2	—"—	37°36'00"	16°21'50"	29°41'00"	59°22'00"	13	6
13.3	—"—	49°07'10"	23°32'00"	36°51'10"	73°42'20"	10	12
13.4	—"—	57°00'20"	32°47'20"	46°06'30"	92°13'00"	5	8
13.5	—"—	62°33'10"	45°03'00"	58°22'10"	116°44'20"	6	8
12.0	19°06'	0°	3°24'40"	16°43'50"	33°27'40"	32	
24.0	—"—	—"—	21°50'20"	35°09'30"	70°19'00"	3	
24.1	—"—	14°42'10"	23°13'30"	36°32'40"	73°05'20"	2	5
12.1	—"—	27°41'20"	8°39'10"	18°58'20"	37°56'40"	9	4
12.2	—"—	46°23'00"	11°20'30"	24°39'40"	49°19'20"	31	22
12.3	—"—	57°34'30"	19°08'20"	32°27'30"	64°55'00"	0,7	7
12.4	—"—	64°31'30"	28°41'20"	42°00'30"	84°01'00"	5	7
12.5	—"—	69°08'10"	40°37'00"	53°56'10"	107°52'20"	0,8	1,2
12.6	—"—	72°22'50"	58°42'30"	72°01'40"	144°03'20"	5	4
23.0	23°25'	0°	15°01'20"	28°20'30"	56°41'00"	8	
23.1	—"—	17°40'00"	16°32'50"	29°52'00"	59°44'00"	20	
23.2	—"—	32°29'50"	20°54'50"	34°14'00"	68°28'00"	4,2	2,8
23.3	—"—	45°41'50"	27°41'00"	41°00'10"	82°00'20"	8	14
23.4	—"—	51°52'20"	36°53'20"	50°12'30"	100°25'00"	4	2
23.5	—"—	57°52'30"	49°50'00"	63°09'10"	126°18'20"	5	9
11.0	30°	0°	−2°23'30'	10°55'40"	21°51'20"	4	
22.0	—"—	—"—	8°51'00"	22°10'10"	44°20'20"	18	
33.0	—"—	—"—	21°09'00"	34°28'10"	68°56'20"	21	
44.0	—"—	—"—	35°39'00"	48°58'10"	97°56'20"	5	
55.0	—"—	—"—	57°15'00"	70°34'10"	141°08'20"	11	
14.1	—"—	11°19'50"	37°00'00"	50°19'10"	100°38'20"	11	
33.1	—"—	14°57'30"	22°30'00"	35°49'10"	71°38'20"	1,1	
22.1	—"—	21°50'20"	10°39'50"	23°59'00"	47°58'00"	3,6	
44.2	—"—	—"—	41°03'00"	54°22'10"	108°44'20"	5	
33.2	—"—	28°07'10"	26°35'00"	39°54'10"	79°48'20"	15	
44.3	—"—	31°00'50"	48°21'00"	61°40'10"	123°20'20"	15	
11.1	—"—	38°42'50"	0°41'00"	14°00'10"	28°00'20"	60	
22.2	—"—	—"—	15°36'00"	28°55'10"	57°50'20"	8	
33.3	—"—	—"—	33°08'00"	46°27'10"	92°54'20"	2	
44.4	—"—	—"—	61°51'00"	75°10'00"	150°20'20"	3	
33.4	—"—	46°54'10"	42°35'10"	55°54'20"	111°48'40"	9	
22.3	—"—	50°15'00"	22°48'00"	36°07'10"	72°14'20"	2	
33.5	—"—	53°11'00"	57°23'00"	70°42'10"	141°24'20"	2	
11.2	—"—	58°02'40"	7°33'00"	20°52'10"	41°44'20"	7	
22.4	—"—	—"—	32°06'30"	45°25'40"	90°51'20"	2	
22.5	—"—	63°28'50"	44°18'00"	57°37'10"	115°14'20"	2,5	
11.3	—"—	67°25'10"	16°06'00"	29°25'10"	58°50'20"	11	
11.4	—"—	72°40'40"	26°00'00"	39°19'10"	78°38'20"	2,5	
11.5	—"—	75°59'30"	37°49'00"	51°08'10"	102°16'20"	7	
11.6	30°	78°15'20"	54°30'00"	67°49'10"	135°38'20"	2,5	

Conclusion

Langasite crystals may have a wide application in television, communication technique, for example, in synchrosignal sources, discriminators, remote control systems and other devices and also in sensors of physical quantities. From the data given it follows that these products have advantageous intermediate characteristics between those of the quartz crystals units and lithium tantalate resonators. We continue investigations of langasite resonators aimed at determination of their precision properties in the range of 3-5 MHz, HF resonators and sensors quantities of physical factorium.

Table 2. Characteristics of langasite resonators.

Frequency, MHz	Resonanse spacing, %	Frequency change over the range 10-+60°C, 10^{-6}	Eqivalent resistance Ohm	Quality factor 10^3
5-30	0.6-1.2	200-300	15-20	10-15

Table 3. Characteristics of langasite filters

Frequency	Passband		Ripple within the passband	Guaranteed attenuation		Attenuation within the passband	Volume
MHz	dB	kHz	dB	dB	kHz	dB	cm^3
5-18	3.0	±(22-50)	1,5-2,0	80	35-125	2.0	1.2-1.3

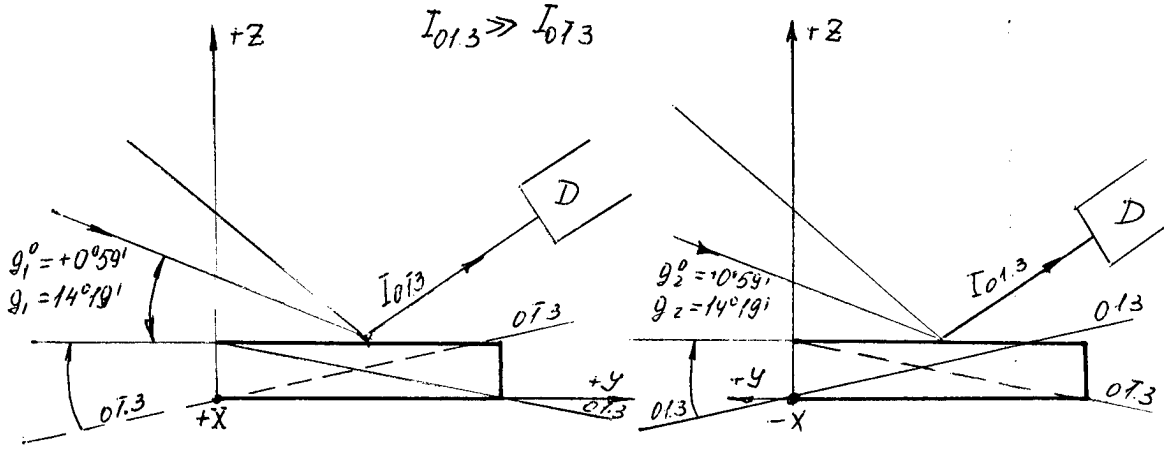


Figure 2. X-ray control of the initial orientation of langasite crystal 00.1-cut



Figure 3. "As grown" langasite boules

Reference

[1] E. Philippot, A. Ibanez et al. "New approach of crystal growth and characterization of a quartz and berlinite isomorph: $GaPO_4$. Proc.IEEE Frequency Control Symposium,U.S A., p.744,1992.

[2] B. V. Mil, A. B. Batachin, et al. "Modified gallates with the structure $Ca_3Ga_2Ge_4O_{14}$," Report of the USSR Academy of Science, vol. 271, no.3, p. 768-771, 1982.

[3] A. A. Kaminsky, B. V. Mil, I. M. Silvestrova, et. al."Acoustic and electric properties of new neorganic material," Al-Union Acoustical Conference, Moscow, 1983, p.20.

[4] V. N. Stassevich "Single crystals technology," Moscow, Radio and Communication, 1990.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PRESENT STAGE OF $La_3Ga_5SiO_{14}$- RESEARCH

I.M.Silvestrova, V.V.Bezdelkin*, P.A.Senyushenkov, Yu.V. Pisarevsky

Institute of Crystallography of the Russian Academy of Science 117333,Moscow, Leninsky pr.59, Russia

*Scientific Research Institute "Fonon", 105023,Moscow,Krasnobogatyrskaya str.44, Russia

Abstract

Refined temperature measurements of elastic end dielectric properties of langasite crystals were made. Based on the data received, the 1st-and the 2nd- order temperature coefficients of frequency and were the electromechanical coupling coefficient were calculated.

Description

As it has been shown in the previous papers [1-3], $La_3Ga_5SiO_{14}$ crystals exhibit strong piezoelectric effect,have thermally stable cuts and low acoustic losses. These crystals can be grown by Czochralski method under relatively low temperatures (their melting point is 1470°C), where the use of a platinum crucible instead of a more expensive iridium one is still possible. These crystals have no phase transitions. They are not water soluble ones.

At present,apart from the research laboratories, $La_3Ga_5SiO_{14}$ is grown in Russia at an industrial scale and is used for manufacturing resonators, filters and other analogous devices.Moreover these crystals are rather promising for the use as SAW devices substrats. These crystals have higher homogeneity than thosed described earlier [1,2]. In the samples with the diameter of 50 mm and the length of 100 mm,no periodic layers have been observed, described in [1,2] and the spread in the sound velocity from region to region was less than 0,3% (in paper [1] it is egual to 0.7-3%).

The grown up crystals slightly differed from the stoichiometric composition of langasite: La_2O_3 - 48%; Ga_2O_3 - 46.1%; SiO_2- 5.9%. In our case it was received: La_2O_3 - 48.6%; Ga_2O_3- 48.9; SiO_2- 3.10%.

We have measured dielectric susceptibilities and ultrasonic velocities over the temperature range from -195°C to +100°C in these crystals as well.

Ultrasonic measurements have been made by phase - pulse method at the frequency of 30 MHz. The relative accuracy of measurements obtained was not worse than 0.03%.

As a result of these measurements corrections for the thermal expansion coefficient have been inserted.

Table 1 lists the data obtained at the room temperature for sound velocities and their 1st - and 2nd - order temperature coefficients together with the data taken from [2].

It is necessary to note here the presence of different signs for the 1st order temperature coefficients for different modes and the low TCV-value for the most piezoelectrucally reactive slow transversal wave propagating along the Y-axis.

Based on these data, the elastic stiffness moduli c_{ik} and the elastic compliance moduli s_{ik} were calculated over

0-7803-0905-7/93 $3.00 © 1993 IEEE

Table 1.

Elastic wave velocities V ($10^3 ms^{-1}$), of langasite crystals and their 1st-order TV(10^{-6} K^{-1}) and 2nd-order T2V($10^{-9}K^{-1}$) at 20°C.

Propagation direction	Mode	V		TV	T2V
		1	*	*	*
X	L	5.784	5.746	-24.9	-24.7
X	FS	3.344	3.305	-53.2	-39.1
X	SS	2.344	2,394	+49.8	-131.0
Y	L	5.792	5.756	-31.0	-42.0
Y	FS	2.971	3.003	-18.8	-57.5
Y	SS	2.767	2.750	+6.6	-89.0
Z	L	6.753	6.826	-50.0	-29.2
Z	S	3.021	3.043	-26.5	-64.9
+45 YZ	L	5.998	5.949	-35.1	-31.2
+45 YZ	FS	3.141	3.281	-52.2	-87.0

In this table L means longitudinal mode, FS- fast shear mode, SS slow shear mode, * data received as a result of this work. the specified temperature range. The same applies to the 1st- and the 2nd- order temperature coefficients.

These data are given in table 2 for the room temperature, the other coustants are also included in this table.

It is necessary to note, that the greatest difference compared to the previously published results was observed for temperature coefficients Tc_{ik} Although for the purpose of this work no exact analysis of composition was made, we cousider, that the difference is connected with the change in the crystal stoichiometry.

The differenoe in the signs of temperature coefficients $T\varepsilon_{11}$ and $T\varepsilon_{33}$ should also be noted, and i.f. the most important for resonators and filters $T\varepsilon_{11}$ is comparably small, ε_{11} increases

Table 2.

Elastic stiffnesses c^E_{ik}(10^{10}N m^{-2}), elastic compliances s^E_{ik} ($10^{-13}m^2$ N^{-1}), dielectric permittivity piezoelectric strain d_{ik} (10^{-12}c N^{-1}) and stress e_{ik} (C m^{-2}) constants, thermal expantion coefficients α and density ρ(g sm^{-3}) of langasite crystals.

	c_{ik}			Tc_{ik}		$T2c_{ik}$
	1	2	*	2	*	*
c_{11}	19.09	19.02	19.01	-47	-52.7	-55.4
c_{12}	10.63	10.63	10.30	-100	-107.4	+48.1
c_{13}	10.42	9.19	9.524	-130	-14.4	+156.8
c_{14}	1.52	1.47	1.57	-370	-176	-10.4
c_{33}	26.19	26.21	26.83	-94	-105	-62.8
c_{44}	5.24	5.382	5.331	-30	-57.6	-133
c_{66}	4.32	4.20	4.35	+36	+12	-178
	ε_{ik}			$T\varepsilon_{IK}$		
ε_{11}	18.99	18.86	18.89	153		
ε_{33}	49.3	48.7	49.0	-760		
	α_{IK}			e_{IK}		
α_{11}	5.11					
α_{33}	3.61					
e_{11}				-0.45	-0.41	
e_{14}				0.07	0.08	
d_{11}				-6.16		
d_{14}				+5.36		
ρ				5.743		

abryptly with temperature decrease. This is, however, linear and has the tendency to flattening at low temperatures.

From the above-mentioned it follows that there exist cuts in $La_3Ga_5SiO_{14}$ with zero TCε, which can find their application in dielectric devices, operating over a wide temperature range, for example in UHF - resonators.

The above data enabled to make the complete theoreti cal analysis of thickness vibration parameters in $La_3Ga_5SiO_{14}$ - resonators parameters depending on the cuorientation xybl/γ/β for compression extensionalmodes (mode A) of the rapid and slow thickness shears (modes B and C, respectively). From the results of calculation obtained it follows that the mode A and the mode B have no directions with the zero, 1st- order temperature coefficient of frequency (TCF), i.e. they do not meet the requirement to the main criterion for resonator cut selection.Mode C of the slow thickness shear is the mostinteresting one from the point of view of the whole complex of temperature characteristics and piezoelectric properties.

Stereographical projetions of the level of electromechanical coupling coefficient (K) lines (figure (a)) and those of the 1st- order (figure (b)) and of the 2nd- order (figure (c)) temperature coefficients of frequency for this mode of vibration are given below.

From the diagramme (b) it is seen that langasite crystal has a zero 1st- order TCF.The existence of a zero, 1st- order TCF is extremely important with small rotation angle around the Y-axis. As one can see from the diagramme shown in figure (a), this enables to have K = 15% with zero TCF. It is extremely important that for this cut the electromechanical coupling coefficients of the A- and B- modes are nearly equal to zero, i.e. this cut is a single mode one. Ufortunately, as it follows from the diagramme (b) the 2nd- order TCF is not yet discovered.

Data are, however, available suggesting that with certain composition change an abrupt lowering of the 2nd- order TCF is possible that allows to expect finding a composition with the cut analogous in TCF to the AT-cut of quartz.

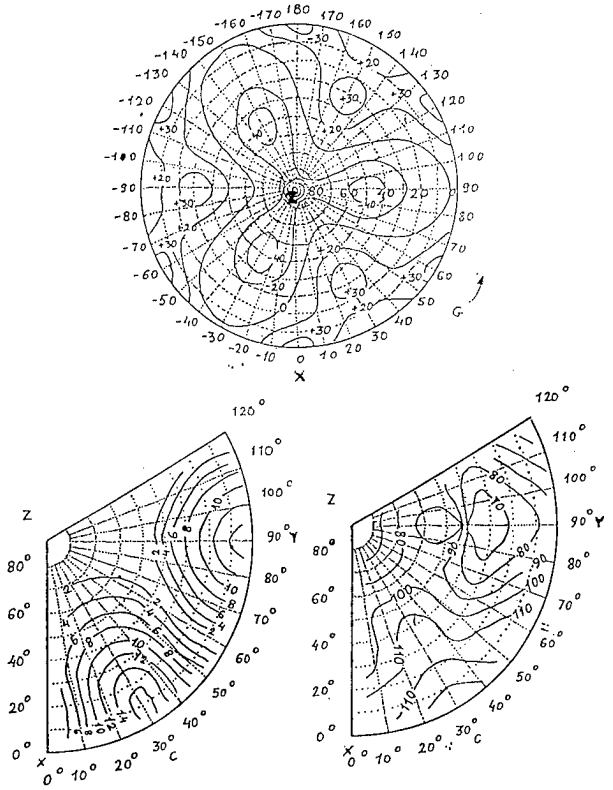


Fig.1. Dependencies of K on the angle of cut (a) of the 1st order (b) and the 2nd order (c) FTCs for the mode C of thickness vibrations in $La_3Ga_5SiO_{14}$.

Aknowledgement
The autors thank Mrs E.Antonova for the help in making calculation.

[1].B.V.Mill,I.M.Silvestrova, A.A.Kaminsky Yu.V.Pisarevsky. Proc.X All Union 7 acoustic conference 1983,p.20-29. Piezoelectrical,dielectrical,elastical and acoustical properties of $La_3Ga_5SiO_{14}$.

[2].I.M.Silvestrova ,Yu.V.Pisarevsky , P.A.Senyshenkov ,A.I.Krupny . Sov. Phys. Solid. Stat 1986, v.28, 9, p.2875-2878. Temperature dependence elastic properties $La_3Ga_5SiO_{14}$.

[3].E.G.Bronnikova,I.M.Larionov,V.A.Isaev, I.M.Silvestrova , Yu.V.Pisarevsky. Souv. Journ. Electronnay technica 2(63),1986,p.83-85. Monolitic filters and resonators based on new piezoelectric lantane gallosilicate.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

NEW PIEZOELECTRIC MATERIALS

I.M.Silvestrova, Yu.V.Pisarevsky, V.V.Bezdelkin*, P.A.Senyushenkov

Institute of Crystallography of the Russian Academy of Science
117333, Moscow, Leninsky pr.59, Fax 330 1956
*
Scientific Research Institute "Phonon",
105023, Moscov, Krasnobogatyrskaya 44, Russia

The Institute of Crystallography of the Russian Academy of Science(RAN) in co-operation with other organizations the investigation projects were implemented aimed at the search, growing and investigation of new materials for different electronic applications: resonators, filters, UHF- acoustic high temperature receivers, temperature, pressure acceleration etc. sensors.

For realization of these project different crystals were grown by the following methods:
1)hydrothermal;
2)Czochralski;
2)from solution in the melt;
4)from solutions.

The grown crystals were investigated by resonance and ultrasonic methods for obtaining information on elastic and piezoelectric properties.

Moreover, dielectric, acoustooptic and UHF-techniques were used for the acoustic loss and dielectric constant measurements.

The perfectness of grown rystals was characterized by different X-ray and optical methods.

The first stage after crystal growth supposed obtaining a sufficently complete information on material constants at the room temperature.

During the last years the following crystals have been investigated:

1)trigonal ones:

$AlPO_4$, $GaPO_4$, $Ga_2Ca_3Ge_4O_{14}$, $Ca_3Ga_5SiO_{14}$, $La_3Ga_5GeO_{14}$, $La_3Ga_{5.5}NbO_{14}$, $Nd_3Ga_5SiO_{14}$, $Sr_2Ga_2Ge_4O_{14}$;

2)tetragonal ones:

$Li_2B_4O_7$, $TeSeO_2$;

3)hexagonal ones:

ZnO, ZnS-MgS;

4)rhombic:

$KTiOPO_4$, $RbTiOPO_4$,

ortho-sulphur-benzoate of cesium;
5)Monoclinic: L-arginine phosphate.

As an example we describe the recently obtained results on L-arginine phosphate.

Table 1.

Elastic moduli $c_{ik}(10^{10}\ N\ m^{-2})$, $s_{ik}(10^{-1}m^2N^{-1})$, piezoelectric moduli $e_{ik}(Kl\ m^{-2})$, $d_{ik}(Kl\ N^{-1})$, and $\varepsilon_{ik}\cdot\varepsilon_o^{-1}$, electromechanical coupling coefficients K_{ik} (%) of L-arginine phosphate.

C_{11}	1.703	S_{11}	10.97	d_{21}	4.585	K_{21}	6.5
C_{22}	3.185	S_{22}	6.46	d_{22}	5.09	K_{22}	9.5
C_{33}	2.971	S_{33}	5.89	d_{23}	3.22	K_{23}	5.8
C_{44}	0.737	S_{44}	13.57	d_{25}	8.60	K_{25}	6.7
C_{55}	0.611	S_{55}	24.32	d_{36}	5.77	K_{36}	6.7
C_{66}	0.981	S_{66}	10.20	d_{14}	7.20	K_{14}	6.8
C_{12}	1.295	S_{12}	-2.98	e_{16}	0.13	K_{16}	22,5
C_{13}	1.390	S_{13}	-3.78	e_{22}	0.11	K_t	9.3
C_{23}	1.585	S_{23}	-1.16	e_{34}	0.02	K_{34}	4.6
C_{15}	0.271	S_{15}	1.63				
C_{25}	0.788	S_{25}	-6.19	ε_{11}^T	4.82	ε_{33}^T	4.37
C_{35}	0.430	S_{35}	-0.98	ε_{22}^T	5.14	ε_{13}^T	-0.38
C_{46}	0.012	S_{44}	-0.17	ρ	1.53		

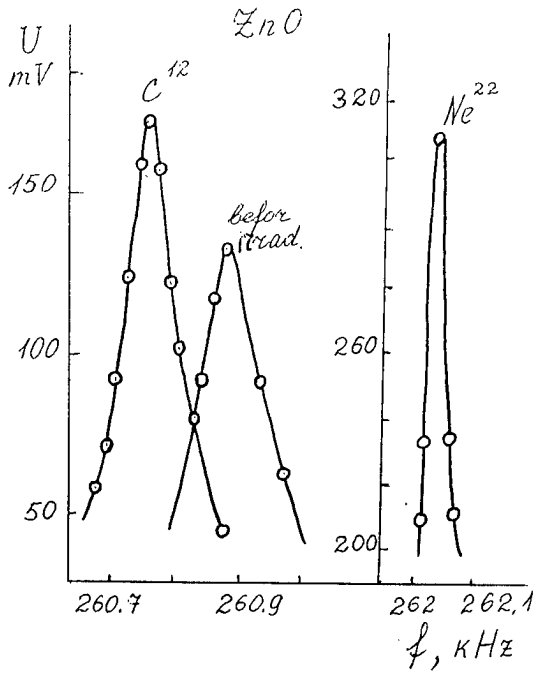
From the table it follows that this crystal has an high piezoelectric activity for shear wave along the X-axis.

0-7803-0905-7/93 $3.00 © 1993 IEEE

This can be useful for excitation of acoustic vibrations in the low impedance medium (Z=3.88 g/cm^{2}s),for example, for use in medical diagnosis.The direction of further investigations depends on the recommended application area.

For example, the ZnO crystal sensor showed a favourable behaviour when used in high temperature ultrasonic defectoscopy at atomic power stations.

In this connection we have a joint work with the Laboratory of Nuclear Reactions (Dubna-city,Russia) and study



Frequency response near main resonance of the ZnO bar;a)before irradiation;b)after irradiation by C^{12}ions(90 Mev,10^{14}ions/sm^2);c)after irradiaton by Ne^{22}ions(160Mev, 10^{15} inns/sm^2

the influence of radiation on piezoelectric and other properties of crystals. It is shown in particular that the Ne-ion radiation improves the acoustic quality of ZnO crystal by several times.

Table 2.
Piezoelectric characteristic of promising materials
$d_{ik}(10^{-12}$ c $N^{-1})$, k (%), T=20°C)

	SiO_2	$La_3Ga_5SiO_{14}$	$Nd_3Ga_5SiO_{14}$	$Ca_3Ga_5Ge_4O_{14}$	$Sr_3Ga_2Ge_4O_{14}$
k_{26}	14	13	7.5	12.5	26.0
k_t	9	8	7.2	10.2	13.0
k_{12}	10	16	10	14.5	26.6
d_{11}	-2.3	-6.16	-4.1	-5.38	-9.4
d_{14}	+0.67	+5.36	+3.2	+2.5	+6.9
$\varepsilon^T_{11}/\varepsilon_0$	4.5	18.99	18.5	15.6	13.8
$\varepsilon^T_{33}/\varepsilon_0$	4.6	49.3	67.0	24.2	18.2
α_{11}	13.5	5.11			
α_{33}	7.3	3.16			

Table 3
Acoustic characteristic of promising materials
V (3.10^5 sm s^{-1}), B (dB cm be at 500 MHz),T=20°C

Compound	Shear wave ,Y dir. Velosity	Attenuation	Logitudinal wave,X dir. Velosity	Attenuation
SiO_2	3.91	1.33	5.75	1.37
$La_3Ga_5SiO_{14}$	2.76	0.44	5.78	0.26
$Nd_3Ga_5SiO_{14}$	2.59	1.01	5.65	0.36
$Ca_3Ga_2Ge_4O_{14}$	2.79	1.09	5.85	0.35
$Sr_3Ga_2Ge_4O_{14}$	2.75	0.77	5.58	0.45
$Y_3Al_5O_{12}$	5.02	0.55	8.56	0.26

The materials with the structure of $Ca_3Ga_2Ge_4O_{14}$ are of great interest.

The best known its representatif is $La_3Ga_5SiO_{14}$ wich will be described in a separate paper.We shall only note here that over 40 compositions had been syntetisized in this group. Eight compositions of them have been grown as large crystals. The results of more or less detailed investigations show considerabe variations of dielectric, piezoelectric and elastic properties of these crystals (tables 2,3) which enables one to choose a material with necessary properties, for piezoelectric engineering for example.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PROPERTIES,PRODUCTION AND APPLICATION OF NEW PIEZOELECTRIC CRYSTAL LITHIUM TETRABORATE $Li_2B_4O_7$

Shi-Ji Fan

Shanghai Institute of Ceramics, Chinese Academy of Sciences
1295 Ding Xi Road,Shanghai 200050,P.R.China

Piezoelectric, elastic, dielectric, thermal, optical, mechanical, chemical, surface acoustic wave(SAW) and bulk acoustic wave(BAW) coupling and propagation properties etc of new piezoelectric crystal lithium tetraborate $Li_2B_4O_7$ (abb.LBO), the growth and production of LBO crystal by the modified Bridgman technique (Fan's process) and its application for SAW and BAW devices have been reviewed in detail in this paper. The features of LBO growth as (1) the critical requirement of stoichiometric purity for raw material, (2) very low crystal growth rate, (3) a convex or plane solid-liquid interface in the growth, (4) a suitable procedure of self-annealing, and the practicality of the BR technique for commercialization of LBO material in large diameter of 82mm and high quality are discussed also.

1. Introduction

In more than ten years, people have been looking for a new piezoelectric material possessing both high electromechanical coupling coefficient and low temperature coefficient of frequency (or delay) in satisfaction to the requirements of surface acoustic wave (SAW) and bulk acoustic wave (BAW) devices. Since R.W.Whatmore et al. discovered that lithium tetraborate crystal is a new temperature-compensated SAW substrate material [1-3], a lot of works on material preparation [4-6], charaterization of properties [5,7-12] and device fabrication [13-19] of LBO crystal have been carried out. T.Shiosaki et al. investigated elastic and piezoelectric etc properties of LBO in detail. Y.Ebata, K.Miwa et al. developed the fabrication of LBO SAW devices into production. But, the Czochralski (CZ) technique for LBO crystal growth met two serious problems: coring and crack of the crystal, specially for large diameter of growth so that the difficulty of preparation of LBO material hindered its industrial application. S-J Fan et al. have grown the large diameter (82mm), core-free, twin-free, no crack, no striation and no scattering particle of LBO crystal successfully and stably by the modified Bridgman (BR) method (Fan's process) for the first time [20-24] and developed this process into industrial technique to realize commercialization of the new piezoelectric material.

Piezoelectric, elastic, dielectric, thermal, optical, chemical, SAW and BAW coupling and propagation, electro-optic and acousto-optic properties of LBO crystal and the comparison between the properties of the CZ-grown and BR-grown LBO crystals; the features of BR growth of LBO crystal and the comparison between the BR and CZ growth methods; application of LBO for high frequency, middle bandwidth, low insertion loss, high stability of frequency and miniaturization of SAW and BAW devices have been reviewed in detail in this paper.

2. Properties of LBO crystal

2-1. Crystallographic data of LBO

$Li_2B_4O_7$ crystal belongs to tetragonal system, point group 4mm, space group $I4_1cd$. The constants of lattice a=9.479Å, c=10.280Å. Two layers of net consisting of $B_4O_7^{-2}$ by three- and four-coordinate of O ions to B ions are formed in its structure which projection along <c> axis is shown in Fig. 1 [25].

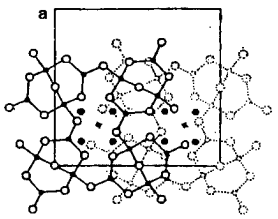


Fig. 1. The projection of $Li_2B_4O_7$ crystal structure along <c> axis.(from J.Krogh-Moe)
○ Oxygen, ● Boron, × Lithium

The density of LBO (2.45 g/cm³) is much lower than that of $LiNbO_3$ (LN:4.64g/cm³) and $LiTaO_3$ (LT:7.45g/cm³), which enhances the yield of Φ3" wafer per Kg of the boules about 2 and 3 times higher than that of LN and LT, respectively.

0-7803-0905-7/93 $3.00 © 1993 IEEE

2-2. Piezoelectric, elastic and dielectric properties.

N.M.Shorrocks et al. and T.Shiosaki et al. measured totally piezoelectric, elastic and dielectric constants of the CZ-grown LBO crystals and we measured all of these constants of the BR-grown LBO crystals shown in Table 1 for comparison. From Table 1, there is no difference in the properties between the CZ- and BR-grown LBO crystals. The errors of the data may be from the practical measurements.

Table 1. Piezoelectric, elastic and dielectric properties of the CZ-grown and BR-grown LBO single crystals

Growth method		CZ		BR
Author		Shorrocks [2]	Shiosaki [9,10]	S-J Fan
Piezoelectric strain const. ($\times 10^{-12}CN^{-1}$)	d_{15}		8.07	8.19
	d_{31}		-2.58	-2.63
	d_{33}		19.4	19.5
Piezoelectric stress const. (Cm^{-2})	e_{15}	0.36	0.472	0.46
	e_{31}	0.19	0.290	0.23
	e_{33}	0.89	0.928	0.91
Elastic stiffness const. ($\times 10^{10}Nm^{-2}$)	C^E_{11}	12.67	13.5	12.84
	C^E_{12}	0.5	0.357	0.46
	C^E_{13}	3.0	3.35	2.97
	C^E_{33}	5.39	5.68	5.42
	C^E_{44}	5.50	5.85	5.78
	C^E_{66}	4.60	4.67	4.75
Elastic compliance const. ($\times 10^{-12}m^2N^{-1}$)	S^E_{11}		8.81	8.82
	S^E_{12}		1.23	1.0
	S^E_{13}		-5.92	-5.5
	S^E_{33}		24.6	24.3
	S^E_{44}		17.1	17.3
	S^E_{66}		21.4	21.1
Dielectric const.	$\epsilon^T_{11}/\epsilon_0$		9.33	9.33
	$\epsilon^T_{33}/\epsilon_0$		9.93	10.1
	$\epsilon^S_{11}/\epsilon_0$	8.97	8.90	8.91
	$\epsilon^S_{33}/\epsilon_0$	8.15	8.07	8.07

T.Shiosaki's group measured also the first and the second normalized temperature coefficients of the above constants [9]. The dielectric constants of LBO is lower than that of LN and LT, which is designable for high frequency SAW devices [5].

2-3. Thermal properties

The congruently melting point of LBO (917°C) is much lower than that of LN (1250°C) and LT (1650°C), there is no phase transition from melting point to room temperature, which indicates LBO crystal growth from melt to be practical. Several incommensurate phases of LBO were discovered at low temperature down to 4.2K which cause its low-temperature elastic anomalies and discontinuity of frequency temperature characteristics and large acoustic loss [8,26,27].

The specific heat C_p of LBO increases from 1.22 to 1.74 $J \cdot g^{-1}K^{-1}$ with enhancing temperature from 100°C to 650°C and the averages of thermal conductivities $\alpha_{x,y}$ 0.0248 and α_z 0.019 $J \cdot cm^{-1}K^{-1}s^{-1}$ are in temperature range 625-975 K [28].

The thermal expansion along <a> axis increases monotonously with increasing temperature while it along <c> axis shows a turnover around 105-120°C [5-9]. The average of thermal expansion coefficients α_{11} 11×10^{-6}/°C, α_{33} -3.7×10^{-6}/°C in the temperature range 20-35°C were also measured [9].

2-4. Optical properties

LBO crystal belongs optically to uniaxial and negative crystal. n_o and n_e are 1.6232 and 1.5649 respectively, and decrease with increasing of λ. The transmission of LBO along <a> axial is shown in Fig. 2. LBO crystal is almost totally transparent in visible region and has good transmission (up to 75%) in ultraviolet region, which makes it possible to use in optics application.

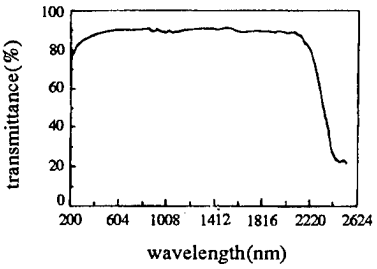


Fig. 2. Transmission of LBO crystal along <a> axis.

2-5. Mechanical properties.

The hardness of LBO is in Mohs 6-7 which changes with different crystallographic face and its compressive strength and bending strength are shown in Table 2.

Table 2. Mechanical properties of LBO crystal

Compressive strength Kg/cm²	∥ <a>	2878
	∥ <c>	5750
	∥ <110>	4496
Bending strength Kg/cm²	⊥ <a>	1946
	⊥ <c>	950
	⊥ <110>	1886

2-6. Chemical stability

LBO crystal dissolves easily in acids, slowly in hot water and alkali solutions and does not dissolve in organic solvents. Because etching of LBO wafer is related directly to the process of device fabrication, H.Abe et al. investigated the solubility of LBO in various solvents (Fig. 3) and pointed out that the dissolution rate can be suppressed below 10 Å/min using alkaline solution of pH=11 as the rinse solution in lift-off process [17]. The organic solvents such as acetone, xylene and so on are used for preparation of LBO wafers and in photolithography.

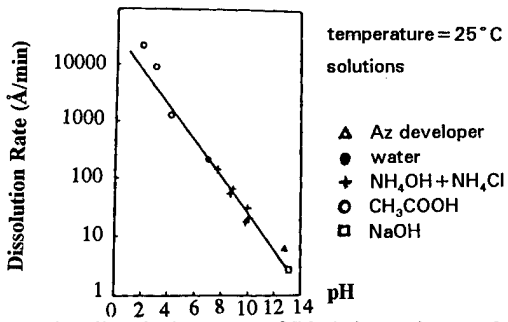


Fig. 3. The dissolution rate of LBO in various solvents with different pH. (From H.Abe[17])

2-7. Bulk and surface acoustic wave properties.

Bulk wave coupling coefficients K_{15},K_{31} and K_{33} were measured as 0.21, 0.094 and 0.42 respectively. The temperature compensated cuts of 32° and 53° rotated Y-cut LBO possess v_L 7600 m/s, K_T 0.18 for the quasi-longitudinal mode and v_s 3400 m/s, K_s 0.27 for the quasi-shear mode, respectively [10]. It was also pointed out that 56°40' X_2 LBO substrate has zero first order temperature coefficient of frequency in thickness shear mode in room temperature and high electromechanical coupling coefficient (K=0.281) as well as small ratio capacity (C_0/C_e=20) [8].

The calculated and measured properties of doubly-rotated LBO resonator and transducers are presented by A.Ballato et al.[12]. The cuts of (yxwl)40°/33° and (yxwl)19°/56° with first- and second-order zero temperature coefficients of frequency were expected for applications in frequency stability with temperature and moderate piezocoupling value. However the predicted values of the second-order temperature coefficients of frequency are less well in agreement with the measured values.

The principal parameters of SAW are SAW velocity (v_s), coupling coefficient (k^2) and temperature coefficient of delay (TCD). They have been calculated by computer over the full substrate cut angles and propagation directions [9]. For Rayleigh SAW properties, the results are v_s:3100--4300 m/s; k^2:0--1.6%; TCD: -20--+110 ppm/°C. Among the orientations with zero TCD, a substrate of (110) cut-[001] propagation (45°X-Z) of LBO is very important and practicable for SAW devices, its v_s 3401 m/s, k^2 0.8%. Around 45° of λ, v_s, k^2 and TCD have little λ angular dependence. The angular dependence of all of the physical properties is also little dependent on all of the Eulerian angles λ,μ and θ [31]. Leaky SAW properties of LBO had also been calculated in detail [10,11]. Its maximum k^2 and minimum TCD are 7.78% and 0 ppm/°C respectively. For (45°, 90°, 67°) cut v=3668.7m/s,TCD=19.2ppm/°C,k^2=4.5%,propagation loss=0.69dB/λ, power flow angle (PFA)=2.47deg.. It is worth to note the propagation loss is equal to zero dB/λ with v 4806.8 m/s, k^2 0.082%, TCD 47.8 ppm/°C and PFA -0.1 deg. for (0°, 90°, 0°) cut.

In comparison with quartz, $LiNbO_3$ and $LiTaO_3$, their two important parameters k^2 and TCD are shown in Fig.4, from which it is obvious that LBO substrate has fine and comprehensive performance for BAW and SAW devices.

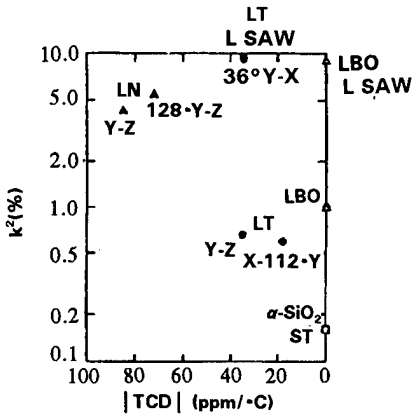


Fig.4. The comparison in k^2 and TCD among four practical piezoelectric crystals.

2-8. Electro-optic and acousto-optic properties

Electro-optic (EO) coefficients of LBO,r_{13} 2.64×10^{-12}, r_{33} 2.55×10^{-12} and r_{51} -1.24×10^{-12} mv^{-1}, are small [29]. There is difference between the data of ours and L.Bohaty's [30]. The acousto-optic (AO) figures of merit M_{13} 0.23×10^{-18},M_{33} 0.28×10^{-18} s^3g^{-1} are also small [8] so it is difficult to use LBO for EO and AO devices.

3. Growth and production of LBO crystal

It is well known that although Φ3" LBO crystal can be grown by CZ method through about ten years of efforts, the CZ method has some difficulties to develop into industrial technique for LBO crystal growth up to date. The intrinsic factors of LBO melt (high viscosity) and LBO crystal (low thermal conductivity and strong anisotropy of thermal expansion) and the characteristics of large temperature gradient and open growth system in CZ method make CZ technique to be difficult for LBO growth. Macrodefects such as core region, crack and striation and so on occur often in CZ-grown LBO crystals. The research of the modified Bridgman method (Fan's process) for LBO started in 1987 by S-J Fan et al. The outline of LBO growth by the BR method is shown in Fig. 5 and the procedure of the growth is as follows: preparing and blocking of LBO raw material--orientating and cutting the seed--processing Pt crucible--assembling of the seed and the blocks of the raw material in Pt crucible--growth in the Bridgman furnace--self-annealing in the furnace--taking the crystal out of the furnace and the crucible. The features of the BR growth of LBO are described below.

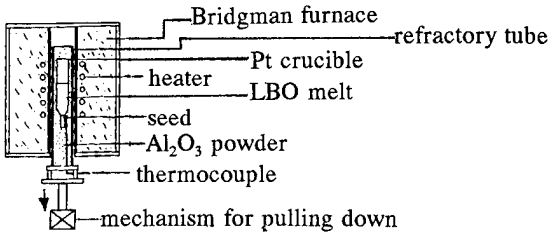


Fig. 5. Scheme of Bridgman growth of LBO

3-1. The critical requirement of stoichiometric purity for LBO raw material.

For LBO crystal grown from its high viscous melt, the stoichiometry of the raw material is an intrinsic factor critically. The raw material for LBO growth is synthesized with 3N purity of H_3BO_3 and Li_2CO_3 or LiOH in accordance to the stoichiometry of LBO. The results indicated impurities in the raw material in Table 3 do not influence obviously the growth and properties of LBO. But it is very important to control the stoichiometry of raw material exactly. The calculated ratio of Li_2O and B_2O_3 by weight in stoichiometry is 0.2146 while the range of 0.2146-0.2125 is acceptable for the growth. Rich Li component in the melt is more harmful to the growth than rich B component in the melt, which was discovered from rich Li component of core region and inclusions in the crystal. A slight deviation from stoichiometry that even if did not be tested sometimes by the conventional chemical analysis can also cause constitutional supercooling inclusions (striations, scattering particles) under a slight temperature vibration.

Table 3. The content of impurities in LBO raw material

element	Ca^{2+}	Al^{3+}	Ba^{2+}	Fe^{3+}	Mg^{2+}	Zn^{2+}	Pb^{2+}
Content(ppm)	11.5	10.3	2.8	1.3	0.5	0.4	0.3

3-2. Very low rate of the crystal growth.

Crystal growth from congruent stoichiometric melt by Bridgman method is, in fact, a phase transition process under exacting conditions. After melting the raw material and the top of seed of LBO, a stable solid-liquid (s-l) interface is kept. When the growth system is pulled down to low temperature region in a certain rate the crystal is grown simultaneously from the seed. Very low growth rate is essential for LBO due to high viscosity of its melt. The mass transfer in high viscous melt is very low, which brings an exact requirement for stoichiometry of melt composition. The BR growth rate of transparent LBO with core-free, no striation and no scattering particle is normally less than 0.3 mm/hr because there is no forced convection in the melt during growth and a small temperature gradient 10-20°C/cm. The crystals in arbitrary orientations are grown by the BR method and there is no difference for their growth rates, which depend only on the rate of pulling down of growth system.

3-3. A convex or plane s-l interface in the growth

In LBO growth a continuous or intermittent opaque region parallel to the growth direction at the central core of the crystal often occurs, that is called "coring". I explained the mechanism of coring that is essentially dendritic growth with growing rifts and constitutional supercooling inclusions at a concave s-l interface [20]. No coring occurs in LBO growth at convex or plane s-l interface so that the control of shape of s-l interface is very important for core-free growth of LBO.

For large diameter LBO growth the concave s-l interface is easily formed because of the large volume of the melt in overheating and heat flow into the seed. The shape of s-l interface must be controlled by adjustment of position of s-l interface and temperature gradient in furnace etc. to ensure growing at convex or plane s-l interface.

3-4. Self-annealing to eliminate cracking by thermal stress

Cracking by thermal stress after LBO growth frequently occurs during cooling down of the crystal because of the strong anisotropy of thermal expansion of LBO. The crystals of different growing orientations show the different behaviours of cracking. The easiness of cracking is in order of <100>, <110> and then <001> growth directions of the crystals. If the residual melt on the top of an as-grown crystal changed into polycrystalline phase by fast cooling crack occurs in the top region of the crystal. A suitable procedure of self-annealing to eliminate crack is that the crystal after growth is kept in the region of small gradient in the furnace for annealing and then cooled down slowly at 20-30 °C/hr.

Because of the above four features of LBO growth,it is inevitable that the BR method can bring its advantages into full play while the CZ method was in an inefficient state.

3-5. The production of three-inch diameter of LBO crystal

According to the requirements of SAW device production line and process of LBO wafer, more than three-inch diameter, two- to four-inch length LBO crystals with core-free, twin-free, no striation, no crack and no scattering particle have been grown stably by the modified Bridgman technique (Fan's process) (Fig.6).

We optimized the growth conditions as furnace temperature 950-1100°C, temperature gradient 10-20 °C/cm at s-l interface and growth rate <0.3 mm/hr in the arbitrary shapes of as-grown crystals and orientations of the seeds. The experimental results demonstrated the practicality of the BR technique for commercial production of LBO crystal material.

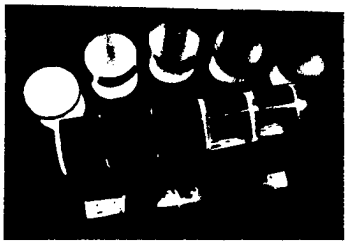


Fig.6.82mm diameter of high quality of BR-LBO boules and its wafers

3-6. The process of Φ3" LBO wafer

Φ3" LBO crystal boule was cut, lapped and polished by the standard procedure without difficulty but it should be reminded that any acids must be avoided to contact the wafer in the procedure. The average roughness of the polished surface of LBO wafer is 11Å and the bow of Φ3" LBO wafer is normally in 20μ without etching process which is better than that of CZ-LBO wafer with etching process [14].

3-7. The comparison between the BR and the CZ methods of LBO growth

For LBO crystal the growth rate of 0.2-1 mm/hr in the CZ growth is faster than that in the BR growth. But the growth rate of LBO can not be fast intrinsically because of very high viscosity of its melt so that this advantage of the CZ technique is limited. On the other hand, the high temperature gradient is often employed in the CZ furnace which is unsuitable for LBO growth because the strong anisotropy of thermal expansion of LBO requires a small gradient for eliminating crack while a small gradient in the BR furnace is beneficial to elimination of crack. The open system in the CZ growth can cause evaporation of the components in LBO melt, which degrades the stoichiometry of LBO melt while the sealed system in the BR growth can maintain the stoichiometry of the melt during long period of growth. In addition, the temperature field in the BR system is very stable and the growing shapes and orientations of LBO can be chosen arbitrarily in the BR technique. It is obvious that the modified BR technique is superior to the CZ technique for LBO growth.

The abundant and inexpensive raw materials, low congruently melting point for saving electric power, high yield of the BR technique, no contamination for the environment in LBO industrial production are also beneficial to the commercialization of the new material.

4. Application of LBO substrate for SAW and BAW devices

On the basis of LBO properties, LBO substrate is suitable for fabrication of very high frequency (VHF) and ultra high frequency (UHF) SAW and BAW devices with moderate bandwidth, high stability of central frequency and miniaturization.

250-900 MHz SAW resonators and resonator filters have been fabricated mainly on 45°X cut, Z propagation LBO substrate by lift-off process [14,15,17,18]. Y.Ebata et al. described the reflection characteristics for aluminium strips on LBO. The SAW reflectivity per strip on LBO is at least 5 times higher than that on quartz, $LiNbO_3$ and $LiTaO_3$ so the size of the devices is reduced [14]. H.Abe and K.Miwa et al. described the design parameters for LBO resonator and resonator filter. For 250-380 MHz LBO resonator filter the insertion loss is very low in so far as 1.6 dB and the image attenuation is high up to 60 dB. For 900 MHz LBO SAW device the central frequency changes 350 ppm and the pass band ripple is less than 1.5 dB in the temperature range from -10°C to 55 °C [18].

The commercial production of LBO SAW devices have been carried out for pager, codeless telephone and cellular phone in Japan [31,32].

C.Emin et al. investigated the characteristics of bulk wave of LBO [7] that indicated the ratio of capacity C_0/C_e is much lower than that of AT-quartz and the maximum bandwidth is 5-12 times larger than that of AT-quartz for different modes. A.Ballato et al. pointed out that LBO possesses high Q ($2\text{-}3\times10^{13}$), large piezocoupling (roughly three times as large as quartz) and zero temperature coefficient of frequency so it has potential advantages over quartz for certain application [12].

J-L Wang fabricated prototype of BAW resonator as voltage controlled oscillator using YZW 54° LBO substrate. This prototype has central frequency f 6.4 MHz, the range of controlled frequency Δf 20 KHz and its $\Delta f/f$ 0.31% (no expansion), the frequency of the resonator changes $\pm$150 ppm in the temperature range of -10 to 55 °C [19]. Obviously, new LBO SAW and BAW devices will be developed continuously along with LBO substrate production and new demands of frequency control devices.

Acknowledgments

The author would like to thank colleagues in his lab for LBO growth and characterization and Prof. T.Shiosaki in Kyoto University for application of LBO material.

References

[1] R.W.Whatmore, N.M.Shorrocks, C.O'Hara, F.W.Ainger and I.M.Young, "Lithium Tetraborate: A New Temperature-compensated SAW Substrate Material", Electronics Letter, vol.17, pp.11-12, January 1981.

[2] N.M.Shorrocks, R.W.Whatmore, F.W.Ainger and I.M.Young, "Lithium Tetraborate: A New

Temperature Compensated Piezoelectric Substrate Material for Surface Acoustic Wave Device", Proc. IEEE Ultrasonics Symp., 1981, pp.337-340.
[3] B.Lewis, N.M.Shorrocks and R.W.Whatmore, "An Assessment of Lithium Tetraborate for SAW Applications", Proc. IEEE Ultrasonics Symp., 1982, pp.389-393.
[4] D.S.Robertson and I.M.Young, "The Growth and Growth Mechanism of Lithium Tetraborate", J. Material Sci., vol.17, pp.1729-1738, 1983.
[5] K.Fukuta, J.Ushizawa, H.Suzuki, Y.Ebata and S.Matsumura, "Growth and Properties of $Li_2B_4O_7$ Single Crystal for SAW Device Application", Jpn.J.Appl.Phys.,vol.22,Suppl.22-2,pp.140-142,1983.
[6] M.Adachi, T.Shiosaki and A.Kawabata, "Crystal Growth of Lithium Tetraborate ($Li_2B_4O_7$)", Jpn.J.Appl.Phys., vol.24, Suppl.24-3, pp.72-75, 1985.
[7] Y.Ebata, H.Suzuki, S.Matsumura and K.Fukuda, "SAW Propagation Characteritics on $Li_2B_4O_7$", Jpn.J.Appl.Phys.,vol.22,Suppl.22-2,pp.160-162,1983.
[8] C.D.J.Emin and J.F.Werner, "The Bulk Acoustic Wave Properties of Lithium Tetraborate", Proc. 37th Annual Frequency Control Symp., 1983, pp.136-143.
[9] T.Shiosaki, M.Adachi, H.Kobayashi, K.Araki and A.Kawabata, "Elastic, Piezoelectric, Acousto-optic and Electro-optic Properties of $Li_2B_4O_7$", Jpn.J.Appl.Phys., vol.24, Suppl.24-1, pp.25-27, 1985.
[10] M.Adachi, T.Shiosaki, H.Kobayashi, O.Ohnishi and A.Kawabata, "Temperature Compensated Piezoelectric Lithium Tetraborate Crystal for High Frequency Surface Acoustic Wave and Bulk Wave Device Applications",Proc.IEEE Ultrasonics Symp., 1985, pp.228-232.
[11] M.Murota and Y.Shimizu, "Characteristics of Leaky Surface Wave Propagating on $Li_2B_4O_7$ Substrate with Arbitrary Cuts", Jpn.J.Acoustics., vol.45, pp.519-526, 1989 (in Japanese).
[12] A.Ballato, J.Kosinski and T.Lukaszek and S.Mallikarjun, "Measured Properties of Doubly-rotated Lithium Tetraborate Resonators and Transducers", Proc. IEEE Ultrasonics Symp., 1991, pp.459-462.
[13] R.C.Peach, C.D.J.Emin, J.F.Werner and S.P.Doherty, "High Coupling Piezoelectric Resonators Using Lithium Tetraborate", Proc.IEEE Ultrasonics Symp., 1983, pp.521-526.
[14] Y.Ebata and M.Koshino, "SAW Resonator and Resonator Filter on $Li_2B_4O_7$ Substrate", Jpn.J.Appl.Phys.,vol.26,Suppl.26-1,pp.123-125,1987.
[15] S.Matsumura, T.Omi, N.Yamaji and Y.Ebata, "A 45°X Cut $Li_2B_4O_7$ Single Crystal Substrate for SAW Resonators", Proc. IEEE Ultrasonics Symp., 1987, pp.247-250.
[16] W.S.Ishak and C.A.Flory, "Acoustic Modes in Lithium Tetraborate", Proc.IEEE Ultrasonic Symp., 1987, pp.241-245.
[17] H.Abe, H.Saitou, M.Ohmura, T.Yamada and K.Miwa, "Lithium Tetraborate ($Li_2B_4O_7$) SAW Resonators", Proc. IEEE Ultrasonics Symp., 1987, pp.91-94.
[18] M.Ohmura, H.Abe, K.Miwa and H.Saito, "A 900MHz SAW Resonator Filter on $Li_2B_4O_7$ Crystal", Proc. IEEE Ultrasonics Symp., 1990, pp.135-138.
[19] J-L Wang, "$Li_2B_4O_7$ Crystal resonator and its application", Proc. Ultrasonics Devices Application Symp., 1990, pp.68-71 (in Chinese).
[20] S-J Fan, G-S Shen, W.Wang, J-L Li and X-H Le, "Bridgman Growth of $Li_2B_4O_7$ Crystals", J.Crystal Growth, vol.99, pp.811-814, 1990.
[21] S-J Fan, W.Wang, J-J Xiang and X-H Le, "Core-free Growth of Three Inch Diameter $Li_2B_4O_7$ Crystal for SAW Devices", Crystal Properties and Preparation , vols 36-38, pp.46-51, 1991.
[22] S-J Fan, W.Wang and T.Shiosaki, Fronties of Materials Research/Electronic and Optical Materials. Elsevier Science Publisher B.V., 1991, pp.575-577.
[23] S-J Fan, W.Wang, J-J Xiang and J-Y Xu, "The Development of Industrial Growth Technology of New Piezoelectric Crystal $Li_2B_4O_7$ Material", Proc. of International Conference on Electronic Components and Materials, 1992, pp.206-209.
[24] R-Y Sun, S-J Fan and Y-B Xu, "Twinning and Dislocation in the Bridgman-grown $Li_2B_4O_7$ Crystals", Presented at Tenth International Conference on Crystal Growth, San Diego, CA, August 16-21,1992.
[25] J.Krogh-Moe, "The Crystal Structure of Lithium Diborate $Li_2O{\cdot}2B_2O_3$", Acta Crystallographica, vol.15, pp.190-193, 1962.
[26] A.A.Sehery and D.J.Somerford, "Low-temperature elastic Anomalies in Lithium Tetraborate", J.Phys.: Condens. Matter, vol.1, pp.2279-2281, 1989.
[27] V.V.Zaretskii and V.A.Zaretskaya-Eliashberg, "New Incommensurate Crystals: $Li_2B_4O_7$ and $Bi_{2.2}Sr_{1.8}CaCuO_{8+x}$", Materials Science Forum, vols 62-64, pp.11-12, 1990.
[28] Y-B Xu, To be published data.
[29] J-Y Xu and S-J Fan, To be published data.
[30] L.Bohaty, S.HaussuHl and J.Liebertz, "Electrooptical Coefficients and Temperature and Pressure derivatives of the Elastic Constants of Tetragonal $Li_2B_4O_7$ ", Crystal Res. Technol, vol.24, pp.1159-1163, 1989.
[31] T. Shiosaki and A. Kawabata, "Future Trends in Piezoelectric Materials and Applications", Ferroelectrics, vol.95, pp.9-14, 1989.
[32] Y.Ebata and H.Satoh, "Current Application and Future Trends for SAW in Asia", Proc. IEEE Ultrasonics Symp., 1988, pp.195-202.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PURE-MODE MEASUREMENTS OF DILITHIUM TETRABORATE MATERIAL PROPERTIES

John A. Kosinski*,+, Arthur Ballato*, and Yicheng Lu+

*US Army Research Laboratory
ATTN: AMSRL-EP-MD
Fort Monmouth, NJ 07703-5601

+Rutgers, The State University of New Jersey
Department of Electrical and Computer Engineering
P.O. Box 909
Piscataway, NJ 08855-0909

Abstract: The full set of elastic, piezoelectric, and dielectric constants of dilithium tetraborate ($Li_2B_4O_7$), along with the mass density, have been measured at room-temperature using the resonator method. The sample orientations used in the measurements were selected to maximize accuracy and precision through (1) the use of an overdetermined data set with critical self consistency checks, and (2) the use of pure-mode vibrations for which exact, analytic expressions for the piezoelectrically stiffened elastic stiffness eigenvalues exist. The experimental program included measurements of the physical dimensions, masses, critical frequencies, and capacitance of the various samples. The set of material constants obtained in this work are, on average, 2x to 10x more accurate in reproducing the measured data when compared to the previously reported sets of material constants.

Introduction

Since the discovery of the piezoelectric effect in 1880 by the brothers Curie, numerous techniques for determining the important physical constants of piezoelectric materials have been put forward. These have included both static and dynamic techniques, with dynamic techniques encompassing both frequency-domain and time-domain measurements of acoustic wave propagation. In this paper, the room-temperature determination of the elastic, piezoelectric, and dielectric constants of dilithium tetraborate using frequency domain measurements of primarily pure-mode vibrations in thin, flat plates is presented. The sample set is chosen in accordance with the pure-mode criteria reported in detail elsewhere, providing an overdetermined sample set with critical self-consistency checks in the measured data [1,2].

Technical Background

Measurements of the elastic, piezoelectric, and dielectric constants of dilithium tetraborate have been made by a number or researchers using a variety of techniques [3-35]; those measurements of greatest interest in the calculation of the room-temperature behavior of dilithium tetraborate piezoelectric resonators are listed in Table 1.

Table 1

Reported room-temperature values of the dielectric, piezoelectric, and elastic constants of dilithium tetraborate in units of ϵ_o, C/m^2, and GPa respectively.

ϵ_{11}^S	ϵ_{33}^S	e_{15}	e_{31}	e_{33}	c_{11}^E	c_{12}^E	c_{13}^E	c_{33}^E	c_{44}^E	c_{66}^E	Ref.
8.97	8.15	.36	.19	.89	126.7	0.5	30	53.9	55.0	46.0	8
8.5	8.2										9
8.90	8.07	.472	.290	.928	135	3.57	33.5	56.8	58.5	46.7	14
9.04	8.81	.278	.10	.77	127.1	0.6	29.4	53.8	53.8	57.4	13
8.88	8.46	.35	.38	.88	135.2	0.8	31.9	54.9	55.9	47.3	21
8.46	7.56	.36	.19	.87	135.3	0.11	31.9	54.8	57.4	47.4	22
					126.8	1.0	23.9	56.2	58.1	45.7	25
					134.2			57.4	47.3		29
8.91	8.13	.39	.24	.93	127.6	8.9	32	54.7	57.5	48.2	30
								55.5	57.4		47
8.91	8.10	.46	.23	.91	128.4	4.6	29.7	54.2	57.8	47.5	35

The first complete set of elastic, piezoelectric, and dielectric constants was reported by Shorrocks, et al., in 1981 based on material grown by the Czochralski technique [8]. The exact composition and purity of the starting materials used in the crystal growth was not specified. Shorrocks, et al., employed three techniques in evaluating the material: pulse-echo overlap measurements of cubes, resonance-antiresonance frequency measurements in bars, and measurements of resonance overtone ratios in plates. The density of the samples was measured as 2451 kg/m^3, which differs from the theoretical density by almost 1%.

0-7803-0905-7/93 $3.00 © 1993 IEEE

The second complete set of constants was reported by Shiosaki, et al., in 1984 based also on Czochralski grown material [10,11,14,15]. The starting material for the samples was 99.995% pure powder, and the measured density of 2439 kg/m^3 was within 0.3% of theoretical. The values of all the elastic and piezoelectric constants as reported by Shiosaki, et al., are slightly larger than those reported by Shorrocks, et al., while the values of the dielectric constants are slightly lower.

A complete set of elastic, piezoelectric, and dielectric constants and the associated first-order temperature coefficients was included in the patent issued to Whatmore, et al., in 1985 for the dilithium tetraborate XZ substrate for SAW devices [13]. The magnitudes of all of the material constants and temperature coefficients reported in the patent were revised from the values reported earlier, and the signs of three of the temperature coefficients were reversed from those reported earlier. In addition, the magnitudes of c^E_{33} and c^E_{44} were reported as being equal, which is strongly suggestive of the presence of at least one error in the values printed in the text of the patent.

In 1989 Sil'vestrova, et al., reported the use of the pulse-echo method in the frequency range of 10-15 MHz to measure Czochralski-grown crystals pulled from a stoichiometric melt of unspecified starting materials [21]. With the exceptions of e_{31} and ϵ^S_{33}, the values reported by Sil'vestrova, et al., fall in or very near to the range of values bounded by Shorrocks, et al., and Shiosaki, et al.

Sil'vestrova, et al., were followed closely by Bohaty, et al., who reported measurements on crystals grown from the melt by a special top-seeding technique [22]. The measured density of the crystals was within 0.1% of theoretical. The values of the elastic constants were obtained using resonance frequency measurements of high order overtones in thick (≈10 mm) plates, while the piezoelectric constants were determined from optical measurements of mechanical deformations. The values reported by Bohaty, et al., fall within the range bounded by Shorrocks, et al., and Shiosaki, et al., with the notable exception of the ϵ^S_{ij}, which are lower than any other reported values.

In 1990, Wang, et al., reported a complete set of constants obtained from measurements of material of unspecified growth technique [30]. The constants were obtained from measurements of resonance and antiresonance frequencies of thin plates, and with the exception of c^E_{12} also fall within the range of values bounded by Shorrocks, et al., and Shiosaki, et al.

The most recently reported complete set of constants is that of Fan on material grown using the modified Bridgman method [34]. The raw material used in the growth contained ≈30 ppm total impurities, and stoichiometry variations up to 0.17% were reported.

Experimental Approach

Single Crystal Boules

Single crystal boules of dilithium tetraborate for use in the material constants determination were grown at INRAD, Inc., of Northvale, New Jersey [36]. The boules were grown using the Czochralski technique in a resistance heated furnace with computer control of the furnace temperature and seed rotation [37]. The crystals were pulled in an air atmosphere at a rate of 0.2 mm/hr along the (100) and (001) directions from a stoichiometric melt of commercially available 99.999% purity dilithium tetraborate powder. The seed rotation rate was 3 rpm with periodic reversals of rotation. The crucible was made of platinum. Chemical analysis of the starting materials revealed typical impurity levels of 1.6 ppm Al, 1 ppm Ca, 1 ppm Fe, 1 ppm K, 0.3 ppm Na, and <5 ppm Si.

The as-grown single crystal boules were typically 40 mm in diameter with good uniformity of diameter. The crystals were colorless and transparent with excellent transmission for wavelengths from 200 nm to 1700 nm as shown in Figure 1. The crystals were free from growth defects with the exception of some bubble inclusions near

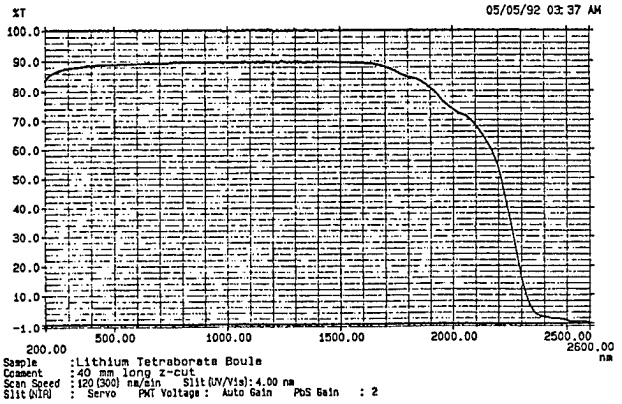


Figure 1. Z-axis transmission spectrum of a 40 mm thick dilithium tetraborate sample as measured at INRAD [37].

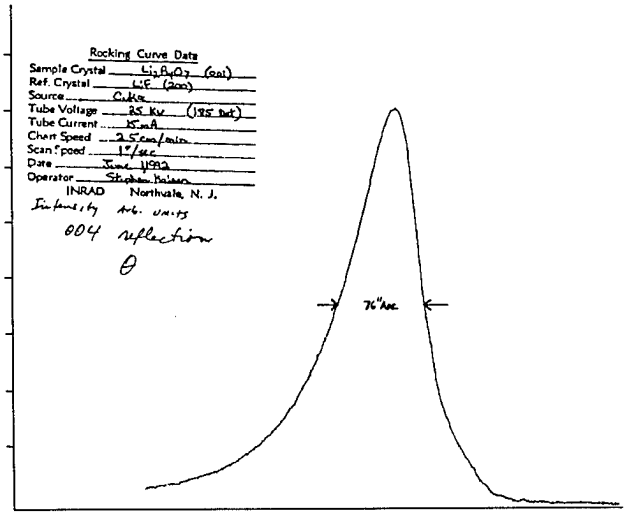


Figure 2. Typical double crystal x-ray diffraction rocking curve data as measured at INRAD [37]. The (004) plane reflection shown has a full-width at half-maximum (FWHM) of 76 arc seconds.

the seed. The degree of crystalline perfection was checked using Cu K-alpha radiation in a double crystal x-ray spectrometer. Typical x-ray data are shown in Figure 2. Chemical analysis of the as-grown material using an inductively coupled plasma mass spectrometer revealed 4.3 ppm Ba as the only detectable impurity.

Oriented Plates

The single crystal boules were processed into a set of thin, flat plates nominally oriented in accordance with the pure-mode sample set summarized in Table 2 and described in detail elsewhere [1,2]. The sample set incorporates seventeen different sample types based on seven different doubly-rotated plate orientations. Both thickness-field excitation and lateral-field excitation are employed in conjunction with each plate orientation. The material constants are accessed directly where possible, and indirectly accessed constants are overdetermined. The material-specific pure-mode loci (YXl)22.7° and (YXwl)45°,32.4° were not known with precision a-priori, and were estimated as (YXl)26.2° and (YXwl)45°,33.9° using the material constants reported by Adachi, et al. [14].

The single crystal boules were oriented and cored, with the cores then sliced to yield 14 mm diameter plates. The plates were lapped and optically polished

Table 2

Dilithium tetraborate pure-mode sample set summary identifying sample orientations and excitations used for measuring various constants/combinations of constants [1,2].

Constant(s)	Orientation	Drive
ϵ_{11}^S	(YXwl)0,0	TE
ϵ_{33}^S	(YXwl)0,90	TE
	(YXwl)0,45	TE
$\epsilon_{11}^S, \epsilon_{33}^S$	(YXwl)45,45	TE
	(YXwl)0,22.7	TE
	(YXwl)45,32.4	TE
e_{15}	(YXwl)0,0	TE
	(YXwl)45,0	TE
e_{33}	(YXwl)0,90	TE
e_{15}, e_{31}, e_{33}	(YXwl)0,22.7	TE
	(YXwlt)0,22.7,90	LE
	(YXwl)45,32.4	TE
	(YXwlt)45,32.4,90	LE
c_{11}^E	(YXwlt)0,0,90	LE
c_{33}^E	(YXwl)0,90	TE
c_{44}^E	(YXwl)0,0	TE
	(YXwlt)0,90,0	LE
	(YXwlt)0,90,90	LE
	(YXwl)45,0	TE
c_{44}^E, c_{66}^E	(YXwlt)0,45,0	LE
	(YXwlt)0,22.7,0	LE
$c_{11}^E, c_{12}^E, c_{44}^E$	(YXwlt)45,45,0	LE
	(YXwlt)45,32.4,0	LE
$c_{11}^E, c_{12}^E, c_{66}^E$	(YXwlt)45,0,90	LE
$c_{11}^E, c_{13}^E, c_{33}^E, c_{44}^E$	(YXwl)0,22.7	TE
	(YXwlt)0,22.7,90	LE
all $c_{\lambda\mu}^E$	(YXwl)45,32.4	TE
	(YXwlt)45,32.4,90	LE

to a nominal thickness of 200 μm, with the actual finished thickness of each processing group measured to 0.2 μm. The plate dimensions of 14 mm diameter and 200 μm thickness were chosen to balance the conflicting requirements for plate thickness (greater thickness yields improved relative thickness measurement accuracy), plate diameter-to-thickness ratio (larger ratio more closely approximates the infinite flat plate), and practical sample sizes (must be compatible with both single crystal boule size and test fixture dimensions). The finished plates are flat to within 1/4 wavelength at 633 nm, parallel to ±3 seconds of arc, and

oriented within ±3 minutes of arc from the nominally specified orientation angles.

Based on inconsistencies in the measured stiffness eigenvalues of the two material-specific pure-mode orientations, it was calculated that the (YXwl)45°,33.9° group had inadvertently been cut to the complementary orientation (YXwl)45°,56.1°, and that the (YXl)26.2° group had inadvertently been cut to (YXl)28.2°. The actual orientations of these two groups were then checked by x-ray which confirmed the misorientations as calculated.

Additional X-cut and Z-cut samples of 10 mm diameter and 1 mm thickness were also obtained and were used for the permittivity determinations.

Measurement Techniques

The antiresonance frequency $f_{Ao}^{(M)}$ of the Mth harmonic of the infinite flat plate piezoelectric resonator with zero mass loading is given by the equation

$$f_{Ao}^{(M)} = \frac{M}{2\,(2h)} \sqrt{\frac{\bar{c}}{\rho}} \qquad (1)$$

wherein M = 1, 3, 5, ... is the harmonic number, 2h is the plate thickness, $\bar{c}$ is the piezoelectrically stiffened elastic stiffness, and ρ is the mass density. In the design of piezoelectric plate resonators and transducers, it is desirable that the critical frequencies calculated using equations such as equation (1) be accurate to a minimum of 1%, with 0.1% accuracy highly desirable. In order to achieve such accuracy in calculating the critical frequencies, the various material constants of importance must be known to a correspondingly high degree of accuracy.

The determination of the various material constants requires the measurement of the plate thicknesses, mass densities, and critical frequencies along with auxiliary measurements of capacitance. The experimental techniques used in the measurement of the various quantities are discussed below.

Thickness 2h: For simple thickness modes in piezoelectric plate resonators, the plate thickness 2h is the frequency determining dimension. The plates used in the material constants determination are nominally 200 μm thick with the actual thicknesses measured to ±0.2 μm during the lapping and polishing process. The plate thicknesses are thus known to within approximately 0.1% as required.

Mass Density ρ: The dilithium tetraborate samples used in the material constants determination are disk-shaped with the major surfaces flat, parallel, and optically polished. Consequently, the mass density ρ of each individual sample is calculated simply as mass per unit volume using mass measured to ±10 μg using a Mettler Model AE 163 balance, thickness as measured by the material supplier, and lateral dimensions measured to ±0.001 mm using a David Mann Co. optical comparator. The nominal mass of a single 14 mm diameter sample is approximately 75,000 μg. The relative mass measurement accuracy is therefore approximately 0.01%, while the relative major surface area measurement accuracy is approximately 0.02%. The overall relative mass density measurement accuracy is thus limited to approximately 0.1% by the relative accuracy of the thickness measurement.

Critical Frequencies $f_{Ro}^{(M)}$ & $f_{Ao}^{(M)}$: The zero mass loading resonance frequencies $f_{Ro}^{(M)}$ and/or antiresonance frequencies $f_{Ao}^{(M)}$ of the unelectroded plates were measured to ±1 kHz using special thickness-field excitation (TE) and lateral-field excitation (LE) air-gap test fixtures [38,39] in conjunction with a network analyzer/balancing bridge system as shown in Figure 3. The frequencies were measured in a transmission mode with the shunt capacitance of the resonator balanced out. The thickness-field excitation resonance frequency measurements were made with the air-gap as close to zero as possible without touching the test sample (≈10 μm), while for the thickness-field excitation antiresonance frequency measurements the typical air-gap was approximately 4.0 mm. For the lateral-field excitation resonance frequency measurements, nearly planar electrodes separated by a 0.6 mm gap were placed as close to the upper surface of the crystal resonator as possible without touching. Due to the design of the test fixtures, both thickness and lateral electric field components are produced in each fixture and care must be taken in interpreting frequency measurements of

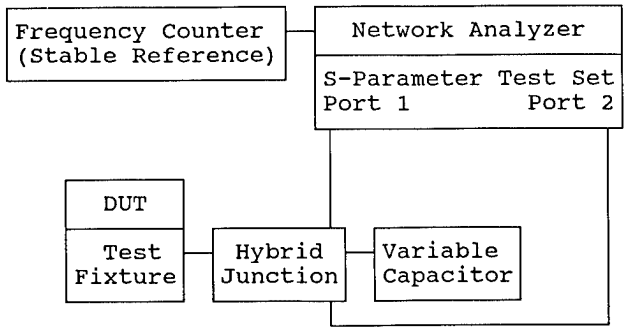


Figure 3. Critical frequency measurement apparatus.

resonators in which both thickness-field excitation piezoelectric coupling k and lateral-field excitation piezoelectric coupling $\underline{k}$ are nonzero.

Multiple harmonics of the thickness-field excitation resonance frequencies, thickness-field excitation antiresonance frequencies, and/or lateral-field excitation resonance frequencies were measured on each crystal. The results were interpreted using the exact transmission line analogs of the infinite flat plate piezoelectric resonator as shown in Figures 4 and 5 for the cases of thickness-field and lateral-field excitation respectively [40-47]. The governing equations for the critical frequencies are found by examination of the resonator impedance or admittance functions using either short-circuited (resonance) or open-circuited (antiresonance) conditions at the electrical port.

The thickness-field excitation resonance frequencies are related as

$$\tan X = \frac{X}{k^2 + \mu X^2} \quad , \quad X = \frac{\pi}{2} \frac{f_{R\mu}^{(M)}}{f_{Ao}^{(1)}} \quad , \qquad (2)$$

while the thickness-field excitation antiresonance frequencies are related as

$$\tan X = \frac{1}{\mu X} \quad , \quad X = \frac{\pi}{2} \frac{f_{A\mu}^{(M)}}{f_{Ao}^{(1)}} \quad , \qquad (3)$$

with $f_{Ao}^{(1)}$ as given by equation (1). The quantity μ in equations (2) and (3) represents the inertial mass loading effect due to the presence of metal electrodes deposited on the major surfaces of the resonator. For the unelectroded plate, the mass loading μ is nominally zero, and the various harmonics of the thickness-field excitation antiresonance frequency should be integer multiples of the fundamental frequency (1st harmonic, M=1).

The lateral-field excitation resonance frequencies are related as

$$\tan X = \frac{1}{\mu X} \quad , \quad X = \frac{\pi}{2} \frac{f_{R\mu}^{(M)}}{f_{Ro}^{(1)}} \quad , \qquad (4)$$

while the lateral-field excitation antiresonance frequencies are related as

$$\tan X = \frac{X}{-\underline{k}^2 + \mu X^2} \quad , \quad X = \frac{\pi}{2} \frac{f_{A\mu}^{(M)}}{f_{Ro}^{(1)}} \quad . \qquad (5)$$

The lateral-field excitation resonance frequencies coincide with the thickness-field excitation antiresonance frequencies, and the zero mass loading fundamental resonance frequency $f_{Ro}^{(1)}$ using lateral-field excitation may also be found using equation (1).

Equations (2) through (5) are used in conjunction with the measured data to yield the zero mass loading, fundamental, thickness-field excitation antiresonance frequencies $f_{Ao}^{(1)}$ of all excitable simple thickness modes.

Capacitance: The parallel plate capacitance values of the 1 mm thick X-cut and Z-cut samples were measured using an Andeen-Hagerling Model 2500A 1 kHz Ultra-Precision Capacitance Bridge and a Hewlett-Packard HP16451B Dielectric Test Fixture. The capacitance values (approximately 1.7 pF) were determined using a three-terminal measurement to a relative accuracy of better than 100 ppm. The nominal relative accuracy of the dielectric test fixture was 2%, limited by the manufacturing tolerance

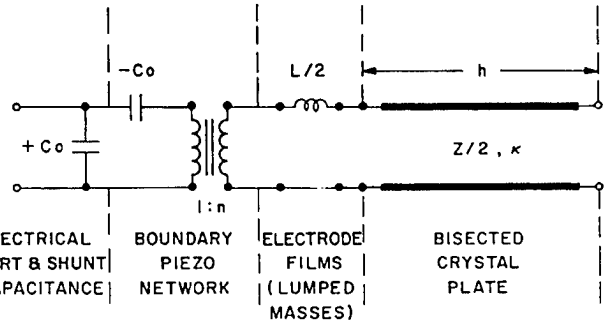


Figure 4. Exact transmission line analog (bisected basis) of a single simple thickness mode in a piezoelectric plate resonator for the case of the infinite, flat plate with thickness-field excitation [40-47].

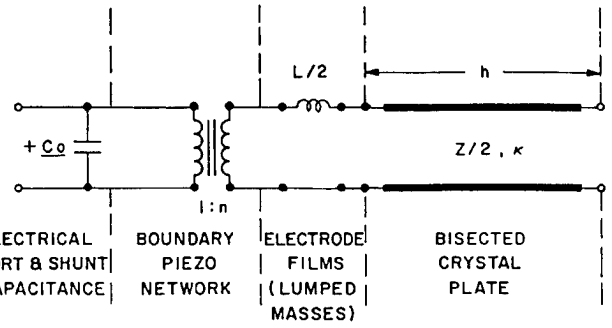


Figure 5. Exact transmission line analog (bisected basis) of a single simple thickness mode in a piezoelectric plate resonator for the case of the infinite, flat plate with lateral-field excitation [40-47].

of the test electrode. Consequently, the actual electrode diameter was measured using a David Mann Co. optical comparator as 5.030 mm ± 0.001 mm, yielding a relative electrode area measurement accuracy of 0.04%.

Experimental Data

Typical amplitude response data as observed during the frequency measurements are illustrated in Figures 6 through 9. For most of the harmonics of the modes that were measured, the amplitude responses were relatively free of spurious modes and the identification of the desired mode was unambiguous.

The measured frequency data fell into three categories of resonator responses, namely 1) modes excitable only by thickness-field excitation, 2) modes excitable only by lateral-field excitation, and 3) modes excitable by both thickness-

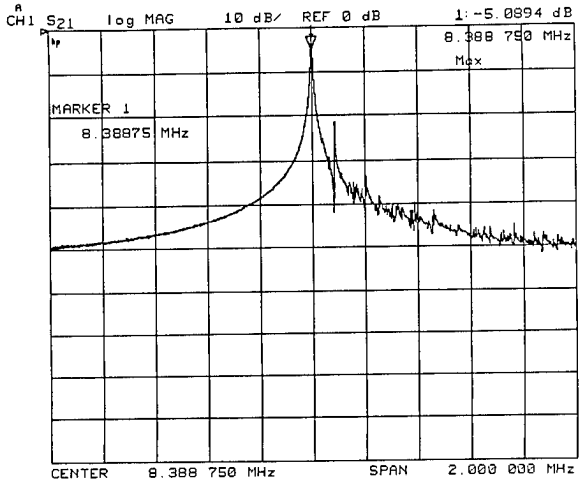


Figure 6. Amplitude response of the fundamental mode (M=1) resonance of the $\bar{c}^{(3)}$ thickness shear mode of a (YXl)45° resonator measured using thickness-field excitation.

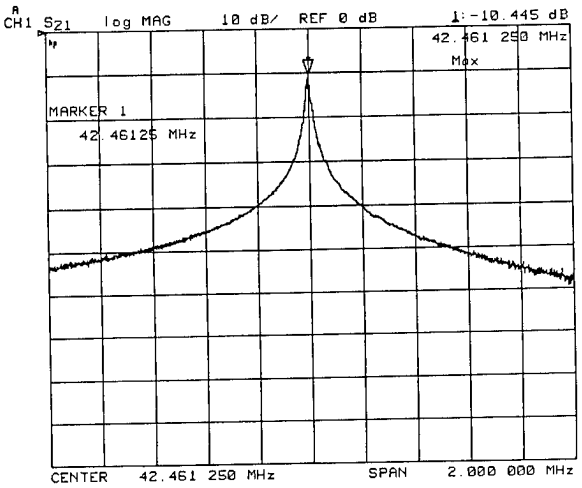


Figure 8. Amplitude response of the fifth overtone (M=5) resonance of the $\bar{c}^{(3)}$ thickness shear mode of a (YXl)45° resonator measured using thickness-field excitation.

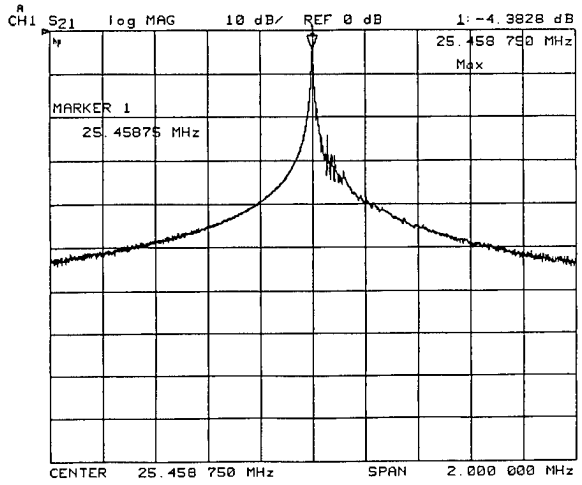


Figure 7. Amplitude response of the third overtone (M=3) resonance of the $\bar{c}^{(3)}$ thickness shear mode of a (YXl)45° resonator measured using thickness-field excitation.

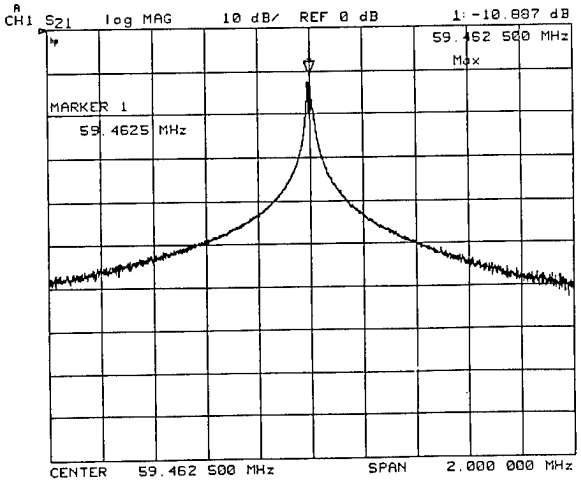


Figure 9. Amplitude response of the seventh overtone (M=7) resonance of the $\bar{c}^{(3)}$ thickness shear mode of a (YXl)45° resonator measured using thickness-field excitation.

Table 3

Summary of stiffness eigenvalue measurements of 67 dilithium tetraborate plate resonators.

φ	Θ	Mode	Drive	N	Stiffness (GPa) Min.	Max.	<Mean>	σ(%)
0	0	$\bar{c}^{(2)}$	LE[90]	8	135.1	136.4	135.9	0.30
		$\bar{c}^{(3)}$	TE		58.75	59.33	59.12	0.29
45	0	$\bar{c}^{(2)}$	LE[90]	7	115.0	115.9	115.4	0.31
		$\bar{c}^{(3)}$	TE		58.75	59.27	58.95	0.31
0	28.2	$\bar{c}^{(1)}$	LE[0]	12	49.50	49.90	49.76	0.26
		$\bar{c}^{(2)}$	LE[90]		137.3	138.4	138.1	0.27
		$\bar{c}^{(2)}$	TE		137.3	138.4	138.0	0.27
		$\bar{c}^{(2)}$	TE		137.4	138.5	138.1	0.28
		$\bar{c}^{(3)}$	LE[90]		40.01	40.34	40.20	0.29
		$\bar{c}^{(3)}$	TE		40.00	40.37	40.20	0.31
		$\bar{c}^{(3)}$	TE		40.01	40.39	40.21	0.31
0	45	$\bar{c}^{(1)}$	LE[0]	8	52.23	52.72	52.49	0.30
		$\bar{c}^{(2)}$	TE		129.9	131.2	130.4	0.33
		$\bar{c}^{(3)}$	TE		28.72	29.00	28.82	0.32
45	45	$\bar{c}^{(1)}$	LE[0]	8	62.30	62.90	62.49	0.32
		$\bar{c}^{(2)}$	TE		122.9	124.1	123.6	0.27
		$\bar{c}^{(3)}$	TE		25.33	25.56	25.46	0.27
45	56.1	$\bar{c}^{(1)}$	LE[0]	12	60.31	60.72	60.49	0.18
		$\bar{c}^{(2)}$	TE		114.2	114.8	114.4	0.17
		$\bar{c}^{(3)}$	TE		24.71	24.91	24.80	0.24
0	90	$\bar{c}^{(1)}$	LE[0]	12	56.83	57.41	57.06	0.27
		$\bar{c}^{(2)}$	TE		64.88	65.57	65.14	0.27
		$\bar{c}^{(2)}$	LE[0]		64.88	65.63	65.12	0.30

field and lateral-field excitations. For the first two cases, the measurements were made in the appropriate thickness-field excitation or lateral-field excitation test fixture and the results were unambiguously interpreted using the corresponding thickness-field excitation or lateral-field excitation transmission line analog. In the third case, measurements were made using both thickness-field excitation and lateral-field excitation test fixtures and care was required in the interpretation of the measured data due to the presence of unintended field components in the test fixtures.

The various categories of resonator frequency responses were analyzed using equations (2) through (5) as appropriate, and the stiffness eigenvalues were obtained from the frequency measurements through the inversion of equation (1). The overall set of stiffness eigenvalue measurements of the 0.2 mm thick plates is summarized in Table 3. For any given set of measurements, the variances of the observed data are on the order of 0.3% or less. No outliers (data points beyond ±3σ of the mean) are found in any set of measurements. The agreement between stiffness eigenvalues obtained from repeated frequency measurements using both thickness-field excitation and lateral-field excitation on the Z-cut and (YXl)28.2° samples is approximately 0.1%.

Table 4

Lattice constants and mass density of dilithium tetraborate.

a(Å)	c(Å)	ρ (kg/m^3) calc.	meas.	Ref(s).
9.47	10.26	2440		4
9.479	10.280			5
9.477	10.286	2430	2440	7
			2451	8
9.475	10.283		2450	9
9.49	10.30	2420	2439	11,14,15
9.47	10.26		2450	12
9.478	10.285			17,19
			2433	22
9.479	10.290			23
9.479	10.340	2430		25
9.474	10.297	2436		25
			2440	29
			2439	31
9.477	10.286			32
9.47	10.26			33
9.4795	10.312	2424		34
		2431	2432	1

Extraction of Material Constants

Mass Density Analysis

The lattice constants of dilithium tetraborate have been measured by a number of researchers as listed in Table 4 [4,5,7-9,11,12,14,15,17,19,22,23,25,29,31-34]. The average values of the reported lattice constants are a = 9.477 Å and c = 10.288 Å, corresponding to a theoretical mass density of 2431 kg/m^3. The average value of the measured mass densities of the 67 samples used here (2432 kg/m^3) was in excellent agreement with the theoretical mass density value.

Dielectric Permittivity Analysis

The relative dielectric permittivities measured at constant stress ϵ^T_{ij} were calculated using the measured 1 kHz capacitance values of the X-cut and Z-cut samples, the corresponding sample thicknesses, and the dielectric test fixture electrode dimensions along with a correction for fringing fields. The sample thicknesses of the nominally 1 mm thick X-cut and Z-cut samples were measured to ±0.001 mm using a Mitsutoyo digital micrometer, for a relative thickness accuracy of 0.1%. The permittivities were calculated as $\epsilon^T_{11}/\epsilon_o$ = 8.966 and $\epsilon^T_{33}/\epsilon_o$ = 9.680. The relative accuracy of the dielectric permittivity values is limited to approximately 0.1% by the relative error in the sample thickness measurements.

Stiffness Eigenvalue Analysis

The procedure used in extracting the material constants from the measured eigenvalue data is shown schematically in Figure 10. The analysis is divided into four steps based on the natural separations in the pure-mode eigenvalue expressions.

In the first step, c_{11}^E, c_{12}^E, c_{44}^E, and c_{66}^E are determined using a linear least-squares approach and the eigenvalue data from those pure-modes which are purely mechanical (unstiffened). The value of e_{15} is also found by incorporating selected stiffened mode data from the Y-cut and (YXw)45° orientations.

In the second step, c_{33}^E, e_{31}, and e_{33} are determined using a "linearized" nonlinear least-squares procedure and a combination of pure-mode and non-pure-mode data. The use of non-pure-mode data was found to be necessary due to the differences between the material-specific pure-mode orientations calculated using the material constants of Adachi, et al., and the actual pure-mode orientations. Although the eigenvalues $\bar{c}^{(2)}$ of the calculated orientations are nearly equal to those at the actual pure-mode loci, the eigenvalues $\bar{c}^{(3)}$ vary rapidly as a function of orientation angle Θ and the measured data do not reliably estimate the actual pure-mode values. In order to use the non-pure-mode data, an expression for the sum of the two non-pure-mode stiffness eigenvalues was derived:

$$\bar{c}^{(2)}+\bar{c}^{(3)} =$$

$$\frac{\begin{aligned}&c_{11}^E[c^2\Theta(1-2s^2\phi c^2\phi)]+c_{12}^E[2c^2\Theta s^2\phi c^2\phi]\\&+c_{33}^E[s^2\Theta]+c_{44}^E+c_{66}^E[4c^2\Theta s^2\phi c^2\phi]\\&+e_{15}^2c^2\Theta+(e_{31}^2+e_{15}e_{31}+e_{15}e_{33})s^2\Theta c^2\Theta+e_{33}^2s^4\Theta\end{aligned}}{\epsilon_{11}^S[c^2\Theta]+\epsilon_{33}^S[s^2\Theta]}$$

(s=sine, c=cosine) (6)

for all values of Θ along ϕ=0° and ϕ=45°. By taking the values from the first step in the analysis along with the current estimates of the ϵ_{ij}^S, equation (6) may be recast in a suitable form to aid in the extraction of c_{33}^E, e_{31}, and e_{33}. The procedure is an iterative "linearization" of an otherwise nonlinear problem, in that an initial estimate of any one of the three constants is entered and the other two are then determined by linear least-squares. The procedure is cycled among the three constants and repeated until convergence is achieved, usually in about six cycles.

In the third step, c_{13}^E is isolated using a nonlinear least-squares equation derived from the value of the product of $(\bar{c}^{(2)}\cdot\bar{c}^{(3)})$. The nonlinear least-squares equation is cubic in c_{13}^E and is solved numerically through computer iteration.

In the fourth and final step, the values of the various $c_{\lambda\mu}^E$ and $e_{i\lambda}$ are used in conjunction with the measured ϵ_{ij}^T to obtain the ϵ_{ij}^S from the constitutive equations of linear piezoelectricity. The entire procedure is repeated until all values converge.

Summary of Results

The room-temperature values of the dielectric, piezoelectric, and elastic constants obtained in this work are summarized in Table 5. All of the various raw data used in the material constants determinations have a relative accuracy of approximately 0.1%, and the constants thus obtained should be reliable to at least three significant figures. Most of the constants are determined from the least-squares analysis of sixty or more eigenvalue measurements. For these

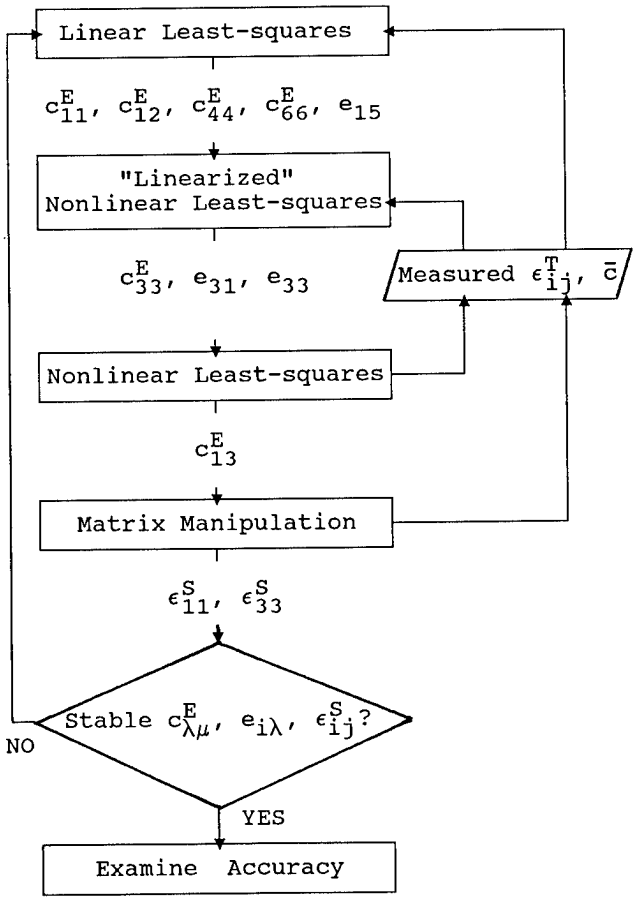


Figure 10. Procedure for extracting material constants from measured data.

Table 5

Summary of experimental results.

Constant	Value	Units
ϵ_{11}^{T}	8.966	ϵ_o
ϵ_{33}^{T}	9.680	ϵ_o
ϵ_{11}^{S}	8.662	ϵ_o
ϵ_{33}^{S}	7.893	ϵ_o
e_{15}	0.3918	C/m^2
e_{31}	0.2544	C/m^2
e_{33}	0.8820	C/m^2
c_{11}^{E}	135.823	GPa
c_{12}^{E}	-0.285	GPa
c_{13}^{E}	31.982	GPa
c_{33}^{E}	54.152	GPa
c_{44}^{E}	57.072	GPa
c_{66}^{E}	47.680	GPa

constants, the number of significant figures may more properly be taken as at least four, and possibly five.

Quality of Material Constants

The extent to which the set of material constants reported here is truly representative of the behavior of dilithium tetraborate depends directly upon the quality of the material samples produced at INRAD and used in the measurements. Both the growth process used and measurements of the as-grown boules argue strongly that the material samples produced at INRAD are of exceptional quality. Factors of note in the growth process include the first ever reported use of 99.999% purity starting materials, the use of slow rotation and pulling rates well below the critical rotation and pulling rates for defect formation, and the maintenance of excellent uniformity of the crystal diameter during growth. The resulting degree of crystalline perfection is reflected not only in the optical and x-ray data measured at INRAD, but also in the measured mass density which is closer to the theoretical mass density than any previously reported value.

The relative merits of the set of material constants reported here vis-à-vis the previously reported sets of material constants may be examined by comparing the

Table 6

Comparison of measured frequency constant data with calculations made using various sets of material constants.

			Frequency Constant N (Hz-m)							
ϕ	θ	Mode	Data	[8]	[13]	[14]	[21]	[22]	[30]	[35]	[1]
0	0	$\bar{c}^{(1)}$	----	2166	2363	2171	2206	2206	2223	2202	2214
		$\bar{c}^{(2)}$	3729	3595	3601	3720	3729	3728	3617	3620	3738
		$\bar{c}^{(3)}$	2461	2403	2420	2507	2431	2465	2468	2484	2465
45	0	$\bar{c}^{(1)}$	----	2537	2540	2595	2629	2635	2468	2513	2646
		$\bar{c}^{(2)}$	3444	3344	3517	3448	3443	3439	3455	3411	3446
		$\bar{c}^{(3)}$	2462	2403	2363	2507	2431	2465	2466	2484	2465
0	28.2	$\bar{c}^{(1)}$	2262	2213	2403	2236	2250	2258	2270	2254	2263
		$\bar{c}^{(2)}$	3767	3643	3618	3794	3759	3763	3700	3698	3773
		$\bar{c}^{(3)}$	2033	1994	1980	2055	2023	2045	2012	2039	2042
0	45	$\bar{c}^{(1)}$	2321	2270	2381	2314	2304	2320	2327	2318	2321
		$\bar{c}^{(2)}$	3653	3546	3500	3705	3644	3651	3622	3615	3660
		$\bar{c}^{(3)}$	1718	1694	1692	1723	1711	1729	1686	1716	1719
45	45	$\bar{c}^{(1)}$	2536	2454	2443	2523	2516	2534	2447	2471	2537
		$\bar{c}^{(2)}$	3569	3466	3471	3616	3550	3557	3571	3549	3565
		$\bar{c}^{(3)}$	1620	1601	1663	1615	1607	1622	1625	1638	1612
45	56.1	$\bar{c}^{(1)}$	2497	2422	2406	2495	2472	2495	2440	2455	2494
		$\bar{c}^{(2)}$	3434	3339	3314	3486	3413	3423	3435	3416	3428
		$\bar{c}^{(3)}$	1599	1592	1634	1602	1588	1608	1609	1618	1592
0	90	$\bar{c}^{(1)}$	2422	2369	2343	2449	2398	2428	2428	2429	2423
		$\bar{c}^{(2)}$	2588	2572	2503	2657	2590	2606	2615	2591	2591
		$\bar{c}^{(3)}$	2422	2369	2343	2449	2398	2428	2428	2429	2423

Table 7

Comparison of average and maximum relative errors between measured frequency constant data and calculations using sets of material constants reported by various researchers.

$\langle\lvert\epsilon\rvert\rangle$ (%)			max $\lvert\epsilon\rvert$ (%)			
$\bar{c}^{(1)}$	$\bar{c}^{(2)}$	$\bar{c}^{(3)}$	$\bar{c}^{(1)}$	$\bar{c}^{(2)}$	$\bar{c}^{(3)}$	Ref.
0.05	0.15	0.25	0.12	0.24	0.49	[1]
0.13	0.28	0.35	0.25	0.70	0.64	[22]
0.81	0.24	0.84	1.00	0.61	1.26	[21]
0.63	1.14	0.95	1.15	2.67	1.87	[14]
1.00	1.14	0.69	2.56	2.92	1.19	[35]
1.33	1.01	0.65	3.51	3.00	1.86	[30]
2.56	2.71	1.70	3.23	3.59	2.40	[8]
3.86	3.32	2.56	6.23	4.19	4.02	[13]

behavior each set predicts for the various orientations with that observed experimentally. Table 6 presents such a comparison based on the mean values of the measured frequency constants. In Table 6, the modes are identified by the associated root of the eigenvalue problem rather than the traditional a-, b-, and c-mode designations ($N_a > N_b \geq N_c$) since this more clearly reflects the quality of the various sets of material constants in calculating a given modal behavior. This point is illustrated in Table 7, wherein the average deviations and maximum deviations of the calculated frequency constants from the measured frequency constants are displayed.

The most accurately reproduced mode is the shear mode associated with $\bar{c}^{(1)}$, which is a pure-mode for all of the orientations studied here. The longitudinal mode associated with $\bar{c}^{(2)}$ and the shear mode associated with $\bar{c}^{(3)}$ are pure-modes for only some of the orientations and are, on average, less accurately reproduced, with the largest errors associated with the non-pure-mode orientations.

For all three simple thickness modes, the constants obtained in the present work provide both the smallest average deviations and smallest maximum deviations from the measured data. Overall, the set of material constants reported in the present work is in excellent agreement with that of Bohaty, et al. [22].

Acknowledgment

The authors wish to thank Mr. Duc Huynh, Mr. Robert Lindenmuth, and Mr. Raymond McGowan for their assistance in various phases of this work.

References

[1] J. Kosinski, "Pure-Mode Loci for Dilithium Tetraborate Resonators and Transducers," Ph.D. Dissertation, Rutgers University, May 1993, 208 pp. Available from University Microfilms, 313 N. First Street, Ann Arbor, MI 48106.

[2] J. Kosinski, Y. Lu, and A. Ballato, "Pure-Mode Loci in Piezoelectric Plate Resonators: Application to Materials Evaluation in Class 4mm," IEEE Trans. UFFC, Vol. 40, No. 3, pp. 258-264, May 1993.

[3] B. S. R. Sastry and F. A. Hummel, "Studies in Lithium Oxide Systems: I, $Li_2O \cdot B_2O_3$-B_2O_3," J. Am. Ceram. Soc., Vol. 41, No. 1, pp. 7-17, Jan. 1958.

[4] J. Krogh-Moe, "The Crystal Structure of Lithium Diborate, $Li_2O \cdot 2B_2O_3$," Acta Crystallographica, Vol. 15, pp. 190-193, 1962.

[5] J. Krogh-Moe, "Refinement of the Crystal Structure of Lithium Diborate, $Li_2O \cdot 2B_2O_3$," Acta Crystallographica, Vol. B24, pt. 2, pp. 179-181, 1968.

[6] J. D. Garrett, M. N. Iyer, and J. E. Greedan, "The Czochralski Growth of $LiBO_2$ and $Li_2B_4O_7$," J. Crystal Growth, Vol. 41, pp. 225-227, 1977.

[7] M. Natarajan, R. Faggiani, and I. D. Brown, "Dilithium Tetraborate, $Li_2B_4O_7$," Crystal Structure Comm., Vol. 8, pp. 367-370, 1979.

[8] N. M. Shorrocks, R. W. Whatmore, F. W. Ainger, and I. M. Young, "Lithium Tetraborate - A New Temperature Compensated Piezoelectric Substrate Material for Surface Acoustic Wave Devices," 1981 IEEE Ultrasonics Symp. Proc., Oct. 1981, pp. 337-340.

[9] K. Fukuta, J. Ushizawa, H. Suzuki, Y. Ebata, and S. Matsumura, "Growth and Properties of $Li_2B_4O_7$ Single Crystal for SAW Device Applications," JJAP, Vol. 22, Supp. 22-2, pp. 140-142, 1983.

[10] T. Shiosaki, M. Adachi, H. Kobayashi, K. Araki, and A. Kawabata, "Elastic, Piezoelectric, Acousto-Optic and Electro-Optic Properties of $Li_2B_4O_7$," Proc. 5th Symp. on Ultrasonic Electronics, Tokyo, 1984, pp. 5-8.

[11] T. Shiosaki, M. Adachi, H. Kobayashi, K. Araki, and A. Kawabata, "Elastic, Piezoelectric, Acousto-Optic and Electro-Optic Properties of $Li_2B_4O_7$," JJAP, Vol. 24, Supp. 24-1, pp. 25-27, 1985.

[12] A. S. Bhalla, L. E. Cross, and R. W. Whatmore, "Pyroelectric and Piezoelectric Properties of Lithium Tetraborate Single Crystal," JJAP, Vol. 24, Supp. 24-2, pp. 727-729, 1985.

[13] R. W. Whatmore and I. M. Young, "Application of Lithium Tetraborate to SAW Devices," U.S. Patent 4,523,119, issued June 11, 1985.

[14] M. Adachi, T. Shiosaki, H. Kobayashi, O. Ohnishi, and A. Kawabata, "Temperature Compensated Piezoelectric Lithium Tetraborate Crystal for High Frequency Surface Acoustic Wave and Bulk Wave Device Applications," 1985 IEEE Ultrasonics Symp. Proc., Oct. 1985, pp. 228-232.

[15] T. Shiosaki, M. Adachi, and A. Kawabata, "Growth and Properties of Piezoelectric Lithium Tetraborate Crystal for BAW and SAW Devices," IEEE ISAF Proc., PA, June 1986, pp. 455-464.

[16] O. T. Antonyak, Ya. V. Burak, I. T. Lyseiko, and Z. A. Khapko, "Luminescence of $Li_2B_4O_7$ Crystals," Optika i Spektroskopiya, Vol. 61, No. 3, pp. 550-553, Sept. 1986.

[17] V. V. Zaretskii and Ya. V. Burak, "Phase Transitions Stimulated by Thermal Cycling," Pis'ma v Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki, Vol. 49, No. 4, pp. 198-201, 25 Feb. 1989.

[18] A. A. Sehery and D. J. Somerford, "Low-Temperature Elastic Anomalies in Lithium Tetraborate," J. Phys.: Condensed Matter, Vol. 1, No. 12, pp. 2279-2281, March 27 1989.

[19] V. V. Zaretskii and Ya. V. Burak, "New

Incommensurate $Li_2B_4O_7$ Crystal," Fizika Tverdogo Tela, Vol. 31, No. 6, pp. 80-84, June 1989.

[20] Ya. V. Burak, B. N. Kopko, I. T. Lyseiko, A. O. Matkovskii, R. R. Slipetskii, and U. A. Ulmanis, "Color Centers in $Li_2B_4O_7$ Single Crystals," Izvestiya Akademii Nauk SSSR, Neorganicheskie Materialy, Vol. 25, No. 7, pp. 1226-1228, July 1989.

[21] I. M. Sil'vestrova, P. A. Senyuschenkov, V. A. Lomonov, and Yu. V. Pisarevskii, "Anomalies of the Temperature Dependences of the Elastic Properties of Lithium Tetraborate Under Thermal Cycling Conditions," Fizika Tverdogo Tela, Vol. 31, No. 10, pp. 311-313, Oct. 1989.

[22] L. Bohaty, S. Haussühl, and J. Liebertz, "Electrooptical Coefficients and Temperature and Pressure Derivatives of the Elastic Constants of Tetragonal $Li_2B_4O_7$," Crystal Res. Tech., Vol. 24, No. 11, pp. 1159-1163, Nov. 1989.

[23] S. F. Radaev, L. A. Muradyan, L. F. Malakhova, Ya. V. Burak, and V. I. Simonov, "Atomic Structure and Electron Density of Lithium Tetraborate $Li_2B_4O_7$," Kristallografiya, Vol. 34, No. 6, pp. 1400-1407, Nov.-Dec. 1989.

[24] K. Ya. Borman and Ya. V. Burak, "Characteristics of the Thermal Expansion of $Li_2B_4O_7$," Izvestiya Akademii Nauk SSSR, Neorganische Materialy, Vol. 26, No. 2, pp. 440-442, Feb. 1990.

[25] H. A. A. Sidek, G. A. Saunders, and B. James, "The Pressure and Temperature Dependences of the Elastic Behavior of Lithium Tetraborate," J. Phys. Chem. Solids, Vol. 51, No. 5, pp. 457-465, 1990.

[26] S. Furusawa, S. Tange, Y. Ishibashi, and K. Miwa, "Raman Scattering Study of Lithium Diborate ($Li_2B_4O_7$) Single Crystal," J. Phys. Soc. Japan, Vol. 59, No. 5, pp. 1825-1830, May 1990.

[27] S. Furusawa, S. Tange, Y. Ishibashi, and K. Miwa, "Dielectric Properties of Lithium Diborate ($Li_2B_4O_7$) Single Crystal," J. Phys. Soc. Japan, Vol. 59, No. 7, pp. 2532-2535, July 1990.

[28] A. É. Aliev, Ya. V. Burak, and I. T. Lyseiko, "Electrophysical Properties of the New Superionic Crystal $Li_2B_4O_7$," Izvestiya Akademii Nauk SSR, Neorganicheskie Materialy, Vol. 26, No. 9, pp. 1991-1993, Sept. 1990.

[29] A. É. Aliev, Ya. V. Burak, V. V. Vorob'ev, I. T. Lyseiko, and E. V. Charnaya, "Absorption of Ultrasound in Ionic Conductor $Li_2B_4O_7$," Fizika Tverdogo Tela, Vol. 32, No. 9, pp. 2826-2828, Sept. 1990.

[30] Wang Jinfen, Zhang Lei, Qin Zikai, Wang Hong, He Dazhen, and Hu Shaoqin, "Measurements of Electro-elastic Constants of Lithium Tetraborate ($Li_2B_4O_7$) Crystal," Acta Acustica, Vol. 15, No. 6, pp. 425-431, Nov. 1990.

[31] A. E. Aliev, A. Sh. Akramov, R. R. Valetov, and P. K. Khabibullaev, "Peculiarities of Ultrasonic Propagation in Piezoactive Superionic Conductors," Solid State Ionics, Diffusion and Reactions, Vol. 46, No. 3-4, pp. 197-203, July 1991.

[32] T. Lukasiewicz and A. Majchrowski, "Czochralski Growth of Lithium Tetraborate Single Crystals," Materials Lett., Vol. 11, No. 9, pp. 281-283, July 1991.

[33] K. Byrappa and K. V. K. Shekar, "Hydrothermal Synthesis and Characterization of Piezoelectric Lithium Tetraborate, $Li_2B_4O_7$, Crystals," J. Mat'ls Chem., Vol. 2, No. 1, pp. 13-18, 1992.

[34] R. M. Catchings, Howard University, unpublished, 1992.

[35] S. Fan, "Properties, Production and Application of New Piezoelectric Crystal Lithium Tetraborate $Li_2B_4O_7$," elsewhere in these Proceeings.

[36] INRAD, Incorporated, 181 Legrand Avenue, Northvale, NJ 07647, (201) 767-1910.

[37] J. Gualtieri, J. Kosinski, W. Wilber, Y. Lu, S. Lin, M. Murray, and W. Ruderman, "Dilithium Tetraborate ($Li_2B_4O_7$) Fabrication Technology," Proc. 1992 IEEE Freq. Cont. Symp., May 1992, pp. 724-731.

[38] G. K. Guttwein, T. J. Lukaszek, and A. Ballato, "Practical Consequences of Modal Parameter Control in Crystal Resonators," Proc. 21st Ann. Freq. Cont. Symp., April 1967, pp. 115-137.

[39] A. Ballato, E. R. Hatch, M. Mizan, B. H. T. Chai, R. S. Tilton, and T. J. Lukaszek, "Lateral-Field Excitation of Berlinite," Proc. 38th Ann. Freq. Cont. Symp., May-June 1984, pp. 206-224.

[40] A. Ballato, H. L. Bertoni, and T. Tamir, "Transmission-Line Analogs for Stacked Crystals with Piezoelectric Excitation," presented at the 83rd meeting of the Acoust. Soc. Am., Buffalo, New York, 18-21 April 1972, paper NN3. Abstract: J. Acoust. Soc. Am., Vol. 52, No. 1 (Part 1), p. 178, July 1972.

[41] A. Ballato, "Transmission-Line Analogs for Stacked Piezoelectric Crystal Devices,"

Proc. 26th Ann. Freq. Cont. Symp., June 1972, pp. 86-91.

[42] A. Ballato, "Transmission-Line Analogs for Piezoelectric Layered Structures," Ph.D. Dissertation, Polytechnic Institute of Brooklyn, June 1972, 264 pp. Available from University Microfilms, 313 N. First Street, Ann Arbor, MI 48106, Order No. 72-28,207.

[43] A. Ballato, H. L. Bertoni, and T. Tamir, "Systematic Network Approach for Piezoelectrically Driven Crystal Plates and Stacks," 1972 IEEE Ultrasonics Symp., 4-7 Oct. 1972, Boston, MA, paper B-2. Abstract: IEEE Trans. Sonics and Ultrasonics, Vol. SU-20, No.1, p. 43, Jan. 1973.

[44] A. Ballato, "Networks for Crossed-Field and In-Line Excitation of Bulk and Surface Acoustic Waves," Proc. Symp. on Optical and Acoustical Micro-Electronics, Polytechnic Institute of New York, April 1974, pp. 599-615.

[45] A. Ballato, "Bulk and Surface Acoustic Wave Excitation and Network Representation," Proc. 28th Ann. Freq. Cont. Symp., May 1974, pp. 270-279.

[46] E. R. Hatch and A. Ballato, "Lateral-Field Excitation of Quartz Plates," 1983 IEEE Ultrasonics Symp. Proc., Oct.-Nov. 1983, pp. 512-515.

[47] A. Ballato, "Doubly Rotated Thickness Mode Plate Vibrators," in Physical Acoustics, Vol. 13 (W. P. Mason and R. N. Thurston, eds.). New York: Academic Press, 1977, pp. 115-181.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A STUDY OF THE TIME-DEPENDENCE OF ELECTRODIFFUSION IN QUARTZ

JOY E. FERRIS, D. W. HART, AND J. J. MARTIN
Department of Physics
Oklahoma State University
Stillwater, OK 74078-0444

ABSTRACT

Electrodiffusion (sweeping) is a high temperature process that affects the dislocation networks with their precipitated impurities that form etch channels and point defects such as the substitutional aluminum with the associated alkali. Quartz is commercially swept to reduce the formation of etch channels and to improve its radiation hardness. The time-dependence of the etch channel reduction and the replacement of the alkali at the aluminum site by hydrogen was measured in AT-cut samples cut from bars of cultured quartz from several sources. A set of AT-cut wafers from each bar were polished, coated with gold and then swept in a hydrogen atmosphere for times ranging from 15 minutes to 200 hours at 500° C. The electric field was applied at room temperature and the current was recorded during the warm-up, hold, and cool-down portions of each run. Low temperature infrared measurements were used to monitor the growth of the Al-OH infrared band versus sweeping time. The Al-OH signal increased rapidly for short sweeping times and reached saturation in 1-2 hours. First order kinetics of the form $(1 - \exp(-k_1 t))$ where t is the time gave a good approximation to the replacement of the alkali by hydrogen at the aluminum. The sweeping current decreased rapidly during the initial part of the run and then became constant. We find that when the current remains constant to within $\pm 2\%$ that the replacement of the interstitial alkali at the aluminum site is complete. After the samples were etched , the etch channels were then counted. The kinetics of the tendency to form etch channels was significantly different than the replacement of the interstitials. Initially, the number of etch channels decayed exponentially with sweeping time. At longer times, the decay shifted to a much slower rate which may also be exponential. These results suggest that there may be several mechanisms for the removal of the tendency to form etch channels or that there are several types of channels.

INTRODUCTION

Electrodiffusion (sweeping) is commercially employed used to reduce the formation of etch channels during processing [1] and to improve the radiation hardness [2]. The process removes neither the dislocation loops which form the etch channels nor the point defects. Instead, the improved performance of swept quartz results from the modification of the defects. For example, substitutional aluminum with an associated interstitial is a point defect present in all quartz. Sweeping in air (with water vapor present) or in hydrogen removes the alkali and replaces it with hydrogen. A similar process must also take place with the impurities trapped in the dislocation networks so that they become less active chemically. Alkalis were first swept into quartz by Verhoogen [3] during his studies of ionic conductivity. King [4] was the first to apply the sweeping process to quartz used for resonator studies. Fraser [5] developed the process for sweeping specific alkalis. Kats [6] studied the effects of alkali and hydrogen sweeping on the infrared spectra of both natural and cultured quartz. Martin [7] has reviewed current electrodiffusion techniques and results. Recently, Asahara, *et al.* [8] reported careful chemical analyses of unswept and swept quartz. They suggest that divalent Ca and Mg are trapped in the dislocation loops that form the channels after etching.

At the present time, several methods are used to determine if the sweeping process has been carried to completion for point defects. Of these, the electron-spin-resonance, ESR, test developed by Halliburton, *et al.* [9] and the measurement of the series resistance of a resonator at high temperatures developed by Koehler [10] are probably the most reliable. ESR evaluation requires the removal of a test sample from the swept

0-7803-0905-7/93 $3.00 © 1993 IEEE

bar and an elaborate radiation procedure. The resistance test requires the fabrication of a resonator blank. No test other than actually etching a samples is available for etch channel reduction by sweeping. Hanson [11] reported that a current peak that appears on warm-up during sweeping was absent in samples that had previously been swept. However, Hart, *et al.* [12] found that while this current peak disappeared for samples that had been cycled to 300° C for short times no alkali removal took place until the sample was taken above 450°C. Therefore, the disappearance of the current peak cannot be taken as a reliable indicator of sweeping. Tests for sweeping effectiveness that do not require elaborate sample preparation or that can be performed during the electrodiffusion process would clearly be useful. Hart, *et al.* found that after the sweeping current became constant to within ±2% the conversion of the Al-alkali centers into Al-OH centers was complete. The dislocation networks that form the etch channels are independent of the point defects and, they also found that the etch channel density was still decreasing when the Al-OH conversion was complete. We report here an investigation of etch channel removal as a function of sweeping time.

EXPERIMENTAL PROCEDURE

Pure z-growth cultured quartz from two sources was used for this investigation. AT-cut slices 1.5 to 2.5 mm thick were used for the sweeping tests. These smaller samples allow the comparison of sweeping effectiveness on material taken from a single quartz bar. Table I gives the aluminum content and etch channel density for the three unswept quartz bars.

TABLE I.
Sweeping sample characteristics.

Bar	Al(ppm)	Channels/cm^2
PQ-C	0.4	610
MH26-39	2.3	110
PQ-H	4.5	1,200

Bar PQ-H has a non-typical large etch channel density; this large density makes it useful for this investigation. Au electrodes were vapor deposited on cleaned polished surfaces. The samples were swept in a hydrogen atmosphere using the system previously described by Martin [7]. Electric fields between 800 and 1000 V/cm were applied at room temperature. A microcomputer was used to log the current and temperature during the sweeping run. Temperature-ramped sweeping was performed [13], i.e., the furnace was slowly brought up to the 500° C sweeping temperature, held there for the selected time, and then ramped back to room temperature.

Low temperature infrared absorption scans were used to monitor the production of the Al-OH center as a function of sweeping time. The scans were made using either a Nicolet 740 FTIR or a Beckman 4240 dispersive spectrophotometer operating at a resolution of 2 cm^{-1}. The samples were mounted in a liquid nitrogen cryostat so that the X-axis was vertical. Under these conditions, the Al-OH concentration is given by

$$C = 3.3\alpha \qquad \text{Eq. 1}$$

where α is the absorption coefficient [14]. After the scans were completed, the samples were etched for 2 hours at 75° C in a saturated ammonium bifluoride solution. The etch channel density was then determined by counting the channels observed using an Olympus Nomarski microscope.

RESULTS AND DISCUSSION

The dashed and solid curves in Fig. 1 show the temperature and current, respectively, versus time for a typical sweeping run. This sample shows a distinct current peak during the warm-up portion of

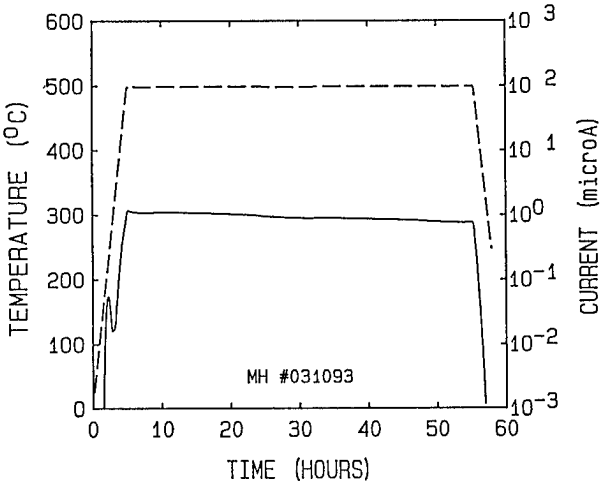


Figure 1. The temperature and current curves for a typical sweeping run are shown.

Hart, *et al.* [12] have shown that this current peak disappears if the sample is heated to about 450° C with an applied field and then cooled back to room

temperature. However, no significant replacement of alkalis takes place until the sample is heated above this temperature or held at temperature for a period of time.

The lower curve in Fig. 2 shows the infrared absorption spectrum of an unswept AT-cut sample from bar PQ-H. The bands at 3300 and 3200 cm^{-1} are intrinsic SiO overtone vibrations [6]. Only the OH-related growth-defect bands at 3581, 3437, 3398, and 3350 cm^{-1} are present in the spectrum for the unswept sample. The upper curve shows the spectrum for a sample from bar PQ-H that had been swept at 500°C for 50 hours.

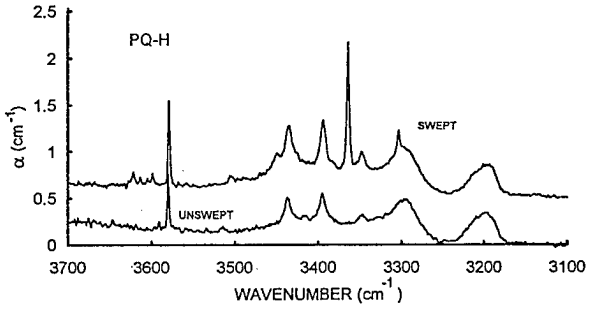


Figure 2. The lower curve shows the low temperature infrared spectrum of an unswept PQ-H sample. The upper curve shows the additional Al-OH bands that were introduced by sweeping.

The Al-OH absorption bands at 3367 and 3306 cm^{-1} reached saturation for sweeping times greater than about one hour for samples from bar PQ-H and from the other bars used in this study. We have found that when the Al-OH concentration becomes constant the sample will test as fully swept according to the ESR sweeping effectiveness test. Because of its lower sensitivity, the infrared test is not as reliable for low aluminum samples as the ESR test.

Figure 3 shows the Al-OH concentration as determined by low temperature infrared absorption measurements versus sweeping time for samples from the three bars used in this study. The curves remained constant for times greater than 5 hours. If we assume that simple first order kinetics apply the Al-OH concentration would be expected to grow as

$$C = C_f(1 - \exp(-k_1 t)) \qquad \text{Eq. 2.}$$

where C_f is the final concentration, k_1 is the rate constant, and t is the sweeping time. The curves shown in Fig. 3 were calculated using Eq. 2. The rate constant, k_1, was adjusted to force the curves through the data points. The rate constants are given in Table II. Lopez, Hwang, and Martin [15] found for unswept quartz that the Al-OH also grows with irradiation dose according to first order kinetics. The reader should remember that for the short times the effective sweeping time is uncertain and that the kinetics involved in the production of the Al-OH center by sweeping may be more complex.

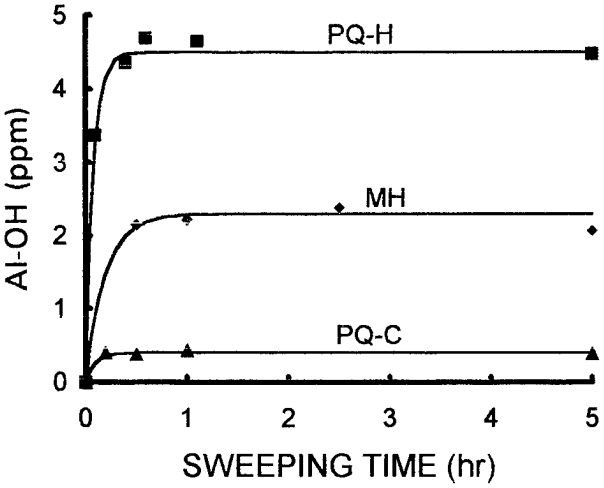


Figure 3. The growth of the Al-OH concentration is shown as a function of sweeping time.

Figure 4 shows the etch channel density versus sweeping time on a log-linear plot for bar PQ-C. The data suggest that there is an initial rapid exponential decay of the tendency to form etch channels followed by a second much slower exponential decay. The curve drawn through the points was found by fitting a double exponential of the form

$$CH = A\exp(-k_2 t) + B\exp(-k_3 t) \qquad \text{Eq. 3}$$

where CH is the number of channels, A and B are constants, and k_2 and k_3 are rate constants. The values found from the fit are given in Table II.

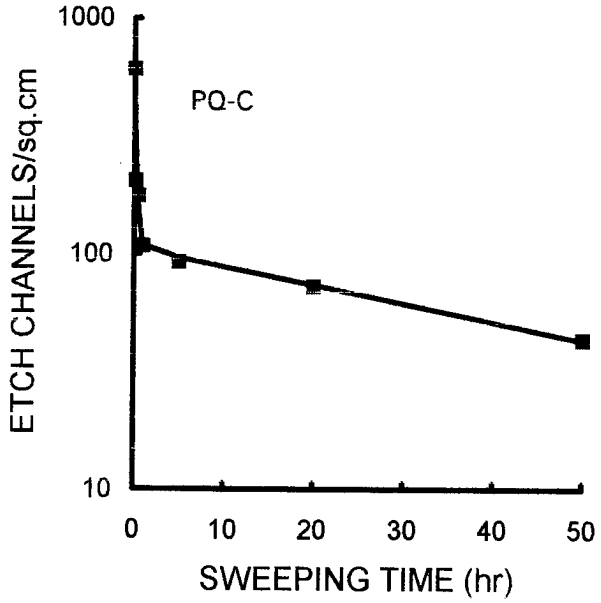


Figure 4. The reduction of the etch channel density versus sweeping time is shown for a series of samples from bar PQ-C. The curve is a fit of Eq. 3 to the data.

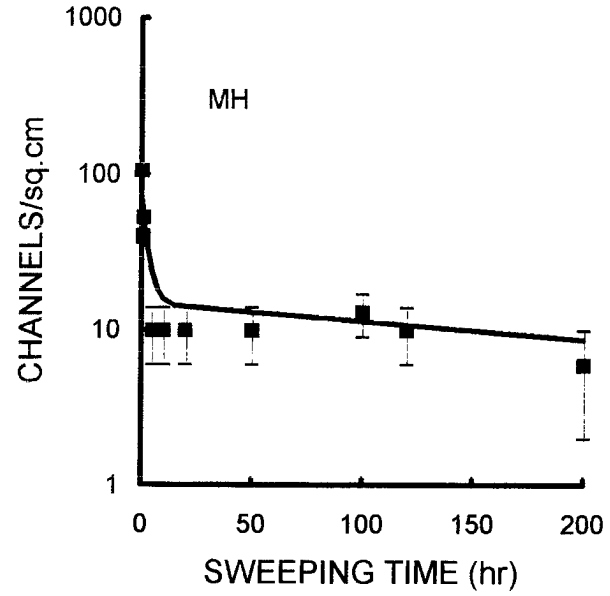


Figure 5. The reduction of the etch channel density versus sweeping time is shown for a series of samples from bar MH. The curve is a fit of Eq. 3 to the data.

Figure 5 shows the etch channel density versus sweeping time on a log-linear plot for bar MH. Balascio and Armington [16] have reported etch channel densities of 410/cm^2 and 21/cm^2 for unswept and swept samples from the same autoclave run (MH26-39). The lower values obtained for unswept and swept samples in our experiment represent the normal variation in etch channel densities for different bars from the same autoclave run. The curve through the data points represents a fit of Eq. 3 to our results. The parameters are given in Table II.

Figure 6 shows the etch channel density versus sweeping time for samples taken from bar PQ-H. The error bars of ± 5 shown in Fig. 5 are hidden by the squares used to indicate the data points on this log-linear plot and on Fig. 4. The unusually large etch channel density in unswept samples from this bar are well above those normally found in commercial quartz [17] and must be considered as non-typical. However, sweeping for 250 hours lowered the density to a value within the normal range for swept quartz. While sweeping has significantly reduced the etch channel density the dislocation networks are still present. The curve shown in Fig. 3 is a fit of Eq. 3 to the data. The parameters are given in Table II.

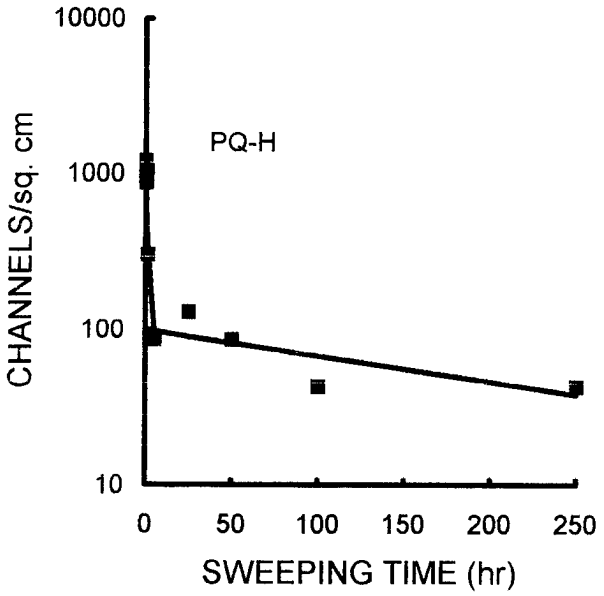


Figure 6. The reduction of the etch channel density versus sweeping time is shown for samples taken from bar PQ-H.

TABLE II.
Rate equation parameters for Al-OH growth and etch channel reduction.

Bar	C_f ppm	k_1 hr^{-1}	A /cm^2	k_2 hr^{-1}	B /cm^2	k_3 hr^{-1}
PQ-C	0.4	10	377	4.2	106	0.018
MH	2.3	5	56	0.3	15	0.0027
PQ-H	4.5	13	1240	1.4	99	0.0038

Etch channel formation goes out with sweeping at a slower rate than the production of the Al-OH center. Since the Al-OH production represents the conversion of point defects and the etch channels are line defects the details of the processes must be different, and different rates are to be expected. As sweeping proceeds the rate of etch channel removal slows perhaps as much as two orders of magnitude. At the present time we are uncertain as to whether the decay in the number of channels will continue to decrease or if they will saturate at some small number. Since the channels tend to lie along the Z-direction [17] sweeping AT-cut plates is probably less efficient than sweeping a full bar were the electric field is along Z. However, the final channel densities that we have obtained seem to be within the range of those reported for air-swept bars [12, 16, 18]. The slower rate that appears after a few hours may be caused by electrode effects; the ions from the channel may build up on (or near to) the electrode and impede the process [12, 19, 20]. If this is the case, grinding off a small layer and then re-sweeping the sample should lead to a lower etch channel density. It is also possible that there are several types of channels or, more likely, channels with differing impurities that are exchanged for hydrogen at different rates. The double exponential decay described by Eq. 3 supports the existence of several types of channels.

CONCLUSIONS

Al-OH conversion proceeds at a significantly faster rate during sweeping than the removal of the tendency to from etch channels. The appearance of a steady sweeping current can be used as an indicator of the completeness of the Al-OH conversion. However, since the point defect current dominates that due to the dislocation networks the sweeping current does not provide a good indication of etch channel removal. The decay of etch channel formation with sweeping time can be approximately described as a fast exponential decay followed by a much slower exponential decay. This result suggests the possibility that there are several types of channels.

ACKNOWLEDGMENTS

The authors would like to thank Alton Armington, Joe Balascio, John Gualtieri, and Larry Halliburton for several helpful discussions. A portion of this work was supported by the National Science Foundation.

REFERENCES

[1] J. R. Vig, J. W. LeBus, and R. L. Filler, "Chemically polished quartz," in Proc. 31st Ann. Symposium on Freq. Control, NTIS, May 1977, pp. 113-143.

[2] J. C. King and D. R. Koehler, "Radiation effects on resonators," in Precision Frequency Control, E.A. Gerber and A. Ballato, eds. Orlando, FL: Academic Press, 1985, vol. 1, pp147-159.

[3] J. Verhoogen, "Ionic diffusion and the properties of quartz," Am. Mineral., vol. 37, pp. 637-655, 1952.

[4] J. C. King, "Electrolysis of synthetic quartz; Effect upon resonator performance," Proc. IEE(London), vol. B109, Suppl. 22, pp295-301, 1962.

[5] D. B. Fraser, "Anelastic effects of alkali ions in crystalline quartz," J. Appl. Phys. , vol 35, pp. 2913-2918, 1964.

[6] A. Kats, "Hydrogen in alpha-quartz," Phillips Res. Repts., vol. 17, pp. 132-279, 1962.

[7] J. J. Martin, "Electrodiffusion (sweeping) of ions in quartz-a review," IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-35, pp. 288-296, 1988.

[8] J. Asahara, K. Nagai, K. Hamaguchi, H. Sone, and S. Taki, Proc. 45th Ann. Symposium on Freq. Control, IEEE, pp. 9-15, May 1991.

[9] L. E. Halliburton, N. Kuomvakalis, M. E. Markes, and J. J. Martin, "Radiation effects in crystalline SiO2: The role of aluminum," J. Appl. Phys. vol. 57, pp. 3565-3674, 1981.

[10] D. R. Koehler, "Radiation induced conductivity and high temperture Q changes in quartz resonators," in Proc. 35th Ann. Symposium on Frequency Control., EIA, pp.322-328, 1981.

[11] W. P. Hanson, "Probable ion signature in quartz electrodiffusion data," Proc. 38th Ann. Symposium on Frequency Control,. IEEE, pp. 38-41, May 1984.

[12] D. W. Hart, J. Frank, D. Smith, J. J. Martin, and J. G. Gualtieri, "A study of the electrodiffusion process in quartz," in Proc. 44th Ann. Symposium on Frequency Control, IEEE, pp. 222-227, May 1990.

[13] J. Gualtieri, "The influence of temperture and electric field effects the etch channel density in swept-cultured quartz," Proc. 39th Ann. Symposium on Freq. Control, IEEE, pp. 247-254, 1985.

[14] J. J. Martin, A. Lopez, A. F. Armington, and J. Balascio, "Evaluation of resonators fabricated from high quality quartz," Proc. 42nd Ann. Symposium on Frequency Control, pp. 162-168, June 1988.

[15] A. R. Lopez, Ho. B. Hwang, and J. J. Martin, "A production study of acoustic loss related defects in quartz," Proc. 42nd Ann. Symposium on Frequency Control, IEEE, pp. 176-183, June 1988.

[16] J. F. Balascio and A. F. Armington, "Developmental results for the production of high quality quartz," Proc. 40th Ann. Symposium on Freq. Control, IEEE, pp. 70-75, 1986.

[17] G. Johnson and R. Irvine, "Etch channels in single crystal cultured quartz," Proc. 41st Ann. Symposium on Freq. Control, IEEE, pp. 175-182, 1987.

[18] J. R. Hunt and R. C. Smythe, "Chemically milled VHF and UHF AT-cut resonators," Proc. 39th Ann. Symposium on Frequency Control, IEEE, pp. 292-300, 1985.

[19] J. Gualtieri, L. Calderon, and R. Lareau, "Update on possible electrode mechanisms in the sweeping of alpha quartz," Proc. 43rd Ann. Symposium on Frequency Control, IEEE, pp. 509-516, 1989.

[20] J. Gualtieri, J. Kosinski, and A. R. Murray, "Nonuniformities in the sweeping of quartz,", Proc. 44th Ann. Symposium on Freq. Control, IEEE, pp. 228-237, May 1990.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

QUARTZ CRYSTALS FOR MONOLITHIC PIEZOELECTRIC FILTERS

I.V.Kabanovich

Scientific Research Institute "Fonon"
Krasnobogatyrskaya Str., 44, Moskow 105023, Russia

High quality quartz crystals on X-cut seeds are grown by precipitation of crystal-forming substance in the negative direction of electric axis X. Screens restricting crystal growth in other directions have been used.

Introduction

In designing and production of monolithic piezoelectric filters, it is very important to ensure both identity of temperature-dependence of frequency and certain difference of frequences for resonators made on the same plate. Tuning of frequency for each resonator can be achieved by changing the thickness of electrode covering, and the identity of temperature-dependence of frequency requires increasing homogeneity of quartz crystal, especially in the direction of electro-mechanical connection between resonators.

To grow quartz of high homogeneity of piezoelectric and other physical properties is a non-trivial problem. It is more complicated if high homogeneity is required in the direction of the optical axis or close to it, that is in the direction of growth (and thickness) for crystals traditionally grown on Z-cut seeds (perpendicular to the optical axis).

Growing quartz crystals of high homogeneity of crystalline structure within grown layer is hampered first of all by continuous changes in growth conditions in autoclave (continuous decrease in quantity and area of dissolution of charge; change in conditions of washing of charge by convective flows; change in character of relief and total growing surface of crystals; change in conditions of delivery of crystal-forming substance); by difficulty of keeping constant during all the growth time the crystallizing temperature and necessary difference between temperatures in the crystal growth chamber and charge solution chamber; by appearing defects of crystalline structure and change in their density and character during crystal growth; by inhomogeneity of physical properties of seed material; and a number of other factors hard to take into account and to control.

In order to grow homogeneous crystals researchers try to exclude or diminish changes in crystal structure induced by unfavourable factors but they face serious difficulties.

For example, it was recommended to solve the problem of improving quartz homogeneity by growing crystals on seeds parallel to rhombohedron surfaces, that is to grow crystals by precipitation of substance on surfaces of the most probable growth of quartz in natural conditions. Crystal elements of many cuts of such crystals prove to be made of a material layer grown at the same time, and therefore they are homogeneous at least within

0-7803-0905-7/93 $3.00 © 1993 IEEE

a plate. With this way of growing quartz crystals, the problem of obtaining large, homogeneous, and free of seed material plates (especially if their orientation coincides with that of growing rhombohedron surface, or is close to it) is facilitated. The disadvantage of this way of growth is that homogeneity of crystals within grown layer is considerably broken because of changes in growth conditions, and therefore it is hard to produce filter resonators with identical characteristics even from crystals of the same autoclave run.

Material of the growth zone <-X>

The problem of obtaining identical homogeneous plates parallel to the optical axis or rotated around electric axis (axis X) is facilitated by growing quartz crystals on seeds perpendicular to axis X.

Material grown in the positive as well as negative direction of axis X can be at certain growth conditions suitable for filter resonators. However, when grown simultaneously on X-cut seed, material of these two growth zones in the same quartz crystal differs a great deal from one zone to the other in quality and properties. Practically used in this case the material grown in the positive direction of axis X only can be.

It is most suitable to achieve high quality of either the positive or the negative growth zone by ensuring growth of only one of them in a run of autoclave.

By results of our research, the most prospective for production of filter resonators is the material growing in the negative direction of axis X (growth zone <-X>). Growing surface of such crystals proves to be smooth almost ideal plane which favours identity of conditions for all parts of growth surface and hampers appearing boundaries that would break homogeneity within it. Another peculiarity of growth of crystal by developing the negative pyramid is that in this case inclusions are catched in significantly less degree and therefore linear defects related to them are appearing in less quantity.

It is true that homogeneity of such crystals can be broken in the thickness of grown layer because of changes in growth conditions, first of all changes in crystallizing temperature. Nevertheless, if plates are made of such crystals parallel to axis X (at any necessary angle to the optical axis) then it is not difficult to ensure identity between resonators of the same plate in the direction of electromechanical connection as well as between all plates made of the same crystal (and of all crystals of the same autoclave batch).

It was before considered as impossible to grow on X-cut seeds crystals with material of growth zone <-X> of high quality suitable for piezoelectronics, at least impossible at growth rates suitable for practice, because of high catch of group OH^- decreasing the Q-factor. That opinion about material of growth zone <-X> was formed due to the fact that samples for examination were made of crystals grown in conditions, when the preferential growth zone was not the growth zone <-X>, but for example the pinacoid zone or the growth zone <+X>; these conditions were chosen to ensure high quality of other growth zones, while quality of the growth zone <+X> and especially that of the growth zone <-X> was comparatively low, unsuitable for practice.

We have established that high quality of material of growth zone <-X> can be achieved for crystals grown at practically suitable rates (0.2-0.4 mm per day). The result was achieved by using conditions favouring preferencial growth of material

of zone <-X>; other factors influencing the quality of grown material (the crystallizing temperature; composition and concentration of solution etc.) were chosen in auspicious combination for ensuring high quality.

Screening of seeds

To ensure growth of crystals on X-cut seeds in the negative direction of axis X only (or preferentially), one can use thin U-shaped screens made of sheet low-carbon steel (fig. 1). A seed plate can be fastened to the screen with narrow bent strips made on this screen. The distance between bent lateral sides of the screen can be equal to the width (Z-dimension) of the seed plate or exceed it somewhat. In the latter case the seed increases its width due to growth on pinacoid surfaces during the first stage of growth (fig. 2). This favours use of material for seeds as well as increases crystal quality due to decreasing the density of linear defects spreading to grown layer from the seed.

With seeds fastened to the screen (which is simultaneously the element of suspension of crystal in the growth chamber), another positive influence of screen on the growing material quality is observed, namely the position of a crystal in the growth chamber is constantly changing due to change in gravity center location caused by non-symmetrical growth of the crystal, and growing surface of crystal is inclining gradually down, so that precipitation of inclusions from solution is hampered.

To diminish expenditures of steel sheet for screens and to place seeds in the growth chamber more compactly, one can change the construction of screen and fasten seeds to both sides of sheet, so that to prevent growth in the positive direction of axis X and restrict growth in

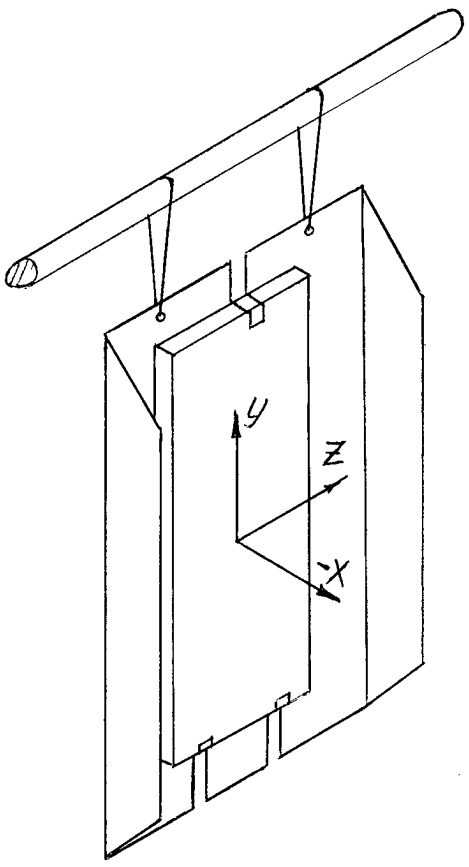


Fig.1. Suspension of an X-cut seed to grow material of the growth zone <-X>.

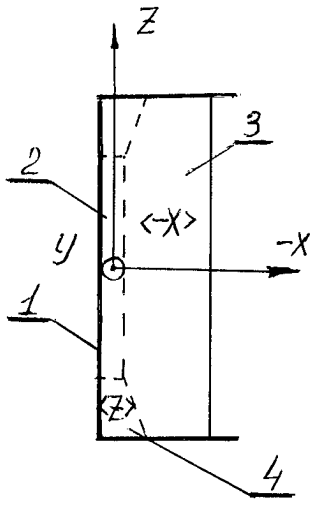


Fig.2. Cross-section of crystal grown on a screened seed in the negative direction of axis X (1 - screen; 2 - seed; 3, 4 - material of growth zones <-X> and <Z>, respectively.

Z-direction. However, in this case growing surfaces of crystal remain vertical during all the time of growth and solid particles (inclusions) can precipitate on them from solution.

Seeds

To produce an initial set of low-dislocational seeds, one can use natural quartz crystals with considerable Y- and Z- dimensions. Grown on such seeds quartz crystals can be cut into bars along axis Y or rotated at some angle around axis X, so that in cross cutting the bar into plates perpendicular to the bar axis, the plates of required orientation be obtained. That will allow make bar cutting with multi-blabe machines more rational.

Results

According to the described method, in large autoclaves we have grown on X-cut seeds crystals with the Q-factor of material of growth zone <-X> in the range of $1.0 \cdot 10^6$ to $2.4 \cdot 10^6$. The crystals were tested in filter resonators (mainly with two pairs of electrodes) and their suitability for these purposes was established.

We have also found another advantage of such quartz, namely its high stability to twinning at sharp changes of temperature. This can broaden applications of quartz of growth zone <-X> and promote developing ways of glass sealing with high-frequency current.

We believe that due to mentioned advantages quartz of growth zone <-X> will find wide applications in manufacture of filter resonators, including monolithic piezoelectric filters.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CONTROLLED DISSOLUTION APPLIED TO BERLINITE AND QUARTZ MATERIALS

M.DELEUZE[1], O.CAMBON[2], A.GOIFFON [1], A.IBANEZ[1], E.PHILIPPOT[1]

1 - Laboratoire de Physicochimie des matériaux solides, UM II, C.N.R.S URA DO407, C 003, Place E.Bataillon, F-34095 Montpellier Cédex 5, FRANCE.
2 - C.E.P.E/ Thomson, 44, Avenue de la Glacière, B.P. 165, F-95100, Argenteuil, FRANCE.

Abstract : Previous studies suggesting that controlled crystal growth and crystal dissolution are governed by the same laws, then, chemical controlled dissolution of berlinite and quartz has been investigated in the medium used for their crystal growth: acid medium for berlinite and basic one for quartz.

Following this concept, conditions of good controlled dissolution of berlinite and quartz have been found for two different plate orientations: AT and SC cuts. Piezoelectric characteristics of resonators made from atraumatic controlled dissolution and from ionic beam etching are compared.

Introduction

Two factors are of prime importance in the manufacture of piezoelectric devices: the crystalline quality of bulk material, and the thinning down of resonators for frequency adjustment.The Q-factor is closely related to both crystalline quality and final surface roughness of the wafer, and lapping must therefore result in a good surface texture, and a thickness which corresponds to the required frequency.

The unconventional and expensive method of ion beam etching (IBE) is used for sophisticated high frequency resonators with a thickness less than 30 µm, as it is not possible to use mechanical lapping. Chemical lapping, however, can be considered as a possible alternative.

Much research has been carried out in this field using fluoride media [1-3], but has not been very successful. HF and saturated NH_4HF_2 baths are very corrosive even with surfactant additions [4-6], and result in a poor quality of final surface state.

Whereas crystal growth has been studied for berlinite $AlPO_4$ [7-10)] there has been no investigation into chemical lapping. We have, therefore, first determined berlinite crystal growth conditions for transfer at an industrial scale [10-11], and then its controlled dissolution in the same solvents [12-14].

Whilst carrying out these experiments we have kept in mind that the two processes of crystal growth and dissolution are closely related through the reverse thermodynamic relations [15-18]. These phenomena are better controlled when experimental conditions are close to the solubility equilibrium: a weak super- or under-saturation of solutions promotes a good controlled crystal growth or dissolution respectively. In the latter case the energy barrier ΔG^*d is high enough to avoid most of etch pit formation during dissolution (figure 1).

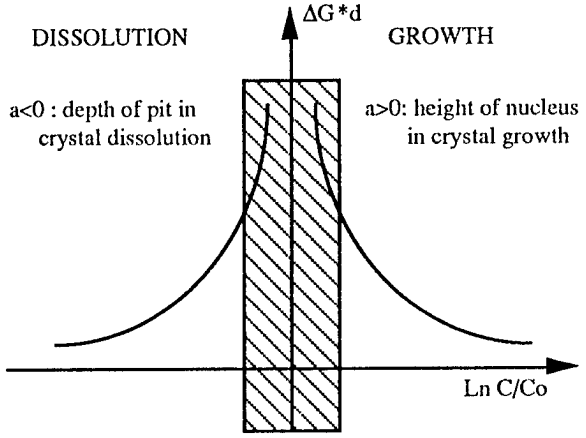


Fig 1 : ΔG^*d energy barrier variation for growth and dissolution of a real crystal in terms of solute concentration, C (Co= solubility at chosen temperature).

Our research differed from the works previously mentioned [1-3], in that we investigated the crystal dissolution of berlinite in acid medium and that of quartz in basic solvents, as these solvents promote the best crystal growth conditions. We have used mainly AT berlinite, and AT and SC quartz cuts for our research.

Experiments

We summarize here the main points developed in a previous report for berlinite and we have adapted the experimental conditions for quartz studies [12]. All the dissolution experiments have been carried out in a closed and thermally insulated apparatus. The samples used are small discs of different sizes. Their thinning down is

0-7803-0905-7/93 $3.00 © 1993 IEEE

followed by frequency measurements from the well known relation:

$$F\,(MHz) = K\,(MHz.\mu m) / e\,(\mu m)$$

where e is the thickness of the wafer. The constant, K, depends on material and its orientation. The values of these constants for the fundamental mode are :

	Berlinite	Quartz AT	Quartz SC
K (MHz.µm)	1480	1670	1820

Dissolution solutions are prepared from high purity commercial acids or sodium and potassium hydroxide pellets and, in the case of berlinite, from aluminum phosphate synthetized in our department. The surface texture is characterized by measuring the roughness parameter, Ra, and by light microcopy observations.

Results and discussion

Controlled dissolution of AT berlinite plates

- Chemical dissolution: Different chemical baths to study berlinite dissolution were used. According to the thermodynamic theory of crystal growth and dissolution, the media are constituted by sulfuric or phosphoric acid solutions which are the best solvents for crystal growth. In order to approach equilibrium conditions, aluminum phosphate is dissolved in these acid solutions. In each medium, the dissolution rate and the evolution of the surface state have been followed. The results, detailed elsewhere [14], are summarized in Table 1 and figure 2.

In all cases, the indicated rate is the intrinsic one, measured after the removal of the surface layer disturbed by sawing and polishing. In pure acid, a much greater dissolution rate is observed in phosphoric than in sulfuric acid, but with the same rate law and activation energy: i.e. it is the same dissolution mechanism in both acids [13]. When aluminum phosphate is added to the medium, dissolution rate is decreased. Then, by changing acid and aluminum phosphate concentrations, as well as temperature, the dissolution process can be monitored.

The most important and interesting feature takes place on the evolution of the surface state. According to thermodynamic predictions, the dissolution of aluminum phosphate markedly improves the surface roughness as can be seen in Figure 2. When pure sulfuric acid is used, the roughness value remains very high, whereas it decreases slowly when berlinite is added to the medium. On the other hand, pure phosphoric acid still improves surface state, probably because of the presence of phosphate anions. When aluminum phosphate is dissolved, the decrease of roughness begins as soon as the dissolution starts and a very good surface state is quickly reached. Some experiments, using polished wafers have led to the preservation of this state.

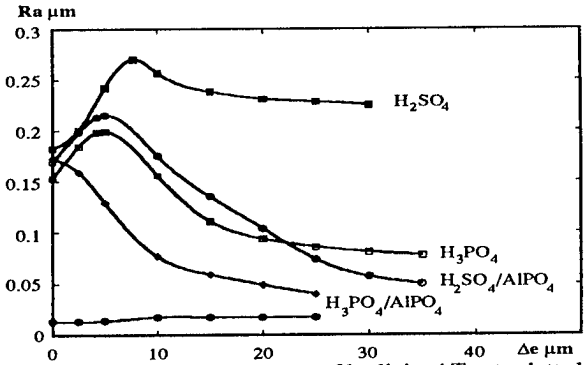


Fig 2 : Roughness parameter of berlinite AT cuts plotted against thinning down for differents solvents

Table 1 : AT berlinite dissolution comparison in different acid media.

Solvent	H_2SO_4	H_3PO_4	H_2SO_4 +$AlPO_4$	H_3PO_4 +$AlPO_4$	
Rate law	$V=(A \exp(-Ea/RT) * [H+]^{1/2})$		$V=(k1[H+]^{\alpha} - k2[Al3+]^{\beta}[H3PO4]^{\gamma})$		
Constants	$A=2.5\ 10^9$	$A=6.25\ 10^{10}$			
Ea kj.mol^{-1}	63.4	63			
T°C	90	90	90	90	
acid mol.l^{-1}	8	9	8	9	10.4
$AlPO_4$ mol.l^{-1}			3.6	2.7	3.5
V µm/h	15.7	83.7	3.4	36.3	11.2

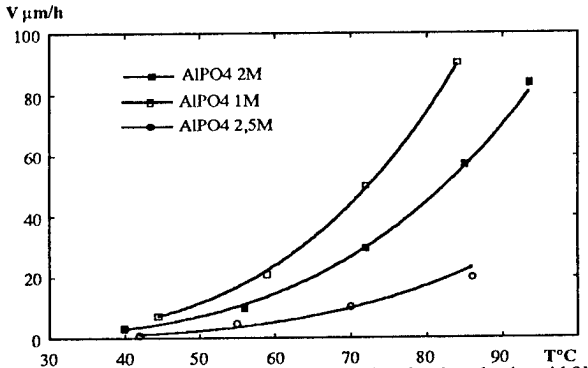


Fig 3 : Dissolution rate of AT berlinite plate in phosphoric acid 9M containing different amounts of dissolved aluminum phosphate

Given these good results, further experiments in phosphoric acid, with dissolved aluminum phosphate, were carried out with different acid and phosphate concentrations. In varying experimental conditions, a large field of dissolution rates from 0.2 to 90 µm.h-1 can be reached, and the desired thickness for a plate can be reached with a very good accuracy. An example of these results is given in figure 3. The quality of surface state increases with the quantity of solute. Thus, a compromise solution, consistent with industrial manufacturing, needs to be found between polishing effect and dissolution rate.

Piezoelectric measurements: In order to compare the influence of chemical thinning down and ion beam etching on piezoelectric performances of resonators, some similar devices have been realized from both techniques. In the case of controlled dissolution process, the experimental conditions are : H_3PO_4 6M, $AlPO_4$ 2M, T=85°C so $V=13\mu m.h^{-1}$, initial thickness : 100 µm and final one ~20 µm corresponding to about 73 MHz. The results are shown in Table 2 in comparison with similar quartz devices.

Values for devices made from chemical dissolution are consistent with those obtained from ion beam etching ones. By comparison with the values for quartz, in a similar frequency range, we note a resistance value for berlinite resonator higher than that of quartz, resulting in a decrease of the Q factor. Nevertheless, for a filter, the lower value of the berlinite self is an advantage with respect to quartz.

Table 2 : Comparison of berlinite resonators made by I.B.E. or controlled dissolution with quartz one.

Material	AT Berlinite		AT Quartz
	Ionic beam etching	Controlled dissolution	
Fr (MHz)	73.317	74.315	73
R (Ohms)	49.1	52	25
L (mH 2089	2089	2041	3000
Q	20000	19000	55000
Q.Fr	1.4 E+12	1.4 E+12	4 E+12

Moreover, the higher AT-cut coupling coefficient of berlinite ($kAlPO_4$=11.4%, $kSiO_2$=8.7 %) enables filters to be attained with 30% wider band pass than for quartz. Figure 4 shows the result of a simulation computation which gives the band width evolution versus frequency for berlinite and quartz, and confirms that the band width is always larger for berlinite than for quartz. Our studies have, therefore, clearly demonstrated the advantages of berlinite, and the interest of controlled dissolution. We thus extended this latter process to quartz, since it has a far greater industial potential.

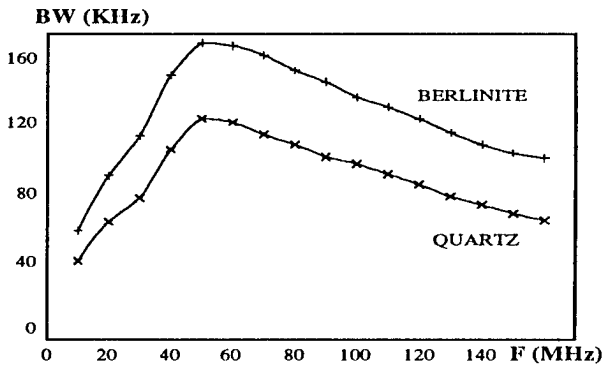


Fig 4 : Band width evolution versus frequency for berlinite and quartz (simulation)

Controlled dissolution of AT and SC quartz plates.

Following the same idea as before, i.e. crystal growth and dissolution are reverse processes, the controlled dissolution of quartz plates was investigated in a basic medium, the growth solvent, which is more suitable than the acid medium already used [1-4]. Thus, the first experiments were carried out in sodium hydroxide solution for AT and SC cuts, at different concentrations and different temperatures.

Controlled dissolution of polished AT plates in $NaOH,xH_2O$.

The initial roughness value of the polished AT plates is Ra=0.004 µm, which corresponds to an optical polish surface state. These first experiments were undertaken to provide us with some knowledge on quartz behavior in strongly basic solutions. The best solution seems to be the most concentrated one, figure 5 . Indeed, for the more diluted medium, $NaOH,3H_2O$, the Ra value increases strongly from 0.004 to 0.09 µm after a thinning down of only 23 µm. Light microscopic observations (x1100) show regular and oriented grooves.

On the other hand, for the most concentrated solvent, $NaOH.H_2O$, the roughness of the plates and their surface texture (x1100) does not differ from the initial values. For this reason, this last concentration was largely used to carry out the investigation of ground AT-cut wafers in basic solutions.

Wafers used in industry generally have a surface texture which is ground, and there is thus great interest in determining conditions which lead to their good controlled dissolution. It is for this reason that chemical lapping has thus been investigated for concentration, temperature and, initial roughness of samples. The dissolution kinetics of AT quartz plates in soda have been determined from the results of these studies.

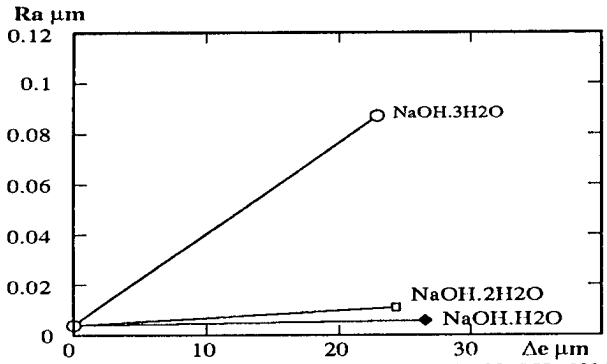


Fig 5: Polished AT plates lapped by NaOH.xH2O

(Fig 5 is only an illustration of these tests, Ra is not a linear function of the decrease in thickness Δe).

Controlled dissolution of ground AT plates.

Concentration influence : Following the results from previous polished samples studies, only concentrated media have been used. Figure 6 summarizes the evolution of Ra parameter plotted against thinning down for the three concentrations : $NaOH,2H_2O$, $NaOH,H_2O$ and $NaOH,0.56H_2O$. We can observe the same behavior for all concentrations with, perhaps, a slightly worse decrease of roughness (or not so good controlled dissolution) for the less concentrated one. Photos 1 show the dissolution pattern of an AT cut in $NaOH,H_2O$.



Photo 1: A:Prior to lapping, B:Δe=24 µm, C: Δe=110 µm. Magnification: x 440

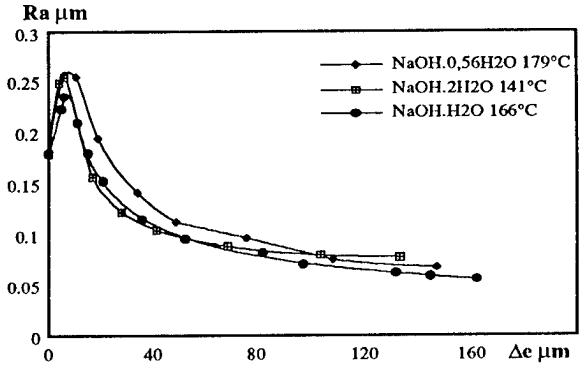


Fig 6 : Roughness parameter plotted against removed depth for several concentrations

On the other hand, no significant difference can be observed between the two more concentrated solutions. Thus, from a stability point of view, the $NaOH,H_2O$ concentration has been chosen for further experiments.

Temperature influence : For this purpose, discs with 6 µm in diameter were used with an initial frequency ~ 10 MHz (e=167µm) and a roughness value in the range 0.16<Ra<0.18 µm. Some evolutions of the thinning down rates against the removed thickness have been plotted in figure 7.

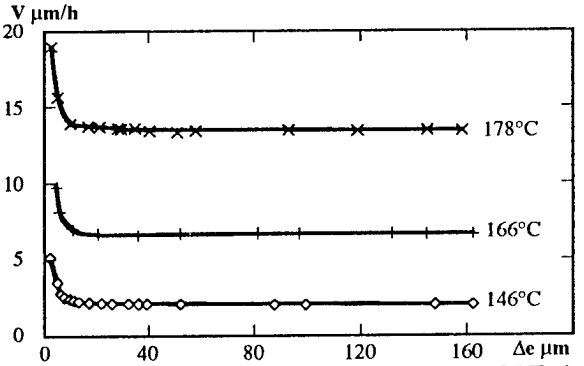


Fig 7 : Dissolution rate at different temperature of AT plates plotted against removed depth in NaOH.H2O

As expected, these dissolution rates increase with temperature. All the curves, figure 7, can be divided in two parts: - the first part, related to a rapid decrease of the dissolution rate, corresponds to the chemical lapping of the surface layer disturbed by cutting and polishing.

- in the second part, the dissolution rate becomes constant and is characteristic of the intrinsic dissolution rate for a given orientation of the material at the chosen temperature.

The first part corresponds to a total dissolution depth of 7 µm, i.e. 3.5 µm per face. Of course, the same result can be found for the roughness parameter, from the curves Ra, plotted against the thinning down depth, Δe,

figure 8. Previous studies on quartz [3] and berlinite [12-13] samples have already shown this behavior.

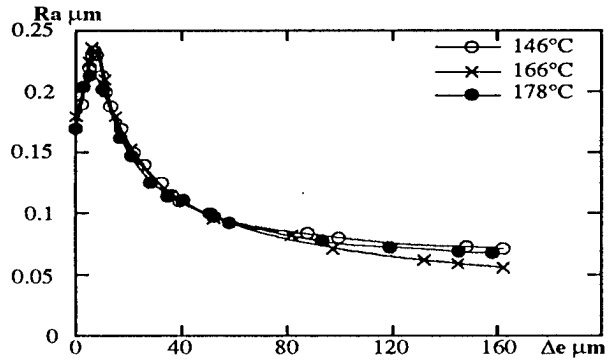


Fig 8 : Roughness parameter Ra plotted against removed depth in NaOH.H2O

Whatever the temperature, the Ra=f(Δe) plots show a similar maximum for a chemical lapping of about 7 µm (for the two faces), and then slowly fall to a low value close to 0.06 µm, which corresponds to a good controlled dissolution of quartz material.

Since the roughness value stabilization was very slow, we tried to define a final Ra value after a longer chemical lapping time. 630 µm were thus removed from a 800 µm thick wafer, and the Ra value decreased from 0.19 to 0.03 µm, and would certainly have continued to converge to a polish surface state. On the other hand, the isotropic effect of this thinning down process should be noted, which is enhanced by no alteration in appearance of the circular shape of the quartz discs, even after removing several hundreds µm.

In conclusion, the temperature of chemical lapping has no effect upon the final roughness value, but only upon the thinning down rate.

Initial roughness (Rai) influence : From previous experiments in NaOH,H_2O medium, it appears that a slight change in the initial roughness of the wafers (0.16<Ra<0.18 µm) corresponds to a slightly different final surface state . It seems that initial roughness influences the stabilization of the surface state in terms of thinning down, and thus several AT plates having different Rai values, Table 3, were dissolved at 166°C.

Table 3 - Characteristics of plates dissolved at 166°C.

	Ø mm	initial frequency MHz	Rai µm
plate 1	8	21	0,15
plate 2	6	10	0,18
plate 3	5	8	0,27
plate 4	5	25	0,004

The roughness evolution of these plates is plotted in figure 9. At first, we noted that the increase of Rai disappeared with the improvement of the surface state from 0.27 to 0.15 µm. We observed, however, that a low Ra value, Ra=0.07 µm, was reached after a chemical lapping depth of only 40 µm for sample 1(Rai=0.15 µm). On the other hand, 160 µm must be removed for sample 2 with worse initial roughness (Rai=0.18 µm) before reaching the same value. These results confirm previous research on quartz sample preparation [19-20] and emphasize its importance for the chemical lapping process.

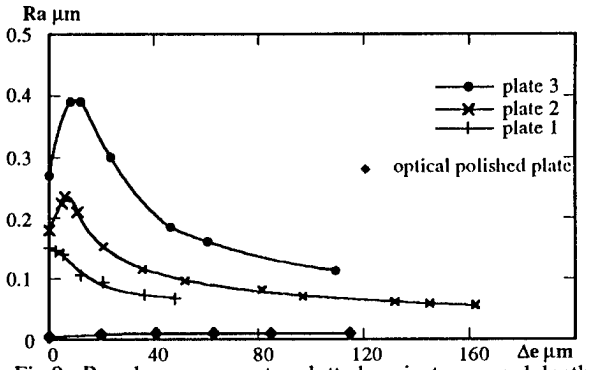


Fig 9 : Roughness parameter plotted against removed depth of differents AT plates in NaOH.H2O at 166°C

The effect of initial roughness is due to the slow polishing effect of the solvent used. Indeed, a high initial roughness corresponds to a thicker disturbed layer and leads to a more irregular bulk surface after the removal of this disturbed layer as schematized in figure 10. The thinning down, to reach the same surface state, must increase with initial roughness value.

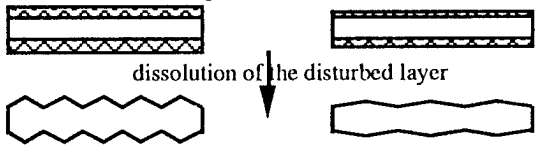


Fig 10 : Crystal states under the disturbed layer

In conclusion, the process of improvement of quartz AT plate roughness is significantly accelerated by a smoother surface state before controlled dissolution.

Dissolution kinetics of AT quartz plates in NaOH,H_2O and NaOH,$2H_2O$: We have plotted in figure 11 the evolution of dissolution rate against temperature. Dissolution rate being strongly influenced by temperature, this parameter must be carefully checked during the adjustment of the resonator frequency. The exponential variation corresponds to a thermically activated mechanism, following Arrhenius's law :

$$k_T = A \exp(-Ea/RT)$$

k_T being directly related to the thinning down rate. Ea can be calculated from :

$$Ln(V) = f(1/T) \quad \text{with} \quad V = A' \exp(-Ea/RT)$$

with V = thinning down rate in µm/h, Ea = activation energy in kJ, and A' = the pre-exponential term including the frequency factor A and the solution concentration.

Indeed, the variation of Ln(V) against 1/T is a straight line. This graph determines the Ea value. Thus the activation energy corresponding to this dissolution process is Ea=91 kJ.mole-1 for NaOH,H_2O and 88 kJ.mole-1 for NaOH,2H_2O.

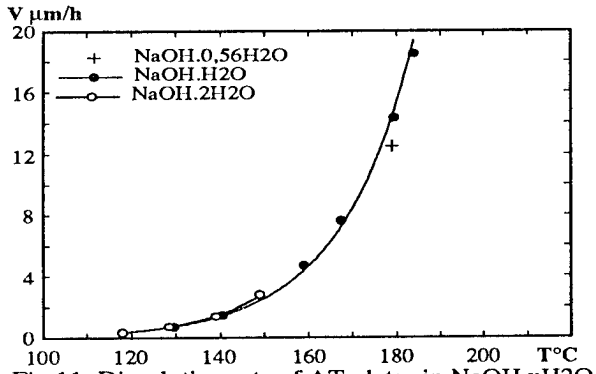


Fig 11: Dissolution rate of AT plates in NaOH,xH2O against temperature

As solvent NaOH,2H_2O cannot be used below its boiling point of 158°C, it has only been possible to measure the dissolution rate in the low temperature part of the graph. Dissolution rate is little affected by temperature in this field, and, it is difficult to determine the slight difference between the two dissolution rates. The margin of error is greater than for NaOH,H_2O, since there are only a few values available for calculating the activation energy.

We can conclude from these results that the best conditions for controlled dissolution of quartz AT plates appear to be in NaOH,H_2O medium. Temperature has no effect upon the final surface state, but only upon the dissolution rate, whereas the initial roughness of sample can strongly affect the improvement of plate roughness.

In the last part of our report, we present plate dissolution processing of an other very used orientation, the SC-cut.

Controlled dissolution of SC plates in NaOH,xH_2O.

Controlled dissolution of polished SC plates in NaOH,xH_2O

Table 4: polished SC plates in NaOH,H_2O

plate	initial Ra µm	Δe µm	final Ra face 1 µm	final Ra face 2 µm
1	0,005	21,9	0,008	0,008
2	"	11,7	0,008	0,008
3	"	9	0,006	0,006
4	"	62	0,008	0,008
5	"	174	0,01	0,01

As with the AT study, first the dissolution process of polished wafers was investigated in soda, the results being summarized in Table 4. All the plates were thinned down at 174 °C and the surface state evolution controlled. The maximum Ra value, obtained after 174 µm removed, was 0.01 µm, which corresponds to an optical polished surface state. Meanwhile, for all tests, etch pits began to appear on one side of the plate after ~60 µm was removed (after that, Ra value did not take into account these etch pits).

Controlled dissolution of ground SC plates in NaOH,H_2O : An example of surface texture evolution is plotted in figure 12 and shows some differences between the two faces of the sample. Nevertheless, the Ra value for both sides converges after removal of about 200 µm . Unfortunately, as with polished discs, some etch pits appeared on one side after a thinning down of 60 µm and made this medium unusable for a good controlled dissolution of SC plates.

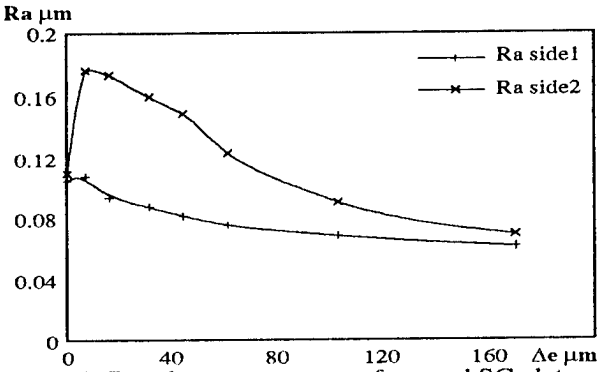


Fig 12: Roughness parameter of ground SC plates plotted against removed depth in NaOH,H2O at 180°C

As a result of these findings, other dissolution conditions were investigated for SC plates, in particular the use of other solvents. Initial results of SC plate dissolution in KOH,xH_2O will be presented.

Controlled dissolution of SC plates in KOH,xH_2O.

As with the other solvents, this study has been carried out in two parts: surface texture and kinetic evolution.

Surface state evolution : SC plates with different initial roughness values have been investigated. When the plate is polished (Rai=0.004 µm), chemical lapping in potassium hydroxide bath leads to the preservation of the surface texture (Ra=0.005 µm on both faces) with an alveolate relief, even after a significant removal of thickness.

If plates are initially ground, as for NaOH.xH_2O, the evolution of the surface roughness is different for the two faces, as can be seen in figure 13. For face 1, with higher initial roughness, the Ra value first increases before the polishing effect takes place. For the other side, face 2, a quicker decrease of roughness is observed. However after removal depending on the initial surface state, the

to 0.03 µm (it must be noted that this value is already close to 0.05 µm after a thinning down of only about 50 µm).

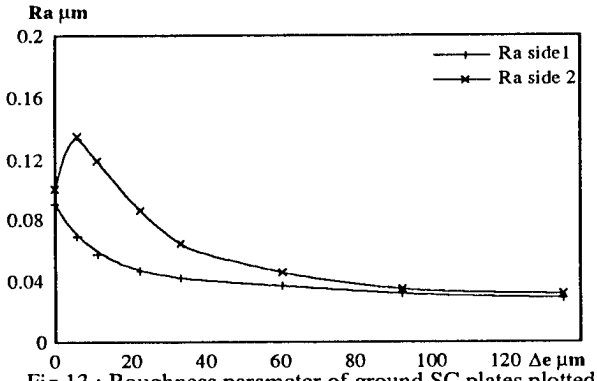


Fig 13 : Roughness parameter of ground SC plates plotted against removed depth in KOH,xH2O at 199°C

In conclusion, potassium hydroxide solvent is very good for SC-cuts, but the roughness evolution is, as already mentioned for AT-cut, closely related to the initial surface state.

Kinetics of dissolution of SC plates : The dissolution rate against the thinning down (Δe) follows the same evolution as that mentioned for AT plates. After removal of the disturbed layer, dissolution rate becomes a constant. The influence of the temperature on this rate is shown in figure 14. The exponential variation due to this parameter corresponds to an activation energy of 93 kJ.mole.l-1, whereas the rate values in the range 1 to 8 µm.h-1 are always lower than those for AT plates in NaOH medium at the same temperature.

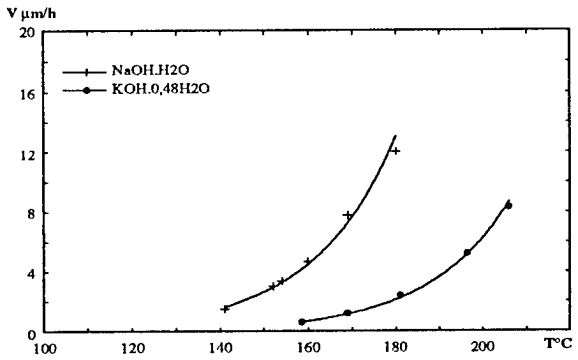


Fig 14: Dissolution rate of SC cuts in different solvents

Comparative conclusion of of AT and SC quartz cuts controlled dissolution.

From these first results, and as expected, the dissolution process is largely affected by the crystallographic orientation of cuts and the sample preparation.

First, for both orientations, the decrease of roughness is closely related to the initial surface texture, and the final smooth figures are different in terms of the crystallographic orientation.

Secondly, the best solvent for a good controlled dissolution seems to differ with crystallographic orientation: NaOH for AT-cuts and KOH for SC-cuts. Work is in progress to provide a satisfactory explanation for this problem.

From a thermodynamic point of view, activation energy is similar for both solvents and both orientations as shown in Table 5.

Table 5 : Activation energy of AT and SC cuts in different solvents

	Ea kJ.mole^{-1}	
Solvents	AT cut	SC cut
NaOH,H2O	91	88
NaOH,2H2O	88	
KOH,xH2O		93

As a result of these studies, we have started to produce and characterize resonators made from chemical lapping of AT cuts. Initial results are given in the next section.

Piezoelectric characterization of quartz resonators made from chemical lapping.

To check the efficiency of the chemical lapping process, piezoelectric devices were produced after controlled dissolution of AT quartz plates. In a first step, some high frequency AT resonators were tested to check their Q-factor and the reproducibility of the process. Then, four resonators were chosen to test a four pole pass-band linear phase filter.

Sample preparation.

The objective was to realize 83 MHz resonators after metallization, i.e. 84.5 ± 0.2 MHz before (in fundamental mode). The final thickness of these samples, ~20 µm, required a stiffening of the external part of the plate by boring a central well. In practice, mechanical lapping first adjusted the resonator frequency to a value close to 25 MHz, (~70µm) and after setting up a safety device on the external part, a 49 MHz central frequency was then reached by I.B.E. process.

At this step, controlled dissolution of these plates was performed in sodium hydroxide solutions at 170°C with a dissolution rate near 10 µm.h-1 for both sides of the wafer. Table 6 summarizes the main results for six different wafers.

Table 6 : Reproducibility of the dissolution rate for different wafers. The final frequency could be achieved with one thinning down step.

Plates	Time mn	initial and final frequency MHz		removed depth µm
1	85	49,31	84,56	14,1
2	85	49,33	84,63	14,1
3	85	49,37	84,63	14,1
4	83	49,71	84,72	13,9
5	85	49,32	84,57	14,1

Piezoelectric characterizations.

Two kinds of measurements were used:

-Air-gap method : The air gap method is a technique well adapted to the chemical lapping process because sample measurements can be taken without adhesive metallic electrodes. Some results from this method are given in Table 6, where all measurements are in the predetermined range, 84.5 ± 0.2 MHz.

-Resonator measurements : For two pole resonators, metallic electrodes are deposited on the two faces of the plates by evaporating technique and connected to a network analyser. Table 7 and figure 15 summarize comparative characteristics of resonators obtained from controlled dissolution and I.B.E. processes. For these first results, the less atraumatic chemical lapping seems to give specifications as least as good as those of ionic beam etching.

In a last step, four pole linear phase filters were made from two chemically lapped resonators and two I.B.E. ones. These filters were characterized by a frequency response which must be chalk up in a gauge (specification) defined by:

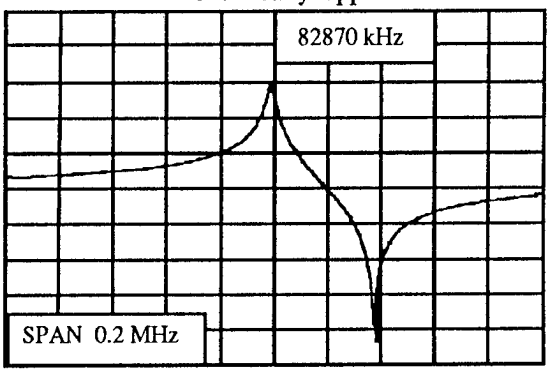
- a bandwidth for a maximum insertion loss (IL) allowed: Amax;
- an attenuated band for a minimum insertion loss allowed: Amin.

Table 7 : Electrical parameters of resonators obtained from controlled dissolution and I.B.E. processes.

	resonators	
	IBE	chemically lapped
Fr KHz	82873	82870
R Ω	28,59	27,02
L mH	2,894	3,3721
Q	52710	65000
Q.Fr	4,37E12	5,38E12

The characteristics of these filters are given by Fc, central frequency, BW, bandwidth at n dB and Ba, attenuated band at n dB. Thus, these linear phase filters are also characterized by a constant propagation time t (delay) in a part of bandwidth. Filter characteristics and specifications are given in Table 8, and frequency response in figure 16 .

Fig 15 : Frequency response of resonator with plates chemically lapped



The characteristics here also confirm the attraction of an atraumatic controlled dissolution process. These piezoelectric results show that the surface state reached is sufficient for BAW devices to be made.

Table 8 : Electrical parameters and specifications of linear phase filters.

			Filters	
		Specifications	Chemically lapped	IBE
Fc KHz		83000	82986	83005
BW (KHz) à 3 dB		180	200	213
IL dB		5,5	3,9	3,0
BA (KHz)	25 dB	Fc+/-200,0	+175,0-183,8	≈+/-180
	40 dB	Fc+/-285,0	+252,9-252,9	≈+/-260
	50 dB	Fc+/-1000,0	+312,5-306,3	≈+/-330
Out of band attenuation dB		60 db at +/- 30,0 MHz	-87,3+89,0	> 80dB

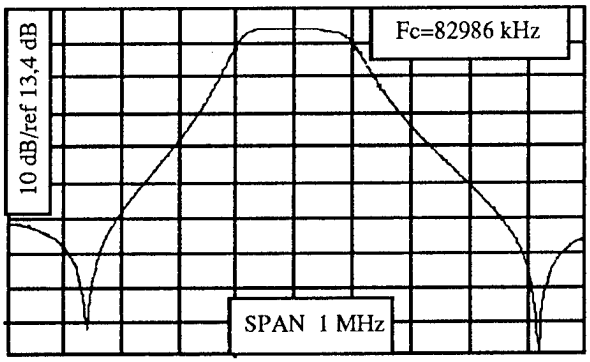


Fig 16 : Linear phase filter manufactured with chemically lapped plates.

Conclusion

Before undertaking experiments on berlinite and quartz controlled dissolution, and taking into account previous failed attempts at quartz lapping, our primary objective was to obtain a good surface texture: without etch pit formation, and with accurate thickness.

Previous studies have indicated that crystal growth and crystal dissolution are controlled by the same laws, and thus the same media must be used in order to avoid all crystal growth and dissolution processes defects.

Care has been taken not use the term "etching" in this report, to avoid any possible confusion with the controlled dissolution. The term "etching", from our point of view, should be reserved for irreversible processes as quartz in fluoride media, whereas controlled dissolution implies that the phenomenon can be reversed, as with quartz in basic solvents. In other terms, solubility and reaction are quite different processes.

Consequently, chemical lapping of berlinite AT-cuts has been developped in acid media, and that of AT- and SC-cuts in basic solvents. Conditions have been specified for both materials which result in a good surface state and the thickness defined by the required frequency [21].

Acknowledgements: The authors wish to thank C.N.R.S for financial support and C.E.P.E/Thomson for providing quartz samples and piezoelectric device characterizations.

References

[1]: K. Bräuer, E. Müller, Crystal & Technol, 101, 1984.
[2]: K.H. Jones, 41st Annual Freq. Control Symposium, 199, 1987.
[3]: C.R. Tellier, Surface Technology, 21, 83, 1984.
[4]: A.J. Bernot, IEEE, 271, 1985.
[5]: R.J. Brandmayr, J.R. Vig, 39th Annual Frequency Control Symposium, 273, 1985.
[6]: R.J. Brandmayr, J.R. Vig, 40th Annual Frequency Control Symposium, 86, 1986.
[7]: E. D. Kolb, R.A. Laudise, J. Cryst. Growth, 43, 313 1978.
[8]: B.H.T. Chai, M.L. Shand, E. Buehler, M.L. Gillco, Proc. IEEE Ultrasonics Symp., 577, 1979.
[9]: J. Détaint, H. Poignant, Y. Toudic, Proc. 34th Freq. Control Symp., 93, 1980.
[10]: E. Philippot, A. Goiffon, M. Maurin, J. Détaint, Schwartzel, Y. Toudic, B. Capelle, A. Zarka, J. Cryst. Growth, 104, 1990.
[11]: E. Philippot, J.C. Jumas, A. Goiffon, M. Maurin, J.C. Doukhan, A. Zarka, B. Capelle, Y. Toudic, J. Détaint, J. Schwartzel, French patent n° 531 149.
[12]: O. Cambon, A. Goiffon, E. Philippot, J. Mat. Sci., 26, 846, 1991.
[13]: O. Cambon, A. Goiffon, A. Ibanez,E. Philippot, Eur. J. Solid State Inorg. Chem., 29, 547, 1992.
[14]: O. Cambon, A. Goiffon, A. Ibanez, E. Philippot, J. Solid State Chem., 103, 240, 1993.
[15]: W.K. Burton, N. Cabrera, F.C. Frank, Phil. Trans. Soc. London, A243, 299, 1951.
[16]: N. Cabrera, M.M. Levine, Phil. Mag., 1, 450, 1955.
[17]: N. Cabrera, M.M. Levine, J.S. Plaskett, Phys. Rev., 96, 1153, 1954.
[18]: W.G. Johnston, Prog. Ceram. Sci., 2, 1, 1962.
[19]: C.R. Tellier, J.L. Waterkowski, Surf. Technol., 15, 275, 1985.
[20]: Y. Sekiguchi, H. Funakubo, J. Mater. Sci., 15, 3066, 1980.
[21]: E. Philippot, O. Cambon, A. Goiffon, A. Ibanez, D. Cachau, Patent n° 91.09170.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ETCHING OF QUARTZ CRYSTAL SPHERES

Roger W. Ward

Quartzdyne, Inc.
1020 Atherton Drive, Bldg. C
Salt Lake City, Utah, USA 84123
801-266-6958

ABSTRACT

Natural quartz crystal spheres were etched in various etchants at room temperature and at 80°C. The resulting quartz pieces were scanning electron micrographed normal to the resulting major plane and at right angles to the plane for the purpose of illustrating the anisotropy of the etch rate of quartz as a function of etchant and temperature.

INTRODUCTION

In 1892, Meyer and Penfield [1] published a treatise on "Results Obtained by Etching a Sphere and Crystals of Quartz with Hydrofluoric Acid," in which they etched a 2.44 cm diameter sphere of natural quartz in hydrofluoric acid (HF). This study revealed that quartz etches very anisotropically. In the past 15-20 years, there has been a revival of interest in the etching of essentially two-dimensional pieces of quartz—for example, quartz resonators, to improve their aging and strength; and for quartz wafers for photolithographic manufacturing of three-dimensional devices: quartz tuning forks and thickness-shear mode strip resonators. The first application is mainly concerned with surface effects/defects, i.e., etch pits and bumps, while the second application is concerned with surface defects and with etch facets on the edges of the finished devices.

The present work studies the macro effects of three-dimensional etching of quartz spheres in commonly used etchants. A quantity of 12 mm diameter polished natural quartz spheres were obtained. Ten spheres were placed into each of several etchants, including 40% HF:60% NH_4F, saturated $NH_4F{\cdot}HF$, and LATTICE ETCH™, at each of two nominal temperatures: 20°C and 80°C. The spheres were removed from each bath one at a time as their minimum diameter was reduced in increments of about 0.8 mm. Representative samples from the resulting ~60 spheres, ranging from nearly spherical to distinctly lenticular in shape, were arranged and photographed from different angles to allow the study of the anisotropic etching of quartz as a function of etchant and etchant temperature.

BACKGROUND

Meyer and Penfield's [1] classic work on the etching of a quartz crystal sphere often has been cited for illustrating the severe anisotropy in the etch rate of quartz. Briefly, Meyer and Penfield etched a 2.44 cm natural quartz sphere from Japan in "rather a strong commercial hydrofluoric acid, such as can be bought in rubber bottles from dealers in chemicals." This would indicate that the HF was probably from 49% to about 75% concentration. The HF was changed daily, with photographs of the sphere made after "a few hours," "about two weeks," and "for about a month longer." The resulting etched sample is shown in Fig. 1. (The specimen was later "deposited in the collection of Professor George J. Brush, New Haven, Conn.")

There are two main applications for chemical etching of quartz crystal resonators: 1) to improve their surface finish and 2) to create three-dimensional quartz devices by chemical micro-machining of a quartz wafer.

John Vig's *et al.* now historic work [2] on the etching of quartz resonators to improve their strength renewed the quartz crystal industry's interest on the effects of chemical etchants for the surface treatment of quartz plates. This work created the current interest in "etch channels," sweeping of quartz to reduce etch channels, and research on various etchants to chemically thin quartz blanks and wafers for various applications.

In addition to the industry's interest in etchant-related surface effects/defects on quartz blanks and wafers, the manufacture of three-dimensional quartz micro-machined devices [3], best typified by the quartz tuning fork for wrist watches [4], has also created a great interest in the effects of various etchants upon quartz, especially for the reduction of etch facets—typically, straight-walled devices are preferred over faceted walls.

Other important papers on the effects of etchants on quartz are by Bond [5] and Frondel [6]. Other papers of interest on the subject of etching and micro-machining of quartz are given in the References [7]-[13]. Of particular interest is the recent (1993) work [13] of Tellier *et al.* on theoretically predicting etch shapes in quartz structures. Three recent papers on three-

0-7803-0905-7/93 $3.00 © 1993 IEEE

dimensional etching of silicon (including the etching of silicon spheres) are also listed [14]-[16].

CURRENT WORK

About 200 natural quartz crystal spheres of 12.0-12.3 mm diameter were obtained from a local supplier [17]. [According to the supplier, the source of the quartz for the spheres was from a 40 ton (36,000 kg) lot of quartz from Minas Gerais, Brazil, via a cutter in China.] The diameter of each sphere varied by less than 0.03 mm. By visual inspection, over 50% of the spheres contained gross (>20% by volume) defects—veils, inclusions, and fractures. The visually best spheres were chosen for the etch studies; often, no visible defects were present.

The etchants for the study were the ones most commonly used today for quartz etching: saturated ammonium bifluoride ($NH_4F{\cdot}HF$) "ABF," commercial grade; 40% HF:60% NH_4F by volume "40%/60%," (standard aqueous HF is 48-51% HF by weight; standard aqueous NH_4F is 40% NH_4F by weight), electronic grade; Lattice Etch™ [18] "LATTICE," fresh sample; and Superwet BOE® (Buffered Oxide Etch) 5:1 [19] "BOE," eight years old. (Several attempts were made to obtain a fresh sample of BOE®; General Chemical only sells $300 minimum lot size. A 1 liter sample was promised but never arrived.) The solutions were placed into covered Teflon® tanks and maintained at either room temperature "RT" (nominally 20-24°C) or at 80±2° C "80C." The Lattice Etch™ was not used hot due to excessive levels of rotten egg smell, even though the etches were kept covered inside a fume hood.

The spheres were cleaned by boiling in a solution of Liqui-Nox® and deionized (DI) water for 5 minutes, rinsed in DI water, and then placed in a solution of 50 ml DI water, 50 ml hydrogen peroxide, and 250 ml sulfuric acid ("Piranha Etch") until the bubbling stopped, followed by a DI rinse. Ten spheres were etched in each etchant at each temperature. The volume of each etchant was approximately 1 liter. The spheres were placed in Teflon® carriers to prevent them touching one another and to raise them off the bottom of the bath (this is especially important in saturated $NH_4F{\cdot}HF$ solution since contact with undissolved $NH_4F{\cdot}HF$ on the bottom of the bath will cause non-uniform etching of the quartz). As the minimum dimension of each sphere was decreased by approximately 0.8 mm, one sphere was removed from each etch bath until all the spheres were removed. The etch solutions were unstirred except for the disturbance caused by removing the spheres for measurement and inspection. The etch solutions were not changed or renewed throughout the course of the experiments.

Two specimens from each bath were selected for scanning electron micrographing, based upon their degree of etching and relative perfection. The majority of the spheres, even though originally selected for their visual perfection, showed moderate to severe imperfections upon deep etching.

EVALUATION

In general, the micrographs speak for themselves. However, a few observations:

- ABF and 40%/60% at 80°C show nearly the same shapes. See Fig. 4(a) and Fig. 5(a).
- All the etchants exhibit distinctly different figures at room temperature versus 80°C (see Fig. 4). This is consistent with earlier reports [20] that crystal blanks etched at room temperature are generally "smoother" (i.e., different) than similar blanks etched in the same etchant at elevated temperatures.
- BOE exhibits a nearly square figure (Fig. 3(b)), as opposed to the trigonal figure of all the other etchants. Is this a result of the age of the solution? It is not clear how this could be caused by changes in the etchant due to time.

NOTES

1. The spheres etched in 40%/60% for >20 days at room temperature showed poor symmetry that did not appear to be caused by crystal defects. No micrographs were taken of these samples.
2. Many of the etched samples exhibited one "rough" +X-axis corner after moderate to deep etching, which appears to be related to crystal defects in the original sphere. It is unknown why so many samples would exhibit this specific defect mechanism.
3. The black circles in a few of the micrographs are caused by electron charging of the non-conductive quartz (i.e., Fig. 5(c)), even though the samples were sputter coated with CrAu to prevent charging.
4. The specimens and the original Polaroid® positives and negatives will be retained by the author indefinitely. They will be loaned to anyone wishing to study them, upon request.

ACKNOWLEDGMENTS

The author thanks Judy Gillen, Quartzdyne, for conducting the etch experiments, and Howard A.

Spencer, Materials Science and Engineering, University of Utah, for taking the scanning electron micrographs.

REFERENCES

[1] O. Meyer and L.L Penfield; "Results Obtained by Etching a Sphere and Crystals of Quartz with Hydrofluoric Acid," *Trans. Conn. Acad., Vol. VIII*, 1888 to 1892, pp. 158-167.

[2] J. R. Vig, J.W. LeBus, and R.L. Filler, "Chemically Polished Quartz," *Proc. of the 31st Annual Symposium on Frequency Control*, 1977, pp. 131-143.

[3] J.H. Staudte, "Micro Resonators in Integrated Circuits," *Proc. of the 22nd Annual Symposium on Frequency Control*, 1968, pp. 226-231.

[4] J.H. Staudte, "Subminiature Quartz Tuning Fork Resonator," *Proc. of the 27th Annual Symposium on Frequency Control*, 1973, pp. 50-54.

[5] W.L. Bond, "Etch Figures of Quartz," *Zeitschritt fur Kristallographie*, Band 99, pp. 488-498, 1938.

[6] C. Frondel, *The System of Mineralogy*, Vol. III Silica Minerals, John Wiley and Sons, NY, 1962.

[7] J.K. Vondeling, "Fluoride-Based Etchants for Quartz," *J. of Materials Science*, 18, 1983, pp. 304-314.

[8] P. Suda, A.E. Zumsteg, and W. Zingg, "Anisotropy of Etching Rate for Quartz in Ammonium Bifluoride," *Proc. of the 33rd Annual Symposium on Frequency Control*, 1979, pp. 359-363.

[9] F. Augustine and D.R. Hale, "Topography and Etch Patterns of Synthetic Quartz," *J. Phys. Chem. Solids*, vol. 13, 1960, pp. 344-346.

[10] J.S. Judge, "A Study of the Dissolution of SiO_2 in Acidic Fluoride Solutions," *J. Electrochem. Soc.*, Nov. 1971, pp. 1772-1775.

[11] O.B. Vol'skaya, "Selective Etchants for Quartz," *Soviet Physics—Crystallography*, vol. 13, no. 4, Jan.-Feb. 1969, pp. 620-622.

[12] M.V. Kilday and E.J. Prosen, "The Enthalpy of Solution of Low Quartz (α-quartz) in Aqueous Hydrofluoric Acid," *J. of Research of the National Bureau of Standards-A. Phys. and Chem.*, vol. 77A, no. 2, Mar. 1973, pp. 205-215.

[13] C.R. Tellier and T. Leblois, "Bi-and Three-Dimensional Prediction of Etching Shapes in Quartz Micromachined Structures," This Symposium, to be published; 1993.

[14] C. Ju and P.J. Hesketh, "Measurements of the Anisotropic Etching of a Single-Crystal Silicon Sphere in Aqueous Cesium Hydroxide," *Sensors and Actuators A*, 33, 1992, pp. 191-196.

[15] C. Ju and P.J. Hesketh, "High Index Plane Selectivity of Silicon Anisotropic Etching in Aqueous Potassium Hydroxide and Cesium Hydroxide," *Thin Solid Films*, 215, 1992, pp. 58-64.

[16] C.H. Séquin, "Computer Simulation of Anisotropic Crystal Etching," *Sensors and Actuators A*, 34, 1992, pp. 225-241.

[17] Allen Dillman, Woodland Treasures Ltd., P.O. Box 417, Fairview, UT 84629; phone 801-427-3377.

[18] Available from XECO, 1579 N. Main Street, Cedar City, UT 84720; phone 801-586-0507.

[19] Superwet BOE® is a product of General Chemical.

[20] The author cannot locate the source of this information.

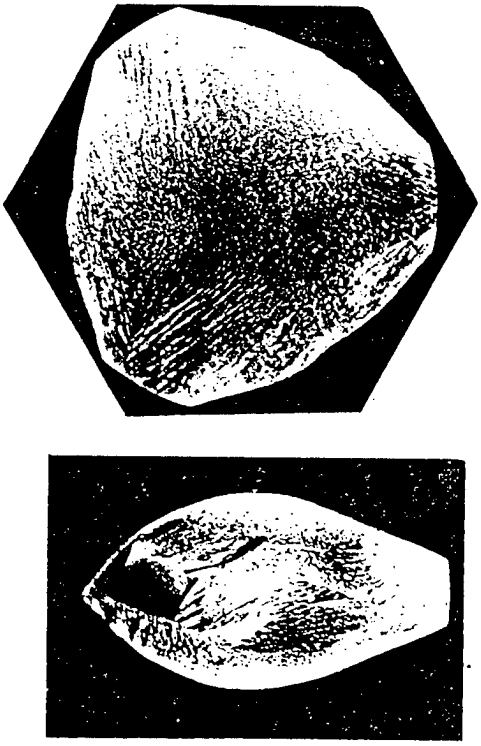


Fig. 1. The quartz sample of Meyer and Penfield [1] after etching in HF at room temperature for about 6 weeks.

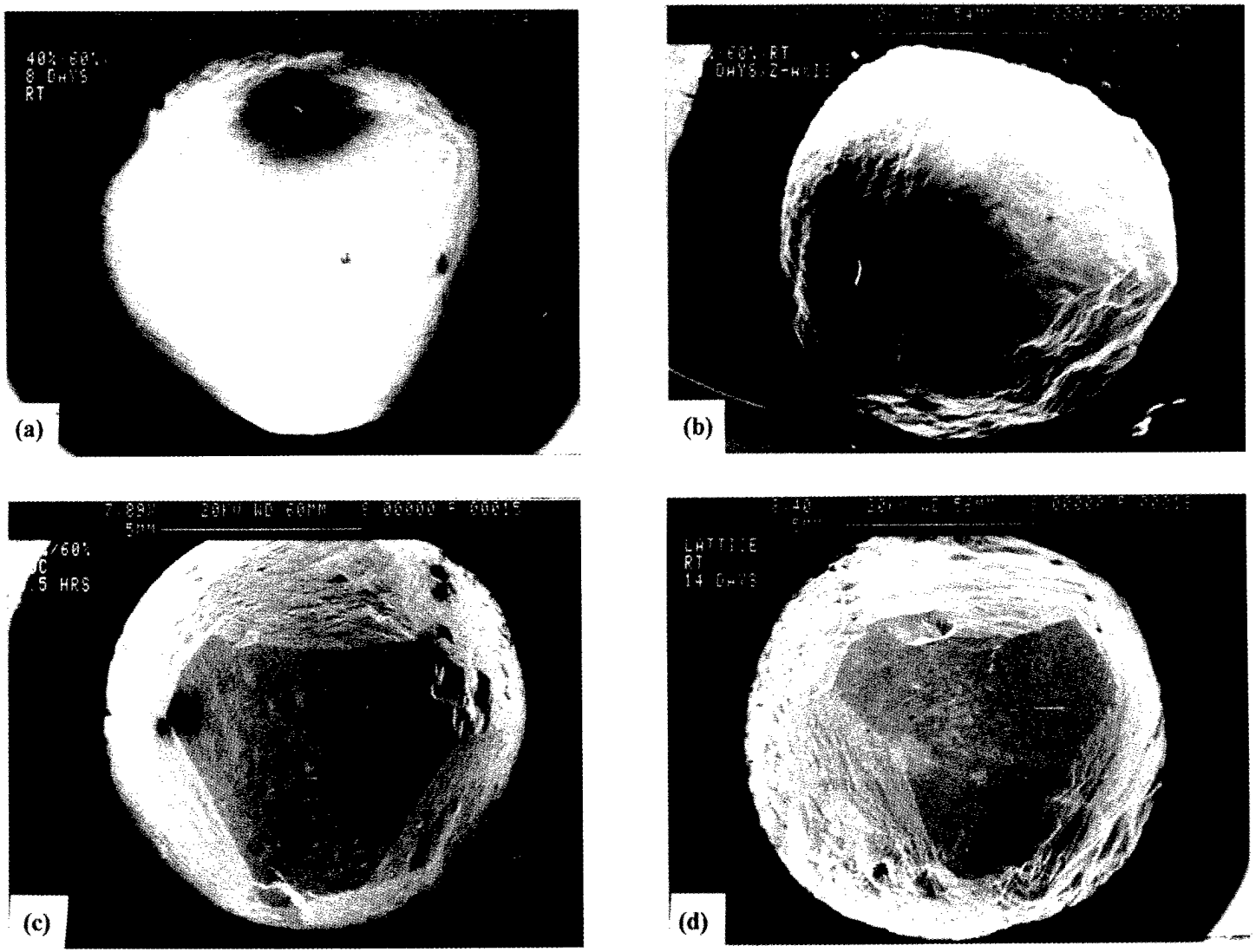


Fig. 2. The above samples are nearly spherical in shape so only one axis was micrographed. The Z-crystallographic axis is out of the page in each micrograph. (a) Etched in 40%/60% (see text) for 8 days at room temperature. (b) Etched in 40%/60% for 20 days at room temperature. (c) Etched in 40%/60% for 8.5 hours at 80°C. (d) Etched in Lattice Etch™ for 14 days at room temperature.

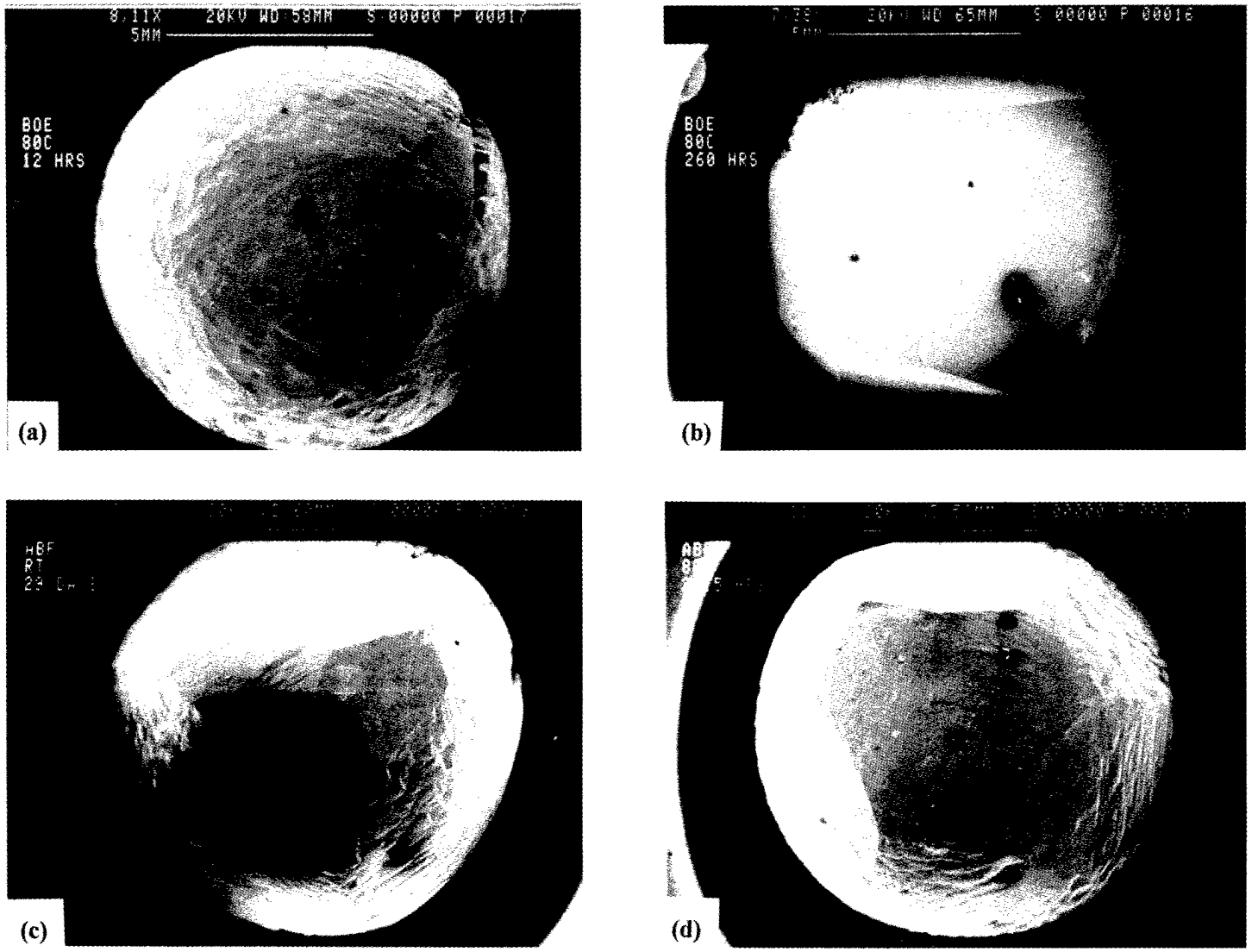


Fig. 3. The above samples are nearly spherical in shape so only one axis was micrographed. The Z-crystallographic axis is out of the page in each micrograph. (a) Etched in BOE (see text) for 12 hours at 80°C. (b) Etched in BOE for 10.5 days at 80°C. Note the nearly "square" shape (all the samples etched in BOE took on this shape) as opposed to the distinctly trigonal shape of all the other samples etched in other etchants. (c) Etched in ABF (see text) for 29 days at room temperature. (d) Etched in ABF for 21.5 hours at 80°C.

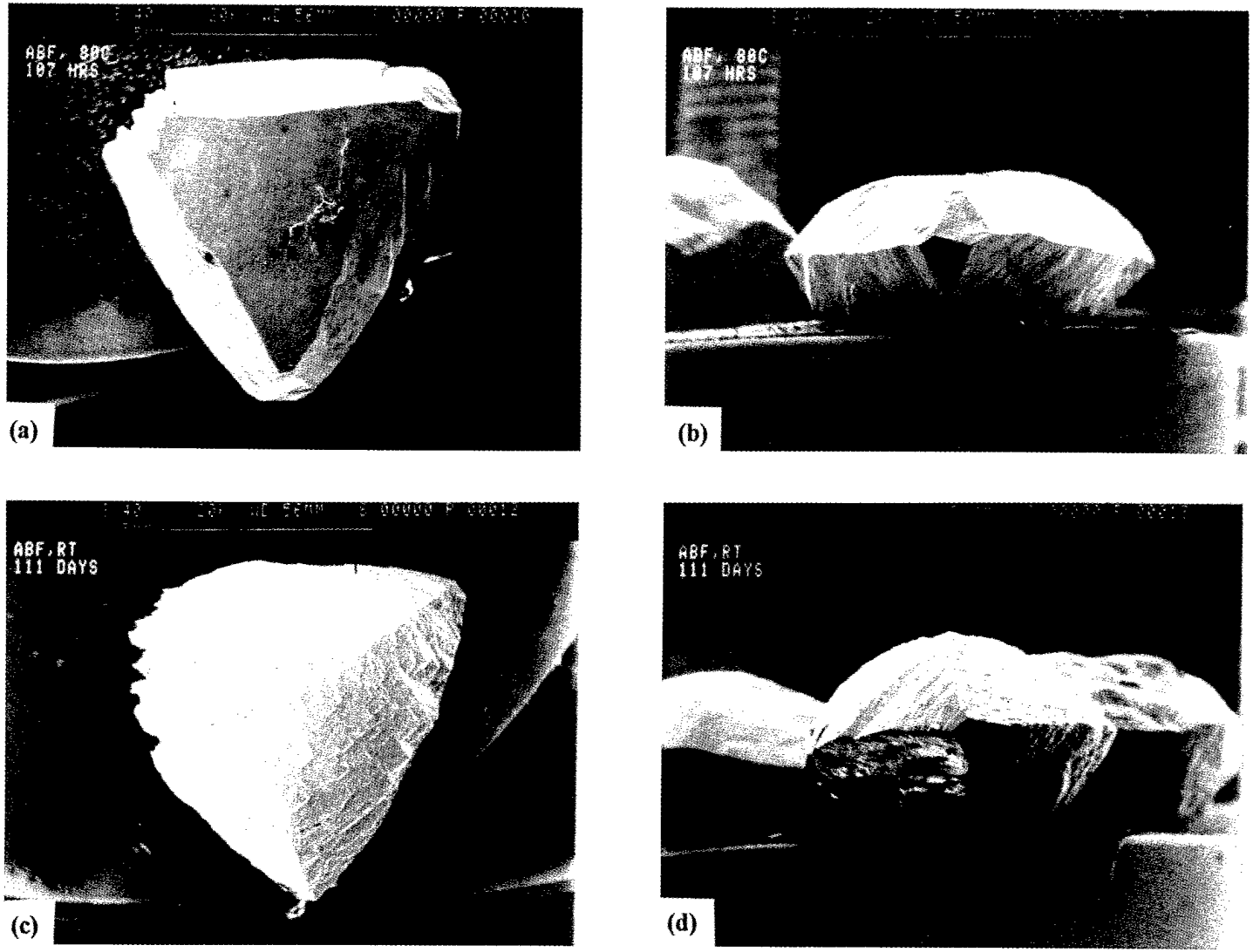


Fig. 4. The samples in (a) and (c) are viewed down the Z-axis; the samples in (b) and (d) are viewed down the +X-axis. (a) Etched in ABF for 4 days at 80°C. (b) Same as (a). (c) Etched in ABF for 111 days at room temperature. (d) Same as (c). Note the marked difference in the shapes of these samples etched in the same etchant at different temperatures. In production situations, etching AT-cuts in ABF at room temperature yields significantly smoother blanks and fewer etch channels than similar blanks etched at 80°C. Note also that one +X-end is etched off each sample (see text).

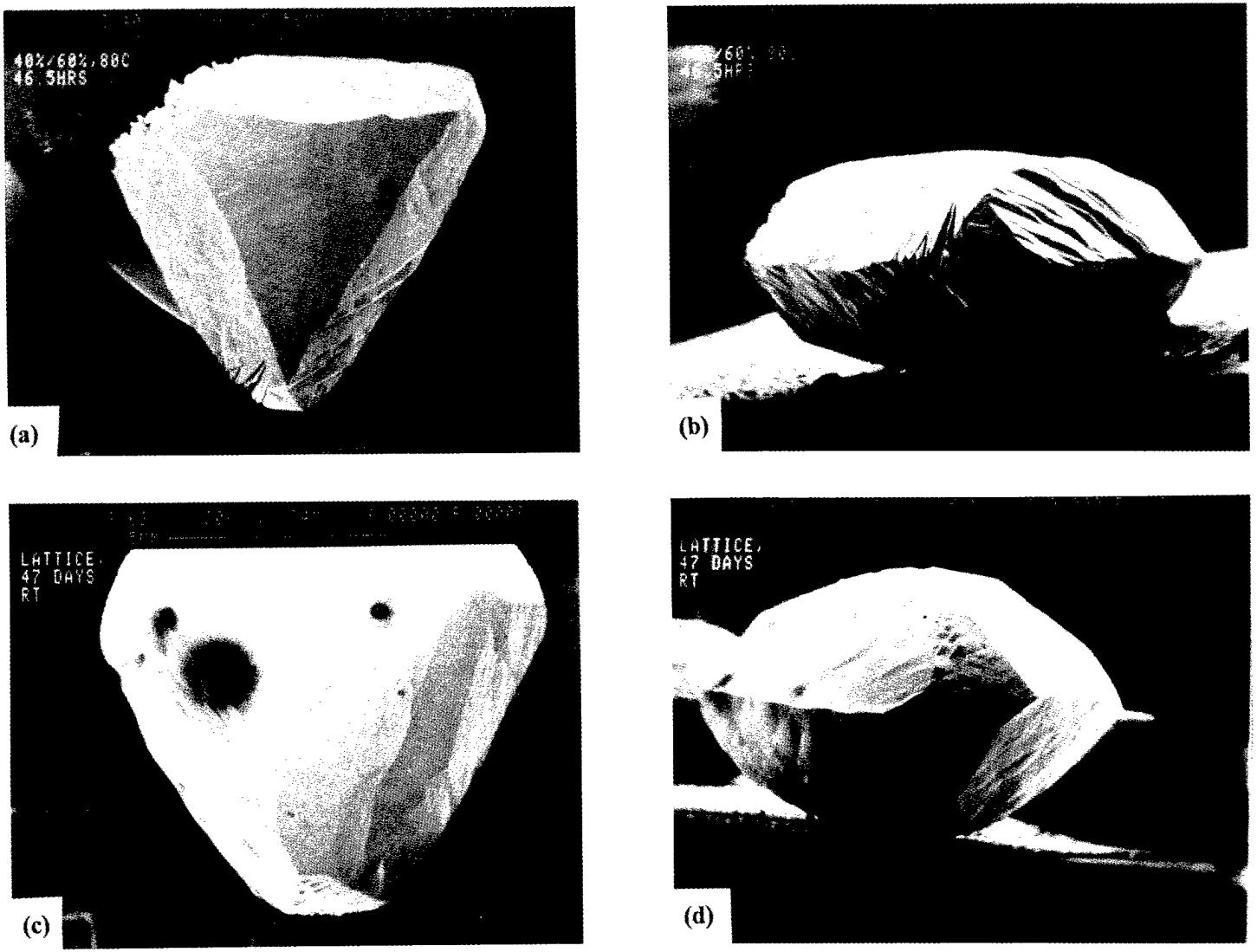


Fig. 5. The samples in (a) and (c) are shown normal to the Z-axis; the samples in (b) and (d) are shown normal to the +X-axis. (a) Etched in 40%/60% (see text) for 2 days at 80°C. (b) Same as (a). (c) Etched in Lattice Etch™ for 47 days at room temperature. (d) Same as (c). Note the similarities between 5(a) & (b) and 4(a) & (b).

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

BI - AND THREE-DIMENSIONAL PREDICTION OF ETCHING SHAPES IN QUARTZ MICROMACHINED STRUCTURES

C. R. TELLIER and T. G. LEBLOIS

LCEP/ENSMM - La Boulole - Route de Gray
25030 BESANCON CEDEX - FRANCE

ABSTRACT

The anisotropic etching behavior of quartz crystal in ammonium bifluoride solution was studied and analyzed according to the tensorial model of dissolution. The dissolution slowness surface was determined from experiments. Bi-dimensional and three-dimensional graphical simulations were presented and the adequation of the proposed slowness surface was discussed.

I - INTRODUCTION

The anisotropic etching of quartz crystals is a widely used process in the fabrication of micromachined mechanical devices which constitute the sensing element of quartz micro-resonators or micro-sensors. Thus the optimum design of the mechanical structure passes through a computer-aided prediction of he exact geometrical etched shapes. Effectively surface geometrical features, dissolution cross-sectional profiles at a mask edge as well as the corner undercutting are of prime importance for the metrological performance of resonators and sensors.

Over the past few years the prediction of etching shapes has often been based on the kinematic model of dissolution proposed by Frank **[1]** which only gives us the necessary tools to construct geometrically limiting etching shapes **[1,2]**. Recently new theoretical concepts have been developed for deriving analytical expressions of the dissolution slowness surface using a tensorial representation **[3]** of the anisotropic etching. Since the trajectory of a moving surface element can be determined from the equation of the slowness surface it becomes possible to construct numerically various etching shapes **[4]**.

All the procedures start with the analytical expression of the slowness surface which requires the experimental determination of a lot of dissolution constants. Thus in this paper we present extensive experimental results related to various quartz plates from which attempts are made to derive the slowness surface describing the chemical etching of quartz crystals in a concentrated $NH_4F.HF$ solution. Numerical procedures are then developed to derive etching shapes encountered in conventional crystal etching (bi-dimensional simulation) and in localized crystal dissolution (three-dimensional simulation for photolithography techniques).

II - THEORETICALS

In a kinematic three-dimensional description of the anisotropic dissolution process the surface of a crystal is decomposed into a succession of planar elements corresponding to crystallographic surfaces. To characterize the anisotropic dissolution process Tellier and co-workers **[3,5,6]** introduced a vector called the dissolution slowness vector **L**. This vector is associated to each surface element, its magnitude, $L(\varphi,\theta)$ is the reciprocal of the dissolution rate and its positive direction coïncides with that of the inward normal unit vector, **n**, to the surface element. It is obvious that as the orientation (φ, θ) of the surface element varies the extremity of **L** generates a representative surface called the dissolution slowness surface (Figure 1a).

If the dissolution process is only governed by orientation the trajectory of a surface element of orientation (φ, θ) follows a straight line **[7]**. Then the displacement of a moving profile element corresponds to a propagation vector **P** (Figure 1b) whose cartesian components can be evaluated from the equation expressing $L(\varphi,\theta)$ **[7]**.

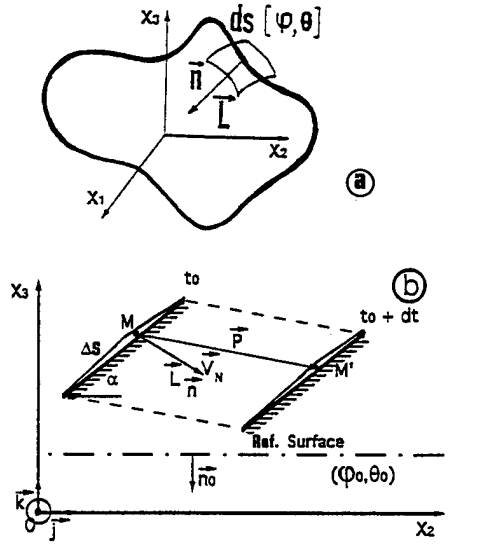


Figure 1 : **(a)** Geometry of the model, **(b)** Definition of the propagation vector **P**.

0-7803-0905-7/93 $3.00 © 1993 IEEE

Moreover the propagation vector **P** lies perpendicular to the plane tangent to the dissolution surface at the point of corresponding orientation. Using this geometrical property it becomes possible to construct geometrically approximate dissolution shapes **[1,2]**. A more precise prediction of etching shapes requires the knowledge of the equation $L(\varphi,\theta)$.

Adopting the convention specified in the IEEE standards on piezoelectricity the cartesian components of **n** are expressed as :

$$n_1 = \sin\varphi\cos\theta,\ n_2 = -\cos\varphi\cos\theta,\ n_3 = -\sin\theta \qquad (1)$$

As a result the magnitude $L(\varphi,\theta)$ can be written in terms of the components D_o, D_i, D_{ij}... of dissolution tensors **[3]**

$$L(\varphi,\theta) = D_o + D_i n_i + D_{ij} n_i n_j + D_{ijk} n_i n_j n_k + \ldots \qquad (2)$$

The number of dissolution constants is reduced by symmetry of the point group 32. For example for tensor of rank 9 we obtain the relations displayed in table 1. Then if we restrict our calculations to tensor of rank 9 only we obtain the final expression :

$$L(\varphi,\theta) = \cos^6\theta\sin\theta\sin3\varphi\cos3\varphi.\left[168\,D_{5(1)\,1(2)\,3(3)}.\sin^2\theta + 24\,D_{7(1)\,1(2)\,1(3)}.\cos^2\theta\right] - \cos^3\theta\sin3\varphi\left[36\,D_{7(1)\,2(3)}\sin^2\theta\cos^4\theta + 126\,D_{5(1)\,4(3)}\sin^4\theta\cos^2\theta + 84D_{3(1)\,6(3)}\sin^6\theta\right] + \cos^9\theta\sin3\varphi\left[4\cos^2 3\varphi - 1\right]D_{9(1)} \qquad (3)$$

$$D_{7(1)\,1(2)\,1(3)} = -3D_{3(1)5(2)1(3)} = -3D_{5(1)3(2)1(1)} = D_{1(1)7(2)1(3)}$$

$$D_{7(1)2(3)} = -7D_{3(1)4(2)2(3)} = -21D_{5(1)2(2)2(3)} = -\frac{7}{3}D_{1(1)6(2)2(3)}$$

$$D_{5(1)1(2)3(3)} = -D_{3(1)3(2)3(3)} = D_{1(1)5(2)3(3)}$$

$$D_{5(1)4(3)} = -5D_{3(1)2(2)4(3)} = \frac{5}{3}D_{1(1)4(2)(3)}$$

$$D_{3(1)6(3)} = -D_{1(1)2(2)6(3)} \quad ; \quad D_{9(1)}$$

Table 1 : Relations between dissolution constants $D_{N(i)\,M(j)\,P(k)}$ (i,j,k = 1,2 or 3) : The case of tensor of rank 9.

In reality to generate a slowness surface whose shape seems sufficiently complex with respect to the observed marked anisotropy we have performed calculations up to tensors of rank 13.

In this case simple but tedious calculations yield the following compact formulation :

$$L(\varphi,\theta) = S_1 + S_2 + S_3 + S_4 \qquad (4)$$

with

$$S_1 = \sum_{j=0} Aj\,(\cos\theta)^{2j}$$

$$S_2 = \sum_{L=0}\sum_{k=1} B_k(\cos\theta)^{2k+1}(\sin 3\varphi)^{3L}$$

$$S_3 = \sum_{M=0}\sum_{l=0} C_l(\cos\theta)^{2l+1}\sin\theta(\cos 3\varphi)^{3M} \qquad (5)$$

$$S_4 = \sum_{m=3} D_m\sin\theta(\cos\theta)^{2m}\cos3\varphi\sin 3\varphi$$

where L and M which are equals to zero for $N_{max} < 9$ appear for $N_{max} \geq 3^{L+1}$ and $N_{max} < 3^{M+1}$ respectively.

The coefficient of series are expressed in terms of independent dissolution constants **[3,6]** that must be determined experimentally. We observe that series S_1 and S_3 are associated with tensors of even rank whereas the series S_2 and S_4 come from tensors of odd rank.

III EXPERIMENTAL RESULTS

3.1 Experimental details

Singly and doubly-rotated quartz plates were mechanically lapped prior to etching. In addition the contour of the initially circular plates was optically polished. The etchant used was ammonium bifluoride with a typical concentration of 10.5 mol l^{-1}. Plates were etched at a constant temperature of about 330°K. The surface texture of etched plates were characterized using an Atomic Force Microscope (AFM). The changes in shapes of starting circular sections were studied using a Talyrond analyser which generates out-of-roundness contours.

3.2 Surface texture of etched plates

In this work we investigated the changes in x_3 (Z") and x_1 (X') surface profilometry traces and in surface topographies with the angles of cut φ and θ. Typical results related to singly rotated plates are illustrated in Figure 2. From the profilometry traces several remarks can be made :

(i) The Z' traces show marked orientation effects with rapid changes in the profile shape as θ varies from 26° to - 49°.

(ii) The evolution of the shape of X traces with decreasing values of θ seems less pronounced but we observe a progressiveenlargment of profiles.

(iii) The etched surface profiles exhibit simple shapes : convex or concave backgrounds as well as profiles of alternate convex-concave or concave-convex shapes result from etching in agreement with other works **[3,5,8]**. Let us recall that these characteristic shapes are correlated to the presence of extrema in $L(\varphi,\theta)$ in the vicinity of the reference surface as previously discussed by Irving **[9]** and Tellier and co-workers **[5,10]**.

The marked anisotropy of the chemical attack is also typified by the AFM micrographs of Figure 2C.

3.3 Cross-sectional out-of-roundness profiles

Various experimental shapes of the etched contours of singly and doubly rotated quartz plates are presented in Figure 3. The out-of-roundness profiles exhibit successive peaks and valleys. Since we are now concerned with planar surface elements which lie tangent to the starting circular section of orientation (φ_o, θ_o) an out-of-roundness profile constitutes a crude image of the polar diagram of **L** in the plane (φ_o, θ_o) **[4]**. In particular for a starting "convex" profile valleys are connected to minima in **L** which cause the development of limiting facets in the final cross-section. Very accentuated minima give rapidly rise to planar facets of large extent. It is the reason why in X and Y sections we observe the formation of relatively "flat" valleys associated with the accentuated minima in **L** corresponding to the Z-cut plate.

IV THE DISSOLUTION SLOWNESS SURFACE

To generate the dissolution slowness surface we have to determine experimentally the unknown dissolution constants. This constitutes a very difficult task because measurements related to cross-sectional

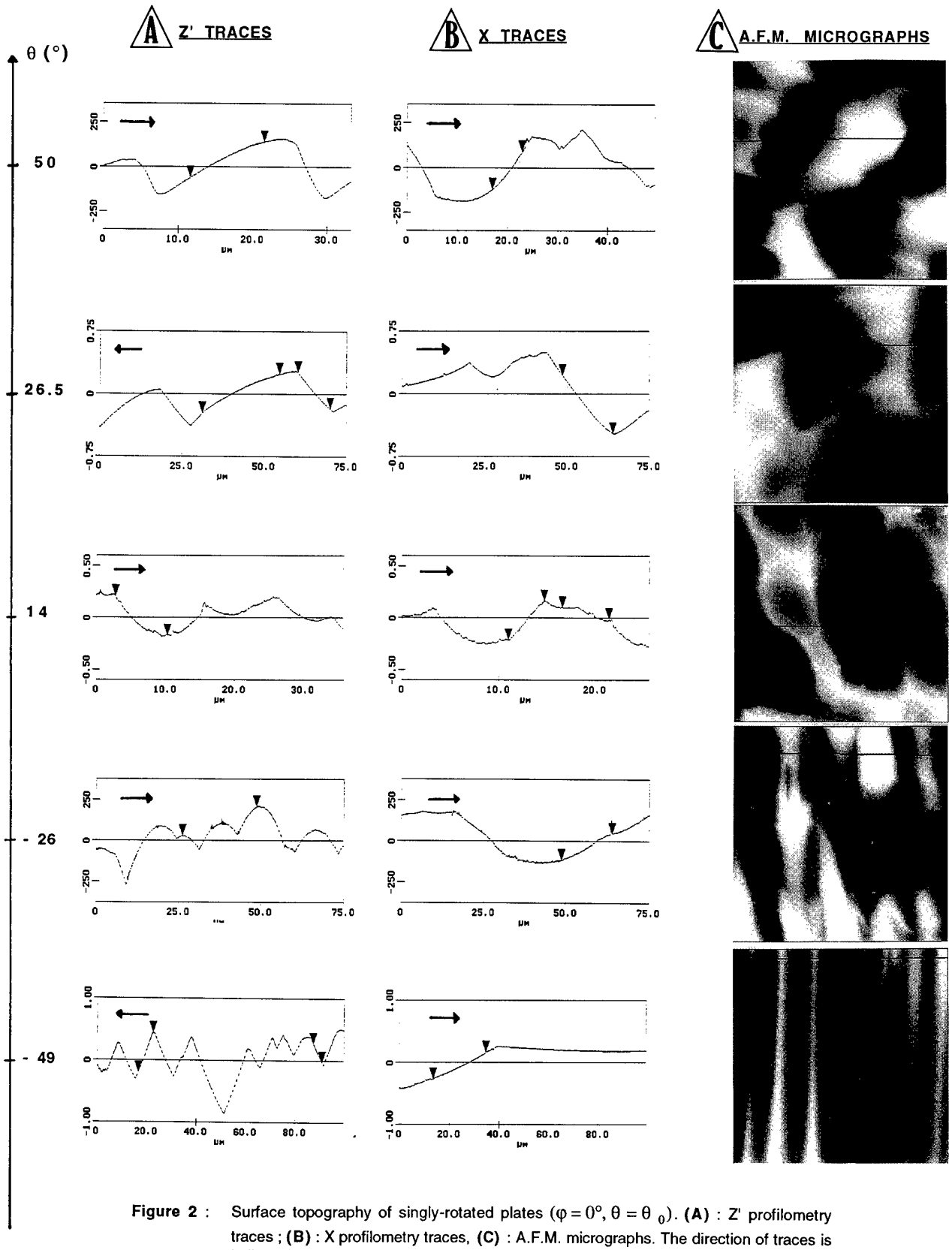


Figure 2 : Surface topography of singly-rotated plates ($\varphi = 0°, \theta = \theta_0$). **(A)** : Z' profilometry traces ; **(B)** : X profilometry traces, **(C)** : A.F.M. micrographs. The direction of traces is indicated by an arrow.

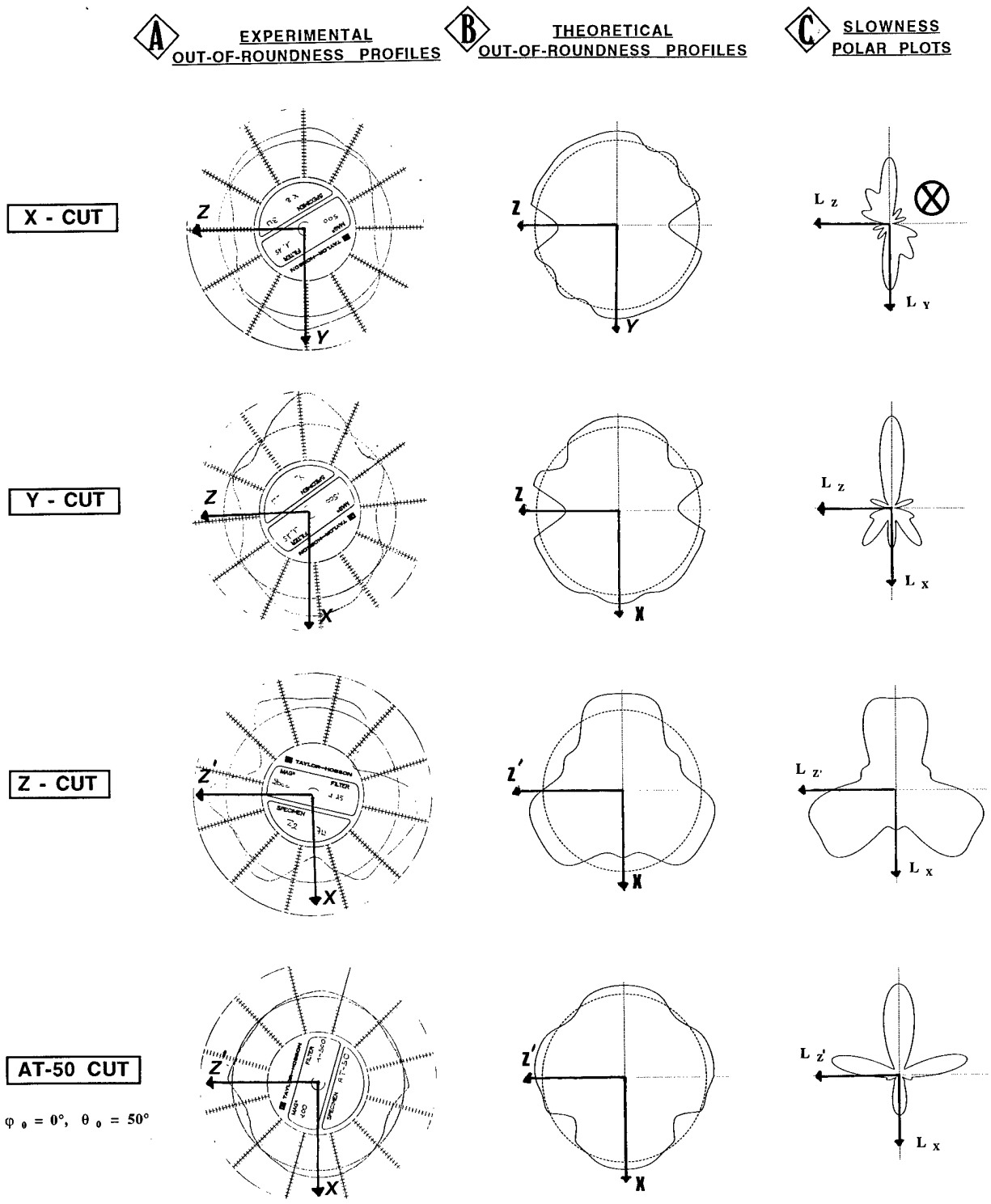


Figure 3 : Cross-sectional out-of-roundness profiles. **(A)** : experimental graphs. **(B)** : predicted graphs derived from a bi-dimensional simulation which works with polar diagrams of $\vec{L}$ **(C)**.

shapes are only reliable if they are associated to extrema in the dissolution slowness [4,10] and because we must restrict the number of experiments together with keeping a satisfactory representation of the slowness surface.
To satisfy this restriction we can simply decide to deal with tensors of higher rank : in this condition, we are sure to obtain a slowness surface with a complex topology together with a limitation of the number of unknown constants. We have finally adopted an iterative procedure to derive the slowness surface. The procedure consists successively in :
(i) Choosing dissolution constants to generate the slowness surface until various theoretical polar diagrams of L resemble to corresponding experimental out-of-roundness profiles.
(ii) Verifying the presence of minor maxima and minima in L by means of a systematic comparison of experimental and theoretical shapes of surface profiles.
A first tentative to apply this procedure to the anisotropic dissolution of quartz crystal has been made using constants related to tensors of rank 9 to 12. As a result, the three-dimensional representation of the slowness surface shown in Figure 4 reflects clearly the symmetry operation 3 associated with the X_3 (Z) axis.

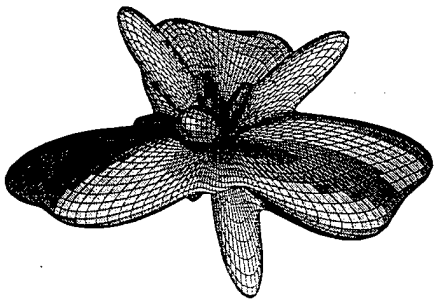


Figure 4 : A three dimensional graphical representation of the dissolution slowness surface.

V BI-DIMENSIONAL SIMULATION OF DISSOLUTION SHAPES

When we are concerned with initially circular sections or with the cross-sectional shapes resulting from etching at a mask edge, it is possible to construct geometrically the approximate dissolution shape using the stability criteria proposed by Irving [9]. The dissolution shape is then formed by the limiting facets for which the dissolution slowness passes through an extremum. In this procedure curved faces which sometimes join together the limiting facets are neglected. Moreover, the stability criteria allow us to predict only the "nature" (convex, concave, ...) of the etched surface profiles rather than the exact geometrical features of profiles [10]. But as soon as the equation of the slowness surface is completely determined better informations on the exact dissolution shapes can be obtained by numerical simulations which involve tests to distinguish between diverging and converging trajectories of moving profile elements composing the starting bi-dimensional shape [3, 4].
Predicted dissolution shapes and corresponding out-of-roundness profiles of various initially circular sections are illustrated in Figure 3B.
The slowness polar diagrams of L from which etched shapes are derived are given in Figure 3C. As expected, we observe the development of "valleys" in the theoretical out-of-roundness profiles directly correlated to the minima in the corresponding polar plots.
It is possible to tract the changes in the shape of surface profiles with the orientation (φ_0, θ_0) of the reference surface. For this purpose we start with a digitalized surface profile as worked by a micro-processor based profilometer. Suppose that the initial distribution of slopes corresponds to angles α ranging from $-\alpha_{min}$ to α_{max}.
To predict the theoretical shapes of various Z' traces made on singly-rotated plates ($\varphi = 0°, \theta_0$) we have just to rotate the polar graph of L (X-section) to θ_0 degrees and to cover a sector ($-\alpha_{min}, \alpha_{max}$) centered on the transformed - Y' axis.
When we are concerned with the graphical simulation of an etched X trace lying in the plane ($\varphi = 0°, \theta_0$) we work with the polar diagram of L corresponding to the singly-rotated plane ($\varphi = 0°, \theta = \theta_0 - 90°$) and we consider the sector centered on the transformed Z' axis (Figure 5C). Figures 5A and 5B show the theoretical Z' and X traces. From examination of the corresponding polar plots we infer that the final shape of etched profiles is correlated to the presence of extrema of L in the region of polar plots involved in the graphical simulation : if a maximum of L occurs for an orientation which coïncides with that of the reference surface we observe that a concave background (Figure 5B, b) develops. The converse situation (minimum in L) gives rise to a convex profile (Figure 5B, d) whereas the presence of two extrema of converse nature results in the formation of profiles with alternate shape.
Bi-dimensional graphical simulation are also used to predict the final cross-sectional shape of etched grooves [4]. It is obvious that the etched section at the mask edge is composed of elements whose orientation varies from 0° to - 180° and that the etched cross-section is potentially concave. Therefore, only maxima in L play a role in the development of limiting elements (faceting). Figure 6 shows various dissolution cross-sectional profiles at the two edges of an infinitely long linear masking pattern covering differently-oriented plates.

VI THREE-DIMENSIONAL SIMULATION OF DISSOLUTION SHAPES

Let us be concerned with the dissolution shapes resulting from the etching of initially circular mesa or hole. We start with doubly-rotated plates and calculate the propagation vector P associated with each potentially present surface element. Now time of calculation increases considerably essentially because we pass very many times in tests for the recognition of the diverging or converging nature of characteristic trajectories. Even if the numerical simulation is three-dimensional the graphical representation remains actually bi-dimensional. Two types of representation are proposed. Firstly, we can draw the constant level contours of etched mesa and diaphragms (Figures 7A,D and 8A, D). In addition, cross-sectional etching patterns (Figures 7C, F and 8B, C, E, F) give quantitative infomations on technical problems caused by underetching . Lateral underetch as well as bordering angle can be easily extracted from the numerical simulation. Turning our attention to the final etched shape of hole, we observe, as expected, the development of facets (facets **1** and **2** Figure 7A for example) associated with pronounced peaks composing parts of the slowness surface (Figure 4).

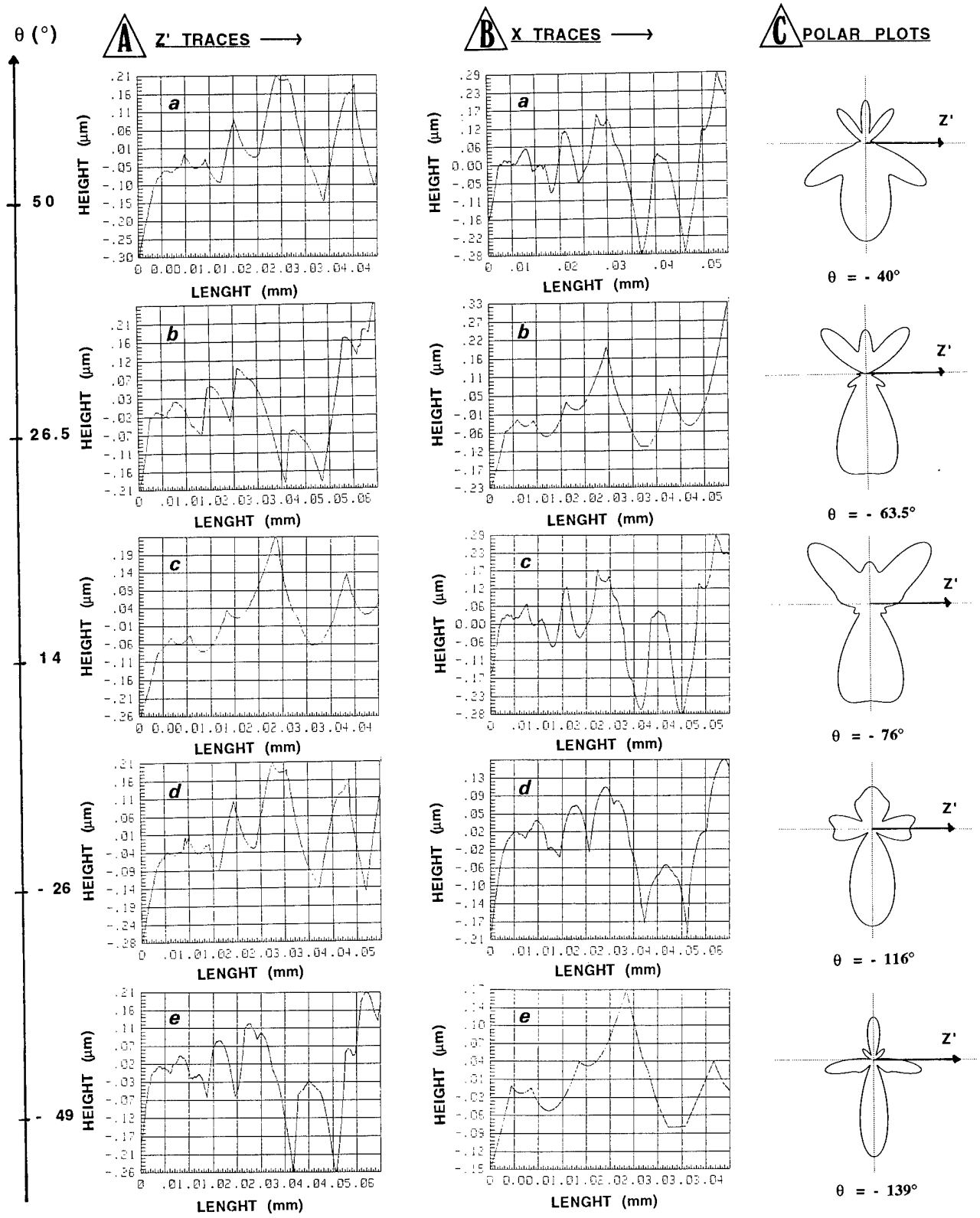


Figure 5 : Graphical simulation of surface profiles : **(A)** Z' traces **(B)** X traces as deduced from the polar graphs shown in **(C)**

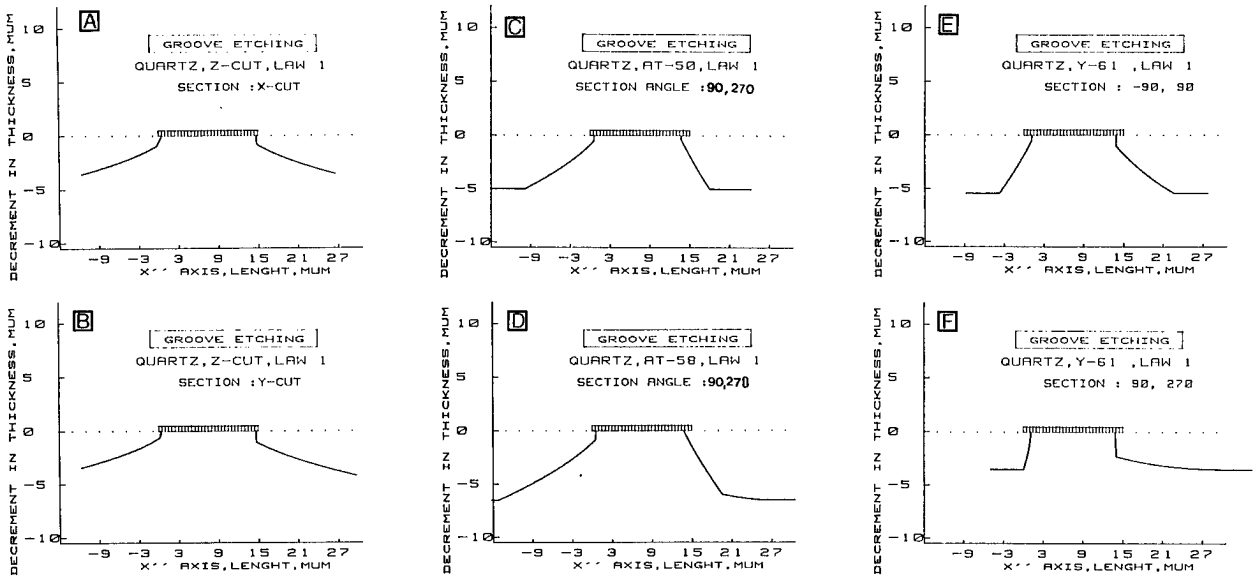


Figure 6 : Graphical simulation of cross-sectional profiles at the edges of an inert mask deposited on Z-cut **(A, B)** AT-50 **(C)**, AT-58 **(D)**, Y-61 **(E, F)** plates. Orientations of masking patterns are indicated in text.

At this point, let us consider a polar representation of **L** and suppose that the two derivatives of **L** with respect to the polar angles determine several maxima maximora for **L** at points M ($\varphi = \varphi_M$, $\theta = \theta_M$).

When etching a membrane we are sure that faceting is connected to these maxima called for convenience principal maxima. Conversely principal minima appear for some specific orientations (φ_m, θ_m). It is clear that the chemical attack of a spherical crystal causes the formation of facets with orientations (φ_m, θ_m). In addition, for some particular orientations (φ^*, θ^*) one derivative corresponds to a maximum and the other to a minimum. The prediction of the etched shape for initially circular mesa constitutes a rather complicated problem because potentially convex and concave profiles are simultaneously present. A cross-sectional profile is undoubtly concave whereas a contour profile lying in a plan parallel to the reference surface is convex **[4]**. Then in this case faceting is apparently produced by two competitive processes. As a consequence, we have to choose what sort of process governs the final dissolution shape. Here we have decided that the dominant process is that which is associated to concave profiles. It is certainly the reason why we cannot give evidence of the role played by principal minima in the "corner undercutting" of mesa. However, we have found that minor "principal maxima" contribute apparently to faceting (facet **4**, Figure 7D for example). Moreover, facets of large extent (facet **2**, Figure 7D) seems to be correlated to orientations (φ^*, θ^*).

VII DISCUSSION AND CONCLUSION

Comparison of theoretical (Figure 3B) and experimental (Figure 3A) out-of-roundness profiles shows a qualitative agreement. At the first sight angular positions of extrema for **L** seem rather good for the out-of-roundness profiles presented here. However, it remains difficult to conclude on quantitative agreements such that corresponding to the relative amplitude of successive extrema. Moreover, less accentuated extrema which give rise to facets of small extent cannot be easily recognized on the experimental graphs (see for example the Z-cut out-of-roundness profile).

Let us now consider the surface profiles. From the comparison of the experimental traces (Figures 2A and 2B) with the predicted traces (Figures 5A and 5B), we observe that the theoretical Z' and X traces fit quite satisfactorely the experimental surface profiles except for the cut with $\theta_0 = -26°$. This disagreement which affects both Z' and X traces can be accounted for as follows :

(i) In the case of the Z' trace the observed convex background (Figure 2A) is associated with a minimum in **L** which necessarily occurs for an orientation corresponding to that of the reference surface **[3, 9]**. Examination of the X-cut polar plot shown in Figure 3C reveals that the minimum in **L** is associated to the singly rotated cut with $\theta_0 \approx -20°$. This angular displacement of about 6° is sufficient to explain the formation of a profile with an alternate convex-concave shape.Effectively, in this case, two extrema of converse nature : a minimum of **L** for $\alpha \approx -6°$ and a maximum for $\alpha \approx 8°$ contribute to the final etched shape.

(ii) In the case of the X trace the departure between the experimental (concave-convexe) and theoretical (convex) shapes may be interpreted in terms of the rapid change of the polar plot in the vicinity of the Z' axis when the angle of cut θ_0 varies from $\theta_0 = 62°$ to $\theta_0 = 66°$ **(Table 2)**.

To discuss the adequation of the proposed slowness surface we can also compare the graphical simulations of Figure 6 with some data previously published in litterature **[11]**.

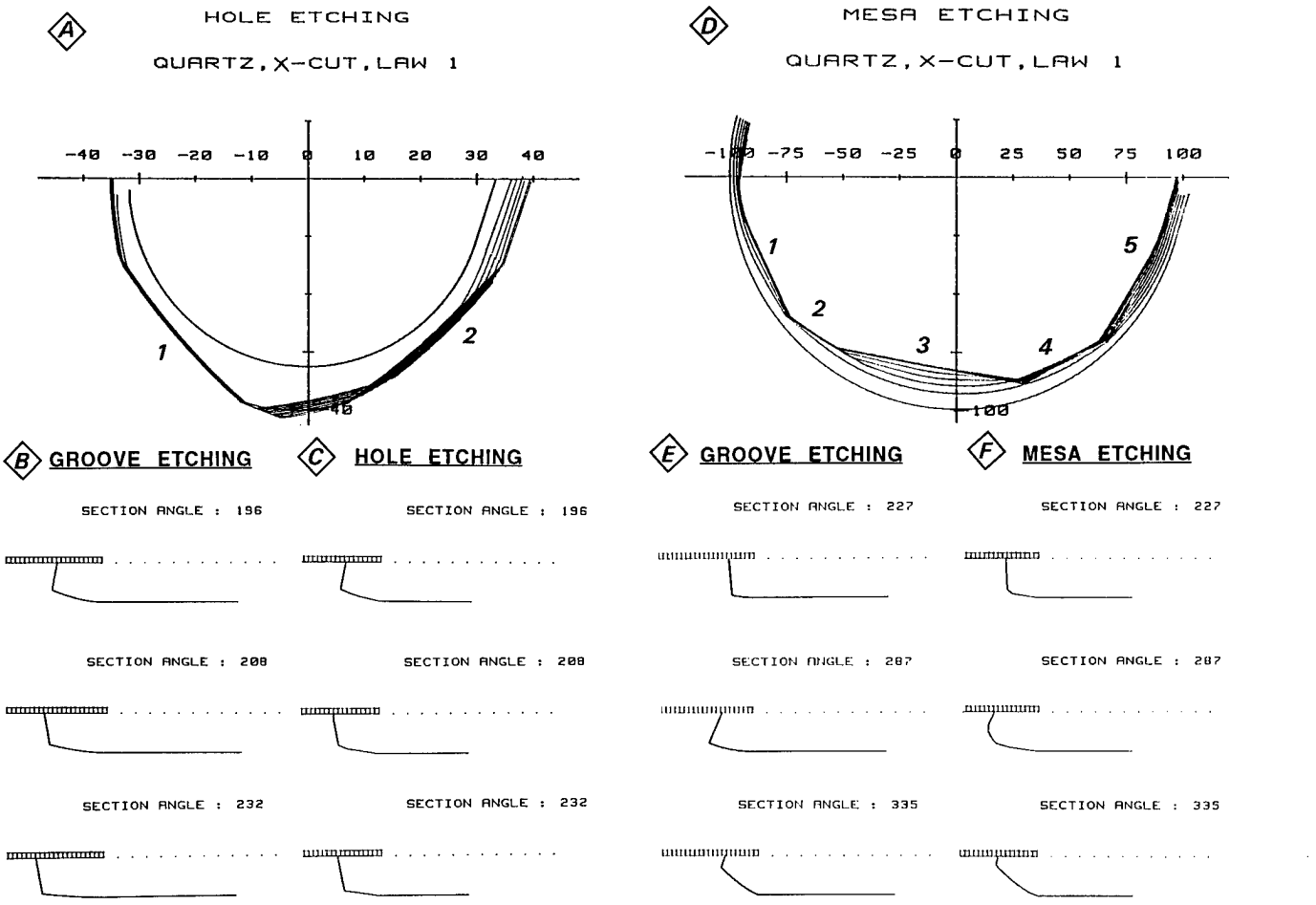


Figure 7 : Three-dimensional numerical simulations : constant-level contours for hole **(A)** and mesa **(D)** etched on an X-cut plate ; corresponding cross-sectional profiles are shown in **(C)** and **(F)**. Some bi-dimensional profiles resulting from the etching of grooves are also given in **(B)** and **(E)**.

θ_0 (°)	62°	64°	66°
Minimum in L	- 5°	- 1°	- 39°
Maxima in L	+ 13	- 10°, + 15°	+ 14°

Table 2 : Orientation, α, of extrema in L for singly-rotated plates with angle θ_0 in the vicinity of 64°. Note that for $\theta_0 = 66°$ the minimum at α = - 39° remains unoperative.

Cut	Z-cut		AT-50	AT-58	Y-61
	X- section	Y-section			
Present work	49°, 84°	≈ 53°, 88°	110°, ...	105°, ...	32°, 60°
Previous data	47°, ≈ 90°	≈ 65°, 90°	120°,125°	105°, 85°	20°,70°

Table 3 : Comparison between theoretical values deduced from the present numerical simulation and previously published values [11].

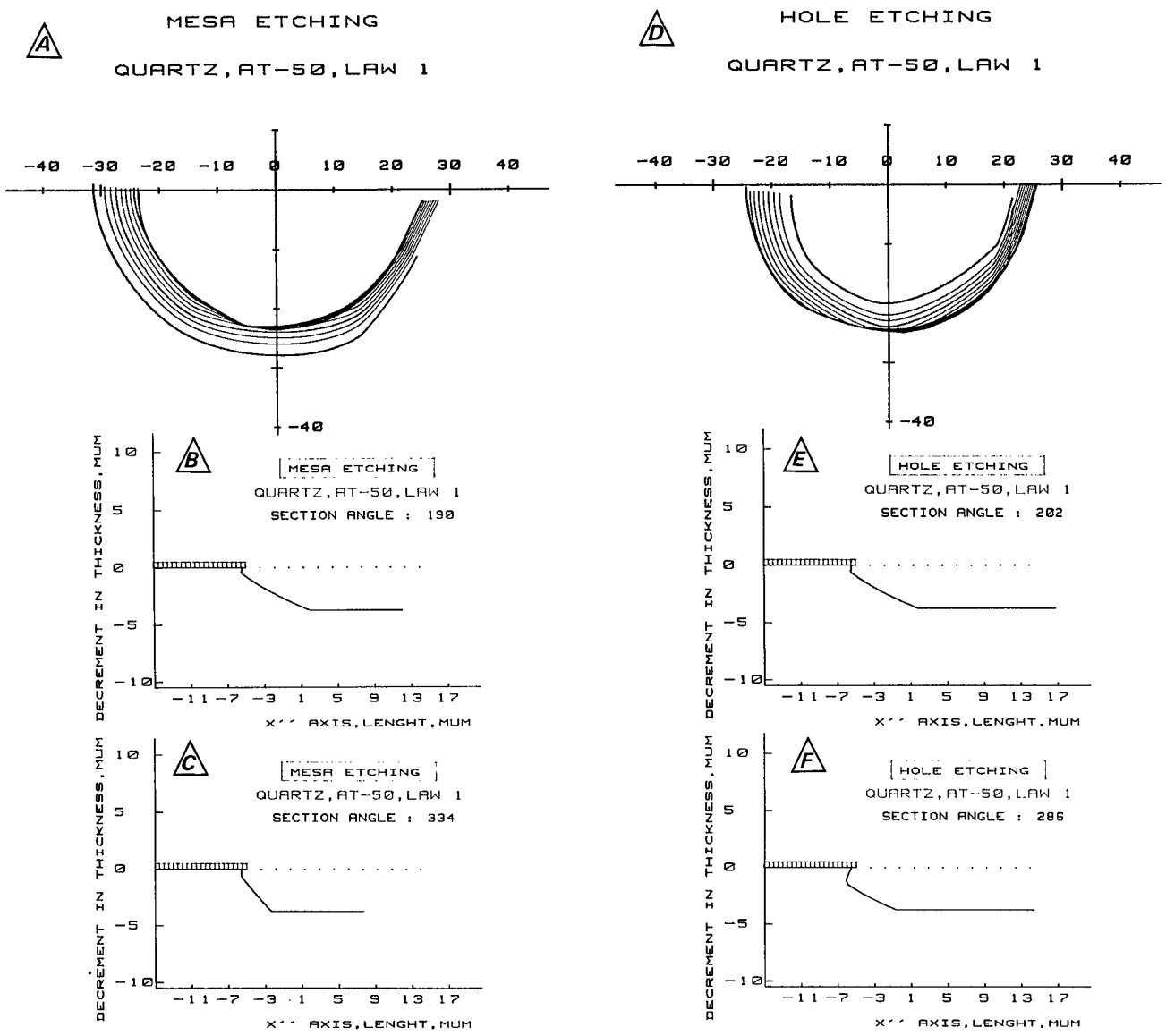


Figure 8 : Three-dimensional numerical simulations : constant level contours and corresponding cross-sectional profiles for a mesa and an hole etched on an AT-50 cut plate ($\varphi = 0°, \theta_0 = 50°$).

Y-cut and X-cut cross-sectional dissolution shapes related to strips etched on a Z-cut plate as well as AT-58 ($\varphi = 0°, \theta = 58°$) and Y-61 ($\varphi = 90°, \theta = 61°$) cuts resemble crudely to dissolution profiles published by Ueda and Al **[11]** . In order to make the comparison more easier some values of typical angles characterizing the geometry of dissolution profiles are reported in **Table 3**. We observe moderate departures between theoretical and experimental values.

This study enables us to make two concluding remarks :

(i) Firstly, it appears that the dissolution shapes derived from the present slowness surface do not depart very much from experimental shapes. Then the proposed slowness surface characterizes crudely the anisotropic dissolution process even if some angular deviations for the extrema of **L** are revealed by a careful examination of etched surface profiles. Thus, some corrections must be made to adjust the theoretical shape of the slowness surface and a special endeavour must be placed on the determination of the true amplitude of extrema.

(ii) In absence of experimental works dealing with the etching of mesa on various quartz plates, it remains difficult to draw definitive conclusions on the three-dimensional graphical simulation of etched mesa. However, the simulation of etched diaphragms appears to be fully appropriate because we effectively generate planar limiting facets associated with principal maxima in **L** .

Finally, taking into account all the possibilities of graphical simulations, it seems very convenient to start with the equation of the slowness surface to predict dissolution shapes.

Acknowlegments :

The authors gratefully acknowledge the Direction des Etudes et Techniques (French Dept. of Defense) for its financial support.

REFERENCES :

1 - **F.C. FRANK and M.B. IVES,** J. Appl. Phys. , 31 (1960) 1996.
2 - **D.W. SHAW**, J. Crystal Growth, 47 (1979) 509.
3 - **C.R. TELLIER and J.L. VATERKOWSKI**, J. Mater. Sci., 24 (1989) 1077.
4 - **C.R. TELLIER, J. Y. AMAUDRUT and A. BRAHIM -BOUNAB**, J. Mater. Sci., 26 (1991) 5595.
5 - **C.R. TELLIER, N. VIALLE and J.L. VATERKOWSKI**, Surf. and Coatings Technol., 34 (1988) 417.
6 - **A. BRAHIM-BOUNAB, J.Y. AMAUDRUT and C.R. TELLIER**, J. Mater. Sci. 26 (1991) 5595.
7 - **C.R. TELLIER**, J. Crystal Growth, 96 (1989) 450.
8 - **A. BRAHIM-BOUNAB and C.R. TELLIER**, 6th European Frequency and Time Forum, Noordwijk, The Netherlands, March 1992, 5 pp.
9 - **B.A. IRVING** in "The Electrochemistry of Semiconductors" edited by P.J. HOLMES (Academic Press, London, 1962) pp 256-289.
10 - **T. LEBLOIS and C.R. TELLIER**, J. Phys. III, 2 (1992) 1259.
11 - **T. UEDA, F. KOHSAKA, T. IINO and D. YAMAZAKI**, Trans. Soc. Instrum. Contr. Eng. , 23 (1987) 1.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DEFECTS IN SURFACE LAYERS OF LAPPED CRYSTAL PLATES AND THEIR INFLUENCE ON Q-FACTORS OF CRYSTAL UNITS

A. S. MOSHKOVSKY

Scientific Research Institute "Fonon"
105023, Moscow, Krasnobogatyrskaya 44, Russia

Defects forming during the lapping cycle in quartz crystal lattice distribute on the surface and around the fractures in the form of 2 layers of the thickness 0.35 and 1.7μm relatively. For a crystal unit the losses in the upper layer reach a value by 28 times dominanting that of the 2nd layer having the same thickess value. The increase in losses due to bigger abrasive grains is caused by defective quartz mass around fractures.

Defects on the surfaces of lapped quartz plates are devided conditionnaly into micro microrouqhnesses, microfractures and a quartz layer containing defects in the crystalline lattice. The report contains results of investigating the last of the mentioned defects. Using the value of the etching velocity change for lapped crystal elements the total value of the quartz layer containing defects in the crystal lattice was defined, two components of this layer were identified, the disposition of these compronents on the lapped qurtz surface was defined and the contribution of each of the components to the loss value in vibrating crystal elements was computed.

Investigations were carried on for crystal elements of AT-cut at 1 Mc vibrating at the main resonance. Crystal elements were excited by using electrodes forming a constant gap with the plate surfaces equal 0,6 mm.

The processing of the spherical surfaces of the crystal elements ended by polishing or lapping using powders of one of the qrain sizes: 5 or 10 or 20 μm. Etching in hydrofluoric acid of either HF-concentraition equal 10, 20, 30 or 48.5% was carried on at $28,2^{\circ}C \pm 0.2^{\circ}C$. The frequency of the crystal elements was measured before and after each etching cycle operation.

In Fig.1 curves are presented characterizing the frequency change versus etching time, based on the experimental results.

Being expressed in logarythmic ordinates this dependence has the form of a straight line (curves 1 and 2) for

0-7803-0905-7/93 $3.00 © 1993 IEEE

polished plates and of a broken line (curves 3..9) for lapped ones. For the last lapping operation any micropowder was used. The broken curves for the lapped samples include three parts: ab, bc and cd. All the ab-parts of the broken curves as well as all the bc-parts have close values of incline angles relative to the ordinate axes.

After the second break at the c-point of the broken curves for lapped samples and of the straight curves for polished ones the incline angles of the both curve types differ insignificantly.

The break points on the curves point to qualitative changes in surface and subsurface qurtz layers and define the borders of these layers. Beginning from the second break c-point on the broken curves for lapped samples as well as on the straight curves for polished ones defects in the crystal lattice are either absent, or just too small for to be defined by the methods used during the experiments.

The first two parts of the broken curves (ab and bc) correspond to the two states of the crystal lattice differing in the extent of the defects and being characterized by an according etching velocity. These defects in the mentioned two layers are named below as defects of I-and II-type accordying to the succession of their disposition.

In Fig.2 curves are given based on the point ordinates of the first and second breaks of the broken lines (Tabl.1)

It is known that the lapping process leads to the origin of microrouqhnesses and microfractures on the quartz surfaces. The formers appear as a result of removing quartz particles which have been separated due to the microfracture origin. Therefore it may be constituted that the structures of the surface layer and of the quartz layer contacting with the microfractures are the same. The thickness of the quartz layer containing microfractures is a linear function of the abrasive grain diameter. For the powder of corundum

$$H=0.5d,$$

where H - the thickness of the layer containing fractures, μm;

d - the abrasive grain diameter, μm.

Hence it follows that with d aproximating 0 the thickess of the fractured layer approximates 0 as well.

At the same time if the curve in the Fig.2 is prolonged up fo the crossing with the ordinate axes then the meaning of ordinate of its crossing with the ordinate axes will be ~0.35μm. The comparison of these data allows to conclude that 1-type defects are distributed (as it is presented in Fig.3) in the surface region formed due to lapping and around the microfractures and constitute a layer of ~0.35μm thikess.

Below this upper layer containing I-type defects is the lower quartz layer containing II-type defects and having the thickess ~1,7 μm, which value was defined from Table 1 data.

The above mentioned assumption relative to the quartz surface layer structure is confirmed by the result of X-ray investigations and of measuring the changes of the thin plate curvature due to polishing and lapping.

A direct method for analising defects in the crystal lattice including the defects formed due to lapping plates is X-ray topography. To investigate the defects in the relative quartz layer we have applied 2-crystal spectrometer using a parallel disposition of the sample being analysed and the monochromator.

For the last processing cycle of AT-samples abrasives with the grain size of 5 and 10μm were used, then the samples were efched in 10% hydrofluoric acid. For X-ray reflection an atom plane was selected which was parallel to the side of the small rhombohedron. On the hasis of these results the curves were plotted (Fig. 4).

After 2-3 etching cycles resulting in removing a quartz layer of 0.1-0.2 μm thikness, an abrupt decreasing of the semi-width of X-ray reflection intensity curves is observed. Further efching does not bring any real changes to the semi-width value.

As it follows from the above mentioned, the increase in the semi-width value of the X-ray reflection intensity curves may be explained by the influence of the 1-type defects.

In the quartz layers having any defects due to lapping, mechanical stresses appear. The lapped surface of thin plates becomes convex, the opposite surface - concave. The results of measuring this flexure for plates of different orientations and of the size 20×20×0.46 mm are qiven in Table 2. The flexure value for plates of equal thickness with using abrasives of the grain size 5 and 20 μm differ insignificantly though the depth of the fractured layers of these samples differ approximately by 4 times.

Etching lapped plates in 10%-hydrofluoric acid for 100-120 sec leads to sharp decreasing this deformation of the sample. During such short-time efching operation a qurtz layer of ~0.1-0.2 μm thickess is dissolved, i.e. a part of the layer having 1-type defects is removed.

The polishing operation insures the removal of the quartz layer containing defects and consequently the accompanying stresses. The flexure of the plates practically vanishes as a result of polishing off the microrelief, for this operation leads to the removal of the quartz layer characterized by defects in

the crystal lattice. Microfractures and the quartz layer with defects contacting with those microfractures influence upon the flexure of the plate insignificantly. It follows that the stresses conditioning the flexure of thin plates are caused by 1-type defects distributed along the surfaces.

Simultaneously with the measurement of the crystal element frequency carried on using technological fixtures and of the frequency changes during the efching operation, the quality factors of the crystal elements were measured as well. During the efching of the samples in hydrofluoric acid the Q-factor of all the polished crystal elements remained practically constant (Fig. 5a), while the quality factor of lapped crystal units drops in the beginning of the efching operation, then raise and on reaching a definite value does not change in further (Fig. 5b). Mean values of the Q-factor for the set of polished crystal elements and for the set of lapped ones subjected to etching off layers having defects in the crystal lattice are close to one another.

As to Fig. 5b it is of interest to select four characteristic points according to: the quality factor of crystal elements just after the lapping operation; the least quality factor during the drop period af the beginning of efching; the quality factor up to the moment of dissolving the quartz layer with 1-type defects, and the quality factor up the moment of dissolving the layer with 2-type defects. For these four points on the curve in Fig. 5b the meanings of the equivalent resistance were found and the summary (total) losses of R_{dist} in the defect-ive layers with defects of lapped crystal elements were defined: for the layer with 1-type defects - R_I, for the layer with 2-type defects - R_{II} and at the point according to the minimum quality factor - R_{etch}. These data are presented in Table 3.

It was computed that the losses for the layer with 1-type defects are by ~ 28 times relative to the losses for the layer of the same thickness but having 2-type defects.

It follows from Table 3 that the component R_{etch} at the point of the minimum quality factor has a meaning close to that of the losses in R_I in the layer with 1-type defects, it depends on the thickness of the fractured layer. This fact allows to make a conclusion that during the first stage of efching some etch products are being kept in microfractures and as result become a source of additional losses in R_{etch}.

Table 1

Dependence of the thickness of quartz layers having crystal lattice defects on the abrasive grain size

Abrasive grain size, μm	The thickness of quartz layers, μm			HF-concentration,%
	Layer1 (I-type defects	Layer2 (II-type defects	Summary thickness	
5	0.57	1.59	2.16	10
5	0.54	1.89	2.43	20
5	0.40	1.78	2.18	30
5	0.44	1.68	2.12	49.5
10	0.97	1.88	2.85	10
10	0.98	1.85	2.80	10
20	3.30	2.20	5.50	20

Table 2

The flexure of various cut-angle quartz plates subjected to lapping by the abrasive of grain size 20 μm (the plate size: 20×20×0.46 mm)

Cut angle	The flexure, μm	
	in the plane of the big ellipse axes	in the plane of the small ellipse axes
Z	6.8	6.83
Y	4.59	5.26
X	4.72	5.80
r	4.18	5.42
R	5.29	5.88
AT	3.91	5.05
BT	4.88	5.80

Table 3

The equivalent resistance components for piezoelements subjected to different initial processing of the surfaces

Abrasive grain size, μm	Meanings of equvivalent resistance components, Ohm						HF-concentration, %
	R_1	R'_1	R_{dist}	R_I	R_{II}	R_{etch}	
Cerium dioxide	7.94	7.94					10
5	13.1	8.37	4.73	4.40	0.33	4.3	10
	16.6	8.87	7.73	6.94	0.80	5.3	20
	18.4	10.3	8.01	6.80	0.88	3.45	49.5
10	19.6	7.53	12.07	10.70	1.37	18.1	10
	26.8	8.13	17.67	16.80	1.87	6.5	10
20	52.3	9.12	42.40	41.20	1.18	45.5	10
	63.5	9.90	54.60	53.40	1.20	58.5	20

R_1 - the piezoelement eqvivalent resistance after mechanical processing (lapping, polishing);

R'_1 - the piezoelement eqvivalent resistance after etching off the quartz layer with crystal lattice distortions;

R_{dist} - the summary component of the resistance for quartz layers with crystal lattice distortions;

R_I, R_{II} - the resistance components for the quartz layers with I-type and II-type defects accordingly;

R_{etch} - the resictance component due to forming etching products in microfractures.

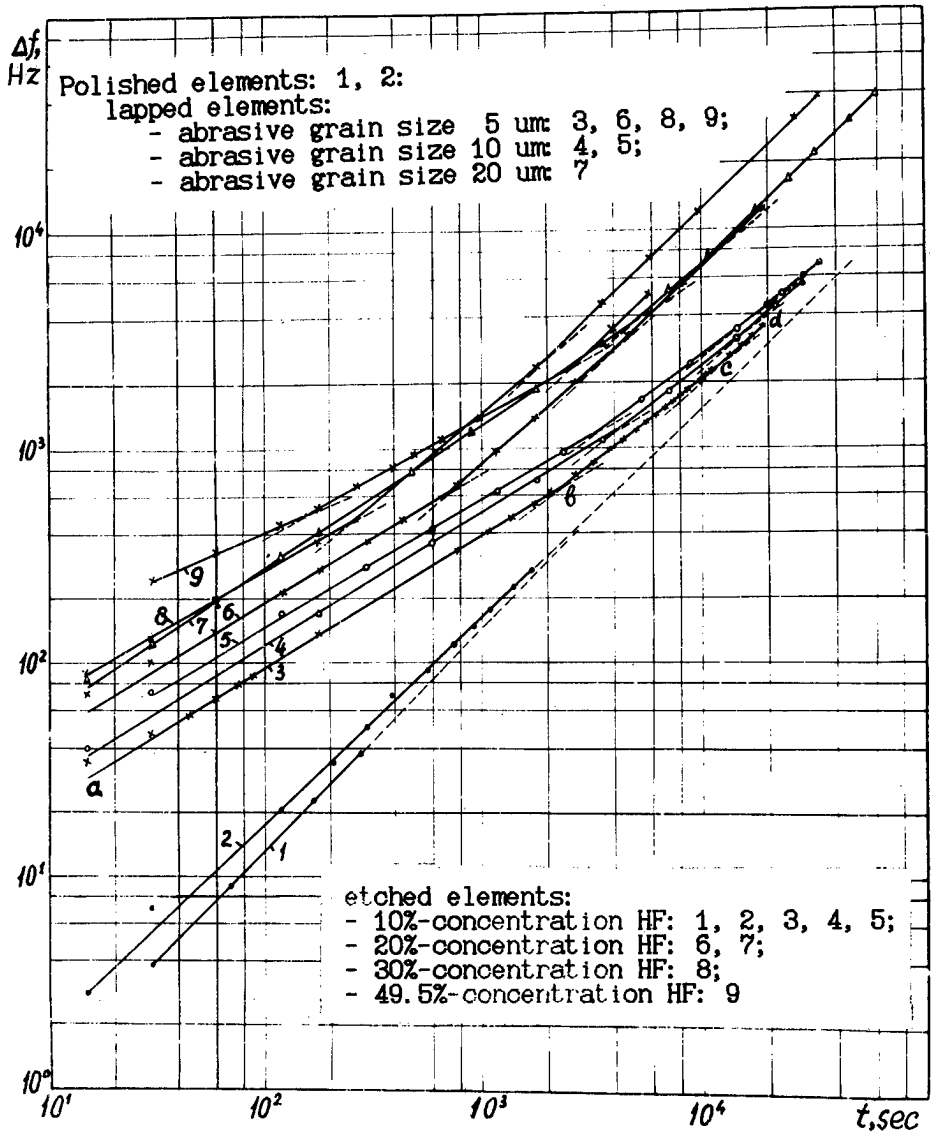


Fig. 1. Crystal element frecuency change versus etching time

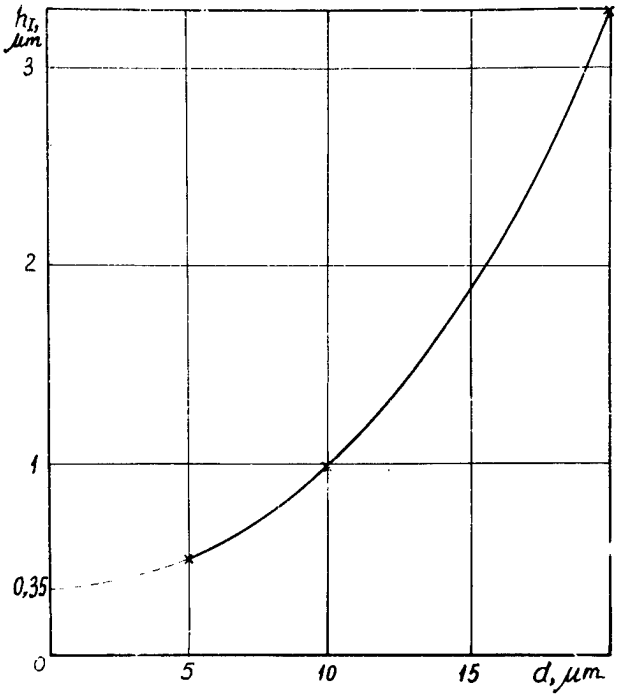


Fig. 2. Total thickness of lapped quartz layer h_I versus abrasive grain size

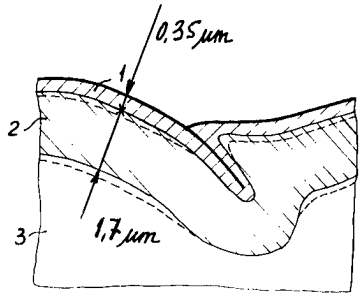


Fig. 3. The distribution of quartz layers having crystal lattice distortions on the lapped plate surface.
1 - type I,
2 - type II,
3 - quartz without distortions

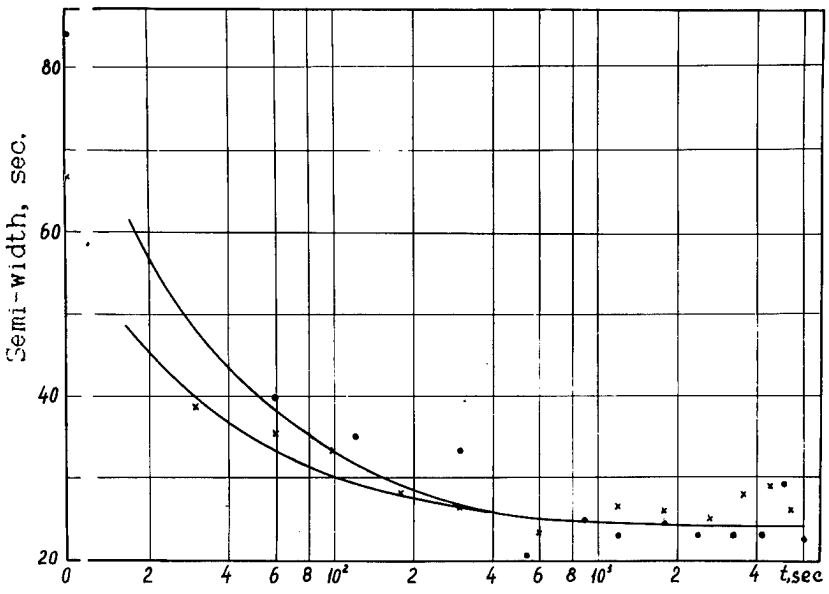
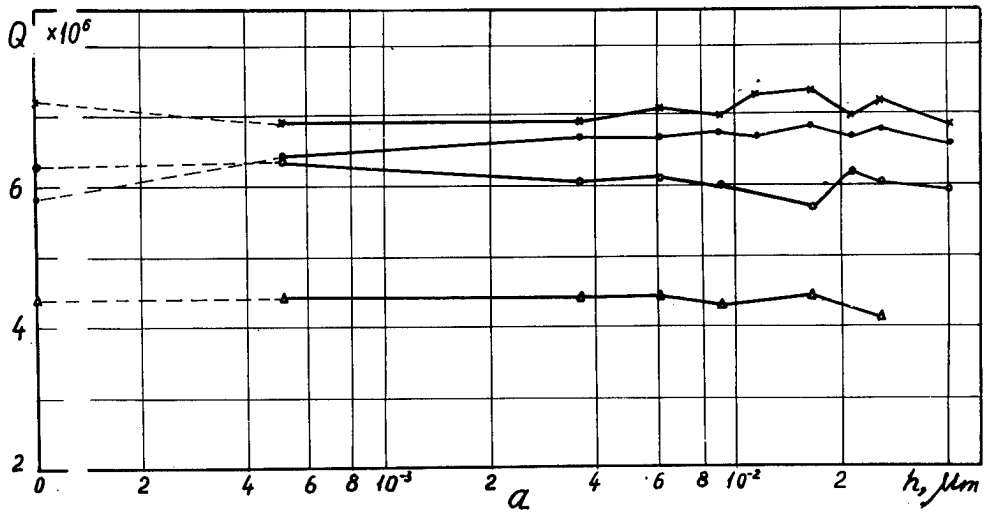


Fig. 4. The semi-width of X-ray reflection intensity curves versus the etching time



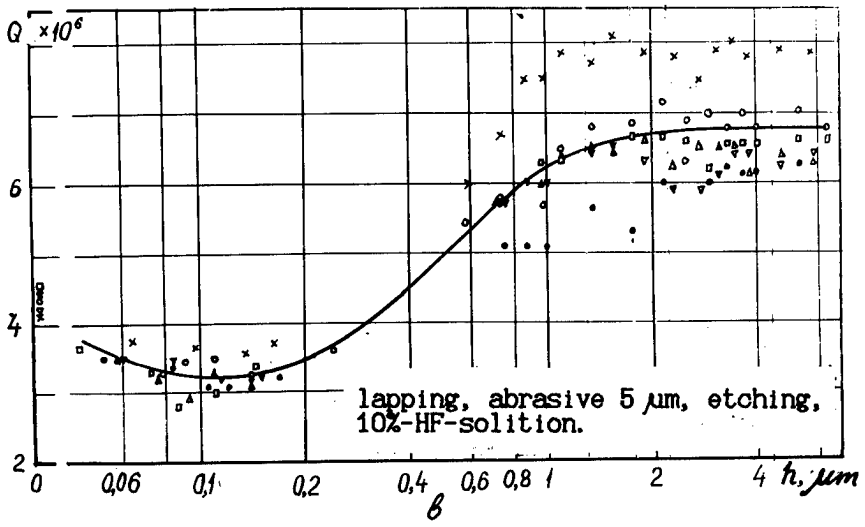


Fig. 5. The Quality-factor of polished (a) and lapped (b) piezoelements versus the thickness of the quartz layer etched off

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE USE OF AN X-RAY THREE BEAM TECHNIQUE TO IMPROVE THE ADJUSTMENT OF BARS FOR CUTTING

Hans Bradaczek, Gerhard Hildebrandt, and Wilhelm Uebach

Institut für Kristallographie, Freie Universität Berlin

Abstract - An adjustment method for bars before cutting will be described which uses the properties of X-ray three beam diffraction. This procedure avoids two types of cutting errors: Miscuttings of the delivered bars, and misalignments of the bars on the holder. Nevertheless remaining errors require sorting after cutting and lapping.

INTRODUCTION

The temperature coefficient determining the frequency stability of quartz oscillators depends strongly on the accuracy of the cutting angle between the surface of the bar and the (01$\bar{1}$1) lattice plane.

Cutting angle deviations of the blanks are caused mainly by four influences:

1.) Nonplanarity of the lattice "plane" due to growing errors of the bars
2.) Miscutting of the delivered bars
3.) Mismounting of the bars on the holder
4.) Cutting errors by the wire saw

These influences can only partly be avoided; therefore the measurements of the distribution of the cutting angles before the final treatment of the blanks are still necessary.

Mainly two methods of measurement of the final cutting angle are in use:

The Θ-scan for rectangular blanks, and the Ω-scan for round ones, cp. Figs. 1a, 1b. Both methods have advantages and disadvantages: The round blank sorting after changing the shape from rectangular to round avoids all further treatments of unsuitable specimens; on the other hand the Θ-scan is less sensitive to XX'-miscuttings. The sorting of quartz blanks is an excellent method to increase the yield; in the case of round blank sorting some 70-90% are reached. But a cutting of still higher precision would be preferred.

With the method described below only errors 2.) and 3.) can be corrected.

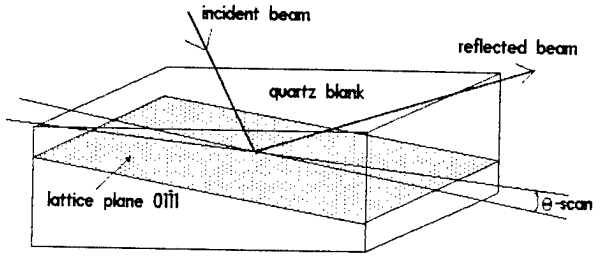


Fig. 1a: Θ-scan for rectangular blanks

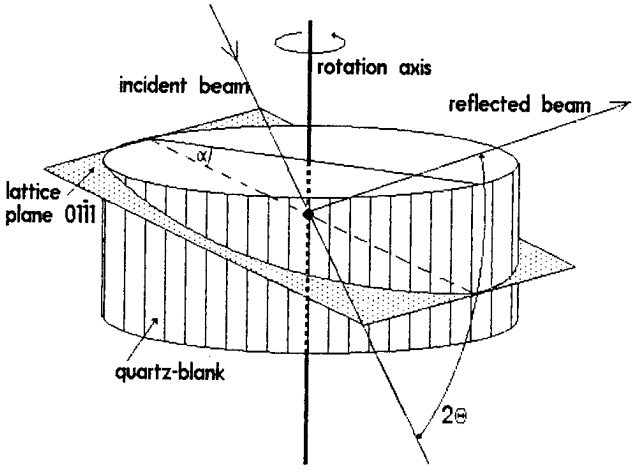


Fig. 1b: Ω-scan for round blanks

0-7803-0905-7/93 $3.00 © 1993 IEEE

CUTTING ERRORS OF THE AT CUT

The angle φ , determining the position of the AT cut, is approximately given by (cp. Fig. 2):

$$\varphi = 38^{\circ}13' - 2^{\circ}\,58' = 35^{\circ}15' ,$$

measured against the XZ-plane of the bar ($38^{\circ}13'$ is the angle between this plane and the lattice plane (011) = ($1\bar{1}1$), which is commonly used in the adjustment procedure). It is assumed that the bar be exactly positioned perpendicular to the X-axis of the crystal lattice.

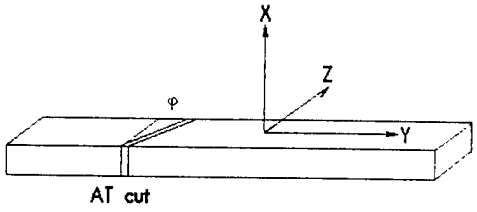


Fig. 2: Coordinates of quartz bar

Cutting errors may therefore originate from two sources: misalignment in φ (error $\Delta\varphi$), and deviation between the actual axis X' of the bar and the X-axis of the quartz lattice (error XX'). The latter is the minor important one, at least for the AT cut, because the temperature coefficient of this cut is less sensitive to deviations XX'.

In order to get an idea about the influence of possible deviations resulting from XX' the following calculation might be useful. Let $\alpha_T = 2^{\circ}58'$ be the expected value for the accurate AT cut, and α_M the actual value due to an XX' error, then we get (if we are using the Ω-scan):

$$\cos \alpha_M = \cos \alpha_T \cdot \cos XX'$$

(cp. Fig. 3), leading to the function reproduced in Fig. 4.

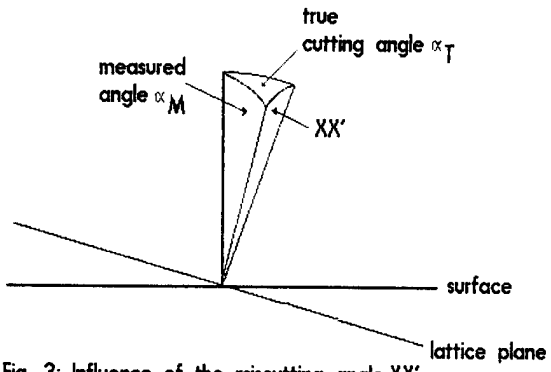


Fig. 3: Influence of the miscutting angle XX'

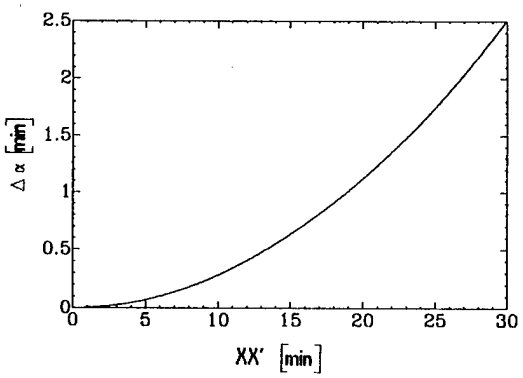


Fig. 4: Influence of XX' upon the cutting angle measurement

X-RAY THREE BEAM DIFFRACTION

In experiments or measurements with X-rays commonly the so-called "two beam case" is applied. In this case the incident beam enters a (low indexed) lattice plane by the Bragg angle and one strongly reflected beam leaves the plane. In this way in a two beam case, besides the incident beam, only one reflection exists, both beams together defining the "plane of incidence": two beam cases are two-dimensional problems. Using the two beam technique there is no chance to find out an unambiguous orientation of the bar.

If the incident beam is kept unchanged and the crystal is rotated around an axis perpendicular to the diffracting lattice plane, the Bragg condition remains maintained, cp.Fig.5.

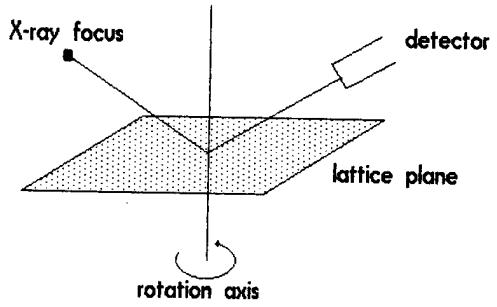


Fig. 5: Two beam technique

But during such a rotation other lattice planes (with other indices) will, at certain azimuths, simultaneously be brought into a position where another Bragg condition is fulfilled by the incident beam: a second strong reflection leaves the crystal, a "three beam case" has been realized, (Fig. 6).

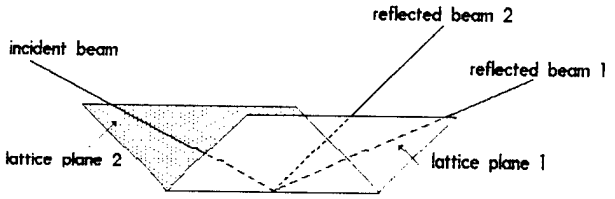


Fig. 6: Three beam technique

Only very rarely are the two reflections together with the primary beam arranged in one plane, establishing a coplanar three beam case. All other three beam cases are three-dimensional problems, the adjustment remaining maintained only if the crystal would be rotated about the direction of the incident beam, the two reflections following this rotation. If, however, two counters with pinhole apertures are fixed in appropriate horizontal positions (which are easily calculated) and a narrow incident pencil ray is used, the geometry is fixed in space in such a way that the crystal can reflect the maximum intensities simultaneously to both counters only if it is correctly positioned in space, within very narrow limits. As a result, all lattice planes in the crystal would be in exactly known positions after just one measurement - at least in principle. In practise the two lattice planes used to establish a convenient three beam case will be different from the lattice plane close to the cutting plane, and strains in the crystal could cause small misalignments. (Cp. influence 4 mentioned in the introduction). It is therefore preferred to select one of the three beam planes close to the cutting plane.

In the following we confine ourselves to the case of simultaneous reflections on $(\bar{2}020)$ and $(\bar{2}200)$ with Bragg angles of $\theta = 21^{\circ}14'$ for CuK_{α} radiation. Those planes intersect in our Z-axis and deviate from the XY-plane by $\pm 30^{\circ}$. Vector algebra then yields an incident beam in the XZ-plane with an angle of $24^{\circ}43'$ to the Z-axis and two reflected beams symmetrical to the XZ-plane, $2\theta = 42^{\circ}28'$ apart and inclined by $12^{\circ}4'$ to the YZ-plane i.e. the crystal surface (see Fig. 7).

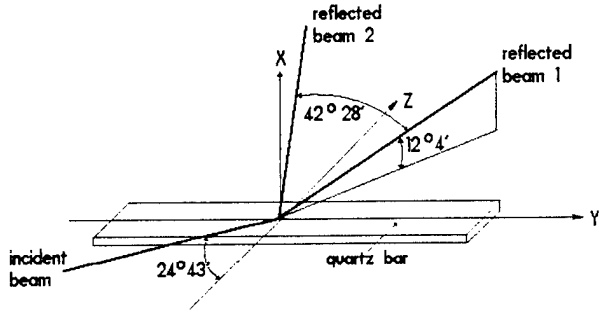


Fig.7: Three beam case

TECHNICAL REALIZATION OF THE ADJUSTMENT

The adjustment system consists of the point focus of an X-ray source, two detectors, three motor driven collimators (each one for the primary beam and the two detectors), and an adjustable barholder.

In principle the adjustment starts with all three collimators widely opened. The adjustment motors, controlled by the main computer, move the bar until both detectors receive a strong signal. The collimators are then stepwise closed down to narrower cross-

sections, while the computer maintaines the detector signals at an optimum height. The procedure stops, when the narrowest collimator apertures are reached.

One may assume that the producers deliver the bars with more or less exactly oriented surfaces. Therefore our precise adjustment will only need movements of the bar within parts of a degree which can be performed by elastical deformation of suited parts of the holder.

The massive bar-holder is constructed in such way that it can be transferred to the bearing of the cutting device without loosing the precise adjustment.

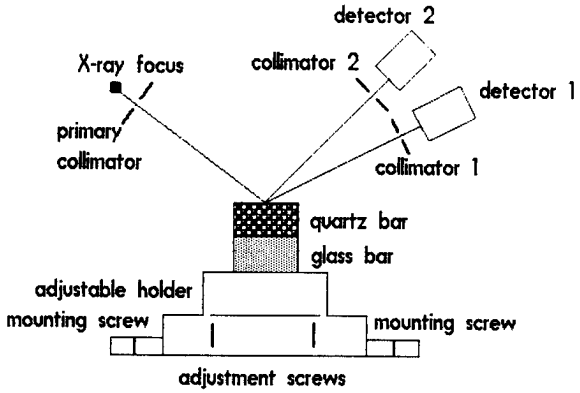


Fig. 8: Adjustable quartz bar holder

In an adjustment procedure especially adapted to three beam cases a narrow incident beam will be used together with wide open counters throughout. The crystal first is rotated around the Z-axis to an angular offset of 30'. If it is then rotated stepwise around the Y-axis, there will be two consecutive reflections measured, one for each lattice plane. This will be repeated for an offset of -30'. From the angular positions for the four reflections, the position for the exact three beam case can be calculated - and eventually confirmed then using narrow counter collimators.

EXTENSION TO A BAR-QUALITY MEASURING DEVICE

In a more sophisticated version the equipment could also be used to check the quality of the bars point by point or in points at given distances. To achieve this the bar has only to be placed on a precision glide. In the simplest mode a fine adjustment could be performed, as described above, in the centre of the bar; afterwards the glide would be moved continuously from one end to the other, simultaneously registering the intensities measured by the two counters. As a matter of course the intensities from a perfect bar should remain constant. In another more time-consuming modification the computer lets the fine adjustment procedure repeat in steps of say every 10 mm along the bar, while in each step the angular positions of the bar are registered and finally evaluated.

Both procedures should make it possible to eliminate "wrong" bars before starting the cutting process.

CONCLUSION

An X-ray adjustment method for bars before cutting has been described, which avoids two of the mentioned cutting errors: Miscuttings of the delivered bars and misalignments of the bars on the holder before the blank cutting. The method cannot avoid growing errors of the quartz and nonregularities of the wire saw. Therefore a sorting before the final treatment remains necessary. But the yield of blanks with the selected cutting angle will increase significantly.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THERMOLUMINESCENCE (TL) RELATED TO $[SiO_4/M]^\circ$ DEFECTS (M = Li or Na) IN QUARTZ

A. Halperin

Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel

Abstract

The dose dependence of the 190 and 200K TL peaks of Al-Li and Al-Na containing quartz (in short TL(M), M=Li or Na respectively) is described. The TL(M) peaks were excited by two x-irradiations at 80K with warming to 250K between the two irradiations. The dose dependence was examined separately as a function of the time of the first irradiation (t_1) and of the second irradiation (t_2). Slopes between 0 and 1.15 were obtained, depending on thermal and radiation pretreatments of the samples. The variations in slope are shown to be related to the filling of deeper traps competing with the $[SiO_4/M]^+$ traps related to the TL(M) peaks.

A first irradiation of only 2 sec was found to release M-ions and produce $[SiO_4/M]^+$ electron traps at a concentration which took about 1h of the second irradiation to fill them. With all competing traps stable at 250K full, and keeping t_1 constant, the dependence of TL(M) on t_2 started with a slope of 1.0. The curves then reached a maximum near t_2=1500 sec and decreased for longer t_2 values. This decrease gives support to a previous hypothesis by which many M-ions get trapped at unidentified X_i defects during long irradiations, and are lost for the production of TL(M).

The TL(M) production involves the migration of the M-ions along the c-axis channels and thus gives an insight into the processes related to the migration of the interstitial M ions in quartz.

Introduction

Quartz generally contains aluminum impurities substitutional for silicon. Charge compensation is then provided by monovalent ions M^+ (mostly Li^+ and Na^+), thus forming $[AlO_4/M]^\circ$, or in short $(Al\text{-}M)^\circ$, centers. Charge compensationcan also be obtained by positive holes thus forming $(Al\text{-}h)^\circ$ centers.

A surprising effect was observed in the thermoluminescence (TL) of quartz. Under suitable conditions of excitation involving a double irradiation, the glow curve of Al-Li containing quartz, or Q(Li), exhibited a very strong TL peak near 190K (at a heating rate of 10K/min), which did not appear at all under conventional single irradiation low temperature excitation [1,2].

The full mechanism of production of the 190K TL peak, or TL (190), was revealed in an investigation combining TL and ESR [3,4]. Positive holes provided during the first irradiation (e.g. by x-rays), when performed below 150K, get trapped at the $(Al\text{-}Li)^\circ$ centers thus forming $(Al\text{-}Li\text{-}h)^+$ centers. Successive warming to, let us say, 250K (in principle to any temperature in the range 150-300K) will release lithium ions from the aluminum centers leaving behind $(Al\text{-}h)^\circ$ centers. The released Li-ions then migrate along the c-axis channels in the quartz lattice and get trapped near regular silicons forming $[SiO_4/Li]^+$, or $(Si\text{-}Li)^+$, centers. Performing the first irradiation, or X1, at a temperature in the range 150-300K will result in trapping of positive holes as above, followed by practically immediate release of the Li-ion and its trapping at a silicon as above.

When the second irradiation, or X2, is performed at a temperature below 190K, electrons get trapped at the efficient $(Si\text{-}Li)^+$ electron traps forming $(Si\text{-}Li)^\circ$ centers. On warming the sample, the $(Si\text{-}Li)^\circ$ centers release the trapped electrons into the conduction band when approaching 190K, and form the 190K TL peak by recombination of the released electrons with holes at luminescence centers.

A peak analogous to the 190K TL peak has been observed in Al-Na containing quartz, or Q(Na), at about 200K [5,6].

Further work has shown that the TL(190) and TL(200) peaks in Q(Li) and Q(Na) respectively can also be obtained by a single excitation, though at

0-7803-0905-7/93 $3.00 © 1993 IEEE

lower intensities. This occurs by irradiation in a limited range of temperatures (about 150-180K) when the M^+ ions are thermally released from the $(Al\text{-}M\text{-}h)^+$ centers, and the formed $(Si\text{-}M)^o$ centers are still stable [7,8].

Recent work in our laboratory has concentrated on the dependence on the radiation dose of the intensity of the TL(190) and TL(200) and on the temperature dependence of the x-ray luminescence of Q(Li) and Q(Na). Some of the results of these investigations give a further insight into the radiation induced drift of the alkali metal-ions in quartz which is known to play a significant role in the radiation induced frequency offsets in quartz controlled oscillators. The present paper concentrates on the dose dependence of Q(Li) and Q(Na), stressing the effects related to the migration of the alkali metal-ions.

Experimental

Samples were cut from the +x-growth section of Y-plates of a Sawyer synthetic electronic grade quartz crystal. The aluminum content in these samples was 30-50 ppm per silicon atom. The unswept samples contained lithium as charge compensator with a negligible content of sodium. In some samples the lithium was replaced by sodium by electrical sweeping between two graphite electrodes pasted with NaCl. The sweeping was carried out at 530-550°C under an electric field of 400 V/cm directed along the c-axis of the crystal. The sweeping was continued until a charge of 4-5 $Coul/cm^3$ was passed through the sample. This was found to replace practically all the lithium by sodium.

Samples 11 x 8 x 2 mm with the main faces perpendicular to the Y-axis were polished to nearly optical flatness, and were clamped into a bore in the cold finger of a metal liquid nitrogen cryostat provided with spectrosil fused silica windows for the optical measurements and a thin (0.3 mm) aluminum window for the x-irradiation. A heater enabled linear heating from 80 to above 750K. The emission from the sample was focused on the photocathode of an E.M.I 6256S photomultiplier by a quartz lens. A Corning 1-69 heat absorbing filter was placed in the optical path between the sample and the photomultiplier to reduce the black body radiation reaching the photomultiplier which was important near or above 700K. The output of the photomultiplier was amplified and recorded on a chart recorder.

When not otherwise specified, the samples were heated at a rate of 10K/min. The arbitrary units used in the figures for the TL intensity are all the same and can therefore be compared with each other.

The x-irradiation was from a tungsten target at 55Kvp, 18ma, when the sample received a dose of about 2 Gy/sec.

Results

Preliminary measurements of the dose dependence of the $(Si\text{-}M)^o$ related TL(190) and TL(200) have shown wide variations in the slopes of the curves [9]. These variations resulted from differences in the conditions of excitation and in the thermal and radiation history of the samples. The experiments described below were therefore carried out under controlled conditions. To obtain information on the release of the M-ions from the Al-centers, including the production of $(Si\text{-}M)^+$ traps, and separately on the trapping of electrons at the $(Si\text{-}M)^+$ traps, the doses (irradiation times) were varied at only one of the two irradiations, keeping the other constant.

Under the conditions of our experiments the intensity of the $(Si\text{-}M)^o$ related TL was proportional to the concentration of the $(Si\text{-}M)^+$ traps, obtained by the release of M-ions from the Al-centers and their trapping at Si sites along the c-channels. It also depended on the concentration of electrons trapped at the $(Si\text{-}M)^+$ traps during the low temperature X2 irradiation, thus forming $(Si\text{-}M)^o$ centers. The actual concentration of the $(Si\text{-}M)^o$ centers depended not only on the dose of X2, but also on the competition between the $(Si\text{-}M)^+$ traps and other empty electron traps for the free electrons formed during the second x-irradiation. The relative concentration of the competing traps is of course affected by preirradiation or preheating. This will be illustrated below.

Fig. 1 shows a few dose dependence curves for TL(190) of a Q(Li) sample. For all the curves in Fig. 1 each of the two x-irradiations was carried out at 80K, with an intermediate warming to 250K between the two irradiations. The dose of X1 was varied as shown on the abscissa, while that of X2 was kept constant (t_2=5sec) for all the points in Fig. 1. The curves in Fig. 1 differ from each other only in the pretreatments. Thus, for Curve 1, or Fig. 1-1, the sample was preheated to 700K, which emptied all the traps unstable at this temperature. For curves 2, 3 and 4 the sample was preirradiated (for 1h or

more) at 400, 300 and 250K respectively, which filled up to saturation all the traps stable above the respective preirradiation temperature. For each point in Fig. 1-1 and Fig. 1-2 the glow curve was taken by warming to 400K, thus letting electrons accumulate from one point to the next along the curve in traps stable at 400 K. For curves 3 and 4 the TL was taken only up to 300 and 250K respectively.

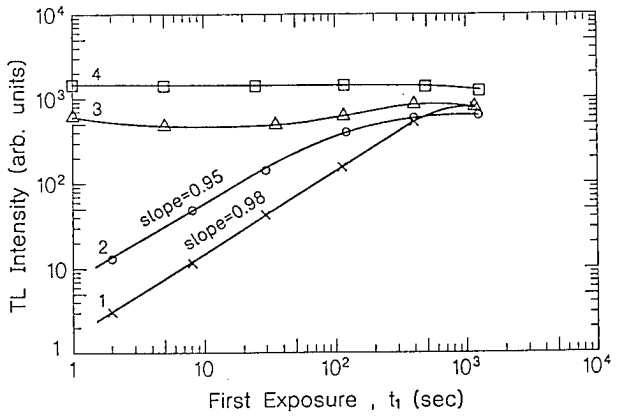


Fig. 1. Dependence of TL(190) on the first irradiation time. Curves 1-4, after filling up to saturation all electron traps stable at 700, 400, 300 and 250K respectively.

Curve 1 rises with a slope close to one and approaches saturation near $t_1=10^3$sec. Further experiments (to be described elsewhere) have shown this rise to be partly due to an increase in concentration of the released Li-ions with t_1, and partly due to a decrease in the competition of the deeper traps which are gradually filled up by the irradiations involved in the building up of Curve 1. Saturation takes place when both the release of Li-ions and the filling of the deeper traps approach equilibrium. The intensity at saturation of the TL(190) gives a measure of the concentration of the $(Si\text{-}Li)^0$ centers. In all the curves of Fig. 1 this was small compared to the concentration of the $(Si\text{-}Li)^+$ traps as only a small fraction of them was filled up by the electrons provided by the $t_2=5$sec second exposure.

Curve 2, also rising with a slope of nearly one, started with intensities 4 times those of Curve 1. This resulted from the lack of competition for electrons by the traps stable above 400K, which are now full to saturation. It starts to saturate at t_1 values near 100 sec, when traps stable between 200 and 400K fill up, an effect which becomes more and more pronounced with further rise in t_1. The TL intensity at saturation is lower in Curve 2 than in Curve 1, an effect which will be dealt with further below.

In Fig. 1-3 the TL at $t_1=1$sec is higher by nearly two orders of magnitude than that in Curve 2. This is mainly due to much lower competition in Curve 3 when all traps stable above 300K are full to saturation. The somewhat lower TL values between $t_1=5$ and $t_1=50$sec is attributed to some emptying of the traps close to 300K when these measurements were taken compared to the case of $t_1=1$sec which was taken soon after the 1 hour irradiation at 300K. At t_1 values of 100 sec or higher, the effect of the long exposure times takes over when the curve rises in a way similar to that in Fig. 1-2. The saturation value is now higher than that in Curve 2. This seems to have been caused by some release of Li-ions during the 300K 1h preirradiation, in addition to those released during the t_1 irradiation at 80K.

In Fig. 1-4, obtained after preirradiation at 250K, there is practically no competition of any traps stable above 250K (which is also the temeprature of the intermediate warming in each measurement). The curve is flat except for a slight decrease near $t_1=10^3$sec. The TL at saturation is now higher than in Curve 3 as a result of the 1h preirradiation at 250K, which is a temperature more efficient in releasing the Li-ions from the Al-centers than the preirradiation at 300K (see References 4 and 6).

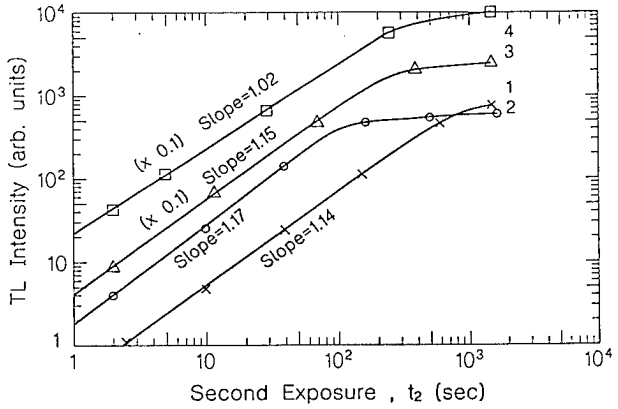


Fig. 2. Dependence of Tl(190) on second irradiation time. Curves 1-4, after filling up to saturation electron traps stable at 700, 400, 310 and 200K respectively.

The curves in Fig. 2, again for TL(190), were obtained with $t_1=2$ sec constant for all the points and with t_2 varying as shown on the abscissa. The

preirradiations, for 1 hour or more, were carried out at 700, 400, 310 and 260K for Curves 1 to 4 respectively. Warming during the glow curve measurements was up to 500, 400, 290 and 250K for the four curves respectively. Slopes of about 1.15 were obtained for Curves 1 to 3, and a slope of 1.0 for Curve 4.

Curves 1 and 2 are similar and nearly equal in intensities to the respective curves in Fig. 1. Curves 3 and 4, (given on a x 0.1 ordinate scale) are now higher compared to curve 2 by factors of about 30 and more than 100 respectively. The saturation value in Fig. 2-4 is higher by nearly two orders of magnitude than that in Fig. 1-4. This shows that 2 sec of x-irradiation releases enough Li-ions to produce the extremely high TL intensities. It is clear that even more Li-ions are released and more $(Si\text{-}Li)^+$ traps are produced by the more than 10^3 sec. of t_1 used at saturation for the curves in Fig. 1. Still the saturation TL in Figs. 1-3 and 1-4 are much lower in intensity than the saturation TL in Fig. 2-4. This is attributed to the short (5 sec) X2 applied for Fig. 1, when the free electrons provided during the irradiation filled up only a small fraction (much less than 1%) of the $(Si\text{-}Li)^+$ traps. In Fig. 2, on the other hand, the long X2 exposures (1500 sec) filled up almost to saturation the $(Si\text{-}Li)^+$ traps produced by the X1 irradiation, which explains the high TL (190) intensities in Fig. 2-4.

The strong dependence of the TL on the X2 dose compared to that on the X1 dose is illustrated in Fig. 3, this time for the TL(200) peak of a Q(Na) sample. All the curves of Fig. 3 were obtained after long preirradiations, filling up to saturation the traps stable at 400K. For Curves 1 and 2, X2 was varied as shown on the abscissa of Fig. 3. X1 was kept at 5sec for Curve 1 and 300 sec for Curve 2. This increase by a factor of 60 in X1 for Curve 2 compared to Curve 1, is seen to have produced an increase by less than a factor of 3 in the TL(200) intensities.

For Figs. 3-3 and 3-4, X1 was varied and X2 was kept constant at 5 and 300 sec for the two curves respectively. This time, increasing X2 by a factor of 60 is seen to give an increase in the TL(200) intensity by nearly a factor of one hundred.

Long preirradiations at 250K gave flat curves for the TL(190) and TL(200) intensities as function of the dose X1, with X2 kept constant. An example of this behavior was given in Fig. 1-4, in which case t_2 was kept at 5 sec. On increasing t_2, TL(190) increased slightly superlinearly with t_2 up to

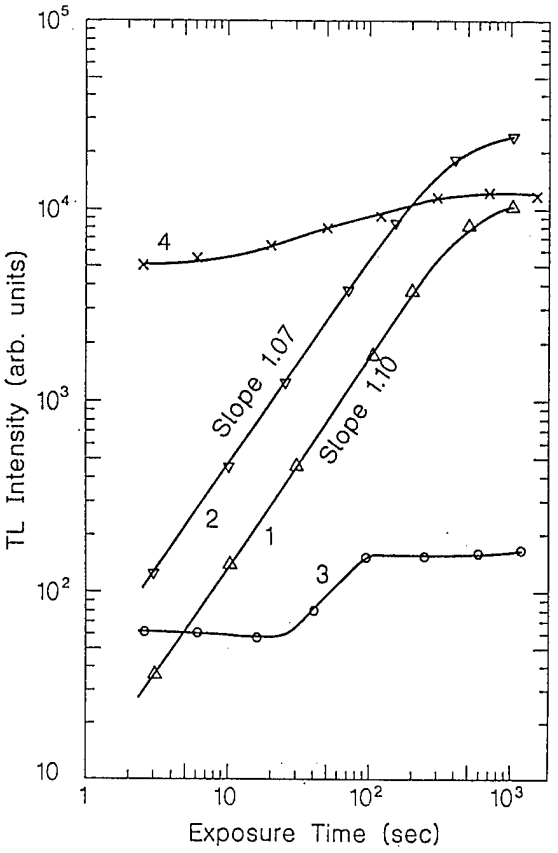


Fig. 3. Curves 1 and 2 - dose dependence of TL(200) on t_2, t_1=5 and 300 sec for Curves 1 and 2 respectively. Curves 3 and 4 - dose dependence of TL(200) on t_1, t_2 = 5 and 300 sec for Curves 3 and 4 respectively. All curves were taken after filling up traps stable above 400 K

about 100 sec. It then reached a maximum near 1500 sec and came down in intensity for larger t_2 values. This behavior is shown in Fig. 4. Fig. 4-1 gives the TL(190) intensity as function of t_2 for t_1 = 2 sec, and Fig. 4-2 gives the same for t_1=1500 sec. Both the curves start rising with a slope slightly above 1. At t_2 values above 20 sec the curves separate and, surprisingly, the curve with the larger X1 exposures (Curve 2) gives the lower TL values. Similar results were obtained for the TL(200) peak of Na-swept samples [9]. This behavior has been attributed to the trapping of M-ions at unidentified X_i defects, when these M-ions get lost for the production of $(Si\text{-}M)^+$, causing the decrease in the TL(190) which is related to the $(Si\text{-}Li)^+$ traps. This explains also the greater reduction in the TL in Fig. 4-2, when the accumulated long t_1 exposures caused

higher losses of the Li-ions which are getting trapped at the X_i defects. This effect is discussed more in detail in Ref. 9.

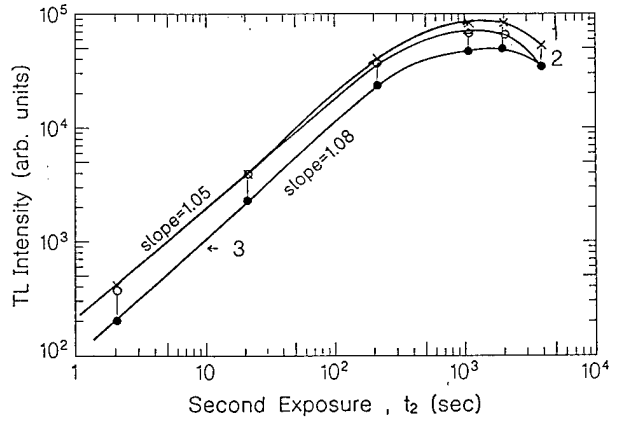


Fig. 4. Dependence of TL(190) on t_2. Curve 1 - t_1 = 2 sec, Curve 2 - t_1 = 1500 sec, Curve 3 - as Curve 1, taken after the exposures involved in Curves 1 and 2. All curves were taken after filling up all traps stable above 250K.

Fig. 4-3 was obtained by exactly the same procedure as for Fig. 4-1, this time after all the exposures involved in Figs. 4-1 and 4-2. The intensities are now about half of those of Curve 1, which, by the above hypothesis means that about 50% of the Li-ions got trapped at the X_i defects. Annealing the sample at 500K or a higher temperature restored the crystal to the original conditions, when the TL(190) gave again the intensities obtained in Figs. 4-1 and 4-2.

The dose dependence of the TL(200) of Q(Na) was in general very similar to that of the TL(190) of Q(Li). More than that, even the TL intensities obtained under similar conditions for Q(Li) and Q(Na) samples were almost equal (within a factor of 2) for the two dopings. It seems, however, that the rate of the release of the alkali-ions from the Al-centers is different for Q(Li) from that for Q(Na). Figs. 5-1 to 5-5 give the dependence of the TL(190) peak on the dose of X2 (on t_2) for t_1 values of 2, 10, 50, 250 and 1250 sec respectively. All the curves were taken after preirradiation at 400K filling up to saturation deeper traps. Up to t_1=50 sec (Curve 3), increasing t_1 by a factor of 5 is seen to give a rise in the TL by about a factor of 4, and an increase by a factor of 10 is obtained by increasing t_1 from 10 sec (curve 2) to 250 sec (curve 4). This is much higher than the increase in TL(200) by nearly a factor of 3 which was obtained by increasing t_1 by a

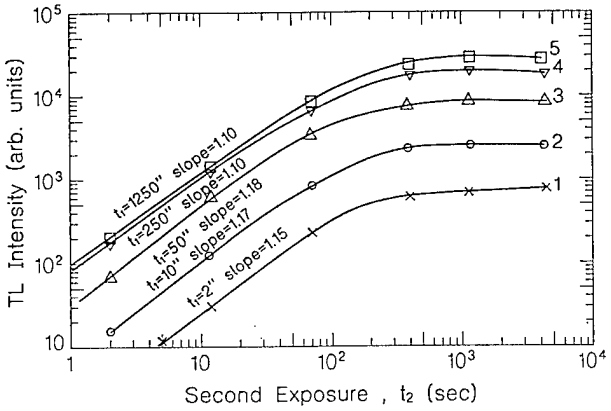


Fig. 5. Dependence of TL(190) on t_2 after long preirradiation at 400K filling up deeper traps. Curves 1 - 5 - after first irradiations (t_1) for 2, 10, 50, 250 and 1250 sec respectively.

factor of 60 (see Figs. 3-1 and 3-2). This difference between Q(Li) and Q(Na) has to be investigated further, which may lead to conclusions on the relative binding of the Li and Na ions at the Al centers.

Discussion and Conclusions

The figures presented in the Results show mainly the dose dependence of the TL(190) peak of Q(Li) samples. The results for the TL(200) peak of Q(Na) samples were very similar to those obtained for the TL(190) peak. We shall therefore make use of the notation TL(M) which will stand for both the TL(190) and the TL(200).

There is one main exception to the above statement. This related to the rate of release of the M-ions from the Al-centers during the X1 irradiation, which was found to be different for the two alkali-metal dopings.

The dose dependence of the TL(M) was observed to change sometimes very strongly with thermal and radiation pretreatments of the samples as shown above (e.g. in Figs. 1 and 2). These changes were shown to result mainly from changes in the competition for electrons between the $(Si\text{-}M)^+$ and other (mainly deeper) empty electron traps. Filling up the deeper traps by preirradiation turned the competition in favor of the empty $(Si\text{-}M)^+$ traps, which resulted in a rise in the TL(M) intensities, and also affected the dose dependence slopes as shown above.

Of special interest are the flat curves describing the TL(M) intensities as function of the X1 exposure time obtained after prolonged irradiations at 250K (Fig. 1-4). Under these conditions all the traps stable above 250K were full to saturation, and so do not compete for electrons during the X2 irradiation. Still, there remains a competition between the $(Si\text{-}M)^+$ traps and the direct recombination of the free electrons with holes at luminescence centers. This later effect is known to produce the very intense x-ray induced luminescence [10]. Now, increasing the dose of X1 (increasing t_1) is accompanied by an increase in the concentration of the $(Si\text{-}M)^+$ traps (e.g. in Fig. 5). One would then expect more and more electrons to get trapped at the $(Si\text{-}M)^+$ traps, with the rise of the TL. This implies, a positive slope in the dose dependence curves, which does not agree with the observed flat curves. It is possible that the very long preirradiations at 250K have released practically to saturation the M-ions from the Al-centers. Under these conditions any additional X1 irradiation will change very little, if at all, the $(Si\text{-}M)^+$ concentration, which explains the flat dose dependence curves.

For all the curves in Fig. 2 t_1 was only 2sec. Curves 1 to 3 give a slope of 1.15. As long as only a small fraction of the $(Si\text{-}M)^+$ traps are full, one expects a linear rise of TL(M) with the dose t_2. Experimentally Curves 1 to 3 rise somewhat superlinearly (slope of 1.15). This superlinearity seems to result from a better filling up of the traps stable above 250K with the increase in t_2. It is somewhat surprising that only 2 sec of X1 produce a concentration of $(Si\text{-}M)^+$ traps which need about an hour of supply of electrons during X2 to have them filled up. This has to do with the relative cross sections for trapping of holes (at the $(Al\text{-}M)^o$ centers), and electrons at the various trapping centers involved during the first and second x-irradiations. An analylsis of these effects is quite complicated, and, is beyond the scope of the present work.

The results give support to a hypothesis put forward previously [11], according to which, under long exposures to x-irradiation, some of the M-ions get trapped at X_i defects. These alkali-metal ions are then lost for the production of $(Si\text{-}M)^o$ centers causing the TL(M) intensities to decrease. The lower saturation intensities in Curves 2 of Figures 1 and 2 compared to Curves 1 in these figures, and the results presented in Fig. 4 give further support to the hypothesis of the trapping of the M-ions at the X_i defects.

In conclusion, the results of the present work explain the various slopes in the dose dependence curves of the TL(M) intensities. In addition, the results give an insight into the processes of the migration of interstitial alkali-ions in quartz, effects known to play a role in the frequency offsets observed in quartz controlled oscillators.

References

[1] D.M. Malik, E.E. Kohnke and W.A. Sibley, "Low Temperature Thermally Stimulated Luminescence of High Quality Quartz", J. Appl. Phys. 52, pp. 3600-3605, 1981.

[2] A. Halperin and S. Katz, "X-ray Transferred Thermoluminescence in Quartz Crystals", J. Luminescence 31, pp.179-132, 1984.

[3] M.G. Jani, L.E. Halliburton and A. Halperin, "Observation of a Simple Lithium Associated Electron Trap in Crystalline SiO_2", Phys. Rev. Letters 56, pp. 1392-94, 1986.

[4] A. Halperin, M.G. Jani and L.E. Halliburton, "Correlated ESR and Thermoluminescence Study of the $[SiO_4/Li]^o$ Center in Quartz, Phys. Rev. 34, pp.5702-5707, 1986.

[5] A. Halperin and S. Katz, "The Thermoluminescence (TSL) Related to the $[SiO_4/Na]^o$ Center in Quartz", Solid State Commun. 63, pp.697-699, 1987.

[6] A. Halperin and S. Katz, "The Thermoluminescence(TSL) of Lithium- and Sodium-Swept Quartz Crystals", Proc. 42nd Freq. Control Symposium, pp. 185-188, 1988.

[7] A. Halperin, "Single Irradiation Excitation of the Thermoluminescence (TSL) Related to the $[SiO_4/Li]^o$ Center in Quartz", J. Phys. Chem. Solids, 51, pp. 303-306, 1990.

[8] A. Halperin and E.W. Sucov, "Single Irradiation Excitation of the Thermoluminescence (TSL) Related to the $[SiO_4/Na]^o$ Center in Quartz", J. Phys. Chem. Solids 52, pp. 1039-1040, 1992.

[9] A. Halperin and E.W. Sucov, "Dose Dependence of the Thermoluminescence (TL) peak at 200K in Al-Na Containing Quartz Crystals", J. Phys. Chem. Solids, in press.

[10] See for example: A. Halperin, "Temperature Dependence of the X-ray Induced Luminescence of Quartz Crystals", J. Luminesc. 48-49, pp. 606-612, 1991.

[11] A. Halperin and S. Katz, "Irradiation Effects on the Low-Temperature Thermo-luminescence (TL) of Quartz Crystals", J. Phys. Chem. Solids 49, pp. 577-583, 1988.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

The determination of internal stress in the quartz crystals

Antonina Pogrebnyak, Stanislav Abdrafikov

The plant "Kristall", Uyzhnouralsk, Russia

Usually the Q-value and inclusion density are measured in the quartz crystals when this crystals are used for production of crystal units. At the last time very often the etch channel density requrements are produced. We put into practice the measurement of the internal stress in addition to above-mensioned for evaluation of quartz quality.

The stresses may be consequence of mechanical treatments or of growth condi sions. For exsample we know the incident when the quartz crystals are cracked spontanelly during of keeping. It is the results of internal stresses destroy actions. The action principle of apparatus for internal stress control is based on measurement of Faraday rotation angle change. Early it was established that stress in quartz influenses to rate of rotation waves diffuce. Faraday rotation angle along optical axis in stress crysLals is have being differed from it in nonstress crystals. The influence of internal stress to value of Faraday rotation angle exchange was established by direct measurements of Faraday rotation angle(FRA) exchange with external mechanical srtain from 0 to 240 kG.

The schema of apparatus is shown in figure 1. For measurements the samples of crystals are enlightened of parallel laser beam along optical axis (Z-axis) The thickness of sample is measured, the FRA is calculated for this case when the crystal is considered as nonstress by formula:

$$\alpha = (d \times \varphi_\lambda - 180 \times k)$$

where: d-thickness of sample in the place where measurement is carried out

φ_λ -the rotaty capacity of crystal for light wave λ, for laser with wave λ =532 nm φ_λ = 27 grad/mm;

k -integer from line 0,1,2,3 ... which is chose for α < 180 grad

Then the polarizer-analyzer is turned at angle α, than at the additional β angle if it is nessessary, so that the transmission of polarizer- crystal-analyzer system reduced to the minimum.

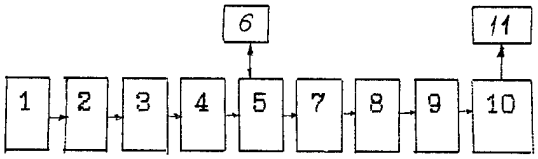


Fig.1 The schema of internal stress control apparatus

1-laser; 2-optical system; 3-aperture; 4-polaroid-polarizer; 5-lens; 6-frosted glass; 7-table for samples; 8-polaroid analyzer; 9-objective; 10-screen; 11-measuring apparatus.

The minimum transmission is registred by measuring apparatus (11). The crystal is

0-7803-0905-7/93 $3.00 © 1993 IEEE

strain if angle $\beta \neq 0$.

For comparison of measuring results for different thicknesses samples the unsizery stress index was calculated by formula:

$$P = \beta / \varphi_{\lambda} \times d$$

where P- internal stress index;

β - additional angle;

The internal stress in the crystals is exactly proportional to P-value.

The investigating samples were 1) Z-bars with dimensions along axes X: Z: Y = 21: 39: 180 mm with seed inside it 2) cube without seed, 20x20 mm sides and 3) plates without seed, 40x100x5 mm. The samples from second grop were of two kinds: the first kind was -the samples which were produced two years ago(cube1) second kind - the samples were produced before measurements (cube2). The influence of the crystal growth rate to P-value was estimated for cube1-samples. The curve of this influence is shown on fig. 2. It was founded that the internal stress increases with increase of growth rate. We noticed that the internal stresses in cube1-samples more than in cube2-samples and decided to ascetain the influence of mechanical treatment as cutting to P-value.

Table 1.

The P-value and estimate in points of crystal quality

P-value, $x10^{-2}$	Estimate point	Comment
0 - 0,2	5	no stresses
0,2 - 0,4	4	stresses are
0,4 - 0,8	3	small
0,8 - 1,2	2	stresses are
> 1,2	1	large

The all samples were divided into five groups by internal stress, P-value according to table 1.

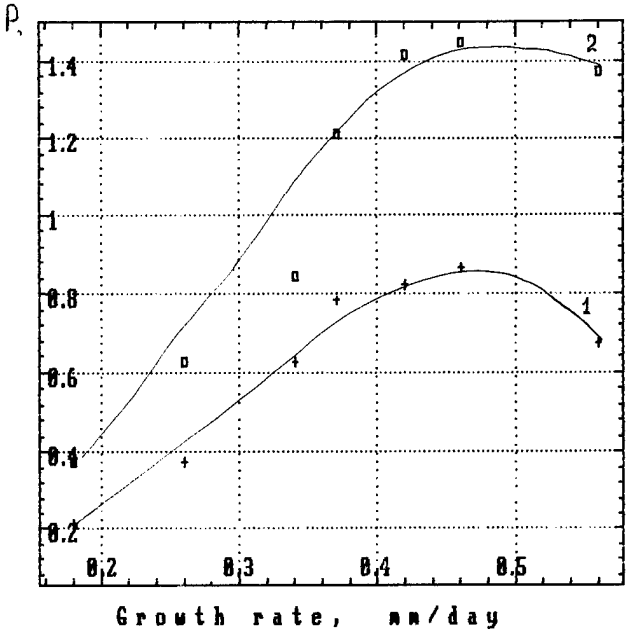


Figure 2. The influence of crystals growth rate to internal stress.

The P-value is the internal pressure along optical axis, it is expressed as kG per square millimetre.

The samples from cube 1-kind were cutting along X axis and internal stress was measured again. The results are shown in the table 2.

The internal stress in the samples increased in 2-10 times after cutting moreover the P-value of crystals which grown more slowly increases greater. When on the same sample a few cuts were made and the internal stress measurements were carried out after each cut, it was found that each cut gives contribution in the internal stress value. The results of the measurements (as shown in the table 3) permit to tell about it.

Table 2.

The results of internal stress measurements

Sample	P-value x 10^{-2}, kG/sq. mm		
	1	2	3
3Aa	0,13	0,84	0,12
3A	0,52	1,41	0,52
3B	0,62	2,16	0,56
7C	1,13	4,31	1,33
8D	0,27	2,22	0,34

1- the internal stress without actions 2- after cutting; 3- after annealing at 500° C, 48 hours.

The increase of the P-value dependes on the quantity, directions and conditions (saws, temperature, liquid) of cuts. And so when you will elaborate the scheme of crystal cutting for the quartz units manufacturing, you should be use as little cuts as possible.

Table 3.

The internal stresses in crystals in dependence on numbers of cuts.

Sample	P-value, x 10^{-2} kg/mm^2		
	one-cut	two-cuts	three-cuts
14C	0,73	0,87	3,6
14D	0,86	1,15	3,4
2Aa	0,23	0,34	0,55

For quick lowering of stress usually the thermal annealing is used. We control three conditions of annealing: 1) 450°C, 2 hours; 2) annealing at 500°C, 8 hours; 3) annealing at 500°C, 48 hours. Results of internal stress measurements before and after annealing are shown in tabl. 2 and 4. The P-value is lowered by third temperature rate. If the first or second conditions of temperature will be use the internal stresses will be redistribute and only partial will be lowering. We noticed that only internal stresses which are appeared after cutting, are lowered by temperature, and the growth internal stresses are not annealed.

We consider that the question about temperature rate of stresses annealing is very important and therefore we will be continued to work at this subject.

Table 4.

The internal stress in crystals after annealing.

Sample	P-value x 10^{-2}, kG/mm^2			
	0	1	2	3
2B	1,62	0,98	0,66	0,62
8B	1,33	1,23	1,24	0,83
1C	0,82	0,68	0,70	0,69
1E	0,60	0,55	0,49	0,47
8E	1,24	0,86	0,87	0,63

0-internal stress before annealing,
1-the first temperature rate;
2-temperature 500°C, exposure 8 hours;
3-the third temperature rate.

We should be studing the influence of samples dimensions, rate of cooling, temperature to the P-value.

We investigated the lowering of internal stress by termoshock. The bars, the samples from first group, were heated in water and than were transfered from warm water to cold water. The temperature difference was 50°. The one part of these samples has no cracks after termoshock; the second part has cracks on the ends of bars; and the third part has cracks both on the ends and inside of bars. For bars, which have no cracks the internal

stress is lowered, see table 5.

Table 5.

The P-value after termoshock.

Sample	P-value x 10^{-2}		Comment
	0	1	
1530-1	1,42	0,44	no cracks
-2	1,33	0,48	-"-
-3	1,32	0,10	-"-
2241-1	0,67	0,40	-"-
-2	2,77	2,13	cracks on
2243-1	0,13	0,55	the ends
-2	0,21	0,67	-"-
2139-1	1,13	0,57	cracks on
-2	0,27	0,55	the ends &
-3	0,23	0,54	inside

0- internal stress before termoshock
1- the measurements after shock.

For bars, that have any cracks the internal stresses are kept or redistributed.

On the samples from group number 3, plates without seed, we determinated the influence of long-term preservation. From crystals of the same mark, but with different date of growth there samples were made. It was founded that the growth internal stresses are lowered by storing of crystals. The large P-value as you'll see in table 6, may be explaining that the samples were made before measuring and it was of cutting influence.

On the small number of samples from this group we tried to establish the exchange of internal stress in crystals with "generation" of seeds. The authories [1] point to this connection. But we have a little results for a simple conclusions were doing.

Table 6.

The internal stress in the different growth date crystals

Sample	Growth date	P-value x 10^{-2}	
		0	1
1	1989	0,39	0,37
2	1990	0,38	0,37
3	1991	0,56	0,42
4	1992	0,90	0,68
5	1993	1,44	0,89

0- the initial internal stress;
1- the internal stress after annealing

So we can drow the following conclusions:

1. They can do the preliminary sorting of the crystals before using by internal stress value by above mention apparatus
2. For the crystals with large P-value (1 or 2 groups) more spare cutting conditions and more long temperatere rate of annealing are required.
3. The crystals after growth have more large P-value than they have after aging by storing.

References

[1] D. Brice, Y. Kylenvain. The quartz materials for resonaters selection. Phillips Technical Review, 1986

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN ANALYSIS OF TRANSVERSELY VARYING THICKNESS MODES IN QUARTZ RESONATORS WITH BEVELLED CYLINDRICAL EDGES

H.F. Tiersten and Y.S. Zhou

Department of Mechanical Engineering, Aeronautical Engineering and Mechanics
Rensselaer Polytechnic Institute, Troy, New York 12180–3590

Abstract

The equation for transversely varying thickness modes in doubly–rotated quartz resonators is applied in the analysis of trapped energy resonators with bevelled cylindrical edges. The coefficients appearing in the planar differential operator are written as a sum of a mean or isotropic part plus a deviation. Asymptotic eigensolutions for the nearby isotropic case are obtained for the cylindrical bevelled resonator. The resonant frequencies for the actual anisotropic case are obtained from an equation for the perturbation in eigenfrequency from the isotropic solution. A lumped parameter representation of the admittance, which is valid in the vicinity of a resonance, is obtained. Calculated results are presented for a few bevelled AT– and SC–cut quartz resonators.

1. Introduction

The equation for transversely varying thickness modes[1] in doubly–rotated quartz plates has been successfully applied in the treatment of flat trapped energy resonators[2], contoured resonators[3] and flat unelectroded circular plates[4]. In this work the equation is applied in the analysis of trapped energy resonators with an electroded flat central region and contoured cylindrical edges. First the planar differential equation for the nth order thickness modes is referred to the coordinate system in which no mixed derivatives occur. Then the transformed coefficients are written as a sum of an isotropic part plus a deviation. Asymptotic eigensolutions for the cylindrical bevelled resonator are obtained for the nearby isotropic case. The resonant frequencies for the actual anisotropic case are obtained from the equation for the perturbation in eigenfrequency from the isotropic solution. However, as in earlier work and for the same reason it turns out that the first perturbation in eigenfrequency vanishes for all modes except those with one angular nodal plane.

A lumped parameter representation of the admittance of the unperturbed isotropic mode shape, which is valid in the vicinity of a resonance, is obtained. Calculated results are presented for bevelled AT– and SC–cut quartz resonators and the influence of the contouring on the mode shape is shown

2. Basic Equations

A plan view and cross–section of the bevelled cylindrical resonator is shown in Fig.1 along with the associated coordinate systems. The x_2–axis points up. In Fig.1, the cylindrical coordinates r and θ are shown, in addition to the particular Cartesian coordinates x_1' and x_3', which are defined in Eqs.(97) and (98) of Ref.1 and will be discussed further later on in this section. Note that θ is measured from the x_3'–axis.

It has been shown[5] that the differential equation for the dominant displacement $u_1^{(n)}$ of the nth harmonic family of transversely varying thickness modes in doubly–rotated quartz plates may be written in the form

$$M_n' \frac{\partial^2 u^n}{\partial x_1'^2} + P_n' \frac{\partial^2 u^n}{\partial x_3'^2} - \frac{n^2\pi^2}{4h^2}\hat{c}^{(1)}u^n - \rho\ddot{u}^n = \frac{\rho\omega^2(-1)^{\frac{n-1}{2}} e_{26} 4V e^{i\omega t}}{c^{(1)}{}_n 2^2\pi^2}, \tag{2.1}$$

where V is the driving voltage applied across surface electrodes and M_n' and P_n' are involved lengthy expressions that are defined in Sec.III of Ref.1 along with x_1' and x_3', which are orthogonal directions in the plane of the plate for which the scalar differential equation (2.1) does not contain mixed derivatives. In (2.1) we have employed the relations

$$u_1^n = u^n(x_1', x_3', t)\sin\frac{n\pi x_2}{2h},$$
$$c^{(1)} = \bar{c}^{(1)}(1 - k_1^2), \tag{2.2}$$
$$\hat{c}^{(1)} = \bar{c}^{(1)}\left[1 - \frac{8k_1^2}{n^2\pi^2} - 2\hat{R}\right],$$

$$k_1^2 = \frac{e_{26}^2}{\bar{c}^{(1)}\epsilon_{22}}, \quad \hat{R} = \frac{2\rho' h'}{\rho h}, \tag{2.3}$$

where $\bar{c}^{(1)}$ is the eigenvalue for the pure thickness mode of the unelectroded plate with nonzero displacement $u_1^{(n)}$, e_{26} is the piezoelectric constant referred to the thickness eigendisplacement direction $u_1^{(n)}$, 2h, 2h′ and ρ, ρ' are the thicknesses and mass densities of the resonator and

0-7803-0905-7/93 $3.00 © 1993 IEEE

electrodes, respectively. The pure thickness frequencies ω_n and $\bar{\omega}_n$ for the unelectroded and electroded regions of the plate, respectively,are given by

$$\omega_n^2 = n^2\pi^2\bar{c}^{(1)}/4h^2\rho, \quad \bar{\omega}_n^2 = n^2\pi^2\hat{c}^{(1)}/4\bar{h}^2\rho, \quad (2.4)$$

where $2\bar{h}$ denotes the thickness of the flat central region and 2h denotes the variable thickness in the contoured region near the edges.

Since the electroded central region and contoured outer region are circular, we write Eq.(2.1) in the form

$$c_n\nabla^2 u^n + A_n - \rho\tilde{\omega}_n^2 u^n + \rho\omega^2 u^n = F_n\,, \qquad (2.5)$$

where

$$A_n = a_n\left[\frac{\partial^2 u^n}{\partial x_1'^2} - \frac{\partial^2 u^n}{\partial x_3'^2}\right], \quad c_n = \frac{M_n' + P_n'}{2}, \quad a_n = \frac{M_n' - P_n'}{2}, \qquad (2.6)$$

in which c_n is the mean or isotropic constant, a_n is the anisotropy constant and A_n is the anisotropy term. In the electroded central region

$$F_n = \frac{\rho\omega^2(-1)^{\frac{n-1}{2}} e_{26} 4V e^{i\omega t}}{c^{(1)}n^2\pi^2}, \quad \tilde{\omega}_n = \bar{\omega}_n\,, \qquad (2.7)$$

and in the unelectroded contoured outer region

$$F_n = 0\,, \quad \tilde{\omega} = \omega_n\,. \qquad (2.8)$$

In cylindrical coordinates r, θ, ∇^2 takes the form

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}\,, \qquad (2.9)$$

and at the edge of the electrode we have the continuity conditions

$$\bar{u}^n = u^n\,, \quad \frac{\partial\bar{u}^n}{\partial r} = \frac{\partial u^n}{\partial r} \text{ at } r = \ell\,. \qquad (2.10)$$

Equations (2.5) – (2.10) define the problem for the anisotropic bevelled resonator. However, before we discuss the solution, we note for later use that the variable thickness 2h in the contoured region and the constant thicknesses $2\bar{h}$ and $2h_d$ in the flat central region and at r = d, which will be defined later, respectively, are given by

$$2h = 2h_o\left[1 - \frac{r^2}{4Rh_o}\right], \quad 2\bar{h} = 2h_o\left[1 - \frac{\ell^2}{4Rh_o}\right],$$

$$2h_d = 2h_o\left[1 - \frac{d^2}{4Rh_o}\right], \qquad (2.11)$$

where R is the radius of a single contour, which is related to the radius of a double contour R′ by R = R′/2.

In the next section we obtain the natural frequencies of the anisotropic bevelled cylindrical resonator by finding the perturbation in eigenfrequencies from the nearby isotropic plate as a result of the anisotropy. To this end we omit A_n in (2.5) for the case of free vibrations and write

$$c_n\nabla^2\bar{v}^\nu + \rho(\omega_\nu^2 - \bar{\omega}_n^2)\,\bar{v}^\nu = 0\,, \quad 0 < r < \ell\,,$$
$$c_n\nabla^2 v^\nu + \rho(\omega_\nu^2 - \omega_n^2)\,v^\nu = 0\,, \quad \ell < r\,, \qquad (2.12)$$

as the equations for the nearby unperturbed isotropic problem, in which $\bar{v}^\nu$ and v^ν are the unperturbed eigendisplacements in the flat electroded central region and contoured outer region, respectively, and ω_ν is the unperturbed eigenfrequency of the nearby isotropic problem. At the interface between the two regions from (2.10), we have the continuity conditions

$$\bar{v} = v\,, \quad \frac{\partial\bar{v}}{\partial r} = \frac{\partial v}{\partial r}\,, \text{ at } r = \ell\,. \qquad (2.13)$$

The equation for the perturbation in eigenfrequency due to the anisotropy is derived at the end of Section I of Ref.4 and is given by

$$\omega = \omega_\nu - \Delta_\nu, \qquad (2.14)$$

where ω is the perturbed frequency and

$$\Delta_\nu = H_\nu/2\omega_\nu N_\nu^2\,, \quad H_\nu = \int_S A_n v^\nu\,ds\,,$$

$$N_\nu^2 = \int_S \rho(v^\nu)^2\,dS\,, \quad A_n = a_n(v^\nu_{,11} - v^\nu_{,33})\,, \qquad (2.15)$$

and the integrations in (2.15) are understood to be over the entire resonator, i.e., the flat central region and bevelled annulus[6].

3. Free Vibrations of Bevelled Cylindrical Plate

We first obtain the unperturbed solution to the isotropic equations (2.12) subject to the continuity conditions (2.13). To this end we write $(2.12)_1$ for the flat central region in cylindrical coordinates, thus

$$\frac{\partial^2 \bar{v}}{\partial r^2} + \frac{1}{r}\frac{\partial \bar{v}}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \bar{v}}{\partial \theta^2} + \alpha_n^2 \bar{v} = 0 , \qquad (3.1)$$

where

$$\alpha_n^2 = \frac{\rho}{c_n}(\omega_\nu^2 - \bar{\omega}_n^2) . \qquad (3.2)$$

The solution satisfying (3.1) may be written in the form

$$\bar{v} = e^{im\theta} f(r) , \quad m = 0 , 1 , 2 , 3 , ... (3.3)$$

the substitution of which in (3.1) yields

$$f'' + \frac{1}{r} f' + \left[\alpha_n^2 - \frac{m^2}{r^2}\right] f = 0 , \qquad (3.4)$$

which is Besssel's differential equation. The solutions satisfying (3.4) and not diverging at r = 0 may be written in the form

$$f = \bar{B} J_m(\alpha_n r). \qquad (3.5)$$

In the contoured outer region the solution decays very rapidly with distance from the edge of the electrode. Consequently, it is possible to make an expansion in $(2.12)_2$ about some distance d slightly larger than ℓ, say 1.1 ℓ, by writing

$$r = d + \zeta , \qquad (3.6)$$

and retaining terms of order ζ/d but not $(\zeta/d)^2$ in $(2.12)_2$ and satisfying the resulting equation[7] to order ζ/d. To this end we substitute (3.6) in $(2.11)_1$ and employ $(2.11)_3$ to obtain

$$2h = 2h_d(1 - \zeta d/2h_d R) , \qquad (3.7)$$

which yields

$$\frac{1}{4h^2} = \frac{1}{4h_d^2}\left[1 + \frac{\zeta d}{h_d R}\right] . \qquad (3.8)$$

Substituting from (3.8) into $(2.12)_2$, we obtain

$$c_n \nabla^2 v - \rho\omega_{nd}^2\left[1 + \frac{\zeta d}{h_d R}\right] v + \rho\omega_\nu^2 v = 0 , \qquad (3.9)$$

where

$$\nabla^2 = \frac{\partial^2}{\partial \zeta^2} + \frac{1}{d}\left[1 - \frac{\zeta}{d}\right]\frac{\partial}{\partial \zeta} + \frac{1}{d^2}\left[1 - \frac{2\zeta}{d}\right]\frac{\partial^2}{\partial \theta^2} ,$$

$$\omega_{nd}^2 = \frac{n^2\pi^2 \bar{c}(1)}{4h_d^2 \rho} , \qquad (3.10)$$

in which $(3.10)_1$ is obtained by employing (3.6) in (2.9) and retaining terms of order ζ/d. The solution satisfying (3.9) may be written in the form

$$v = w(\zeta) e^{im\theta} , \quad m = 0 , 1,2,3 ..., \qquad (3.11)$$

the substitution of which in (3.9) yields

$$\frac{\partial^2 w}{\partial \zeta^2} + \frac{1}{d}\left[1 - \frac{\zeta}{d}\right]\frac{\partial w}{\partial \zeta} - \frac{m^2}{d^2}\left[1 - \frac{2\zeta}{d}\right] w$$

$$- \gamma_n^2 \frac{\zeta d}{h_d R} w + \beta_n^2 w = 0 , \qquad (3.12)$$

where

$$\gamma_n^2 = \rho\omega_{nd}^2/c_n , \beta_n^2 = (\rho/c_n)(\omega_\nu^2 - \omega_{nd}^2). \qquad (3.13)$$

We write the asymptotic solution of (3.12) in the form

$$w = Be^{-k_0\zeta - k_1\zeta^2} , \qquad (3.14)$$

in which the real parts of k_0 and k_1 must be positive. Substituting from (3.14) into (3.12), we obtain

$$\left[k_0^2 - 2k_1 - \frac{k_0}{d} - \frac{m^2}{d^2} + \beta_n^2 + \left[4k_0k_1 - \frac{2k_1}{d} + \frac{k_0}{d^2} + \frac{2m^2}{d^3}\right.\right.$$

$$\left.\left. - \frac{\gamma_n^2 d}{h_d R}\right]\zeta + \left[4k_1^2 + \frac{2k_1}{d^2}\right]\zeta^2\right] Be^{-k_0\zeta - k_1\zeta^2} = 0 , (3.15)$$

and we now note, in accordance with the earlier discussion, that ζ^2/d^2 is negligible when $|w|$ is not negligible and that $|w| \approx 0$ when ζ^2/d^2 is large. Accordingly, we ignore[8] the coefficient of the ζ^2 term in (3.15) and take the constant term and the coefficient of the ζ term to vanish, which after some manipulation yields

$$4k_0^3 - \frac{6k_0^2}{d} + 4\left[\frac{1}{d^2} - N\right] k_0 - 2\left[S - \frac{N}{d}\right] = 0 , (3.16)$$

where

$$N = \frac{m^2}{d^2} - \beta_n^2 , \quad S = \frac{\gamma_n^2 d}{h_d R} - \frac{2m^2}{d^3} . \qquad (3.17)$$

Equation (3.16) is a cubic, which always yields one real positive root k_0 and a corresponding positive k_1, from either of the two equations which led to (3.16).

At this point the solution consists of (3.3) with (3.5) in the flat electroded central region and (3.11) with (3.14) in the outer bevelled annulus. The unperturbed isotropic solution must satisfy the edge conditions (2.13), which with (3.3), (3.11) and (3.6) enables us to write

$$w = f , \quad \partial w/\partial \zeta = \partial f/\partial r \text{ at } r = \ell \quad (\zeta = -g) . (3.18)$$

The substitution of (3.5) and (3.14) in (3.18) yields

$$\bar{B}J_m(\alpha_n \ell) - B\, e^{k_o g - k_1 g^2} = 0 ,$$

$$\bar{B}\alpha_n J_m'(\alpha_n \ell) + B(k_o - 2k_1 g)\, e^{k_o g - k_1 g^2} = 0 , \tag{3.19}$$

which constitute two homogeneous linear algebraic equations in B and $\bar{B}$. For a nontrivial solution the determinant of the coefficients must vanish, which yields the transcendental equation

$$(k_o - 2k_1 g)J_m(\alpha_n \ell) + \alpha_n J_m'(\alpha_n \ell) = 0 , \tag{3.20}$$

since $e^{k_o g - k_1 g^2} \neq 0$. Satisfaction of (3.20) along with (3.16) yields the unperturbed eigenfrequencies ω_{nmp}, where p represents the order of the root α_{nmp} of (3.20).

In order to determine the perturbation in eigenfrequency resulting from the anisotropy we must evaluate H_ν, A_n and N_ν^2 in (2.15). The analysis for doing this in a simpler case is given in Section III of Ref.4, which, although simpler, is nearly identical with what is done here[9]. In that way as for Eq.(52) of Ref.4, we obtain

$$H_{nmp} = a_n \int_0^{2\pi} d\theta \int_0^{\infty} r\, dr \left[\left[-\frac{\partial^2 w}{\partial r^2} + \frac{1}{r}\frac{\partial w}{\partial r} - m^2 \frac{w}{r^2} \right] \cos 2\theta \begin{bmatrix} \cos m\theta \\ \sin m\theta \end{bmatrix} \pm 2m \left[\frac{w}{r^2} - \frac{1}{r}\frac{\partial w}{\partial r} \right] \cdot \sin 2\theta \begin{bmatrix} \sin m\theta \\ \cos m\theta \end{bmatrix} \right] w \begin{bmatrix} \cos m\theta \\ \sin m\theta \end{bmatrix} , \tag{3.21}$$

where from the solution in (3.5) and (3.14), we have

$$w = \bar{B}_{nmp} J_m(\alpha_{np} r) , \quad 0 < r < \ell ,$$

$$w = B_{nmp} e^{-k_o^{nmp}\zeta - k_1^{nmp}\zeta^2} , \quad \ell < r , \tag{3.22}$$

in which the upper terms in (3.21) correspond to the real part of $e^{im\theta}$ and the lower terms, to the imaginary part of $e^{im\theta}$ in the solution functions in (3.3) and (3.11). As a consequence of the periodicity of the trigonometric functions, evaluation of the angular integrals in (3.21) reveals that

$$H_{nmp} = 0 \ \text{ unless } m = 1 , \tag{3.23}$$

while for m = 1, we obtain

$$H_{nlp} = \pm \frac{\pi}{2} a_n \alpha_{nlp}^2 \left[\frac{\ell^2}{2} (J_1^2(\alpha_{nlp}\ell) - J_o(\alpha_{nlp}\ell)J_2(\alpha_{nlp}\ell)) \right] \pm \frac{\pi}{2} a_n \frac{J_1^2(\alpha_{nlp}\ell)}{4k_1^{nlp}} \left[-\beta_{nlp}^2 \left[\sqrt{2\pi k_1^{nlp}}\, e^{W^{nlp}} \left[d - \frac{k_o^{nlp}}{2k_1^{nlp}} \right] \cdot (1 - \operatorname{erf} z^{nlp}) + 1 \right] + \frac{\gamma_{nlp}^2 d}{h_d R} \left[d - g - \frac{k_o^{nlp}}{2k_1^{nlp}} + e^{W^{nlp}} \sqrt{\frac{2\pi k_1^{nlp}}{4k_1^{nlp}}} \left[1 - 2k_o^{nlp} d + \frac{(k_o^{nlp})^2}{k_1^{nlp}} \right] (1 - \operatorname{erf} z^{nlp}) \right] \right] \tag{3.24}$$

where

$$W^{nlp} = \frac{(k_o^{nlp})^2}{2k_1^{nlp}} - 2k_o^{nlp} g + 2k_1^{nlp} g^2 ,$$

$$z^{nlp} = \frac{k_o^{nlp} - 2k_1^{nlp} g}{\sqrt{2k_1^{nlp}}} . \tag{3.25}$$

Clearly, the normalized integral N_ν^2 in (2.15) can be written in the form

$$N_{nmp}^2 = \rho L_{nmp} , \tag{3.26}$$

where

$$L_{nmp} = \int_0^{2\pi} d\theta \left[\int_0^{\ell} r(\bar{v}^{nmp})^2\, dr + \int_{\ell}^{\infty} r(v^{nmp})^2\, dr \right] . \tag{3.27}$$

Since only m = 1 is of interest for perturbation purposes, substitution of (3.3) with (3.5) and (3.11) with (3.14) into (3.27) and evaluation for m = 1 yields

$$L_{nlp} = \frac{\pi}{2} \ell^2 \left[J_1^2(\alpha_{nlp}\ell) - J_o(\alpha_{nlp}\ell)J_2(\alpha_{nlp}\ell) \right] + \frac{\pi}{4k_1^{nlp}} J_1^2(\alpha_{nlp}\ell) \left[1 + \sqrt{2k_1^{nlp}}\,\pi \left[d - \frac{k_o^{nlp}}{2k_1^{nlp}} \right] \cdot e^{W^{nlp}} (1 - \operatorname{erf} z^{nlp}) \right] . \tag{3.28}$$

The perturbed eigenfrequencies for m = 1, may now be obtained from (2.14) and (2.15), which enables us to write

$$\Delta_{nlp} = \frac{H_{nlp}}{2\omega_{nlp}\rho L_{nlp}} , \quad \omega = \omega_{nlp} - \Delta_{nlp} , \tag{3.29}$$

while for $m \neq 1$, there is no perturbation in frequency and we have

$$\omega = \omega_{nmp} \quad m \neq 1 . \tag{3.30}$$

4. Forced Vibrations of Bevelled Cylindrical Plate

Now that we have the free vibrational modes of the bevelled cylindrical plate, we can obtain the forced vibrational solution by expanding in a series of the free vibrational solutions. However, before we do this we must make an important observation. Although we obtained the natural frequencies of free vibration of the anisotropic bevelled cylindrical plate, we did not obtain the mode shape for the actual anisotropic case[10]. As a consequence, we neglect the influence of A_n in the forced vibration analysis[11]. Accordingly, from (2.5) we write the equation for forced vibrations in the electroded region in the isotropic form

$$c_n \nabla^2 \bar{u}^n - \rho \bar{\omega}_n^2 \bar{u}^n + \rho \omega^2 \bar{u}^n = \bar{F}_n e^{i\omega t} , \tag{4.1}$$

where from (2.7), we have

$$\bar{F}_n = \rho \omega^2 (-1)^{(n-1)/2} e_{26} 4V / c^{(1)}{}_n{}^2 \pi^2 \tag{4.2}$$

and from (2.5), (2.8), (3.8) and $(3.10)_2$ we may write the equation for forced vibrations in the unelectroded region in the form

$$c_n \nabla^2 u^n - \rho \omega_{nd}^2 (1 + \zeta d/h_d R) u^n + \rho \omega^2 u^n = 0 . \tag{4.3}$$

We now write the steady–state solution of (4.1) and (4.3) as a sum of the eigensolutions obtained in Section 3, thus

$$\bar{u}^n = e^{i\omega t} \sum_{mp} K^{nmp} \bar{v}^{nmp}, \quad u^n = e^{i\omega t} \sum_{mp} K^{mp} v^{nmp} \tag{4.4}$$

where

$$\bar{v}^{nmp} = f^{nmp} e^{im\theta} , \; v^{nmp} = w^{nmp} e^{im\theta} , \tag{4.5}$$

and f^{nmp} and w^{nmp} denote the r and ζ dependent solution functions obtained in Section 3. Since no energy is lost from the eigensolutions for the bevelled resonator obtained in Section 3, they satisfy an orthogonality condition, which may be written in the form[9]

$$\int_0^{2\pi} d\theta \left[\int_0^{\ell} r \bar{v}^{nmp} \bar{v}^{rst} dr + \int_{\ell}^{\infty} r v^{nmp} v^{nst} \, dr \right] =$$

$$L_{n(m)(p)} \delta_{ms} \delta_{pt} . \tag{4.6}$$

Substituting from (4.4) into (4.1) and (4.3) and employing the homogeneous forms in (3.1) with (3.2) and (3.9), we obtain

$$\sum_{mp} K^{nmp} \rho (\omega^2 - \omega_{nmp}^2) \bar{v}^{nmp} = \bar{F}_n ,$$

$$\sum_{mp} K^{nmp} \rho (\omega^2 - \omega_{nmp}^2) v^{nmp} = 0 , \tag{4.7}$$

in the electroded and bevelled regions, respectively. From (4.7) we form

$$\sum_{mp} K^{nmp} \rho (\omega^2 - \omega_{nmp}^2) \int_0^{2\pi} d\theta \left[\int_0^{\ell} r \bar{v}^{nmp} \bar{v}^{nst} \, dr + \right.$$

$$\left. \int_{\ell}^{\infty} r v^{nmp} v^{nst} \, dr \right] = \int_0^{2\pi} d\theta \int_0^{\ell} \bar{F}_n \bar{v}^{nst} r \, dr, \tag{4.8}$$

and with the aid of the orthogonality condition in (4.6), we obtain

$$K^{nst} = \bar{F}_n \mathcal{J}_{nst} / \rho (\omega^2 - \omega_{nst}^2) L_{nst} , \tag{4.9}$$

$$\mathcal{J}_{nmp} = \int_0^{2\pi} d\theta \int_0^{\ell} r \bar{v}^{nmp} \, dr , \tag{4.10}$$

$$L_{nmp} = \int_0^{2\pi} d\theta \left[\int_0^{\ell} r (\bar{v}^{nmp})^2 \, dr + \int_{\ell}^{\infty} r (v^{nmp})^2 \, dr \right] . \tag{4.11}$$

The substitution of (4.2) into (4.9) and (3.3) with (3.5) for $\bar{B} = 1$ into (4.10), which is then integrated, enables us to write

$$K^{nmp} = \frac{(-1)^{\frac{n-1}{2}} e_{26} 4V \; \mathcal{J}_{nmp}}{c^{(1)}{}_n{}^2 \pi^2 (1 - \omega_{nmp}^2 / \omega^2) L_{nmp}} , \tag{4.12}$$

$$\mathcal{J}_{nmp} = 0 \text{ for } m \neq 0 , \; \mathcal{J}_{nop} = (\ell / \alpha_{nop}) J_1 (\alpha_{nop} \ell). \tag{4.13}$$

On account of the resonance denominator in (4.12), in the vicinity of a resonance one term in the sums in (4.4) dominates the others, and in view of the transformations for electroded regions in Eqs.(87) and (88) of Ref.1, the solution in the electroded flat central region may be written in the form

$$\bar{u} = e^{i\omega t} K^{nmp} \sin \frac{n\pi x_2}{2\bar{h}} \bar{v}^{nmp} - \frac{e_{26} V x_2}{c^{(1)} 2\bar{h}} e^{i\omega t} ,$$

$$\bar{\varphi} = \frac{Vx_2}{2\bar{h}} e^{i\omega t} + \frac{e_{26}}{\epsilon_{22}} \left[\sin \frac{n\pi x_2}{2\bar{h}} - (-1)^{\frac{n-1}{2}} \frac{x_2}{\bar{h}} \right] K^{nmp} \bar{v}^{nmp} e^{i\omega t}, \quad (4.14)$$

where, as usual, ω_{nmp} in Eqs.(4.14) with (4.12) is to be replaced by

$$\hat{\omega}_{nmp} = \omega_{nmp} + i\omega_{nmp}/2Q_{nmp}, \quad (4.15)$$

in which Q_{nmp} is the unloaded quality factor in the nmpth mode. The admittancee Y of the nopth mode of the bevelled resonator is obtained by first substituting from (4.14) into[12]

$$\bar{D}_2 = e_{26}\bar{u}_{,2} - \epsilon_{22}\bar{\varphi}_{,2} \quad (4.16)$$

and then substituting from (4.16) into

$$I = -\int_0^{2\pi} d\theta \int_0^{\ell} \dot{\bar{D}}_2 r \, dr, \quad (4.17)$$

with the result

$$Y = \frac{I}{V} = i\omega \frac{\pi \ell^2}{2\bar{h}} \epsilon_{22}(1 + \hat{k}_1^2) - \frac{i\omega\epsilon_{22}\hat{k}_1^2 4 \mathcal{J}_{nop}^2}{\bar{h}n^2\pi^2(1 - \hat{\omega}_n^2/\omega^2)L_{nop}}, \quad (4.18)$$

which by defining the static and motional capacitances C_o and $\hat{C}_{nop}$, respectively, may be written in the form

$$Y = i\omega[C_o + \hat{C}_{nop}/[(\hat{\omega}_{nop}^2/\omega^2) - 1], \quad (4.19)$$

where the expressions for C_o and $\hat{C}_{nop}$ are clear from (4.18) and (4.19). Finally, the substitution of (3.3) with (3.5) and (3.11) with (3.14) into (4.11) and evaluation for m = 0 yields

$$L_{nop} = \pi\ell^2[J_o^2(\alpha_{nop}\ell) + J_1^2(\alpha_{nop}\ell)] + \frac{\pi}{2k_1^{nop}} J_o^2(\alpha_{nop}\ell) \cdot \left[1 + \sqrt{2\pi k_1^{nop}} \left[d - \frac{k_o^{nop}}{2k_1^{nop}}\right] e^{W^{nop}} (1 - \text{erf } z^{nop})\right]. \quad (4.20)$$

5. Calculated Results

The results of calculations based on the analyses presented in Secs.3 and 4 are presented in this section. However, we first indicate the degree of anisotropy present in the families of anharmonics for the first, third and fifth harmonic for both the AT and SC cuts. In Table I we show the values of c_n, a_n and a_n/c_n for AT– and SC–cut quartz. The table shows that for the third and the fifth families of anharmonics both cuts are very nearly isotropic in the $x_1 - x_3$ plane, whereas for the fundamental both cuts are quite anisotropic. Since the perturbation analysis in Sec.3 shows that for $m \neq 1$ there is no perturbation in eigenfrequency linear in the anisotropy a_n/c_n, the results should be very accurate even for the quite anisotropic fundamentals for $m \neq 1$ because the first correction to the eigenfrequency for the nearby isotropic plate would be proportional to $(a_n/c_n)^2$, which from the last column in Table I is always quite small. Since for m = 1 the perturbation in frequency is calculated to order a_n/c_n, the results should be accurate even for m = 1 in the case of the fundamental families. However, since the anisotropy is so small for the third and fifth families of anharmonics, the perturbation in frequency for m = 1 should be essentially negligible for these families. The calculations indicate that this is indeed the case.

Calculations have been performed for a fundamental, 5 MHz AT–cut quartz resonator with an electrode diameter of 4.6 mm and a radius of curvature R for a single outer contour of 2.94 cm, which corresponds to a biconvex outer annulus with radii of curvature R′ = 2R = 5.88 cm. The results are shown in Table II for m = 0 and 1 and p = 0 – 3. The perturbation in frequency labeled delta is shown for m = 1 and it can be seen that it is not that large even for this highly anisotropic case. It is considerably smaller than the spacings in frequency between the anharmonics. The calculated motional capacitance in farads is shown for m = 0.

It is not shown for m = 1 because, as already noted, in this description it vanishes for $m \neq 0$. Table III shows the calculated results for a third overtone, 10 MHz AT–cut resonator for the geometric configuration data shown at the top of the table. It can be seen that the calculated perturbation in frequencies for m = 1 is considerably smaller for the third overtone family than for the fundamental and is, in fact, negligible. As already noted, this is a consequence of the near isotropy in the $x_1' - x_3'$ plane of the third overtone family of modes. Tables IV and V show the same type of calculated results for bevelled SC–cut resonators as Tables II and III, respectively, show for bevelled AT–cut resonators.

Figure 2 shows the radial variation in mode shape for the fundamental, 5 MHz, AT–cut bevelled resonator for different values of the radius of curvature R of the contour. The curves show how the mode becomes more sharply confined to the center of the plate with decreasing R. The curve labeled A is for the bevelled resonator with m = 0, p = 0 considered in Table II and the curve labelled E may be considered the mode shape for the flat plate because increasing R further does not change the mode shape in the scale of the figure. The energy ratio de, which is defined as the ratio of the energy in the calculated mode from the edge of the plate to ∞ divided by the total energy in the calculated mode to ∞, is also shown in the figure. The curves and the values of de show that the energy in the mode at the edge of the plate is indeed negligible in the

bevelled resonator [A] and is quite large in the flat plate. Figure 3 shows the radial variation in mode shape for the same resonator with m = 0, p = 1 as a function of the radius of the contour. In this case the curve labeled D represents the R of the flat plate.

Acknowledgements

We wish to thank R.C. Smythe of Piezo Technology, Inc. for suggesting the problem, for a valuable comment and providing the relevant geometric data for both AT–cut resonators treated in this work.

This work was supported in part by the Army Research Office under Grant No.DAAL03–91–G–0120.

REFERENCES

1. D.S. Stevens and H.F. Tiersten, "An Analysis of Doubly–Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," J. Acoust. Soc. Am., 79, 1811 (1986)

2. H.F. Tiersten, "Analysis of Trapped Energy Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist," J. Acoust. Soc. Am., 59, 879 (1976)

3. H.F. Tiersten and R.C. Smythe, "An Analysis of Contoured Crystal Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist," J. Acoust. Soc. Am., 65, 1455 (1979).

4. H.F. Tiersten and R.C. Smythe, "Coupled Thickness–Shear and Thickness–Twist Vibrations of Unelectroded AT–Cut Quartz Plates," J. Acoust. Soc. Am., 78, 1684 (1985).

5. Ref.1, Eq.(99).

6. The derivation of the perturbation relations in (2.15) is similar to that in Sec.I of Ref.4 with the additional complication that the integrals over the central region and bevelled annulus are added to form the equivalent of Eq.(16) of Ref.4. The terms arising at the edge of the electrodes as a result of the integration by parts vanish on account of (2.10) and (2.13), and since the solution essentially vanishes before the outside edge of the plate is reached, no edge integrations appear in $(2.18)_2$. For more detail see Ref.9.

7. We could take $d = \ell$ in $(2.12)_2$, but a small error is made by neglecting terms of order $(\zeta/d)^2$. By taking d slightly larger than ℓ, we can make this error smaller. The solution is not sensitive to the exact value of d in this small range.

8. The procedure could be carried to higher order, i.e., ζ^2/d^2, by including $e^{-k_2\zeta^3}$ in the solution and retaining terms of order ζ^2/d^2 everywhere. This would increase the order of Eq.(3.16) and is not deemed to be purposeful in this work.

9. For more detail see H.F. Tiersten and Y.S. Zhou, "On Transversely Varying Thickness Modes in Quartz Resonators with Bevelled Cylindrical Edges," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.

10. We found the eigenfrequencies by obtaining the perturbation in frequency from the nearby isotropic solution resulting from the anisotropy. However, since the perturbation in frequency vanishes for all $m \neq 1$, the mean (or isotropic) mode shape is quite accurate for all $m \neq 1$. Moreover, for n = 3 and 5, the perturbation in frequency is very small for m = 1, as shown in Tables II–V. Consequently, for n = 3 and 5 the mean mode shape is accurate even for m = 1. However, on account of this approximation, although the motional capacitances for all nop modes, which are the largest ones and of most interest, can be calculated very accurately, the very small motional capacitances of all the modes with $m \neq 0$ resulting from the small anisotropy of the actual mode shape cannot be calculated from this description.

11. If A_n were included in the forced vibration analysis, the orthogonality of the isotropic solution functions would be lost and the solution would become complicated and not purposeful.

12. Equation (135) of Ref.1. As already noted, all other Y_{nmp} $(m \neq 0)$ vanish by virtue of (4.13).

TABLE I

PLANAR ISOTROPIC AND ANISOTROPIC COEFFICIENTS FOR TRANSVERSELY VARYING THICKNESS MODES IN AT- AND SC-QUARTZ RESONATORS

n	Cut	c_n 10^{10} N/m^2	a_n 10^{10} N/m^2	a_n/c_n
1	AT	8.9222	2.063	0.231
	SC	7.716	2.043	0.265
3	AT	7.213	0.355	0.049
	SC	5.877	0.516	0.088
5	AT	7.939	1.080	0.136
	SC	7.495	0.411	0.055

TABLE II

CALCULATED RESONANT FREQUENCIES AND MOTIONAL CAPACITANCES OF 5 MHz, FUNDAMENTAL, BEVELLED AT-CUT RESONATOR

AT-cut N = 1 f_{100} = 5 MHz

(R = 2.94 cm, 21 = 4.6 mm, R = R'/2)

First Nodal Axis	Indices m	p	Frequency kHz	Delta kHz	Motional Capacitance Farads
	0	0	5000.002		1.07470×10^{-14}
	0	1	5256.575		7.33360×10^{-16}
	0	2	5684.942		7.57117×10^{-21}
	0	3	6294.308		7.63139×10^{-17}
x_1	1	0	5076.988	17.738	
	1	1	5399.399	41.259	
	1	2	5895.843	59.179	
	1	3	6571.499	84.907	
x_3	1	0	5112.463	-17.738	
	1	1	5481.916	-41.259	
	1	2	6014.200	-59.179	
	1	3	6741.313	-84.907	

TABLE III

CALCULATED RESONANT FREQUENCIES AND MOTIONAL CAPACITANCES OF 10 MHz, THIRD OVERTONE, BEVELLED AT-CUT RESONATOR

AT-Cut N = 3 f_{300} = 10 MHz
(R = 6.625 cm, 2l = 6.4 mm, R = R'/2)

First Nodal Axis	Indices m	p	Frequency kHz	Delta kHz	Motional Capacitance Farads
	0	0	10000.018		1.47260×10^{-15}
	0	1	10065.876		2.06633×10^{-16}
	0	2	10182.045		4.16363×10^{-17}
	0	3	10344.866		3.62606×10^{-18}
x_1	1	0	10022.857	1.010	
	1	1	10112.148	3.139	
	1	2	10249.547	5.830	
	1	3	10431.875	8.436	
x_3	1	0	10024.878	-1.010	
	1	1	10118.426	-3.139	
	1	2	10261.206	-5.830	
	1	3	10448.748	-8.436	

TABLE IV

CALCULATED RESONANT FREQUENCIES AND MOTIONAL CAPACITANCES OF 5 MHz, FUNDAMENTAL, BEVELLED SC-CUT RESONATOR

SC-cut N = 1 f_{100} = 5 MHz
(R = 3.0 cm, 2l = 5.0 mm, R = R'/2)

First Nodal Axis	Indices m	p	Frequency kHz	Delta kHz	Motional Capacitance Farads
	0	0	5000.000		3.71774×10^{-15}
	0	1	5197.476		3.17181×10^{-16}
	0	2	5528.666		5.13496×10^{18}
	0	3	5985.008		1.98911×10^{-17}
x_1	1	0	5056.670	15.854	
	1	1	5300.229	40.271	
	1	2	5674.207	58.285	
	1	3	6183.874	74.628	
x_3	1	0	5088.377	-15.854	
	1	1	5380.771	-40.271	
	1	2	5790.777	-58.285	
	1	3	6333.130	-74.628	

TABLE V

CALCULATED RESONANT FREQUENCIES AND MOTIONAL CAPACITANCES OF
10 MHz, THIRD OVERTONE, BEVELLED SC-CUT RESONATOR

SC-Cut N = 3 f_{300} = 10 MHz
(R = 6.0 cm, 2l = 6.8 mm, R = R'/2)

Nodal Axis	Indices m	p	Frequency kHz	Delta kHz	Motional Capacitance Farads
	0	0	10000.004		4.81910×10^{-16}
	0	1	10049.316		7.36993×10^{-17}
	0	2	10136.912		1.89057×10^{-17}
	0	3	10260.877		3.98826×10^{-18}
x_1	1	0	10016.493	1.339	
	1	1	10082.197	4.272	
	1	2	10184.283	8.304	
	1	3	10321.098	12.719	
x_3	1	0	10019.170	- 1.339	
	1	1	10090.745	- 4.272	
	1	2	10200.891	- 8.304	
	1	3	10346.536	-12.719	

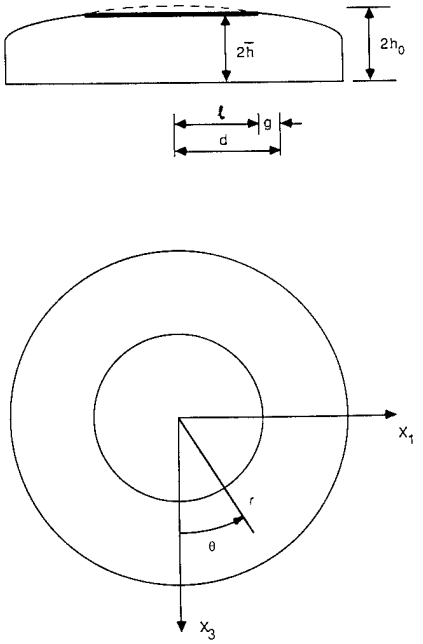


Figure 1 Cross-Section and Plan View of Bevelled Resonator

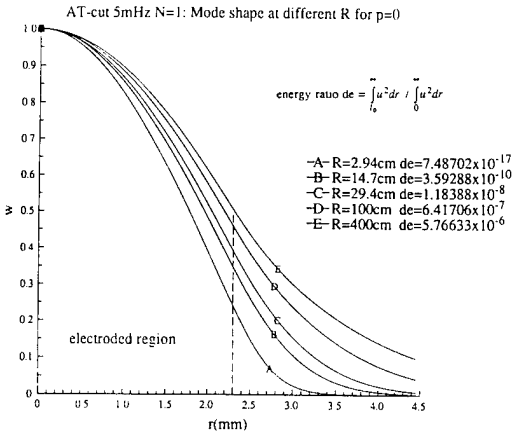


Figure 2 Radial Dependence of Mode Shape for the Fundamental, 5 MHz, AT-Cut Bevelled Resonator for Different Radii of Curvature of the Contour

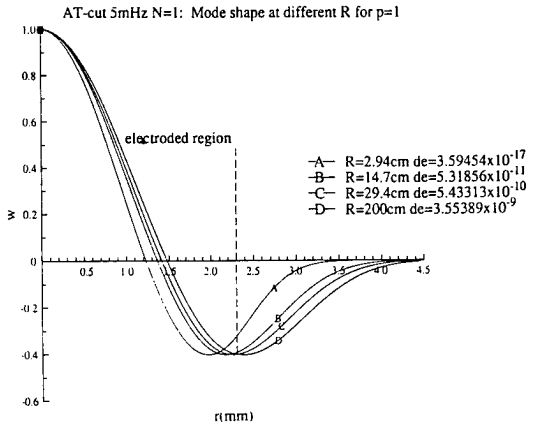


Figure 3 Radial Dependence of Mode Shape for the First Radial Anharmonic for the Fundamental, 5 MHZ, AT-Cut Bevelled Resonator for Different Radii of Curvature of the Contour

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CALCULATED ORIENTATIONS AND ASPECT RATIOS OF STIFFENED RECTANGULAR SUPPORT SYSTEMS OF SC-CUT QUARTZ RESONATORS WHICH MINIMIZE THE INFLUENCE OF FABRICATION IMPERFECTIONS ON ACCELERATION SENSITIVITY

Y.S. Zhou and H.F. Tiersten

Department of Mechanical Engineering, Aeronautical Engineering and Mechanics
Rensselaer Polytechnic Institute, Troy, New York 12180-3590

Abstract

Existing computer programs are employed in the calculation of rectangular support orientations and aspect ratios to minimize the influence of fabrication imperfections on the resultant acceleration sensitivity of stiffened rectangular support systems of SC–cut quartz resonators. Both contoured and flat plate trapped energy resonators are treated. The shorter distance between the sidewalls is established by a criterion associated with the energy in the mode shape for each particular harmonic. Results are presented for a range of energy criteria and a number of contoured and flat plate resonators. Although the square plate never yields good results, we obtain optimal and near optimal results for aspect ratios considerably less than 2.0

1. Introduction

It has been shown[1,2] that the resultant acceleration sensitivity of contoured quartz resonators vanishes for perfect symmetry of the resonator with respect to the support system. It has further been shown[1] that any loss of symmetry in the combined resonator plus support configuration results in a linear increase in the acceleration sensitivity. Since it is impossible to construct a perfectly symmetric resonator plus support configuration in practice, a stiffened structure has been under investigation. The structure consists of the active biconvex quartz resonator attached to identical top and bottom quartz cover plates by means of small sidewalls around the periphery. The mounting clips are attached to the top and bottom cover plates without touching the active plate.

In recent work analyses of both the normal[3] and in–plane[4] acceleration sensitivities of quartz resonators stiffened by rectangular quartz cover plates supported by clips were performed. Calculations based on these analyses have been presented and it was shown[3,4] that for a particular rectangular area of the active resonator and orientation of the rectangular support system, the influence of fabrication imperfections on the acceleration sensitivity was minimized for an aspect ratio of 2.0.

In this work the existing computer programs[3,4] are used in the calculation of rectangular support orientations and aspect ratios to minimize the influence of fabrication imperfections on the acceleration sensitivity of a number of SC–cut quartz resonators. Both contoured and flat plate trapped energy resonators are treated. The shorter distance between the sidewalls is established by a criterion associated with the energy in the mode shape for each particular harmonic. Results are presented for a range of energy criteria and a number of contoured and flat plate resonators. Although the square plate never yields good results[5], we obtain optimal and near optimal results for aspect ratios considerably less than 2.0. Indeed, it is shown that an aspect ratio of 1.5 consistently yields optimum or near optimum results. However, the orientation of the support rectangle varies significantly from one resonator to another.

2. Preliminary Considerations

For purely elastic nonlinearities the equation for the perturbation in eigenfrequency[6] may be written in the form given in Eq.(2.1) of Ref.4, which is the form

$$\Delta_\mu = H_\mu/2\omega_\mu\,, \quad \omega = \omega_\mu - \Delta_\mu\,, \tag{2.1}$$

where ω_μ and ω are the unperturbed and perturbed eigenfrequencies, respectively. It has further been shown that H_μ may be written in the form given in Eq.(2.9) of Ref.4, which is the form

$$H_\mu = -\int_V \hat{c}_{L\gamma M\alpha} g^\mu_{\gamma,L} g^\mu_{\alpha,M}\, dV\,, \tag{2.2}$$

where $\hat{c}_{L\gamma M\alpha}$ is defined in Eq.(2.4) of Ref.4 and all other quantities are defined in Sec.2 of Ref.4. Since g^μ_α denotes the normalized mode shape in the resonator and from Eq.(2.4) of Ref.4 $\hat{c}_{L\gamma M\alpha}$ depends on the biasing state, H_μ can be evaluated when the mode shape in the resonator and biasing state are known.

A schematic diagram of a contoured crystal resonator along with the associated coordinate system is shown in Fig.1. It has been shown that the eigensolution for the dominant displacement of the harmonic modes in a contoured quartz resonator can be written in the form[7]

$$u_{1n} = \sin\frac{n\pi X_2}{2h}\, u_{noo}\, e^{i\omega_{noo}t}\,,$$

$$n = 1,3,5,\ldots\,, \tag{2.3}$$

0-7803-0905-7/93 $3.00 © 1993 IEEE

where

$$u_{noo} = e^{-\alpha_n \frac{\hat{X}_1^2}{2}} e^{-\beta_n \frac{\hat{X}_3^2}{2}} , \quad (2.4)$$

in which

$$\alpha_n^2 = \frac{n^2\pi^2\hat{c}(1)}{8Rh_o^3 M_n'} , \beta_n^2 = \frac{n^2\pi^2\hat{c}(1)}{8Rh_o^3 P_n'} , \quad (2.5)$$

and M_n' and P_n' are involved lengthy expressions that are defined in Sec.II of Ref.7 along with $\hat{X}_1$ and $\hat{X}_3$, which are orthogonal directions in the plane of the plate for which the scalar differential equation for the nth anharmonic family of modes does not contain mixed derivatives and R denotes the radius of the spherical contour. The eigenfrequencies corresponding to the eigensolutions for the harmonic modes are given by

$$\omega_{noo}^2 = \frac{n^2\pi^2\hat{c}^{(1)}}{4h_o^2\rho}\left[1 + \frac{1}{n\pi}\left[\frac{2h_o}{R}\right]^{1/2} \cdot \right.$$

$$\left. \cdot \left[\sqrt{\frac{M_n'}{\hat{c}(1)}} + \sqrt{\frac{P_n'}{\hat{c}(1)}}\right]\right] , \quad (2.6)$$

where $\hat{c}^{(1)}$ is defined in Eq.(78) of Ref.7.

In the in–plane case only the dominant displacement field in (2.3) is needed. However, in the normal case the equivalent trapped energy mode shape, which is much more complicated, is needed. Since this is discussed in Sec.VI of Ref.3 along with Sec.5 of Ref.8 and the analyses and computer programs from Refs.3 and 4 are simply used in this work, we do not present the equivalent trapped energy mode shape here. In addition to the contoured resonators, which were treated in all the other references, in this work we also treat the flat trapped energy resonator with rectangular electrodes. However, the mode shape is the same as the equivalent trapped energy mode shape discussed in detail in Sec.V of Ref.8, with the exception that Eqs.(63), (64) and (66), (67) of Ref.8 are replaced by the real trapped energy mode equations

$$\hat{\xi}\tan\hat{\xi}\ell = \xi^{SR} , \quad \hat{\nu}\tan\hat{\nu}\ell = \nu^{TM} , \quad (2.7)$$

$$\xi^{SR} = [(k_n/M_n')\Delta_n - \hat{\xi}^2]^{1/2} ,$$

$$\nu^{TM} = [(k_n/P_n')\Delta_n - \hat{\nu}^2]^{1/2} , \quad (2.8)$$

in which 2ℓ is the dimension of the square electrode and where

$$k_n = (\bar{c}^{(1)}\rho)^{1/2}\frac{n\pi}{h} ,$$

$$\Delta_n = \frac{n\pi}{2h}\left[\frac{\bar{c}^{(1)}}{\rho}\right]^{1/2}\left[\frac{4k_1^2}{n^2\pi^2} + \hat{R}\right] , \quad (2.9)$$

and $\bar{c}^{(1)}$, k_1^2 and $\hat{R}$ are defined in Eqs.$(2.2)_2$ and $(2.3)_{2,3}$ of the previous paper. Equations (2.7) – (2.9), respectively, correspond to Eqs.(2.12), (2.10), (2.23) and (2.24) of Ref.9. Since the normal and in–plane biasing states are known from Sec.IV of Ref.3 and Sec.6 of Ref.4, respectively, and the mode shapes are known from the discussion earlier in this section, the $\hat{c}_{L\gamma M\alpha}$ and g_γ^μ are known and the perturbation integral for H_μ in (2.2) can be evaluated.

3. Small Offset of Mode Center

A plan view and cross–section of the stiffened structure is shown in Fig.2 along with the coordinate systems. The origin of the unprimed coordinate system is located at the center of the supports and the origin of the primed coordinate system is located at the center of the mode shape, which we permit to be displaced with respect to each other by an arbitrary small distance $\underset{\sim}{d}$ in an arbitrary direction θ. Hence, $\underset{\sim}{d}$ has arbitrary components d_1 and d_3. From Fig.2 it is clear that we may write

$$X_A = d_A + X_A' , \quad A = 1,3, \neq 2 , \quad (3.1)$$

which may then be substituted in the expression for the biasing state, as in Eqs.(8.2)–(8.4) of Ref.4. Since the perturbation integral in (2.2) vanishes for $d_A = 0$, substitution of the above–mentioned expression and expansion to first order in d_A yields, with the aid of (2.1),

$$\Delta\omega/\omega)^L = \Gamma^L = C_A^L d_A , \quad (3.2)$$

and we note that the C_A^L are complicated expressions which are not terribly revealing and have been calculated.

Since we are interested in the maximum value of the magnitude of the resultant acceleration sensitivity Γ for a given $|\underset{\sim}{d}|$, we write

$$\Gamma^2 = \Gamma^L\Gamma^L = C_A^L d_A C_B^L d_B = C_A^L C_B^L \; Q_A Q_B d^2 \quad (3.3)$$

where

$$Q_1 = \cos\theta , \quad Q_3 = \sin\theta , \quad (3.4)$$

and θ is the angle between the X_1– axis and $\underset{\sim}{d}$. Since the maximum value of the resultant sensitivity per unit error d is the quantity of interest, we define γ by

$$\gamma^2 = \Gamma^2/d^2 , \quad (3.5)$$

which may be maximized by

$$\partial\gamma^2/\partial\theta = 0\,, \quad \partial^2\gamma^2/\partial\theta^2 < 0\,, \tag{3.6}$$

to obtain

$$\gamma^2)_{max} = \tfrac{1}{2}\,[C_A^L C_A^L + \Delta]\,, \tag{3.7}$$

where

$$\Delta = \sqrt{(C_1^L C_1^L - C_3^L C_3^L) + 4(C_1^L C_3^L)^2}\,. \tag{3.8}$$

Since the direction of the mispositioning of any clip is fixed, the C_f compose vectorially, i.e.,

$$C_f^2 = C_f^L C_f^L\,, \tag{3.9}$$

where analogous to (3.2), we have

$$(\Delta\omega/\omega)_{f_i}^L = C_{f_i}^L f_i\,, \quad (\Delta\omega/\omega)_{f_i} = C_{f_i} f_i\,. \tag{3.10}$$

4. Calculated Results

From the discussion in Sec.2 we now know g_γ^μ and from Refs.3 and 4, we know $\hat{c}_{L\gamma M\alpha}$ for normal and in–plane acceleration, respectively. Hence, we can now evaluate H_μ in Eq.(2.2). Since the perturbation integral H_μ vanishes for a perfectly symmetric mode–shape plus support configuration, i.e., when both d_A and f_i vanish, in this work we calculate the $\gamma)_{max}$ in (3.7) and C_{f_i} defined in (3.10). As noted earlier the f_i denote the mispositioning of a clip and are scalars, while the d_A denote the components of the planar displacement of the center of the mode shape from the center of the rectangle. In the earlier work[1,2] C was referred to as an offset coefficient and in the more recent work[3,4] the name was changed to a mispositioning coefficient. The symbol $\gamma)_{max}$ is sometimes represented by C, which denotes the mispositioning coefficient for the mode center, and C_{f_i} denotes the mispositioning coefficient for the ith clip shown in Fig.2. Since for a perfectly symmetric system the resultant acceleration sensitivity vanishes, the C and the C_{f_i} are the important quantities to study because when they are minimized, the influence of an error in fabrication is minimized. Since C and C_{f_i} will be plotted in this work and the significance of their value is not familiar, the relation between the values of C and the acceleration sensitivity for a range of mispositioning dimensions is shown by the straight lines in Fig.3.

Calculations of the different C and C_{f_i} have been performed using the known values of the second order[10] and the third order[11] elastic constants of quartz. Since the earlier work in Refs.3 and 4 has shown that low values of C are consistently achieved for ratios of the thickness of the cover plates to the active plate of 2.0, this ratio is adhered to in all calculations presented in this work. Furthermore, since it was shown in Ref.4 that much larger mounting clips on the large side than on the small side resulted in much smaller mispositioning coefficients, we have found that using two clips on each of the large sides and one clip on each of the small sides, as shown in Fig.2, consistently yields smaller mispositioning coefficients. Consequently, we use two clips on the large side and one on the small side in all calculations presented in this work.

In Fig.4 we have plotted $\gamma)_{max} \equiv C$ and the $|C_{f_i}|$ as a function of the orientation of the support rectangle β for an aspect ratio a/b = 1.5 for a 5 MHz, 3rd overtone SC–cut contoured quartz resonator. It can be seen from the figure that the $|C_{f_i}|$ are relatively flat and much lower than C. Since this is consistently the case the $|C_{f_i}|$ are not as important parameters as C. The figure also shows that the optimum value of β is around –45°. In Fig.5 we have plotted C vs β for an aspect ratio a/b = 2.0 for a 5 MHz, 3rd overtone SC–cut contoured resonator for three different values of the energy ratio de defined in the figure. The energy ratio indicates the proportion of the energy in the mode beyond the long edge of the support configuration, i.e., in the short direction, and is used to determine the short dimension of the resonator. The figure clearly shows that there is not much dependence on this ratio in the range shown, and we consistently use de = 10^{-6} in all other calculations in this work.

Figure 6 shows C vs β for different aspect ratios for a 1.5 MHz, fundamental SC–cut contoured resonator. The figure clearly shows the sharp minimum in C around $\beta = -50°$ for aspect ratios a/b $\geq$ 1.3. It also shows that the C vs β curves are relatively flat for a/b = 1.0 and a/b = 1.2. This implies that a circle is not a good support shape for minimizing the influence of fabrication imperfections on acceleration sensitivity. The figure also shows that the lowest value of C is obtained for the aspect ratio a/b = 1.5. Figure 7 shows the optimal values of C vs aspect ratio a/b obtained from Fig.6. The figure shows that C is relatively flat for a/b > 1.5, but becomes appreciably larger for a/b < 1.5. The figure also shows the values of β associated with the optimal values of C. It can be seen that the curve is quite flat for a/b > 1.5.

Figure 8 shows the optimal values of C vs aspect ratio a/b for a number of different SC–cut resonators. The curve labeled A was shown in Fig.7. The curve labeled B, which has the lowest values of C, is for a 5 MHz, 3rd overtone contoured SC–cut resonator. The remaining curves are for flat plate trapped energy resonators in which the size of the electrodes determines the size of the mode. With the exception of the curves labeled C and E, the curves seem to indicate that the smaller the spot size, the less the sensitivity to fabrication imperfections. This is essentially a result of the fact that in the case of the 10 MHz flat plate resonator the spot size changes by only about 10 percent for the larger and smaller electrodes, while in the case of the 20 MHz resonator the spot size changes by 30 percent. The curves also indicate that the optimal or near optimal aspect ratio a/b is always

1.5. Figure 9 shows the values of β associated with the optimal values of C shown in Fig.8. The figure clearly indicates that the optimal orientation of the support rectangle depends significantly on the mode.

In closing it should be noted that the symmetric positioning of the two clips on each of the large sides can be adjusted to reduce the mispositioning coefficients. This should be studied further along with the influence of small differences in the stiffnesses of the clips.

Acknowledgements

We wish to thank R.C. Smythe of Piezo Technology, Inc. for suggesting the energy procedure for determining the shorter dimension between sidewalls and that we investigate a number of different resonators.

This work was supported in part by the Army Research Office under Grant No.DAAL03–91–G–0120.

References

1. Y.S. Zhou and H.F. Tiersten, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators with the Mode Shape Displaced with Respect to Rectangular Supports," J. Appl. Phys., 69, 2862 (1991).

2. H.F. Tiersten and Y.S. Zhou, "On the In–Plane Acceleration Sensitivity of Contoured Quartz Resonators Supported Along Rectangular Edges," J. Appl. Phys., 70, 4708 (1991).

3. Y.S. Zhou and H.F. Tiersten, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Quartz Cover Plates Supported by Clips," J. Appl. Phys., 72, 1244 (1992).

4. Y.S. Zhou and H.F. Tiersten, "An Analysis of the In–Plane Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Identical Top and Bottom Quartz Cover Plates Supported by Clips," Proceedings of the 1992 Frequency Control Symposium, IEEE Cat. No.92CH 3083–3, Institute of Electrical and Electronics Engineers, New York 614 (1992).

5. This indicates that the circular plate is not a good shape for keeping the influence of fabrication imperfections on acceleration sensitivity small.

6. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 832 (1978).

7. D.S. Stevens and H.F. Tiersten, "An Analysis of Doubly–Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," J. Acoust. Soc. Am., 79, 1811 (1986).

8. H.F. Tiersten and D.V. Shick, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along Rectangular Edges," J. Appl. Phys., 67, 60 (1990).

9. H.F. Tiersten, "Analysis of Trapped Energy Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist," J. Acoust. Soc. Am., 59, 879 (1976).

10. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha–Quartz, Phys. Rev., 110, 1060 (1958).

11. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

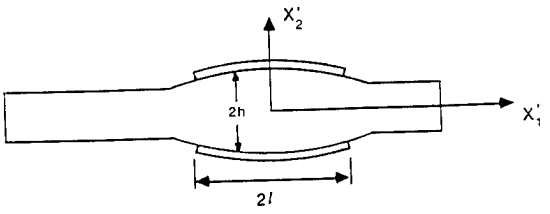


Figure 1 Cross-Section of Biconvex Resonator

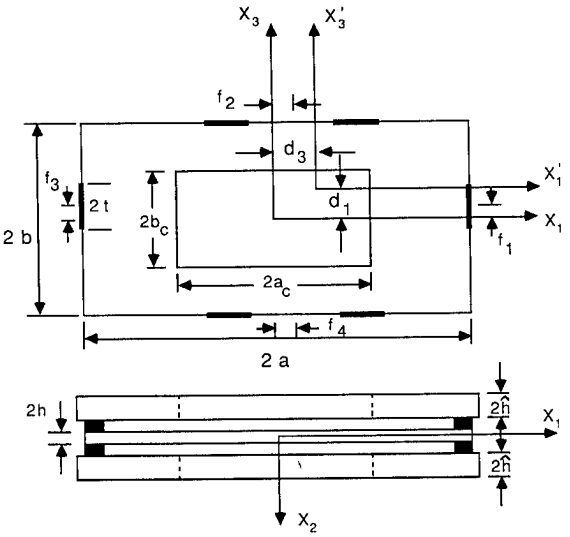


Figure 2 Plan View and Cross-Section of Box Structure Showing Active Plate, Cover Plates, Clips and Sidewalls and Displaced Location of Mode Center

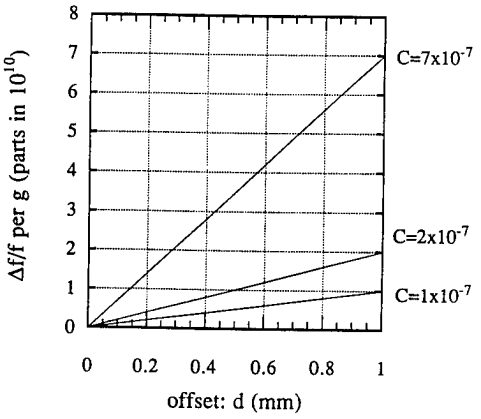


Figure 3 Acceleration Sensitivity Versus Mispositioning Dimension for Different Values of the Mispositioning Coefficient

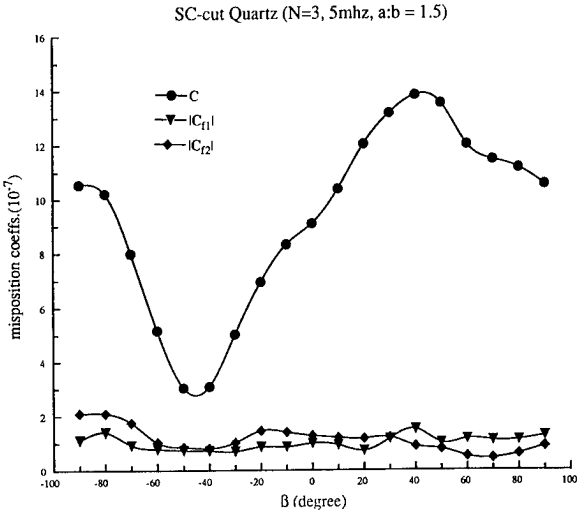


Figure 4 Mispositioning Coefficients Versus Orientation of Support Rectangle for 5 MHz, Third Overtone SC-cut Contoured Resonator. a/b = 1.5.

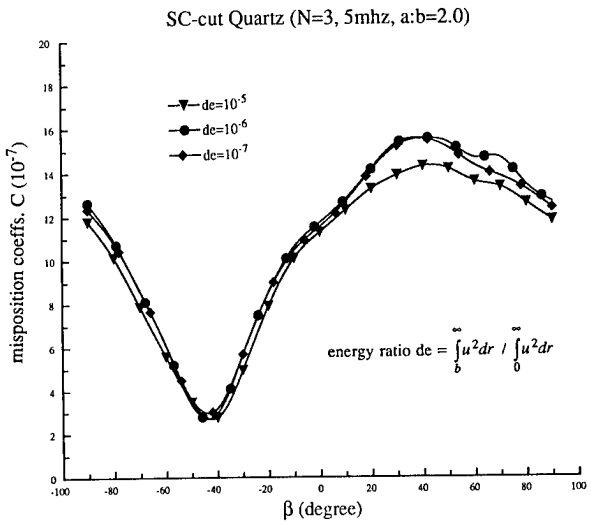


Figure 5 Mispositioning Coefficients Versus Orientation of Support Rectangle for 5 MHz, Third Overtone SC-cut Contoured Resonator for Different de. a/b = 2.0.

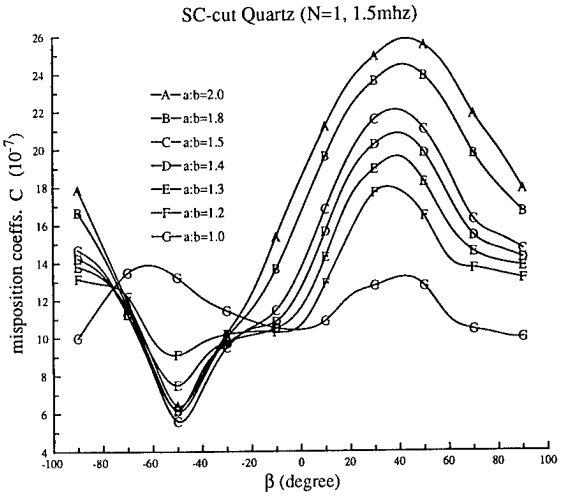


Figure 6 Mispositioning Coefficients Versus Orientation for Different Aspect Ratios for a 1.5 MHz, Fundamental SC-cut Contoured Resonator.

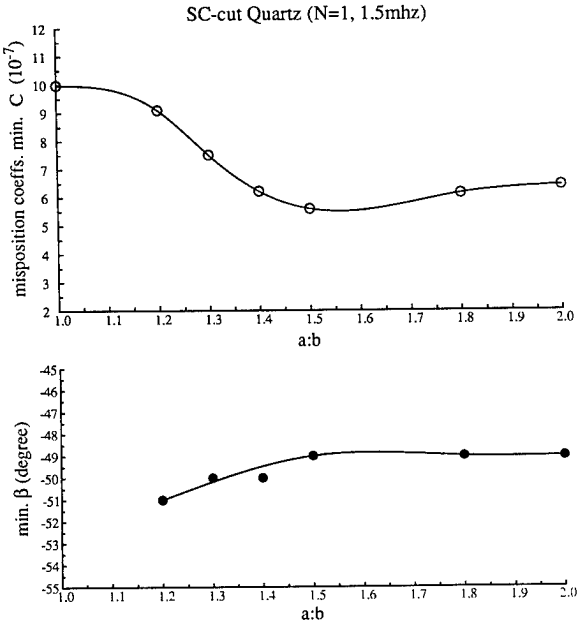


Figure 7 Optimal Mispositioning Coefficients and Associated Orientation of Support Rectangle Versus Aspect Ratio Obtained from Figure 6

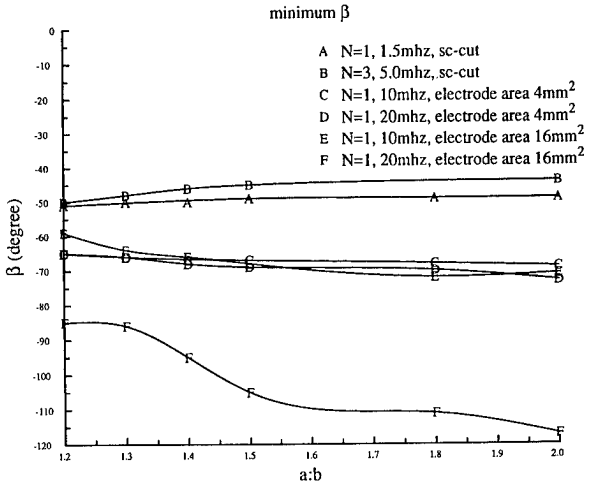


Figure 9 Orientation of Support Rectangle Associated with Optimal Mispositioning Coefficients in Figure 8

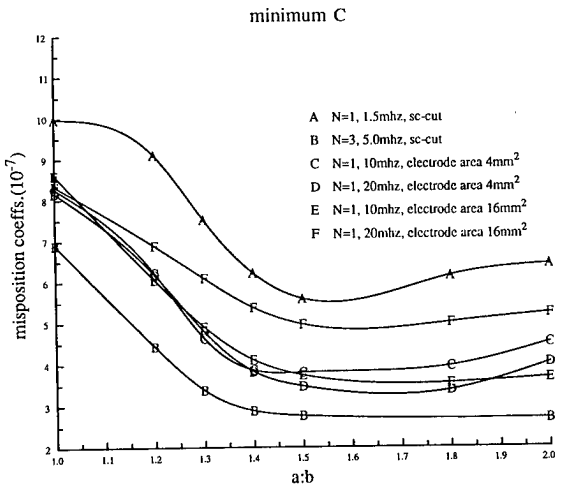


Figure 8 Optimal Mispositioning Coefficients Versus Aspect Ratio for a Number of SC-cut Contoured Resonators and Flat-Plate Trapped Energy Resonators

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A PERTURBATION ANALYSIS ON CONTOURED CRYSTAL PLATES

Weiping Zhang

Component Products Division, Motorola, Inc. Schaumburg, IL 60196

Jiashi Yang

Department of Civil Engineering and Operation Research
Princeton University, Princeton, NJ 08544

Abstract

The present paper presents a perturbation analysis on the eigenvalue problem for the resonance of crystal strip resonators with variable thickness of general shapes. Mindlin's coupled thickness shear and flexure modes model for AT-Cut crystal plates is employed. In the analysis, a small parameter ε, representing the thickness variation, is chosen as the perturbation parameter. The frequencies and mode shapes of contoured plates are then expressed in two parts: $\omega^2 = \omega_0^2 + \varepsilon\omega_1$ and $u_i^{(n)} = u_{i0}^{(n)} + \varepsilon u_{i1}^{(n)}$, where the zero-th order terms correspond to resonance and mode shapes for uniform thickness plates and the first order terms are the corrections due to thickness variations, respectively. Governing equations for the zero-th and first order problem are derived. The Mindlin equations for plates of uniform thickness are naturally recovered as the zero-th order equations. Conditions for the existence of nontrivial solutions to the first order problem is established for strip resonators. With these conditions, the frequency corrections and mode shape corrections are calculated. A numerical example of the approach is given in the paper. A thickness variation of cosine shape is employed in the example. Frequency spectra are constructed and the results are discussed.

I. Introduction

Due to energy trapping, crystal resonators with variable thickness have been studied for a long time. Mindlin [5] studied the uncoupled thickness shear-flexure motions of wedged crystal strip with finite width in x_1 and unbounded in x_3. Bleustein [1] and Pearman [8] did theoretical and experimental studies on wedged strip with infinite x_1 dimension. Thickness twist modes were analyzed in their work. Lee [4] studied quartz plate with x_1 contouring of exponential form. Tiersten [9] published the results of the study on infinite round blank with spherical contouring. Dorwsky [2] studied wedged rectangular blanks to determine the deterioration in the resonator performance due to non-parallel plate surface. The above works are characterized by that only specific contouring shapes are studied so that closed form solutions can be obtained. To the authors' best knowledge, there have not been results on the frequency behavior and mode shape of rectangular crystal plates (strips) with general surface contouring.

From resonator design point of view, Mindlin's pioneer work on the construction of frequency spectrum for crystal plates of constant thickness operating at high frequency provides insight on the crystal behavior and has been used as guidance in the design of such resonators. It is therefore desired to have a frequency spectrum, in parallel with Mindlin's, for contoured plates.

Perturbation method has been effectively used in analyzing static behavior of isotropic elastic plate of variable thickness [10]. Lee[3] used the method to study the effect of temperature change on crystal plate behavior. In the present paper, a perturbation method is employed to analyze vibrations of AT-Cut crystal strips with a small x_1-contouring of general shape. The first order expansion of the solutions are obtained. Frequency spectra, in form of Ω(normalized frequency) vs a/b_0

0-7803-0905-7/93 $3.00 © 1993 IEEE

(width/edge-thickness), are constructed. Corrections of mode shape, due to the contouring, are also derived.

II. Plate Equations

The plate is referred to Cartesian coordinate system as shown in Fig. 1, with the thickness of the plate in x_2

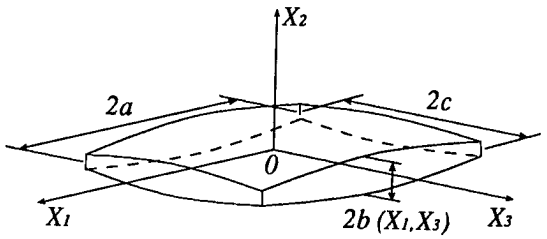


Fig.1 Contoured Crystal Plate

direction and the middle plane in x_1, x_3-plane. The thickness of the plate varies across the plate. The variation is symmetric with respect to the middle plane of the plate. Therefore, the upper and bottom surface of the plate can be described by

$$\begin{aligned} b_{top}(x_1,x_3) &= b(x_1,x_3) \\ b_{bottom}(x_1,x_3) &= -b(x_1,x_3) \end{aligned} \tag{1}$$

When coupled thickness-shear and flexural motions of the plate, with their thickness-twist overtones, are under consideration, the stress equations of motion are, from [6],

$$\begin{aligned} Q_{1,1} + M_{5,13} &= -2b\rho\omega^2 u \\ M_{1,1} + M_{5,3} - Q_1 &= -\frac{2}{3}b^3\rho\omega^2\psi \end{aligned} \tag{2}$$

where time harmonic solutions of the motion with frequency ω are to be sought. Note that we have replaced u_2, ψ_1 used in [6] by u and ψ to facilitate the present discussion.

The stress-displacement relations are:

$$\begin{aligned} Q_1 &= 2b\kappa_1^2 C_{66}(u_{,1} + \psi) \\ M_1 &= \frac{2}{3}b^3\gamma_{11}\psi_{,1} \\ M_5 &= -\frac{2}{3}b^3\gamma_{55}(u_{,13} - \psi_{,3}) \end{aligned} \tag{3}$$

where constants are defined by

$$\begin{aligned} \kappa_1^2 &= \frac{\pi^2}{12} \\ \gamma_{11} &= \frac{c_{11} - c_{12}^2}{c_{22}} - \frac{(c_{14} - c_{12}c_{24}/c_{22})^2}{c_{44} - c_{24}^2/c_{22}} \\ \gamma_{55} &= c_{55} - c_{55}^2/c_{66} \end{aligned} \tag{4}$$

The boundary conditions are

At $x_1 = \pm a$:

$$\begin{aligned} Q_1 + M_{5,3} &= 0 \\ M_1 &= 0 \end{aligned} \tag{5}$$

At $x_3 = \pm c$:

$$M_5 = 0 \tag{6}$$

III. Perturbed Equations for Contoured Plate

Consider the case when the variation of the plate thickness can be described as

$$\begin{aligned} b_{top}(x_1,x_3) &= b_0\{1+\varepsilon f(x_1,x_2)\} \\ b_{bottom}(x_1,x_3) &= -b_0\{1+\varepsilon f(x_1,x_2)\} \end{aligned} \tag{7}$$

where b_0(=constant) is the edge thickness of the plate, ε is a small parameter and, $f(x_1,x_3)$ is function of plate coordinates which characterizes the shape of the contouring.

We will seek for a first order expansion of the solution to the problem defined by (2), (5) and (6) with variable plate thickness given by (7). To this end, let u, ψ and ω be expanded in series form valid for small parameter ε

$$\begin{aligned} u(x_1,x_3) &= u_0(x_1,x_3) + \varepsilon u_1(x_1,x_3) + \cdots \\ \psi(x_1,x_3) &= \psi_0(x_1,x_3) + \varepsilon\psi_1(x_1,x_3) + \cdots \\ \omega^2 &= \omega_0^2 + \varepsilon\omega_1 + \cdots \end{aligned} \tag{8}$$

Substitution of (7) and (8) into (3), then (2), retaining only terms of zero-th and first order of ε, we obtain the perturbed displacement equations of motion

$$\begin{aligned}
&\kappa_1^2 c_{66}(1+\varepsilon f)(u_{0,11}+\varepsilon u_{1,11}+\psi_{0,1}+\varepsilon\psi_{1,1})\\
&+\kappa_1^2 c_{66}\varepsilon f_{,1}(u_{0,1}+\varepsilon u_{1,1}+\psi_0+\varepsilon\psi_1)\\
&-\frac{1}{3}\gamma_{55}b_0^2(1+3\varepsilon f)\\
&(u_{0,1133}+\varepsilon u_{1,1133}-\psi_{0,133}-\varepsilon\psi_{1,133})\\
&-\gamma_{55}b_0^2\varepsilon f_{,1}(u_{0,133}+\varepsilon u_{1,133}-\psi_{0,33}-\varepsilon\psi_{1,33})\\
&-\gamma_{55}b_0^2\varepsilon f_{,3}(u_{0,113}+\varepsilon u_{1,113}-\psi_{0,13}-\varepsilon\psi_{1,13})\\
&-\gamma_{55}b_0^2\varepsilon f_{,13}(u_{0,13}+\varepsilon u_{1,13}-\psi_{0,3}-\varepsilon\psi_{1,3})\\
&+\rho(1+\varepsilon f)(\omega_0^2+\varepsilon\omega_1)(u_0+\varepsilon u_1)\\
&=0
\end{aligned}\tag{9}$$

$$\begin{aligned}
&\frac{1}{3}\gamma_{11}b_0^2(1+3\varepsilon f)(\psi_{0,11}+\varepsilon\psi_{1,11})\\
&+\gamma_{11}b_0^2\varepsilon f_{,1}(\psi_{0,1}+\varepsilon\psi_{1,1})\\
&-\frac{1}{3}\gamma_{55}b_0^2(1+3\varepsilon f)\\
&(u_{0,133}+\varepsilon u_{1,133}-\psi_{0,33}-\varepsilon\psi_{1,33})\\
&\gamma_{55}b_0^2\varepsilon f_{,3}(u_{0,13}+\varepsilon u_{1,13}-\psi_{0,3}-\varepsilon\psi_{1,3})\\
&-\kappa_1^2 c_{66}(1+\varepsilon f)(u_{0,1}+\varepsilon u_{1,1}+\psi_{0,}+\varepsilon\psi_{1,})\\
&+\frac{1}{3}\rho b_0^2(1+3\varepsilon f)(\omega_0^2+\varepsilon\omega_1)(\psi_0+\varepsilon\psi_1)\\
&=0
\end{aligned}\tag{10}$$

Equating the coefficients of like powers of ε in the above two equations, we have the governing equations for zero-th and first order problem, respectively

Zero-th order problem:

$$\begin{aligned}
&\kappa_1^2 c_{66}(u_{0,11}+\psi_{0,1})\\
&-\frac{1}{3}\gamma_{55}b_0^2(u_{0,1133}-\psi_{0,133})+\rho\omega_0^2 u_0=0\\
&\gamma_{11}\psi_{0,11}-\gamma_{55}(u_{0,133}-\psi_{0,33})\\
&-3b_0^{-2}\kappa_1^2 c_{66}(u_{0,1}+\psi_0)+\rho\omega_0^2\psi_0=0
\end{aligned}\tag{11}$$

First order problem:

$$\begin{aligned}
&\kappa_1^2 c_{66}(u_{1,11}+\psi_{1,1})\\
&-\frac{1}{3}\gamma_{55}b_0^2(u_{1,1133}-\psi_{1,133})+\rho\omega_0^2 u_1=G_1(x_1,x_3)\\
&\gamma_{11}\psi_{1,11}-\gamma_{55}(u_{1,133}-\psi_{1,33})\\
&-3b_0^{-2}\kappa_1^2 c_{66}(u_{1,1}+\psi_1)+\rho\omega_0^2\psi_1=G_2(x_1,x_3)
\end{aligned}\tag{12}$$

where in (12)

$$\begin{aligned}
G_1(x_1,x_3)=-\{&\kappa_1^2 c_{66}f(u_{0,11}+\psi_{0,1})\\
&+\kappa_1^2 c_{66}f_{,1}(u_{0,1}+\psi_0)\\
&-\gamma_{55}b_0^2 f(u_{0,1133}-\psi_{0,133})\\
&-\gamma_{55}b_0^2[f_{,1}(u_{0,133}-\psi_{0,33})+f_{,3}(u_{0,113}-\psi_{0,13})]\\
&-\gamma_{55}b_0^2 f_{,13}(u_{0,13}-\psi_{0,3})+\rho u_0(\omega_0^2 f+\omega_1)\}
\end{aligned}$$

$$\begin{aligned}
G_2(x_1,x_3)=-\frac{3}{b_0^2}\{&\gamma_{11}b_0^2(f\psi_{0,11}+f_{,1}\psi_{0,1})\\
&-\gamma_{55}b_0^2 f(u_{0,133}-\psi_{0,33})\\
&-\gamma_{55}b_0^2 f_{,3}(u_{0,13}-\psi_{0,3})-\kappa_1^2 c_{66}f(u_{0,1}+\psi_0)\\
&+\frac{\rho}{3}b_0^2\psi_0(3\omega_0^2 f+\omega_1)\}
\end{aligned}\tag{13}$$

The perturbed boundary conditions, up to first order of ε, can be obtained by inserting (7) and (8) into (3), then (5) and (6), with the results given as

At $x_1=\pm a$:

$$\begin{aligned}
&\kappa_1^2 c_{66}(1+\varepsilon f)(u_{0,1}+\varepsilon u_{1,1}+\psi_0+\varepsilon\psi_1)\\
&-\frac{1}{3}\gamma_{55}b_0^2(1+3\varepsilon f)\\
&(u_{0,133}+\varepsilon u_{1,133}-\psi_{0,33}-\varepsilon\psi_{1,33})\\
&-\gamma_{55}b_0^2\varepsilon f_{,3}(u_{0,13}+\varepsilon u_{1,13}-\psi_{0,3}-\varepsilon\psi_{1,3})\\
&=0
\end{aligned}\tag{14}$$

$$(1+3\varepsilon f)(\psi_{0,1}+\varepsilon\psi_{1,1})=0$$

At $x_3=\pm c$

$$(1+3\varepsilon f)(u_{0,13}+\varepsilon u_{1,13}-\psi_{0,3}-\varepsilon\psi_{1,3})=0\tag{15}$$

Again, equating the coefficients of like powers of ε in (14) and (15), we have boundary conditions of zero-th order and first order problem

Zero-th order problem

At $x_1 = \pm a$:

$$\kappa_1^2 c_{66}(u_{0,1} + \psi_0) - \frac{1}{3} b_0^2 \gamma_{55}(u_{0,133} - \psi_{0,33}) = 0$$
$$\psi_{0,1} = 0 \qquad (16)$$

At $x_3 = \pm c$:

$$u_{0,13} - \psi_{0,3} = 0 \qquad (17)$$

First order problem:

At $x_1 = \pm a$:

$$\kappa_1^2 c_{66}(u_{1,1} + \psi_1) \qquad (18)$$
$$-\frac{1}{3} b_0^2 \gamma_{55}(u_{1,133} - \psi_{1,33}) = F(x_1, x_3)$$
$$\psi_{1,1} = 0$$

At $x_3 = \pm c$:

$$u_{1,13} - \psi_{1,3} = 0 \qquad (19)$$

where in (18)

$$F(x_1, x_3) = \frac{2}{3}\gamma_{55} b_0^2 f(u_{0,133} - \psi_{0,33}) + \gamma_{55} b_0^2 f_{,3}(u_{0,13} - \psi_{0,3})$$

In obtaining (18) and $F(x_1,x_3)$, (16) was used.

Summary of equations:

Problem of order ε^0

Governing equations: (11)

Boundary conditions: (16) and (17)

Problem of order ε^1

Governing equations: (12)

Boundary conditions: (18) and (19)

IV. Rotated At-Cut Quartz Strip

In this section, we consider an AT-Cut strip infinite in x_3 direction. The motion of interest is also x_3-independent. The contouring occurs only in x_1 direction.

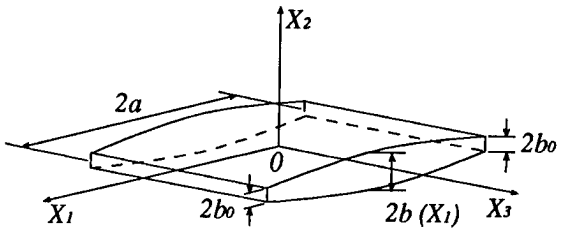


Fig.2 A Segment of an X_1-Contoured Strip Infinite in X_3-direction

The strip configuration is shown in Fig.2. Due to the restriction of x_3 independence of the problem, (11), (12), (16), (17), (18), (19) reduce to, respectively,

Zero-th order problem

Governing equations:

$$\kappa_1^2 c_{66}(u_{0,11} + \psi_{0,1}) + \rho\omega_0^2 u_0 = 0$$
$$\gamma_{11}\psi_{0,11} - \frac{3}{b_0^2}\kappa_1^2 c_{66}(u_{0,1} + \psi_0) + \rho\omega_0^2\psi_0 = 0 \qquad (20)$$

Boundary conditions:

At $x_1 = \pm a$

$$u_{0,1} + \psi_0 = 0$$
$$\psi_{0,1} = 0 \qquad (21)$$

First order problem

Governing equations:

$$\kappa_1^2 c_{66}(u_{1,11} + \psi_{1,1}) + \rho\omega_0^2 u_1 = g_1(x_1)$$
$$\gamma_{11}\psi_{1,11} - \frac{3}{b_0^2}\kappa_1^2 c_{66}(u_{1,1} + \psi_1) + \rho\omega_0^2\psi_1 = g_2(x_1) \qquad (22)$$

Boundary conditions:

At $x_1 = \pm a$

$$u_{1,1} + \psi_1 = 0$$
$$\psi_{1,1} = 0 \qquad (23)$$

where in (22) $g_1(x_1)$ and $g_2(x_1)$ are given by, respectively

$$g_1(x_1) = -\kappa_1^2 c_{66} f(u_{0,11} + \psi_{0,1}) - \kappa_1^2 c_{66} f_{,1}(u_{0,1} + \psi_0) - \rho u_0 (f\omega_0^2 + \omega_1)$$

$$g_2(x_1) = -\frac{3}{b_0^2}[\gamma_{11} b_0^2 (f\psi_{0,11} + f_{,1}\psi_{0,1}) - \kappa_1^2 c_{66} f(u_{0,1} + \psi_0) + \frac{b_0^2}{3}\rho\psi_0(3f\omega_0^2 + \omega_1)] \qquad (24)$$

Due to the unbound feature in x_3 direction, boundary conditions on $x_3 = \pm c$ are omitted.

IV.1 Determination of ω_1

The zero-th order problem defined by (20) and (21) has been solved by Mindlin[6] which characterizes vibrations of a plate with uniform thickness. Since the present paper concerns primarily with the corrections to u, ψ, and ω due to the plate contouring, we will not repeat the solution process for u_0, ψ_0, and ω_0. Interested readers may be referred to [6].

Since the homogeneous parts of (22) and (23) have nontrivial solutions, i.e. u_0 and ψ_0, (22) and (23) have a solution only if a solvability condition is satisfied [7]. This solvability condition can then be used to determine ω_1, the correction to ω_0^2.

To derive the solvability conditions for (22) and (23), we multiply $(22)_1$ and $(22)_2$ by u_0 and ψ_0, respectively, integrate over (-a, a) and obtain

$$\int_{-a}^{a} [\kappa_1^2 c_{66}(u_{1,11} + \psi_{1,1})u_0 + \rho\omega_0^2 u_1 u_0]\,dx_1 = \int_{-a}^{a} g_1(x_1)u_0\,dx_1$$

$$\int_{-a}^{a} [\gamma_{11}\psi_{1,11}\psi_0 - \frac{3}{b_0^2}\kappa_1^2 c_{66}(u_{1,1} + \psi_1)\psi_0 + \rho\omega_0^2\psi_1\psi_0]\,dx_1 = \int_{-a}^{a} g_2(x_1)\psi_0\,dx_1 \qquad (25)$$

which, upon integration by parts, becomes

$$\kappa_1^2 c_{66}[u_0(u_{1,1} + \psi_1)]\Big|_{-a}^{a} - \kappa_1^2 c_{66} u_1 u_{0,1}\Big|_{-a}^{a} + \int_{-a}^{a} u_1(\kappa_1^2 c_{66} u_{0,11} + \rho\omega_0^2 u_0)\,dx_1 - \kappa_1^2 c_{66}\int_{-a}^{a}\psi_1 u_{0,1}\,dx_1 = \int_{-a}^{a} g_1(x_1)u_0\,dx_1$$

$$-\frac{3}{b_0^2}\kappa_1^2 c_{66} u_1\psi_0\Big|_{-a}^{a} + -\gamma_{11}\psi_0\psi_{1,1}\Big|_{-a}^{a} - \gamma_{11}\psi_1\psi_{0,1}\Big|_{-a}^{a} + \gamma_{11}\int_{-a}^{a}\psi_1\psi_{0,11}\,dx_1 + \frac{3}{b_0^2}\kappa_1^2 c_{66}\int_{-a}^{a} u_1\psi_{0,1}\,dx_1 - \frac{3}{b_0^2}\kappa_1^2 c_{66}\int_{-a}^{a}\psi_1\psi_0\,dx_1 + \rho\omega_0^2\int_{-a}^{a}\psi_1\psi_0\,dx_1 = \int_{-a}^{a} g_2(x_1)\psi_0\,dx_1 \qquad (26)$$

Next, we multiply $(26)_1$ by $3b_0^{-2}$, add it to $(26)_2$ and obtain

$$-\frac{3}{b_0^2}\kappa_1^2 c_{66} u_1[u_{0,1} + \psi_0]\Big|_{-a}^{a} - \gamma_{11}\psi_1\psi_{0,1}\Big|_{-a}^{a} - \frac{3}{b_0^2}\kappa_1^2 c_{66} u_0[u_{1,1} + \psi_1]\Big|_{-a}^{a} - \gamma_{11}\psi_0\psi_{1,1}\Big|_{-a}^{a} + \frac{3}{b_0^2}\int_{-a}^{a} u_1[\kappa_1^2 c_{66}(u_{0,11} + \psi_{0,1}) + \rho\omega_0^2 u_0]\,dx_1 + \int_{-a}^{a}\psi_1[\gamma_{11}\psi_{0,11} - \frac{3}{b_0^2}\kappa_1^2 c_{66}(u_{0,1} + \psi_0) + \rho\omega_0^2\psi_0]\,dx_1 = \int_{-a}^{a}[\frac{3}{b_0^2}g_1(x_1)u_0 + g_2(x_1)\psi_0]\,dx_1 \qquad (27)$$

The left hand of (27) vanishes identically, given (21), (23) and (20). The desired solvability condition is then

$$\int_{-a}^{a}[\frac{3}{b_0^2}g_1(x_1)u_0 + g_2(x_1)\psi_0]\,dx_1 = 0 \qquad (28)$$

In the following, we will consider specially the m-th resonance frequency $\omega_m^2 = \omega_{0m}^2 + \varepsilon\,\omega_{1m} + \cdots$ and the

associated mode $u^{(m)}$ and $\psi^{(m)}$. Substituting $g_1(x_1)$ and $g_2(x_1)$ of (24) into (28), ω_1, the correction to ω_0^2, can be solved

$$\int_{-a}^{a} u_0^{(m)} \{\kappa_1^2 c_{66}[f(u_{0,11}^{(m)} + \psi_{0,1}^{(m)}) + f_{,1}(u_{0,1}^{(m)} + \psi_0^{(m)})] + \rho\omega_{0m}^2 f u_0^{(m)}\} dx_1$$
$$+ \int_{-a}^{a} \psi_0^{(m)} \{b_0^2 \gamma_{11}[f\psi_{0,11}^{(m)} + f_{,1}\psi_{0,1}^{(m)}] - \kappa_1^2 c_{66} f(u_{0,1}^{(m)} + \psi_0^{(m)}) + b_0^2 \rho\omega_{0m}^2 f\psi_0^{(m)}\} dx_1$$
$$= -\omega_{1m} \int_{-a}^{a} \rho[(u_0^{(m)})^2 + \frac{b_0^2}{3}(\psi_0^{(m)})^2] dx_1$$

or in quotient form

$$\omega_{1m} = -\frac{A^{(m)}}{B^{(m)}}, \qquad m = 1,2,3,\cdots \tag{29}$$

where $A^{(m)}$ is the left hand side of (29), $B^{(m)}$ is the integral in the right hand side of (29).

IV.2. Determination of u_1 and ψ_1

In analyzing certain electrical behavior of a resonator, vibration modes u and ψ are needed. On the other hand, to improve the approximation of ω, ω_2, second order correction to ω_0, may be desired. The determination of ω_2 requires that u_1 and ψ_1 be known in advance. It is therefore natural that u_1 and ψ_1 be solved. To this end, we first present the orthogonality condition of the solutions to the problem defined by (2), (3), (4) (5) and (6). Again we consider the case of no x_3 dependence. With this restriction, the above 4 groups of equations are reduced to

$$\begin{aligned} Q_{1,1} &= -2b(x_1)\rho\omega^2 u \\ M_{1,1} - Q_1 &= -\frac{2}{3}b^3(x_1)\rho\omega^2\psi \end{aligned} \tag{30}$$

$$\begin{aligned} Q_1 &= 2b(x_1)\kappa_1^2 C_{66}(u_{,1} + \psi) \\ M_1 &= \frac{2}{3}b(x_1)^3\gamma_{11}\psi_{,1} \\ M_5 &\equiv 0 \end{aligned} \tag{31}$$

and at $x_1 = \pm a$

$$\begin{aligned} Q_1 &= 0 \\ M_1 &= 0 \end{aligned} \tag{32}$$

In writing (30), (31) and (32) we have particularly stressed the fact that b, the plate thickness, is function of x_1. And again, boundary conditions at $x_3 = \pm c$ are omitted.

Consider two sets of solutions

$$u^{(i)}, \psi^{(i)}, Q_1^{(i)}, M_1^{(i)} \text{ and } u^{(j)}, \psi^{(j)}, Q_1^{(j)}, M_1^{(j)}$$

such that (32) and equations

$$\begin{aligned} Q_{1,1}^{(i)} &= -2b(x_1)\rho\omega_i^2 u^{(i)} \\ M_{1,1}^{(i)} - Q_1^{(i)} &= -\frac{2}{3}b^3(x_1)\rho\omega_i^2\psi^{(i)} \\ Q_{1,1}^{(j)} &= -2b\ (x_1)\rho\omega_j^2 u^{(j)} \\ M_{1,1}^{(j)} - Q_1^{(j)} &= -\frac{2}{3}b^3(x_1)\rho\omega_j^2\psi^{(j)} \end{aligned}$$

are satisfied. Multiplying these equations by $u^{(j)}, \psi^{(j)}$, $-u^{(i)}, -\psi^{(i)}$, respectively, adding and integrating, we have

$$\int_{-a}^{a} [Q_{1,1}^{(i)}u^{(j)} - Q_{1,1}^{(j)}u^{(i)} + M_{1,1}^{(i)}u^{(j)} - M_{1,1}^{(j)}u^{(i)} - (Q_1^{(i)}\psi^{(j)} - Q_1^{(j)}\psi^{(i)})] dx_1$$
$$= 2\rho(\omega_i^2 - \omega_j^2)\int_{-a}^{a} [b(x_1)u^{(i)}u^{(j)} + \frac{b^3(x_1)}{3}\psi^{(i)}\psi^{(j)}] dx_1 \tag{33}$$

Upon integrating by parts, with the use of displacement-stress relation (31), the left hand side of (33) is transformed to

$$[Q_1^{(i)}u^{(j)}+Q_1^{(j)}u^{(i)}]_{-a}^{a}+[M_1^{(i)}\psi^{(j)}+M_1^{(j)}\psi^{(i)}]_{-a}^{a}$$

$$-2\kappa_1^2 c_{66}\int_{-a}^{a} b(x_1)[(u_{,1}^{(i)}+\psi^{(i)})u_{,1}^{(j)}-(u_{,1}^{(j)}+\psi^{(j)})u_{,1}^{(i)}]\,dx_1$$

$$-2\kappa_1^2 c_{66}\int_{-a}^{a} b(x_1)[(u_{,1}^{(i)}+\psi^{(i)})\psi^{(j)}-(u_{,1}^{(j)}+\psi^{(j)})\psi^{(i)}]\,dx_1$$

$$-\frac{2}{3}\gamma_{11}\int_{-a}^{a} b^3(x_1)(\psi_{,1}^{(i)}\psi_{,1}^{(j)}-\psi_{,1}^{(j)}\psi_{,1}^{(i)})\,dx_1$$

The first two terms vanish due to (32). The integrals vanish identically after regrouping the terms. Then, we have the orthogonality condition

$$\int_{-a}^{a}[b(x_1)u^{(i)}u^{(j)}+\frac{b^3(x_1)}{3}\psi^{(i)}\psi^{(j)}]\,dx_1=\delta_{ij} \qquad (34)$$

where in writing (34) a normalization operation is implied, i.e.

$$\int_{-a}^{a}[b(x_1)(u^{(i)})^2+\frac{b^3(x_1)}{3}(\psi^{(i)})^2]\,dx_1=1 \qquad (35)$$

With this orthogonal feature of the solutions, we can proceed to determine $u_1^{(m)}$ and $\psi_1^{(m)}$ (m=1,2,3,...). Note that index m indicates that the m-th eigenvalue (resonance frequency) and its associated eigenfunctions are under discussion. First we express $g_1^{(m)}$ and $g_2^{(m)}$, the inhomogeneous part of (22), in terms of eigenfunctions $u_0^{(i)}$ and $\psi_0^{(i)}$

$$g_1^{(m)}=\sum_i^{\infty}\alpha_i^{(m)}u_0^{(i)}$$
$$g_2^{(m)}=\sum_i^{\infty}\alpha_i^{(m)}\psi_0^{(i)} \qquad (36)$$

The coefficients $\alpha_i^{(m)}$ (for each m, i=1,2,3,..) are determined by multiplying $(36)_1$ and $(36)_2$ by $b_0u_0^{(j)}$ and $\frac{b_0^3}{3}\psi_0^{(j)}$, adding and integrating both sides over (-a,a)

$$\sum_i^{\infty}\alpha_i^{(m)}\int_{-a}^{a}[b_0u_0^{(i)}u_0^{(j)}+\frac{b_0^3}{3}\psi_0^{(i)}\psi_0^{(j)}]\,dx_1$$
$$=\int_{-a}^{a}[b_0g_1^{(m)}u_0^{(j)}+\frac{b_0^3}{3}g_2^{(m)}\psi_0^{(j)}]\,dx_1$$

The use of the orthogonality condition (34) in this equation yields

$$\alpha_i^{(m)}=\int_{-a}^{a}[b_0g_1^{(m)}u_0^{(i)}+\frac{b_0^3}{3}g_2^{(m)}\psi_0^{(i)}]\,dx_1$$
(i=1,2,3,..) (37)

Next, we express $u_1^{(m)}$ and $\psi_1^{(m)}$ in series form in terms of $u_0^{(i)}$ and $\psi_0^{(i)}$, separately,

$$u_1^{(m)}=\sum_i^{\infty}\beta_i^{(m)}u_0^{(i)}$$
$$\psi_1^{(m)}=\sum_i^{\infty}\beta_i^{(m)}\psi_0^{(i)} \qquad (38)$$

To determine coefficients $\beta_i^{(m)}$ (for each m, i=1,2,3,..) we substitute from (38) into (22) and have

$$\kappa_1^2 c_{66}\sum_i^{\infty}\beta_i^{(m)}(u_{0,11}^{(i)}+\psi_{0,1}^{(i)})+\rho\omega_{0m}^2\sum_i^{\infty}\beta_i^{(m)}u_0^{(i)}$$
$$=\sum_i^{\infty}\alpha_i^{(m)}u_0^{(i)}$$
$$\gamma_{11}\sum_i^{\infty}\beta_i^{(m)}\psi_{0,11}^{(i)}-\frac{3}{b_0^2}\kappa_1^2 c_{66}\sum_i^{\infty}\beta_i^{(m)}(u_{0,1}^{(i)}+\psi_0^{(i)}) \qquad (39)$$
$$+\rho\omega_{0m}^2\sum_i^{\infty}\beta_i^{(m)}\psi_0^{(i)}=\sum_i^{\infty}\alpha_i^{(m)}\psi_0^{(i)}$$

With (20), we may rewrite $(39)_1$ to

$$\sum_i^{\infty}\beta_i^{(m)}(-\rho\omega_{0i}^2u_0^{(i)})+\rho\omega_{0m}^2\sum_i^{\infty}\beta_i^{(m)}u_0^{(i)}=\sum_i^{\infty}\alpha_i^{(m)}u_0^{(i)}$$

or

$$\sum_i^{\infty}u_0^{(i)}[\beta_i^{(m)}\rho(\omega_{0m}^2-\omega_{0i}^2)-\alpha_i^{(m)}]=0 \qquad (40)$$

Since eigenfunctions $u_0^{(i)}$ are linearly independent, the above equation holds only if all coefficients of $u_0^{(i)}$ vanish. From this condition, $\beta_i^{(m)}$ are solved

$$\beta_i^{(m)} = \frac{\alpha_i^{(m)}}{\rho(\omega_{0m}^2 - \omega_{0i}^2)} \quad for\ i = 1,2,3,\cdots;\ and\ i \neq m \tag{41}$$

In the case of i=m, Equation (40) identically holds due to solvability condition (28). It can be readily shown that (41) can be identically obtained if the above procedures are repeated on $(39)_2$.

Special attention must be paid to the determination of $\beta_m^{(m)}$ since when i=m (41) becomes indefinite. To determine $\beta_m^{(m)}$, we resort to normalization condition (35). Using (8), we have for $b(x_1)$, $u_1^{(m)}$ and $\psi_1^{(m)}$

$$\begin{aligned} b(x_1) &= b_0 + \varepsilon f(x_1) \\ u^{(m)} &= u_0^{(m)} + \varepsilon u_1^{(m)} \\ \psi^{(m)} &= \psi_0^{(m)} + \varepsilon \psi_1^{(m)} \end{aligned} \tag{42}$$

We substitute from (42) into (35), letting i=j=m, and obtain,

$$\int_{-a}^{a} [(b_0 + \varepsilon f)(u_0^{(m)} + \varepsilon u_1^{(m)})^2 + \frac{(b_0 + \varepsilon f)^3}{3}(\psi_0^{(m)} + \varepsilon \psi_1^{(m)})^2]\, dx_1 = 1 \tag{43}$$

By a process similar to that used in obtaining equations for the first order problem, i.e. expanding (43) and equating the coefficients of ε^1, we have

$$2\varepsilon \int_{-a}^{a} [b_0 u_0^{(m)} u_1^{(m)} + \frac{b_0^3}{3}\psi_0^{(m)}\psi_1^{(m)} + \frac{1}{2} f b_0 (u_0^{(m)})^2 + \frac{1}{2} b_0^3 f (\psi_0^{(m)})^2]\, dx_1 = 0 \tag{44}$$

or

$$\int_{-a}^{a} [b_0 u_0^{(m)} u_1^{(m)} + \frac{b_0^3}{3}\psi_0^{(m)}\psi_1^{(m)}]\, dx_1 = -\frac{1}{2}\int_{-a}^{a} f b_0 [(u_0^{(m)})^2 + b_0^2 (\psi_0^{(m)})^2]\, dx_1$$

Substitution of (38) into the left side of (44) results in

$$\begin{aligned} &\int_{-a}^{a} \left[b_0 u_0^{(m)} \sum_i^{\infty} \beta_i^{(m)} u_0^{(i)} + \frac{b_0^3}{3}\psi_0^{(m)} \sum_i^{\infty} \beta_i^{(m)} \psi_0^{(i)} \right] dx_1 \\ &= \sum_i^{\infty} \beta_i^{(m)} \int_{-a}^{a} (b_0 u_0^{(m)} u_0^{(i)} + \frac{b_0^3}{3}\psi_0^{(m)}\psi_0^{(i)})\, dx_1 \\ &= \sum_i^{\infty} \beta_i^{(m)} \delta_{im} \\ &= \beta_m^{(m)} \end{aligned} \tag{45}$$

When (45) is inserted back to (44), we obtain the expression of $\beta_m^{(m)}$

$$\beta_m^{(m)} = -\frac{1}{2}\int_{-a}^{a} b_0 f [(u_0^{(m)})^2 + f b_0^2 (\psi_0^{(m)})^2]\, dx_1 \tag{46}$$

IV.3 Summary of Results

We write here the explicit expressions of ω, u, and ψ in form of their expansion of first order

$$\begin{aligned} \omega_m^2 &= \omega_{0m}^2 + \varepsilon \omega_{1m} \\ &= \omega_0^2 - \varepsilon \frac{A^{(m)}}{B^{(m)}} \end{aligned} \tag{47}$$

where $A^{(m)}$ and $B^{(m)}$ are given in (29);

$$u^{(m)} = u_0^{(m)} + \varepsilon u_1^{(m)}$$

$$= u_0^{(m)} + \varepsilon \frac{b_0}{2}\left[-\int_{-a}^{a} f\left(\left(u_0^{(m)}\right)^2 + b_0^2 \left(\psi_0^{(m)}\right)^2 \right) dx_1 \right] u_0^{(m)}$$

$$+\varepsilon \sum_{\substack{i \\ i \neq m}}^{\infty} \frac{\alpha_i^{(m)}}{\rho(\omega_{0m}^2 - \omega_{0i}^2)} u_0^{(i)}$$

(48)

$$\psi^{(m)} = \psi_0^{(m)} + \varepsilon \psi_1^{(m)}$$

$$= \psi_0^{(m)} + \varepsilon \left[-\frac{b_0}{2} \int_{-a}^{a} f\left(\left(u_0^{(m)}\right)^2 + b_0^2 \left(\psi_0^{(m)}\right)^2 \right) dx_1 \right] \psi_0^{(m)}$$

$$+\varepsilon \sum_{\substack{i \\ i \neq m}}^{\infty} \frac{\alpha_i}{\rho(\omega_{0m}^2 - \omega_{0i}^2)} \psi_0^{(i)}$$

(49)

m=1,2,3,...

V. Frequency Behavior of Contoured X_1-Length Strips

Frequency behavior of contoured, x_1-length strips is studied using the presented method. To serve as a numerical example while not losing generality, a cosine type of contouring is imposed on the strips' surface, i.e., $f(x_1, x_3) = f(x_1) = \cos(\frac{\pi x_1}{2a})$ in (7). Frequecny spectra are constructed using (47) for cases of ε = 0, 0.01, 0.05, 0.1, respectively. The calculated spectra are given in Fig.3 - 7 in the form of normalized frequency Ω vs a/b_0, where: $\Omega = \omega/\omega_0$, $\omega_0 = \frac{\pi}{2b_0}\sqrt{\frac{c_{66}}{\rho}}$; a, b_0 = width and edge thickness of the strip. Among these frequnecy spectra, ε=0 corresponds to the uniform thickness case and was obtained by Mindlin [6]. The zero-th order solutions, $u_0^{(m)}$, $\psi_0^{(m)}$ and ω_{0m}, used in evaluating ω_{0m} in (29), are also from the same work. The specific forms of these solutions are given in the appendix.

The following observations can be made on the calculated spectra:

(1) Small surface contouring retains the flat segments in the spectra which represent the fundamental TSH mode and its anharmonic overtones.

(2) The transition regions between flexure branches and TSH branches are significantly distorted by the presence of the surface contouring. Mindlin [6] noted in his work on wedged quartz plates that "while the thickness-shear motion is concentrated near the center of the wedge, the flexural motion is concentrated near the edges". Since the calculated frequency in these examples are normalized by the frequency of uniform plate with the thickness equal to the edge thickness of its contoured counterpart, we may imagine that the contouring "grows" from a constant thickness plate with the edge thickness unchanged. According to Mindlin, the TSH frequency will decrease at a much higher rate than the flexure frequency does. This may qualitatively explain the tips observed in those transition regions.

(3) The distance between overtones of TSH increases noticeably with amount of surface contouring. The similar phenomenon was observed by Bleusten [1] for thickness-twist modes of wedged x_3-strip.

Appendix Zero-th Order Soltuions Of $u(X_1,X_3)$ and $\psi(X_1,X_3)$

From [6], zero-th order solutions of $u(x_1,x_3)$ and $\psi(x_1,x_3)$ of the m-th resonance, $u_0^{(m)}$, $\psi_0^{(m)}$, are of the form

$$u_0^{(m)} = b_0 [\alpha_1^{(m)} \beta^{(m)} \sin(\bar{\xi}_1^{(m)} \frac{x_1}{b_0}) + \alpha_2^{(m)} \sin(\bar{\xi}_2^{(m)} \frac{x_1}{b_0})]$$

$$\psi_0^{(m)} = \beta^{(m)} \cos(\bar{\xi}_1^{(m)} \frac{x_1}{b_0}) + \cos(\bar{\xi}_2^{(m)} \frac{x_1}{b_0})]$$

where

$$\alpha_i^{(m)} = \frac{\bar{\xi}_i^{(m)}}{(\bar{\xi}_i^{(m)})^2 - 3\Omega_{(m)}^2} \qquad i = 1,2$$

$$\beta^{(m)} = -\frac{\bar{\xi}_2^{(m)} \sin(\bar{\xi}_2^{(m)} a)}{\bar{\xi}_1^{(m)} \sin(\bar{\xi}_1^{(m)} a)}$$

$\bar{\xi}_1^{(m)}$ and $\bar{\xi}_2^{(m)}$ = normalized wave numbers along x_1 axis, and

$\Omega_{(m)}$ = normalized resonance frequency of the m-th mode.

REFERENCES

(1) Bleustein, J. L.(1966), "Thickness-Twist and Face-Shear Vibrations of a Contoured Crystal Plate", Int. J. Solids Structures, Vol.2, P.351-360

(2) Dworsky, L.N. (1983) "Properties of AT Quartz Resonators on Wedgy Plates", IEEE

(3) Lee, P.C.Y. and Chen, S.(1969), "Vibrations of Contoured and Partially Plated, Contoured, Rectangular, AT-Cut Quartz Plates", J. Acoust. Amer., Vol.46, No.5, P.1193-1202

(4) Lee, P.C.Y. (1986), "Frequency-Temperature Behavior of Thickness Vibrations of Doubly Rotated Quartz Plates Affected by Plate Dimensions and Orientations", J. Appl. Phys., Vol.60, No.7, P.2327-2342

(5) Mindlin, R. D. & Forray, M.(1954), "Thickness-Shear and Flexural Vibrations of Contoured Crystal Plates", J. Appl. Phys., Vol.25, No.1, P.12-20, 1954

(6) Mindlin, R. D. & Spencer, W. J.(1967), "Anharmonic, Thickness-Twist Overtones of Thickness-Shear and Flexural Vibrations of Rectangular, AT-Cut Quartz Plates", J. Acous. Soc. America, Vol. 42, No.6, P.1268-1277

(7) Nayfeh, A. L.(1981), "Introduction to Perturbation Technique", Wiley, New York

(8) Pearman, G. T.(1968), "Thickness-Twist Vibrations in Beveled AT-Cut Quartz Plates", J. Acous. Soc. America, Vol. 45, No.4, P.928-934

(9) Tiersten, H.F. & Smyth, R.C. (1979) "An Analysis of Contoured Crystal Resonators operating in Overtones of Coupled Thickness Shear and Thickness Twist", J. Acoust. Amer., Vol.65, No.6, P.1455-1460

(10) Yang, J.S. & Xie, Z (1984) "Perturbation Method in the Problem of Large Deflections of Circular Plates with Nonuniform Thickness", Appl. Math. Mech.(China), Vol.5, No.2, P.1238-1242

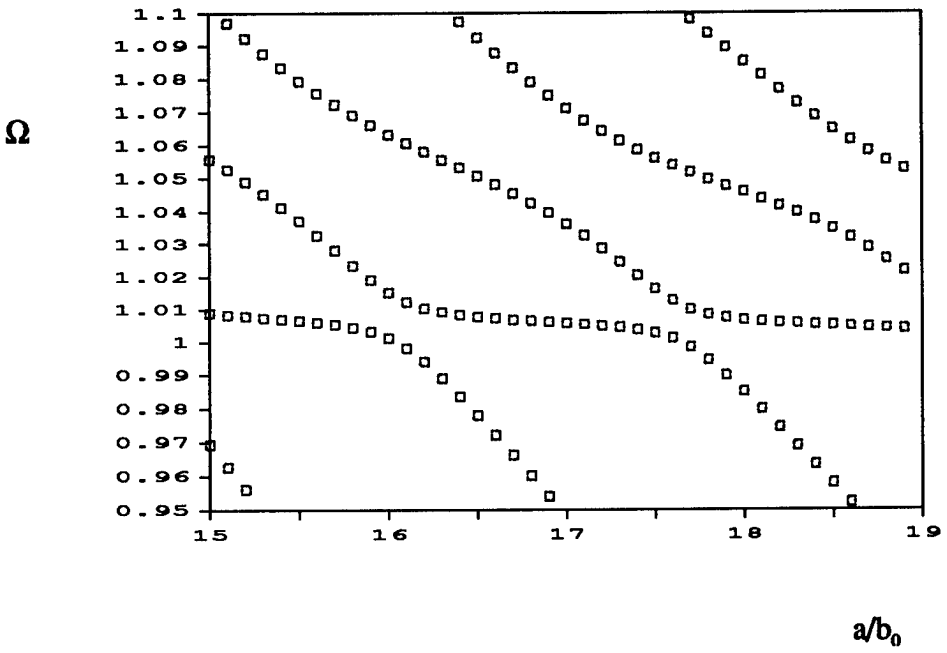


Fig. 3 Normalized Frequency (Ω) vs Length/edge-thickness a/b_0: Case $\varepsilon = 0$

(Obtained by Mindlin, 1967 [6])

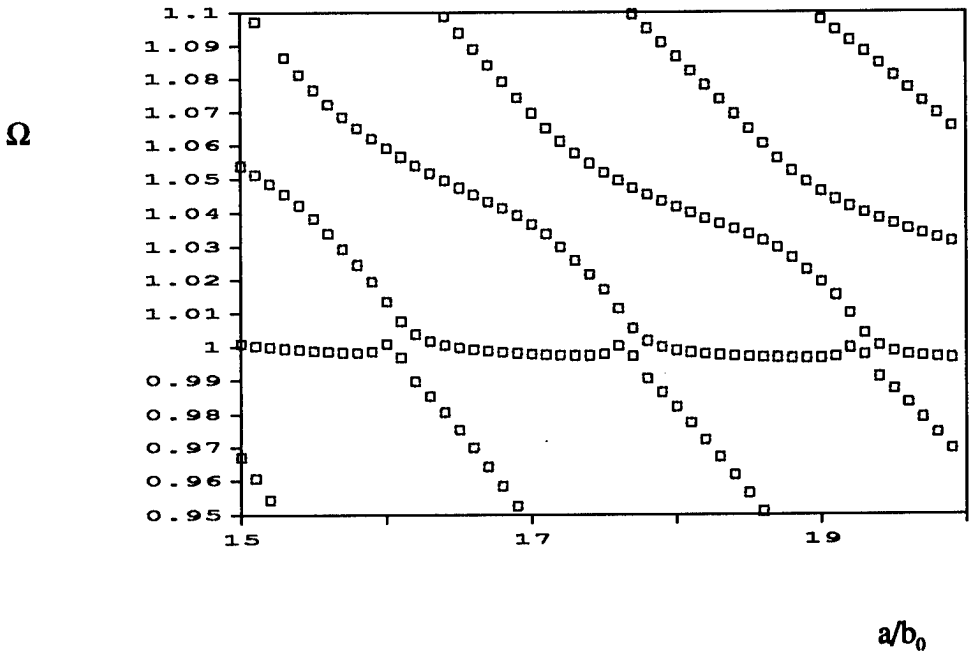


Fig. 4 Normalized Frequency (Ω) vs Length/edge-thickness a/b_0: Case $\varepsilon = 0.01$

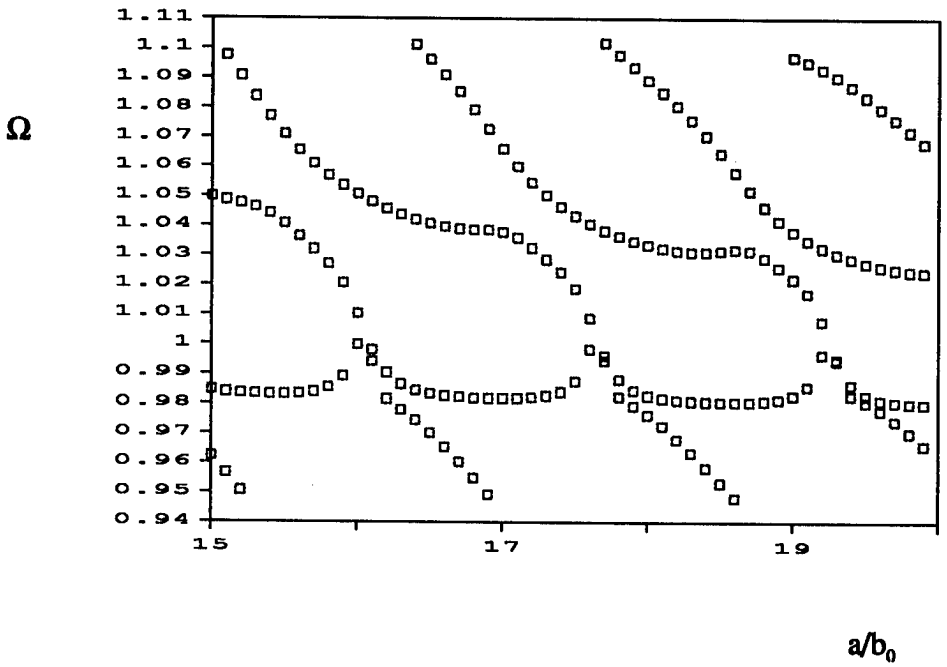


Fig. 5 Normalized Frequency (Ω) vs Length/edge-thickness a/b_0: Case $\varepsilon = 0.03$

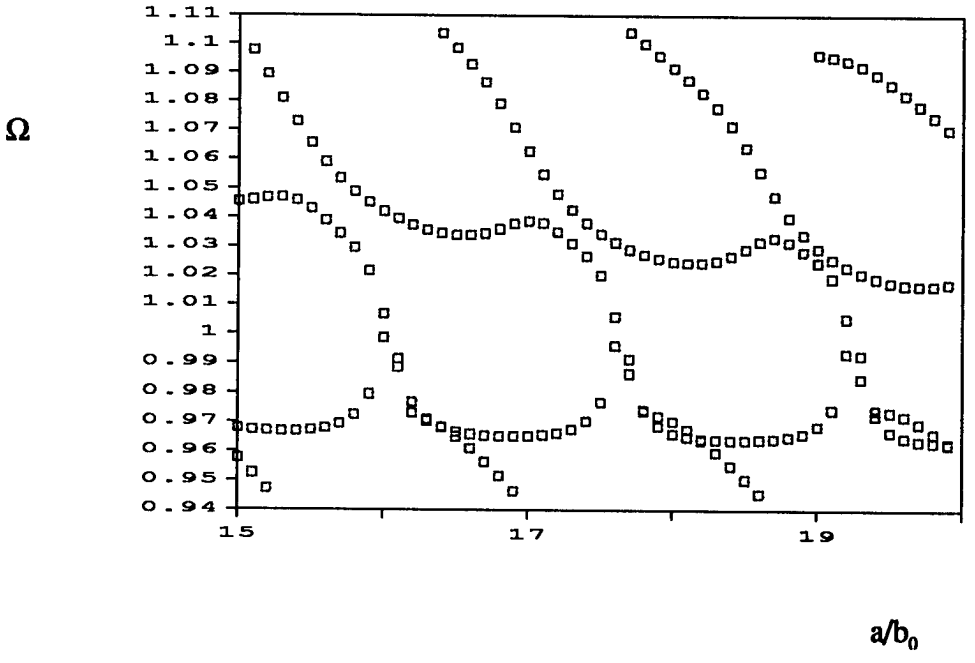


Fig. 6 Normalized Frequency (Ω) vs Length/edge-thickness a/b_0: Case $\varepsilon = 0.05$

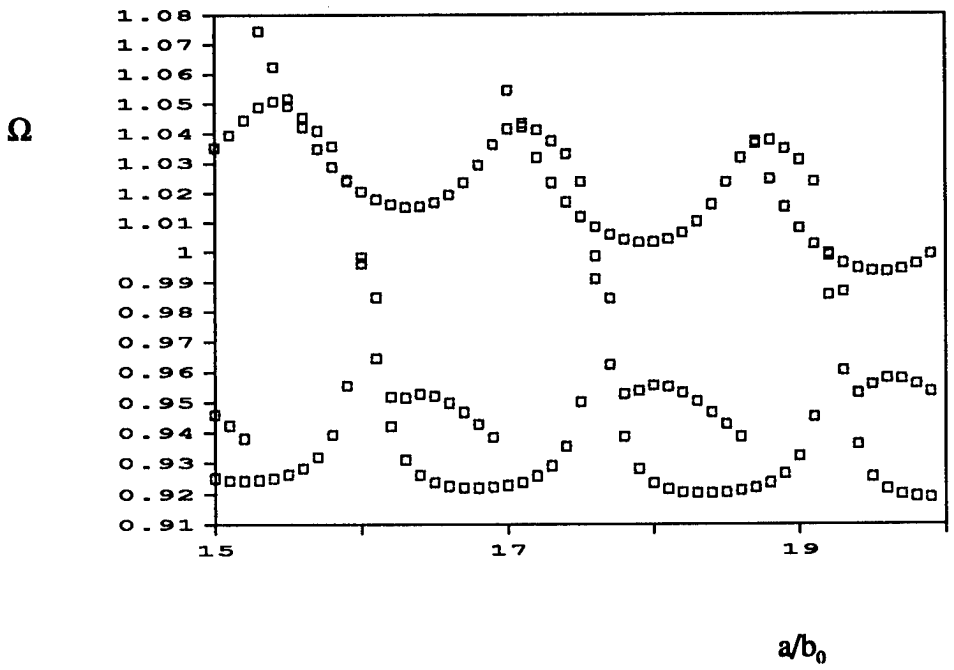


Fig.7 Normalized Frequency (Ω) vs Length/edge-thickness a/b_0: Case $\varepsilon = 0.10$

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Effect of Stress on Guided EM Waves in Anisotropic Dielectric Plates

P.C.Y. Lee, J.D. Yu*, and A. Ballato***

**Department of Civil Engineering & Operation Research*
Princeton University, Princeton, N.J. 08544

***U.S. Army Research Laboratory*
Ft. Monmouth, N.J. 07703-5601

Abstract

The changes in the frequencies or velocities of guided EM waves in infinite and anisotropic dielectric plates affected by the uniform and thickness-dependent stress fields are studied.

For the uniformly applied stresses, the changes of frequencies as a function of wave number are computed based on the recently obtained exact solutions of the 3-D Maxwell's equations.

For the thickness-dependent stress fields as those caused by the steady accelerations in plates attached to a rigid base, frequency changes are computed based on the variational techniques and the Rayleigh–Ritz method.

For numerical computations, a dielectric plate of lithium niobate with thickness $2b = 3.27mm$ is considered. The relative frequency shifts as a function of normalized wave number are computed for the TE and TM waves of various orders.

It's found that the maximum frequency shift is about $(1 \sim 2) \times 10^{-5}$ for the uniformly applied stress of $1MPa$ (or $10^6 N/m^2$). For steady acceleration on plates attached to a rigid base, the maximum frequency shift is about $(3 \sim 15) \times 10^{-10}/g$ for $2b = 3.27mm$. In the present supporting condition, the acceleration sensitivity is linearly proportional to the thickness of the plate.

1. Introduction

When an elastic dielectric waveguide is subjected to steady forces or accelerations, strains are produced. The strains induce changes in the relative dielectric impermeability tensor through the nonlinear piezooptical effect which, in turn, affects the velocity or frequency of the electromagnetic waves and hence causes frequency shift.

In the general formulation for elastic dielectrics, the mechanical fields of stress, strain, and displacements are governed by the equations of elasticity and the EM field by Maxwell's equations. These two sets of field equations are coupled by the constitutive equations through the linear piezoelectric effect and the nonlinear piezooptic effect[1,2]. We note that in the studies of *elastic vibrations in acoustic resonators*, the piezooptic effect has mostly been ignored in the equations of piezoelectricity (quasi-electrostatic approximation)[3,4] and even in the equations of piezoelectromagnetism[5,6]. However, in the present study of stress effect on the *electromagnetic vibrations in dielectric resonators*, the piezooptic effect is of fundamental importance and must be retained. On the other hand, the effect of EM field on the changes of stresses in the dielectric resonators due to the piezoelectric effect may be regarded unimportant and ignored.

Hence, the present foumulation reduces to a one-way coupled problem, i.e., the stress and strain fields are firstly solved from the uncoupled equations of elasticity and, then, the EM fields are obtained from Maxwell's equations in which the impermeability tensor is regarded dependent on strains of known functions of space.

Based on the exact solutions of Maxwell's equations for the guided EM waves in infinite and anisotropic dielectric plates[7], the effect of uniformly applied normal and shearing stresses of the magnitude of $1MPa$ on the frequency changes is studied as described in Sections 2 to 5.

When a dielectric plate is supported by a rigid base and subjected to steady accelerations, the acceleration induced stresses in the plate are thickness-dependent. The impermeability tensor of the dielectric plate also becomes thickness-dependent for which no exact solutions of guided waves are available. To compute ap-

0-7803-0905-7/93 $3.00 © 1993 IEEE

proximately the frequency changes due to thickness-dependent stresses, the variational techniques and the Rayleigh-Ritz method are employed as described in Sections 6 to 10. To check the accuracy, frequency changes under uniform stress field are computed by the variational method. The results are compared with the exact ones (Section 5) with very close agreement.

2. 3-D Equations and Continuity Conditions

Referring to a rectangular coordinate system x_i, we denote the middle plane of an infinite dielectric crystal plate by the x_1x_3 plane and the faces of the plate by $x_2 = \pm b$, where the thickness of the plate is $2b$ as shown in Fig.1.

For the study of the stress effect on the guided EM waves, the piezooptical effect is of fundamental importance. The effect of EM field on the changes of stresses due to piezoelectric effect may be regarded unimportant and ignored by setting the piezoelectric constants to zero. Hence the governing equations reduce to the following one-way coupled system:

For EM fields in the dielectric plate:

$$\begin{aligned} &\epsilon_{ijk}E_{k,j} = -B_{i,t} \ , \quad B_{i,i} = 0 \ , \\ &\epsilon_{ijk}H_{k,j} = D_{i,t} \ , \quad D_{i,i} = 0 \ , \\ &E_i = \epsilon_0^{-1}\beta_{ij}D_j \ , \quad B_i = \mu H_i . \end{aligned} \tag{1}$$

For piezooptical effect:

$$\begin{aligned} &\beta_{ij} = \beta_{ij}^0 + \Delta\beta_{ij} \ , \\ &\Delta\beta_{ij} = p_{ijkl}S_{kl} = p_{ijmn}(c_{mnkl})^{-1}T_{kl} \ . \end{aligned} \tag{2}$$

We note that (1) consists of the Maxwell's equations and constitutive equations, where E_i and H_i are the electric and magnetic field intensities, D_i and B_i the electric and magnetic flux densities, ϵ_{ijk} the unit alternating tensor, μ the magnetic permeability of the dielectric plate, ϵ_0 the dielectric permittivity of free space, and β_{ij} the relative dielectric impermeability tensor. In (2), β_{ij}^0 is the relative impermeability tensor when the dielectric is in a stress-free state, $\Delta\beta_{ij}$ the perturbation of β_{ij} caused by strains or stresses due to the piezooptical effect, p_{ijkl} the piezooptical tensor, and c_{ijkl} the elastic stiffness tensor.

For the surrounding vacuum, we designate $\bar{E}_i$, $\bar{D}_i$, $\bar{H}_i$, and $\bar{B}_i$ as the field vectors in the free space. The governing equations of EM field in free space may be obtained from (1) by letting $\beta_{ij} = \delta_{ij}$ and $\mu = \mu_0$, where μ_0 is the magnetic permeability of free space.

The continuity conditions to be satisfied at $x_2 = \pm b$, the interfaces between the plate and free space, are

$$\begin{aligned} &E_a(x_2 = \pm b) = \bar{E}_a(x_2 = \pm b) \ , \\ &B_2(x_2 = \pm b) = \bar{B}_2(x_2 = \pm b) \ , \\ &H_a(x_2 = \pm b) = \bar{H}_a(x_2 = \pm b) \ , \\ &D_2(x_2 = \pm b) = \bar{D}_2(x_2 = \pm b) \ . \end{aligned} \tag{3}$$

In the present paper, we consider first the effect of uniformly applied normal and shear stresses on the changes of frequencies of guided EM waves. It can be seen from (2) that, for uniform stress fields, the relative impermeability tensor β_{ij} is homogeneous throughout the dielectric crystal plate. In an earlier work[7], exact solutions for the guided EM waves propagating in an anisotropic but homogeneous plate were obtained from the three-dimensional Maxwell's equations. By using the exact solutions, at a given wave number, we can calculate f_0 and f, the frequencies of the EM waves for the plate without and with stresses, respectively, and, hence, obtain the frequency shift caused by the stress field.

By letting ω be the circular frequency and ξ the wave number, we summarize the exact solutions in the next section.

3. Exact Solutions[7]

Case I : β_{ij} is diagonal

$$[\beta_{ij}] = \begin{bmatrix} \beta_{11} & 0 & 0 \\ 0 & \beta_{22} & 0 \\ 0 & 0 & \beta_{33} \end{bmatrix} . \tag{4}$$

The propagating EM wave can be separated into TE wave and TM wave, and either of them can be separated further into symmetric and antisymmetric waves.

A. TE WAVES

The frequency equations are

$$\tan(\xi\lambda_1 b) = \pm\left(\hat{m}^2\frac{\bar{\lambda}}{\lambda_1}\right)^{\pm 1} , \tag{5}$$

where $+$ sign is for the symmetric E_3 modes and $-$ sign for the anti-symmetric E_3 modes and

$$\begin{aligned} &\lambda_1 = \sqrt{(V^2 - \beta_{33})/\beta_{33}} \ , \quad \bar{\lambda} = \sqrt{1 - v^2/c_0^2} \ , \\ &v = w/\xi \ , \quad V = \hat{m}v/c_0 \ , \quad \hat{m}^2 = \mu/\mu_0 \ . \end{aligned} \tag{6}$$

It can be shown that (5) has no roots for $V \le \beta_{33}^{1/2}$.

B. TM WAVES

The frequency equations are

$$\tan(\xi\lambda_1 b) = \pm\left(\frac{\bar{\lambda}}{\beta_{11}\lambda_1}\right)^{\pm 1} , \tag{7}$$

where + sign is for the symmetric H_3 modes and − sign for the anti-symmetric H_3 modes, and

$$\lambda_1 = \sqrt{(V^2 - \beta_{22})/\beta_{11}} \,. \tag{8}$$

Case II : $\beta_{11}, \beta_{22}, \beta_{33}, \beta_{12} \neq 0$

$$[\beta_{ij}] = \begin{bmatrix} \beta_{11} & \beta_{12} & 0 \\ \beta_{21} & \beta_{22} & 0 \\ 0 & 0 & \beta_{33} \end{bmatrix}. \tag{9}$$

A. TE WAVES

The solution and dispersion relations are identical to those given for the TE waves in Case I.

B. TM WAVES

In the present case, the frequency equation is not separable and it is given by

$$\begin{aligned} &[4\bar{\lambda}^2 - \beta_{11}^2(\lambda_1 - \lambda_2)^2] \sin[\xi b(\lambda_1 - \lambda_2)] \\ &\quad + 4\bar{\lambda}\beta_{11}(\lambda_1 - \lambda_2)\cos[\xi b(\lambda_1 - \lambda_2)] = 0 \,, \end{aligned} \tag{10}$$

where

$$\begin{aligned} \lambda_1 &= \frac{\beta_{12} + \sqrt{\beta_{12}^2 + \beta_{11}(V^2 - \beta_{22})}}{\beta_{11}}, \\ \lambda_2 &= \frac{\beta_{12} - \sqrt{\beta_{12}^2 + \beta_{11}(V^2 - \beta_{22})}}{\beta_{11}}. \end{aligned} \tag{11}$$

We note that (10) reduces to the two equations in (7) if $\beta_{12} = 0$.

Case III : $\beta_{11}, \beta_{22}, \beta_{33}, \beta_{23} \neq 0$

$$[\beta_{ij}] = \begin{bmatrix} \beta_{11} & 0 & 0 \\ 0 & \beta_{22} & \beta_{23} \\ 0 & \beta_{32} & \beta_{33} \end{bmatrix}. \tag{12}$$

In the present case, frequency equations still can be separated into symmetric (in E_3, H_3) and anti-symmetric (in E_3, H_3) cases. They are given below, for $1 > V > V_{b2}$

A. Symmetric case

$$\det \begin{bmatrix} 1 & 0 & -\beta_{11}\alpha_{11}\sin(\xi\lambda_1 b) \\ 0 & 1 & -V^2\frac{\alpha_{31}}{1+\lambda_1^2}\cos(\xi\lambda_1 b) \\ 0 & \hat{m}^2\bar{\lambda} & -V^2\frac{\alpha_{31}\lambda_1}{1+\lambda_1^2}\sin(\xi\lambda_1 b) \\ 1/\bar{\lambda} & 0 & \alpha_{21}\cos(\xi\lambda_1 b) \\ & & -\beta_{11}\alpha_{13}\sin(\xi\lambda_3 b) \\ & & -V^2\frac{\alpha_{33}}{1+\lambda_3^2}\cos(\xi\lambda_3 b) \\ & & -V^2\frac{\alpha_{33}\lambda_3}{1+\lambda_3^2}\sin(\xi\lambda_3 b) \\ & & \alpha_{23}\cos(\xi\lambda_3 b) \end{bmatrix} = 0, \tag{13}$$

B. Anti-symmetric case

$$\det \begin{bmatrix} 1 & 0 & -\beta_{11}\alpha_{11}\cos(\xi\lambda_1 b) \\ 0 & 1 & V^2\frac{\alpha_{31}}{1+\lambda_1^2}\sin(\xi\lambda_1 b) \\ 0 & \hat{m}^2\bar{\lambda} & -V^2\frac{\alpha_{31}\lambda_1}{1+\lambda_1^2}\cos(\xi\lambda_1 b) \\ -1/\bar{\lambda} & 0 & \alpha_{21}\sin(\xi\lambda_1 b) \\ & & -\beta_{11}\alpha_{13}\cos(\xi\lambda_3 b) \\ & & V^2\frac{\alpha_{33}}{1+\lambda_3^2}\sin(\xi\lambda_3 b) \\ & & -V^2\frac{\alpha_{33}\lambda_3}{1+\lambda_3^2}\cos(\xi\lambda_3 b) \\ & & \alpha_{23}\sin(\xi\lambda_3 b) \end{bmatrix} = 0, \tag{14}$$

where

$$\begin{aligned} \lambda_1 &= \left(\frac{G_1 + G_2}{2\beta_{11}\beta_{33}}\right)^{1/2}, \qquad \lambda_2 = -\lambda_1 \,, \\ \lambda_3 &= \left(\frac{G_1 - G_2}{2\beta_{11}\beta_{33}}\right)^{1/2}, \qquad \lambda_4 = -\lambda_3 \,, \end{aligned} \tag{15}$$

and

$$\begin{aligned} G_1 &= (\beta_{11} + \beta_{33})V^2 + \beta_{23}^2 - \beta_{22}\beta_{33} - \beta_{11}\beta_{33} \,, \\ G_2 &= \Big\{ \left[(\beta_{11} - \beta_{33})\,V^2 + \beta_{22}\beta_{33} - \beta_{23}^2 - \beta_{11}\beta_{33}\right]^2 \\ &\qquad + 4\beta_{11}\beta_{23}^2 V^2 \Big\}^{1/2}, \end{aligned} \tag{16}$$

and α_{ij}, $j = 1, 2, 3, 4$ must satisfy

$$\begin{bmatrix} -\beta_{11}\lambda_j^2 + V^2 & \beta_{22}\lambda_j & \beta_{23}\lambda_j \\ \beta_{11}\lambda_j & -\beta_{22} + V^2 & -\beta_{23} \\ 0 & \beta_{32}(1+\lambda_j^2) & \beta_{33}(1+\lambda_j^2) - V^2 \end{bmatrix} \begin{Bmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \alpha_{3j} \end{Bmatrix} = 0. \tag{17}$$

Frequency equations at $V = V_{b2}$ and for $V_{b2} > V > V_{b1}$ can similarly be obtained, but they will not be given here.

Case IV : All $\beta_{ij} \neq 0$

The frequency equation for the guided EM waves is, for $1 > V > V_{b2}$

$$\det \begin{bmatrix} Q_{11}\mathcal{C}_1 - \bar{\lambda}\alpha_{21}\mathcal{S}_1 & Q_{12}\mathcal{C}_2 - \bar{\lambda}\alpha_{22}\mathcal{S}_2 \\ Q_{11}\mathcal{S}_1 + \bar{\lambda}\alpha_{21}\mathcal{C}_1 & Q_{12}\mathcal{S}_2 + \bar{\lambda}\alpha_{22}\mathcal{C}_2 \\ Q_{31}(\bar{\lambda}\mathcal{C}_1 - \lambda_1\mathcal{S}_1) & Q_{32}(\bar{\lambda}\mathcal{C}_2 - \lambda_2\mathcal{S}_2) \\ Q_{31}(\bar{\lambda}\mathcal{S}_1 + \lambda_1\mathcal{C}_1) & Q_{32}(\bar{\lambda}\mathcal{S}_2 + \lambda_2\mathcal{C}_2) \\ Q_{13}\mathcal{C}_3 - \bar{\lambda}\alpha_{23}\mathcal{S}_3 & Q_{14}\mathcal{C}_4 - \bar{\lambda}\alpha_{24}\mathcal{S}_4 \\ Q_{13}\mathcal{S}_3 + \bar{\lambda}\alpha_{23}\mathcal{C}_3 & Q_{14}\mathcal{S}_4 + \bar{\lambda}\alpha_{24}\mathcal{C}_4 \\ Q_{33}(\bar{\lambda}\mathcal{C}_3 - \lambda_3\mathcal{S}_3) & Q_{34}(\bar{\lambda}\mathcal{C}_4 - \lambda_4\mathcal{S}_4) \\ Q_{33}(\bar{\lambda}\mathcal{S}_3 + \lambda_3\mathcal{C}_3) & Q_{34}(\bar{\lambda}\mathcal{S}_4 + \lambda_4\mathcal{C}_4) \end{bmatrix} = 0, \tag{18}$$

where

$$\begin{aligned} &Q_{kj} = \textstyle\sum_{l=1}^{3} \beta_{kl}\alpha_{lj} \,, \quad \mathcal{C}_j = \cos(\xi\lambda_j b) \,, \\ &\mathcal{S}_j = \sin(\xi\lambda_j b) \,, \quad k = 1, 3 \,, \quad j = 1, 2, 3, 4 \,. \end{aligned} \tag{19}$$

and α_{ij}, $j = 1, 2, 3, 4$ must satisfy

$$\begin{bmatrix} \beta_{21}\lambda_j - \beta_{11}\lambda_j^2 + V^2 & \beta_{22}\lambda_j - \beta_{12}\lambda_j^2 & \beta_{23}\lambda_j - \beta_{13}\lambda_j^2 \\ \beta_{11}\lambda_j - \beta_{21} & \beta_{12}\lambda_j - \beta_{22} + V^2 & \beta_{13}\lambda_j - \beta_{23} \\ \beta_{31}(1+\lambda_j^2) & \beta_{32}(1+\lambda_j^2) & \beta_{33}(1+\lambda_j^2) - V^2 \end{bmatrix} \begin{Bmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \alpha_{3j} \end{Bmatrix} = 0 \ . \tag{20}$$

Frequency equations at $V = V_{b2}$ and for $V_{b2} > V > V_{b1}$ are also obtained, but they will not be given here.

4. Material Constants

For numerical computations, the material constants of lithium niobate listed in Landolt-Börnstein[8] are used. Let x_1, x_2, x_3 be the principal axes of the crystal and x_1 be the diagonal axis, the stiffness tensor, relative impermeability tensor, and piezooptical tensor are respectively,

$$[c_{rs}] = \begin{bmatrix} 203 & 53 & 75 & 9 & 0 & 0 \\ 53 & 203 & 75 & -9 & 0 & 0 \\ 75 & 75 & 245 & 0 & 0 & 0 \\ 9 & -9 & 0 & 60 & 0 & 0 \\ 0 & 0 & 0 & 0 & 60 & 18 \\ 0 & 0 & 0 & 0 & 18 & 75 \end{bmatrix} \times 10^9 \ N/m^2 \ , \tag{21}$$

$$[\beta_{ij}^0] = \begin{bmatrix} .0227273 & 0 & 0 \\ 0 & .0227273 & 0 \\ 0 & 0 & .0344828 \end{bmatrix} \ , \tag{22}$$

$$[p_{rs}] = \begin{bmatrix} -.026 & .09 & .13 & -.075 & 0 & 0 \\ .09 & -.026 & .13 & .075 & 0 & 0 \\ .18 & .18 & .07 & 0 & 0 & 0 \\ -.15 & .15 & 0 & .146 & 0 & 0 \\ 0 & 0 & 0 & 0 & .146 & -.15 \\ 0 & 0 & 0 & 0 & -.3 & -.058 \end{bmatrix} . \tag{23}$$

In (21) and (23), the contracted forms are used. With these material constants, we have

$$[p_{rq}(c_{qs})^{-1}] = \begin{bmatrix} -.3545 & .2803 & .5533 \\ .2803 & -.3545 & .5533 \\ .7548 & .7548 & -.1764 \\ -1.167 & 1.167 & -6. \times 10^{-17} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1.1548 & 0 & 0 \\ 1.1548 & 0 & 0 \\ -3. \times 10^{-18} & 0 & 0 \\ 2.7834 & 0 & 0 \\ 0 & 3.2687 & -2.7845 \\ 0 & -5.1379 & 0.4598 \end{bmatrix} \times 10^{-12} m^2/N \ . \tag{24}$$

Since β_{ij}^0 is diagonal, the propagating EM wave in the stress-free infinite dielectric plate can be seperated into TE and TM waves, and either of them can be separated further into symmetric and antisymmetric waves. For propagating EM wave in uniformly stressed dielectric plate, the coupling of TE and TM waves depends on the nonzero component of the applied stress. Four cases are studied in the next section.

5. Stress Sensitivity

With the dimensionless wave number defined by $X = \xi/(\frac{\pi}{2b})$, the stress sensitivity of guided EM waves is shown in this section in terms of the relative frequency shift $(f - f^0)/f^0$ vs. X relations, where f_0 and f are the frequencies of the EM waves in the unstressed and the stressed dielectric plates, respectively.

Case 1 : $T_{11} = -1. \times 10^6 \ N/m^2$

From (2) and (24), the perturbation in the relative impermeability tensor caused by the given normal stress is

$$[\Delta\beta_{ij}] = \begin{bmatrix} 3.54489 & 0 & 0 \\ 0 & -2.80271 & 11.67006 \\ 0 & 11.67006 & -7.54809 \end{bmatrix} \times 10^{-7}. \tag{25}$$

The resulting β_{ij} has the form as that given in (12), hence, the EM waves in the stressed plate can still be separated into symmetric and antisymmetric waves.

The relative frequency shift is plotted as a function of the dimensionless wave number in Fig.2, where $n = 1, 2$ denote the lowest and second lowest branches of the TE (or TM) waves. In fact, the lowest branch of TE (or TM) is a symmetric wave and the second lowest branch is an antisymmetric wave.

We can see that the relative frequency shift caused by the uniform normal stress $T_{11} = -1. \times 10^6 N/m^2$ is in the order of 10^{-6}.

Case 2 : $T_{22} = -1. \times 10^6 \ N/m^2$

The change in the relative impermeability tensor is

$$[\Delta\beta_{ij}] = \begin{bmatrix} -2.80271 & 0 & 0 \\ 0 & 3.54489 & -11.67006 \\ 0 & -11.67006 & -7.54809 \end{bmatrix} \times 10^{-7}. \tag{26}$$

The resulting β_{ij} has the same form as the previous case. The relative frequency shift is plotted in Fig.3.

Comparing Fig.2 and Fig.3, we find that the relative frequency shifts of TM waves in the two cases have very close magnitude but different sign, while the relative frequency shifts of TE waves have almost the same trend and magnitude. This is because, in (26) and (25), $\Delta\beta_{11}, \Delta\beta_{22}$ have opposite signs and $\Delta\beta_{33}$

have the same signs and magnitudes. $\Delta\beta_{11}$ and $\Delta\beta_{22}$ have the predominant effect on the frequency shifts of TM waves while $\Delta\beta_{33}$ has predominant effect on the frequency changes of TE waves.

Case 3 : $T_{33} = -1. \times 10^6\ N/m^2$

The change in the relative impermeability tensor is

$$[\Delta\beta_{ij}] = \begin{bmatrix} -5.53332 & 0 & 0 \\ 0 & -5.53332 & 6. \times 10^{-16} \\ 0 & 6. \times 10^{-16} & 1.76413 \end{bmatrix} \times 10^{-7}. \tag{27}$$

The resulting β_{ij} has the same form as that for the previous two cases. The relative frequency shift is plotted in Fig.4.

Case 4 : $T_{23} = -1. \times 10^6\ N/m^2$

The perturbation in the relative impermeability tensor caused by the uniform shear stress is

$$[\Delta\beta_{ij}] = \begin{bmatrix} 11.54786 & 0 & 0 \\ 0 & -11.54786 & -27.83435 \\ 0 & -27.83435 & -3. \times 10^{-17} \end{bmatrix} \times 10^{-7}. \tag{28}$$

We note that the values of $\Delta\beta_{11}, \Delta\beta_{22}$ in (28) are much larger than those in (25), (26), and (27). The $(f - f_0)/f_0$ vs. X curves are plotted in Fig.5.

From Fig.5, we see that, owing to the larger values of $\Delta\beta_{11}, \Delta\beta_{22}$, the frequency shifts of TM waves are larger than those in the Cases 1, 2, and 3. The frequency shifts for TE waves are extremely small since $\Delta\beta_{33}$ is almost zero.

6. Variational Principle for Frequency

Since no exact solutions are available for guided EM waves in anisotropic dielectric plates with thickness-dependent impermeability tensors, the variational technique and Rayleigh–Ritz method are employed to study the effect of thickness-dependent stresses on frequency changes. These methods of approximtion had been employed to calculate the resonance frequencies of a cavity filled with inhomogeneous or anisotropic substances by Berk[9] and to analyse the normal modes in anisotropic thin-film waveguides by Yamamotto et al[10].

In this section, a variational principle is given for an anisotropic dielectric surrounded by the free space. The accompanying orthogonality relation, which can much simplify the final formulation of Rayleigh–Ritz approximation, will be derived in the next section.

Let the electric and magnetic field intensities E_i, H_i be time harmonic, i.e.

$$\begin{aligned} E_{i,t} &= -i\omega E_i\ , \\ H_{i,t} &= -i\omega H_i\ . \end{aligned} \tag{29}$$

For a dielectric crystal of volume V bounded by S which separates V from the outer vacuum $\bar{V}$, as shown in Fig.7, the Maxwell's equations can be written as

$$\begin{aligned} \epsilon_{ijk}E_{k,j} &= i\omega\mu H_i\ , \\ \epsilon_{ijk}H_{k,j} &= -i\omega\epsilon_{ij}E_j\ , \end{aligned} \quad \text{in } V$$
$$\begin{aligned} \epsilon_{ijk}\bar{E}_{k,j} &= i\omega\mu_0\bar{H}_i\ , \\ \epsilon_{ijk}\bar{H}_{k,j} &= -i\omega\epsilon_0\bar{E}_i\ , \end{aligned} \quad \text{in } \bar{V} \tag{30}$$

and the continuity conditions on S are

$$\begin{aligned} \epsilon_{ijk}n_j(H_k - \bar{H}_k) &= 0\ , \\ \epsilon_{ijk}n_j(E_k - \bar{E}_k) &= 0\ . \end{aligned} \tag{31}$$

It's noticeable that the divergence equations are omitted in (30) because they are implied by these curl equations.

For solutions which are dimishing from the dielectric, we propose the following variational principle

$$\begin{aligned} \omega = i\Big\{ &\int_V (E_i^*\epsilon_{ijk}H_{k,j} - H_i^*\epsilon_{ijk}E_{k,j})\,dv \\ &+ \int_{\bar{V}} (\bar{E}_i^*\epsilon_{ijk}\bar{H}_{k,j} - \bar{H}_i^*\epsilon_{ijk}\bar{E}_{k,j})\,dv \Big\} \\ \times \Big\{ &\int_V (E_i^*\epsilon_{ij}E_j + H_i^*\mu H_i)\,dv \\ &+ \int_{\bar{V}} (\bar{E}_i^*\epsilon_0\bar{E}_i + \bar{H}_i^*\mu_0\bar{H}_i)\,dv \Big\}^{-1}\ . \end{aligned} \tag{32}$$

To prove that the stationarity of the above equation gives the two Maxwell's equations and appropriate boundary conditions, we take the variation of the equation

$$\begin{aligned} &\delta\omega \Big\{ \int_V (E_i^*\epsilon_{ij}E_j + H_i^*\mu H_i)\,dv \\ &\quad + \int_{\bar{V}} (\bar{E}_i^*\epsilon_0\bar{E}_i + \bar{H}_i^*\mu_0\bar{H}_i)\,dv \Big\} \\ &= i\int_V (\delta E_i^*\epsilon_{ijk}H_{k,j} + E_i^*\epsilon_{ijk}\delta H_{k,j} \\ &\qquad -\delta H_i^*\epsilon_{ijk}E_{k,j} - H_i^*\epsilon_{ijk}\delta E_{k,j})\,dv \\ &\quad -\omega\int_V (\delta E_i^*\epsilon_{ij}E_j + E_i^*\epsilon_{ij}\delta E_j \\ &\qquad +\delta H_i^*\mu H_i + H_i^*\mu\delta H_i)\,dv \\ &\quad +i\int_{\bar{V}} (\delta\bar{E}_i^*\epsilon_{ijk}\bar{H}_{k,j} + \bar{E}_i^*\epsilon_{ijk}\delta\bar{H}_{k,j} \\ &\qquad -\delta\bar{H}_i^*\epsilon_{ijk}\bar{E}_{k,j} - \bar{H}_i^*\epsilon_{ijk}\delta\bar{E}_{k,j})\,dv \\ &\quad -\omega\int_{\bar{V}} (\delta\bar{E}_i^*\epsilon_0\bar{E}_i + \bar{E}_i^*\epsilon_0\delta\bar{E}_i \\ &\qquad +\delta\bar{H}_i^*\mu_0\bar{H}_i + \bar{H}_i^*\mu_0\delta\bar{H}_i)\,dv\ . \end{aligned} \tag{33}$$

With the following relations

$$\begin{aligned}
\int_V E_i^* \epsilon_{ijk} \delta H_{k,j} dv &= -\int_S \delta H_i \epsilon_{ijk} n_j E_k^* \, ds \\
&\quad + \int_V \delta H_i \epsilon_{ijk} E_{k,j}^* dv \, , \\
\int_V H_i^* \epsilon_{ijk} \delta E_{k,j} dv &= -\int_S \delta E_i \epsilon_{ijk} n_j H_k^* \, ds \\
&\quad + \int_V \delta E_i \epsilon_{ijk} H_{k,j}^* dv \, , \\
\int_{\bar{V}} \bar{E}_i^* \epsilon_{ijk} \delta \bar{H}_{k,j} dv &= -\int_{S+S_\infty} \delta \bar{H}_i \epsilon_{ijk} \bar{n}_j \bar{E}_k^* ds \\
&\quad + \int_{\bar{V}} \delta \bar{H}_i \epsilon_{ijk} \bar{E}_{k,j}^* dv \, , \\
\int_{\bar{V}} \bar{H}_i^* \epsilon_{ijk} \delta \bar{E}_{k,j} dv &= -\int_{S+S_\infty} \delta \bar{E}_i \epsilon_{ijk} \bar{n}_j \bar{H}_k^* ds \\
&\quad + \int_{\bar{V}} \delta \bar{E}_i \epsilon_{ijk} \bar{H}_{k,j}^* dv \, ,
\end{aligned} \tag{34}$$

where $n_i, \bar{n}_i$ are the unit normals as shown in Fig.2, we have

$$\begin{aligned}
0 = i \int_V dv \, \{ & \delta E_i^* \left(\epsilon_{ijk} H_{k,j} + i\omega \epsilon_{ij} E_j \right) \\
& - \delta H_i^* \left(\epsilon_{ijk} E_{k,j} - i\omega \mu H_i \right) \\
& - \delta E_i \left(\epsilon_{ijk} H_{k,j}^* - i\omega \epsilon_{ij} E_j^* \right) \\
& + \delta H_i \left(\epsilon_{ijk} E_{k,j}^* + i\omega \mu H_i^* \right) \} \\
+ i \int_{\bar{V}} dv \, \{ & \delta \bar{E}_i^* \left(\epsilon_{ijk} \bar{H}_{k,j} + i\omega \epsilon_{ij} \bar{E}_j \right) \\
& - \delta \bar{H}_i^* \left(\epsilon_{ijk} \bar{E}_{k,j} - i\omega \mu \bar{H}_i \right) \\
& - \delta \bar{E}_i \left(\epsilon_{ijk} \bar{H}_{k,j}^* - i\omega \epsilon_{ij} \bar{E}_j^* \right) \\
& + \delta \bar{H}_i \left(\epsilon_{ijk} \bar{E}_{k,j}^* + i\omega \mu \bar{H}_i^* \right) \} \\
+ i \int_S ds \, \{ & \delta E_i \epsilon_{ijk} n_j (H_k^* - \bar{H}_k^*) \\
& - \delta H_i \epsilon_{ijk} n_j (E_k^* - \bar{E}_k^*) \} \\
+ i \int_{S_\infty} ds \, \{ & \delta \bar{H}_i \epsilon_{ijk} \bar{n}_j \bar{E}_k^* - \delta \bar{E}_i \epsilon_{ijk} \bar{n}_j \bar{H}_k^* \} \, .
\end{aligned} \tag{35}$$

Because $\delta \bar{E}_i, \delta \bar{H}_i$ are diminishing when far from the dielectric crystal, the last surface integral is zero when the radius of S_∞ approaches infinity. For $\delta E_i, \delta H_i$ are arbitrary, we obtain the Maxwell's equations (30) and the complex conjugate of the continuity conditions (31).

For guided EM waves propagating along the x_1 direction of an thickness inhomogeneous wave guide, the electric and magnetic field intensity vectors have the following form

$$\begin{aligned}
E_i &= \mathcal{E}_i(x_2) e^{i(\xi x_1 - \omega t)} \, , \\
H_i &= \mathcal{H}_i(x_2) e^{i(\xi x_1 - \omega t)} \, , \\
\bar{E}_i &= \bar{\mathcal{E}}_i(x_2) e^{i(\xi x_1 - \omega t)} \, , \\
\bar{H}_i &= \bar{\mathcal{H}}_i(x_2) e^{i(\xi x_1 - \omega t)} \, .
\end{aligned} \tag{36}$$

For such case, the volume integral in (32) should be replaced with a line integral in x_2 direction from $-\infty$ to ∞. The curls of these vectors are

$$\begin{aligned}
\epsilon_{ijk} E_{k,j} &= \left(\epsilon_{ijk} \mathcal{E}_{k,j} + i\xi \epsilon_{ijk} \delta_{j1} \mathcal{E}_k \right) e^{i(\xi x_1 - \omega t)} \, , \\
\epsilon_{ijk} H_{k,j} &= \left(\epsilon_{ijk} \mathcal{H}_{k,j} + i\xi \epsilon_{ijk} \delta_{j1} \mathcal{H}_k \right) e^{i(\xi x_1 - \omega t)} \, , \\
\epsilon_{ijk} \bar{E}_{k,j} &= \left(\epsilon_{ijk} \bar{\mathcal{E}}_{k,j} + i\xi \epsilon_{ijk} \delta_{j1} \bar{\mathcal{E}}_k \right) e^{i(\xi x_1 - \omega t)} \, , \\
\epsilon_{ijk} \bar{H}_{k,j} &= \left(\epsilon_{ijk} \bar{\mathcal{H}}_{k,j} + i\xi \epsilon_{ijk} \delta_{j1} \bar{\mathcal{H}}_k \right) e^{i(\xi x_1 - \omega t)} \, ,
\end{aligned} \tag{37}$$

form which the variation principle is reduced to be

$$\begin{aligned}
\omega = \Bigg\{ & \int_{-b}^{b} dx_2 \left(i\mathcal{E}_i^* \epsilon_{ijk} \mathcal{H}_{k,j} - \xi \mathcal{E}_i^* \epsilon_{ijk} \delta_{j1} \mathcal{H}_k \right. \\
& \left. - i\mathcal{H}_i^* \epsilon_{ijk} \mathcal{E}_{k,j} + \xi \mathcal{H}_i^* \epsilon_{ijk} \delta_{j1} \mathcal{E}_k \right) \\
& + \int_{-\infty}^{b} dx_2 \left(i\bar{\mathcal{E}}_i^* \epsilon_{ijk} \bar{\mathcal{H}}_{k,j} - \xi \bar{\mathcal{E}}_i^* \epsilon_{ijk} \delta_{j1} \bar{\mathcal{H}}_k \right. \\
& \left. - i\bar{\mathcal{H}}_i^* \epsilon_{ijk} \bar{\mathcal{E}}_{k,j} + \xi \bar{\mathcal{H}}_i^* \epsilon_{ijk} \delta_{j1} \bar{\mathcal{E}}_k \right) \\
& + \int_{b}^{\infty} dx_2 \left(i\bar{\mathcal{E}}_i^* \epsilon_{ijk} \bar{\mathcal{H}}_{k,j} - \xi \bar{\mathcal{E}}_i^* \epsilon_{ijk} \delta_{j1} \bar{\mathcal{H}}_k \right. \\
& \left. - i\bar{\mathcal{H}}_i^* \epsilon_{ijk} \bar{\mathcal{E}}_{k,j} + \xi \bar{\mathcal{H}}_i^* \epsilon_{ijk} \delta_{j1} \bar{\mathcal{E}}_k \right) \Bigg\} \\
\times \Bigg\{ & \int_{-b}^{b} dx_2 \left(\mathcal{E}_i^* \epsilon_{ij} \mathcal{E}_j + \mathcal{H}_i^* \mu \mathcal{H}_i \right) \\
& + \int_{-\infty}^{b} dx_2 \left(\bar{\mathcal{E}}_i^* \epsilon_0 \bar{\mathcal{E}}_i + \bar{\mathcal{H}}_i^* \mu_0 \bar{\mathcal{H}}_i \right) \\
& + \int_{b}^{\infty} dx_2 \left(\bar{\mathcal{E}}_i^* \epsilon_0 \bar{\mathcal{E}}_i + \bar{\mathcal{H}}_i^* \mu_0 \bar{\mathcal{H}}_i \right) \Bigg\}^{-1} \, .
\end{aligned} \tag{38}$$

It can be shown that the stationality of (38) gives the Maxwell's curl equations, and the continuity conditions at $x_2 = \pm b$.

7. Orthogonality of Eigenfunctions

Let $E_i^{(p)}, H_i^{(p)}, \bar{E}_i^{(p)}, \bar{H}_i^{(p)}$ and $E_i^{(q)}, H_i^{(q)}, \bar{E}_i^{(q)}, \bar{H}_i^{(q)}$ be the EM fields of the dielectric crystal and the surrounding vacuum, as shown in Fig.1, corresponding to frequencies ω_p and ω_q respectively, i.e.

$$\begin{aligned}
\epsilon_{ijk} E_{k,j}^{(p)} - i\omega_p \mu H_i^{(p)} &= 0 \, , \\
\epsilon_{ijk} H_{k,j}^{(p)} + i\omega_p \epsilon_{ij} E_j^{(p)} &= 0 \, , \\
\epsilon_{ijk} \bar{E}_{k,j}^{(p)} - i\omega_p \mu_0 \bar{H}_i^{(p)} &= 0 \, , \\
\epsilon_{ijk} \bar{H}_{k,j}^{(p)} + i\omega_p \epsilon_0 \bar{E}_i^{(p)} &= 0 \, ,
\end{aligned} \tag{39}$$

$$\begin{aligned}
\epsilon_{ijk} E_{k,j}^{(q)*} + i\omega_p \mu H_i^{(q)*} &= 0 \, , \\
\epsilon_{ijk} H_{k,j}^{(q)*} - i\omega_p \epsilon_{ij} E_j^{(q)*} &= 0 \, , \\
\epsilon_{ijk} \bar{E}_{k,j}^{(q)*} + i\omega_p \mu_0 \bar{H}_i^{(q)*} &= 0 \, , \\
\epsilon_{ijk} \bar{H}_{k,j}^{(q)*} - i\omega_p \epsilon_0 \bar{E}_i^{(q)*} &= 0 \, .
\end{aligned} \tag{40}$$

Multiplying $(39)_1$, $(39)_2$, $(39)_3$, $(39)_4$, $(40)_1$, $(40)_2$, $(40)_3$, and $(40)_4$ respectively with $H_i^{(q)*}$, $-E_i^{(q)*}$,

$\bar{H}_i^{(q)*}$, $-\bar{E}_i^{(q)*}$, $H_i^{(p)}$, and $-E_i^{(p)}$, $\bar{H}_i^{(p)}$, and $-\bar{E}_i^{(p)}$, summing them, and integrating , we obtain

$$\begin{aligned}
&\int_V dv \Big[H_i^{(q)*}\epsilon_{ijk}E_{k,j}^{(p)} - E_i^{(q)*}\epsilon_{ijk}H_{k,j}^{(p)} \\
&\qquad -i\omega_p(H_i^{(q)*}\mu H_i^{(p)} + E_i^{(q)*}\epsilon_{ij}E_j^{(p)})\Big] \\
&+\int_{\bar{V}} dv \Big[\bar{H}_i^{(q)*}\epsilon_{ijk}\bar{E}_{k,j}^{(p)} - \bar{E}_i^{(q)*}\epsilon_{ijk}\bar{H}_{k,j}^{(p)} \\
&\qquad -i\omega_p(\bar{H}_i^{(q)*}\mu_0\bar{H}_i^{(p)} + \bar{E}_i^{(q)*}\epsilon_0\bar{E}_i^{(p)})\Big] \\
&+\int_V dv \Big[H_i^{(p)}\epsilon_{ijk}E_{k,j}^{(q)*} - E_i^{(p)}\epsilon_{ijk}H_{k,j}^{(q)*} \\
&\qquad +i\omega_q(H_i^{(p)}\mu H_i^{(q)*} + E_i^{(p)}\epsilon_{ij}E_j^{(q)*})\Big] \\
&+\int_{\bar{V}} dv \Big[\bar{H}_i^{(p)}\epsilon_{ijk}\bar{E}_{k,j}^{(q)*} - \bar{E}_i^{(p)}\epsilon_{ijk}\bar{H}_{k,j}^{(q)*} \\
&\qquad +i\omega_q(\bar{H}_i^{(p)}\mu_0\bar{H}_i^{(q)*} + \bar{E}_i^{(p)}\epsilon_0\bar{E}_i^{(q)*})\Big] = 0 \ .
\end{aligned} \tag{41}$$

Using divergence theorem, continuity conditions, and the decaying condition, the above equation reduces to

$$\begin{aligned}
&i(\omega_p-\omega_q)\Big\{\int_V dv\left(H_i^{(q)*}\mu H_i^{(p)} + E_i^{(q)*}\epsilon_{ij}E_j^{(p)}\right) \\
&\quad +\int_{\bar{V}} dv\left(\bar{H}_i^{(q)*}\mu_0\bar{H}_i^{(p)} + \bar{E}_i^{(q)*}\epsilon_0\bar{E}_i^{(p)}\right)\Big\} = 0 \ .
\end{aligned} \tag{42}$$

Hence, for different frequency,

$$\begin{aligned}
&\int_V dv\left(H_i^{(q)*}\mu H_i^{(p)} + E_i^{(q)*}\epsilon_{ij}E_j^{(p)}\right) \\
&\quad +\int_{\bar{V}} dv\left(\bar{H}_i^{(q)*}\mu_0\bar{H}_i^{(p)} + \bar{E}_i^{(q)*}\epsilon_0\bar{E}_i^{(p)}\right) = 0 \ .
\end{aligned} \tag{43}$$

Replacing the volume integral with the line integral in x_2, we can reduce the general orthogonality relation to the following form for the guided waves in the infinite dielectric crystal plate

$$\begin{aligned}
&\int_{-b}^{b} dx_2\left(\mathcal{H}_i^{(q)*}\mu\mathcal{H}_i^{(p)} + \mathcal{E}_i^{(q)*}\epsilon_{ij}\mathcal{E}_j^{(p)}\right) \\
&+\int_{-\infty}^{-b} dx_2\left(\bar{\mathcal{H}}_i^{(q)*}\mu_0\bar{\mathcal{H}}_i^{(p)} + \bar{\mathcal{E}}_i^{(q)*}\epsilon_0\bar{\mathcal{E}}_i^{(p)}\right) \\
&+\int_{b}^{\infty} dx_2\left(\bar{\mathcal{H}}_i^{(q)*}\mu_0\bar{\mathcal{H}}_i^{(p)} + \bar{\mathcal{E}}_i^{(q)*}\epsilon_0\bar{\mathcal{E}}_i^{(p)}\right) = 0 \ .
\end{aligned} \tag{44}$$

8. Rayleigh–Ritz Approximation

In an anisotropic dielectric plate subjected to transversely inhomogeneous stress field, the dielectric tensor can be written as

$$\epsilon_{ij} = \epsilon_{ij}^0 + \Delta\epsilon_{ij} \ , \tag{45}$$

where ϵ_{ij}^0 is the dielectric tensor before the stress field is applied and $\Delta\epsilon_{ij} = \Delta\epsilon_{ij}(x_2)$ the change in dielectric tensor caused by stress. Because the elements in $\Delta\epsilon_{ij}$ are usually very small, $\Delta\epsilon_{ij}$ is related to the perturbation in relative dielectric impermeability tensor $\Delta\beta_{ij}$ by

$$\Delta\epsilon_{ij} = -\epsilon_0^{-1}\epsilon_{ik}^0(\Delta\beta_{kl})\epsilon_{lj}^0 \ . \tag{46}$$

The perturbation in β_{ij} is caused by the strain field S_{ij} through the piezooptical effect

$$\Delta\beta_{ij} = p_{ijkl}S_{kl} \ . \tag{47}$$

Let $\mathcal{E}_i^{(p)}e^{i(\xi x_1-\omega_p t)}$, $\mathcal{H}_i^{(p)}e^{i(\xi x_1-\omega_p t)}$ be the field components corresponding to frequency w_p^0 for the unstressed plate at a given wave number ξ and

$$\mathcal{E}_i = \sum_{p=1}^{n} m_p\mathcal{E}_i^{(p)} \ , \ \mathcal{H}_i = \sum_{p=1}^{n} m_p\mathcal{H}_i^{(p)} \ , \tag{48}$$

be the trial field components corresponding to frequency ω for the stressed plate. The m_p in (48) are variational parameters which can be considered as expansion coefficients of the electromagnetic fields of the stressed plate. Substitution of (48) into (38) leads to

$$\omega(m_1,m_2,\cdots,m_n) = \frac{\sum_{p,q=1}^{n} m_p m_q N_{pq}}{\sum_{p,q=1}^{n} m_p m_q Q_{pq}} \ , \tag{49}$$

where

$$\begin{aligned}
N_{pq} &= \int_{-b}^{b} dx_2\left(i\mathcal{E}_i^{(p)*}\epsilon_{ijk}\mathcal{H}_{k,j}^{(q)} - \xi\mathcal{E}_i^{(p)*}\epsilon_{ijk}\delta_{j1}\mathcal{H}_k^{(q)}\right. \\
&\qquad \left. -i\mathcal{H}_i^{(p)*}\epsilon_{ijk}\mathcal{E}_{k,j}^{(q)} + \xi\mathcal{H}_i^{(p)*}\epsilon_{ijk}\delta_{j1}\mathcal{E}_k^{(q)}\right) \\
&\quad +\int_{-\infty}^{-b} dx_2\left(i\bar{\mathcal{E}}_i^{(p)*}\epsilon_{ijk}\bar{\mathcal{H}}_{k,j}^{(q)} - \xi\bar{\mathcal{E}}_i^{(p)*}\epsilon_{ijk}\delta_{j1}\bar{\mathcal{H}}_k^{(q)}\right. \\
&\qquad \left. -i\bar{\mathcal{H}}_i^{(p)*}\epsilon_{ijk}\bar{\mathcal{E}}_{k,j}^{(q)} + \xi\bar{\mathcal{H}}_i^{(p)*}\epsilon_{ijk}\delta_{j1}\bar{\mathcal{E}}_k^{(q)}\right) \\
&\quad +\int_{b}^{\infty} dx_2\left(i\bar{\mathcal{E}}_i^{(p)*}\epsilon_{ijk}\bar{\mathcal{H}}_{k,j}^{(q)} - \xi\bar{\mathcal{E}}_i^{(p)*}\epsilon_{ijk}\delta_{j1}\bar{\mathcal{H}}_k^{(q)}\right. \\
&\qquad \left. -i\bar{\mathcal{H}}_i^{(p)*}\epsilon_{ijk}\bar{\mathcal{E}}_{k,j}^{(q)} + \xi\bar{\mathcal{H}}_i^{(p)*}\epsilon_{ijk}\delta_{j1}\bar{\mathcal{E}}_k^{(q)}\right) \\
Q_{pq} &= \int_{-b}^{b} dx_2\left(\mathcal{E}_i^{(p)*}\epsilon_{ij}\mathcal{E}_j^{(q)} + \mathcal{H}_i^{(p)*}\mu\mathcal{H}_i^{(q)}\right) \\
&\quad +\int_{-\infty}^{-b} dx_2\left(\bar{\mathcal{E}}_i^{(p)*}\epsilon_0\bar{\mathcal{E}}_i^{(q)} + \bar{\mathcal{H}}_i^{(p)*}\mu_0\bar{\mathcal{H}}_i^{(q)}\right) \\
&\quad +\int_{b}^{\infty} dx_2\left(\bar{\mathcal{E}}_i^{(p)*}\epsilon_0\bar{\mathcal{E}}_i^{(q)} + \bar{\mathcal{H}}_i^{(p)*}\mu_0\bar{\mathcal{H}}_i^{(q)}\right) \ .
\end{aligned} \tag{50}$$

With the orthogonality relation (44), we can normalize

$\mathcal{E}_i^{(p)}, \mathcal{H}_i^{(p)}$ such that

$$Q_{pq}^0 = \delta_{pq} = \int_{-b}^{b} dx_2 \left(\mathcal{E}_i^{(p)*} \epsilon_{ij}^0 \mathcal{E}_j^{(q)} + \mathcal{H}_i^{(p)*} \mu \mathcal{H}_i^{(q)} \right) + \int_{-\infty}^{-b} dx_2 \left(\bar{\mathcal{E}}_i^{(p)*} \epsilon_0 \bar{\mathcal{E}}_i^{(q)} + \bar{\mathcal{H}}_i^{(p)*} \mu_0 \bar{\mathcal{H}}_i^{(q)} \right) + \int_{b}^{\infty} dx_2 \left(\bar{\mathcal{E}}_i^{(p)*} \epsilon_0 \bar{\mathcal{E}}_i^{(q)} + \bar{\mathcal{H}}_i^{(p)*} \mu_0 \bar{\mathcal{H}}_i^{(q)} \right) . \tag{51}$$

Hence (50) reduces to

$$N_{pq} = \begin{cases} \omega_p^0 & \text{if } q = p \\ 0 & \text{otherwise} \end{cases} , \tag{52}$$

$$Q_{pq} = \delta_{pq} + \int_{-b}^{b} dx_2 \left(\mathcal{E}_i^{(p)*} \Delta\epsilon_{ij} \mathcal{E}_j^{(q)} \right) . \tag{53}$$

It's noticeable that N_{pq} is diagonal and Q_{pq} is Hermittian.

To get the stationary value of ω, the derivative of ω with respect to $m_p, p = 1, \ldots, n$ should be zero, which gives

$$\left[\omega_p^0 \delta_{pq} - \omega \Re(Q_{pq}) \right] m_q = 0 . \tag{54}$$

So the determination of frequency reduces to an algebraic eigenvalue problem.

For one term approximation, with the field components in the unstressed plate at frequency ω^0 being expressed as $\mathcal{E}_i^0, \mathcal{H}_j^0$, (54) reduces to

$$\omega = \frac{\omega^0}{Q} . \tag{55}$$

where

$$Q = 1 + \int_{-b}^{b} dx_2 \left(\mathcal{E}_i^{0*} \Delta\epsilon_{ij} \mathcal{E}_j^0 \right) . \tag{56}$$

It's noticeable that we omit the "$\Re$" in front ot Q because we see from (56) Q is real. The relative frequency shift can be written as

$$\frac{\omega - \omega^0}{\omega^0} = \frac{f - f^0}{f^0} = \frac{1 - Q}{Q} = \frac{-\int_{-b}^{b} \mathcal{E}_i^{0*} \Delta\epsilon_{ij} \mathcal{E}_j^0 dx_2}{Q} . \tag{57}$$

9. Special Case – Diagonal β_{ij}^0

We consider in this section the special case that β_{ij}^0 (and hence ϵ_{ij}^0) is diagonal. From equation (46), we have

$$\Delta\epsilon_{ij} = -\epsilon_0^{-1} \epsilon_{ii}^0 \left(\Delta\beta_{ij} \right) \epsilon_{jj}^0 \quad \text{(no sum)}. \tag{58}$$

Using the expressions of $\mathcal{E}_i^0$ for diagonal β_{ij}^0 from [7] and substituting (58) into (56), we obtain the following results.

For TE waves:

Symmetric modes:

$$\begin{aligned} &\mathcal{E}_1^0 = \mathcal{E}_2^0 = 0 , \\ &\mathcal{E}_3^0 = \frac{\beta_{33}^0}{\epsilon_0} A_1' \cos(\xi\lambda_1 x_2) , \\ &Q = 1 - \frac{{A_1'}^2}{\epsilon_0} \int_{-b}^{b} \Delta\beta_{33} \cos^2(\xi\lambda_1 x_2) dx_2 . \end{aligned} \tag{59}$$

Antisymmetric modes:

$$\begin{aligned} &\mathcal{E}_1^0 = \mathcal{E}_2^0 = 0 , \\ &\mathcal{E}_3^0 = \frac{\beta_{33}^0}{\epsilon_0} A_2' \sin(\xi\lambda_1 x_2) , \\ &Q = 1 - \frac{{A_2'}^2}{\epsilon_0} \int_{-b}^{b} \Delta\beta_{33} \sin^2(\xi\lambda_1 x_2) dx_2 . \end{aligned} \tag{60}$$

For TM waves:

Symmetric modes:

$$\begin{aligned} &\mathcal{E}_1^0 = \frac{\beta_{11}^0}{\epsilon_0} A_2' \sin(\xi\lambda_1 x_2) , \\ &\mathcal{E}_2^0 = \frac{i\beta_{22}^0}{\epsilon_0 \lambda_1} A_2' \cos(\xi\lambda_1 x_2) , \\ &\mathcal{E}_3^0 = 0 , \\ &Q = 1 - \frac{{A_2'}^2}{\epsilon_0} \int_{-b}^{b} \left\{ \Delta\beta_{11} \sin^2(\xi\lambda_1 x_2) + \Delta\beta_{22} \cos^2(\xi\lambda_1 x_2) \right\} dx_2 . \end{aligned} \tag{61}$$

Antisymmetric modes:

$$\begin{aligned} &\mathcal{E}_1^0 = \frac{\beta_{11}^0}{\epsilon_0} A_1' \cos(\xi\lambda_1 x_2) , \\ &\mathcal{E}_2^0 = \frac{-i\beta_{22}^0}{\epsilon_0 \lambda_1} A_1' \sin(\xi\lambda_1 x_2) , \\ &\mathcal{E}_3^0 = 0 , \\ &Q = 1 - \frac{{A_1'}^2}{\epsilon_0^2} \int_{-b}^{b} \left\{ \Delta\beta_{11} \cos^2(\xi\lambda_1 x_2) + \Delta\beta_{22} \sin^2(\xi\lambda_1 x_2) \right\} dx_2 . \end{aligned} \tag{62}$$

In (59), (60), (61), and (62), A_1' and A_2' are constants to be determined from the normalization relation (51) and λ_1 is the wave no. ratio given in [7].

By comparing (57) and (59) $\sim$ (62), we can see that, for diagonal β_{ij}^0, the relative frequency shifts of TE waves depend mostly on the perturbation $\Delta\beta_{33}$ and those of TM waves depend mostly on $\Delta\beta_{11}$ and $\Delta\beta_{22}$.

10. Numerical Results From Variational Method

The effect of stress fields is studied for three cases of loading on the infinite lithium niobate crystal plate with material constants given in section 4.

1. Under uniform stress

The perturbation in β_{ij} from the uniform compression $T_{11} = -1. \times 10^6$ is given in (25). The relative frequency shifts from one term approximation are computed from (59), (60), (61), (62), and (57) and plotted in Fig.6. By comparing Fig.2 and Fig.6, we can see that the results from exact solution and variational method with one term approximation are almost the same.

2. Under vertical acceleration

For an anisotropic elastic plate under a steady acceleration of magnitude G in the x_2 direction, the stress and strain fields are given by

$$\begin{aligned} T_{22} &= -\rho G b \left(1 - \frac{x_2}{b}\right) , \qquad \text{all other } T_{ij} = 0 , \\ S_{ij} &= (c_{ijkl})^{-1} T_{kl} = S_{ij}^0 \left(1 - \frac{x_2}{b}\right) , \end{aligned} \tag{63}$$

where

$$S_{ij}^0 = -(\rho G b)(c_{ijkl})^{-1} \delta_{k2}\delta_{l2} , \tag{64}$$

is the strain field at the middle plane of the plate. The perturbation in β_{ij} caused by the strain field can be obtained from (47)

$$\Delta\beta_{ij} = p_{ijkl} S_{kl}^0 \left(1 - \frac{x_2}{b}\right) . \tag{65}$$

The relative frequency shifts for $G = 9.81 m/sec^2$ are plotted in Fig.7.

3. Under horizontal acceleration

For an anisotropic elastic plate under a steady horizontal acceleration G which makes an angle ϕ with the x_1 axis, the stress and strain fields are given by

$$\begin{aligned} T_{23} &= \rho G b \left(1 - \frac{x_2}{b}\right) \sin\phi , \\ T_{21} &= -\rho G b \left(1 - \frac{x_2}{b}\right) \cos\phi , \\ T_{11} &= T_{22} = T_{33} = T_{13} = 0 , \\ S_{ij} &= (c_{ijkl})^{-1} T_{kl} = S_{ij}^0 \left(1 - \frac{x_2}{b}\right) , \end{aligned} \tag{66}$$

where

$$S_{ij}^0 = (\rho G b)(c_{ijkl})^{-1} (\delta_{k2}\delta_{l3} - \delta_{k2}\delta_{l1}) , \tag{67}$$

is the strain field at the middle plane. The relative frequency shifts for $\phi = 90°$ and $G = 9.81 m/sec^2$ are calculated and plotted in Fig.8.

Acknowledgements

This work was supported by a grant (Contract No. DAAL 03-90-G-0079) from the Army Research Office.

References

[1] J.F. Nye, "Physical properties of crystals," Oxford University Press, London, 1957.

[2] W.P. Mason, "Crystal physics of interaction process," Academic Press, N.Y., 1966.

[3] H.F. Tierstein, "Linear Piezoelectric Plate Vibrations," Plenum Press, New York, 1969.

[4] P.C.Y. Lee, "Electromagnetic radiation from an AT-cut quartz plate under lateral-field excitation," *J. Appl. Phys.* **65**(4), pp.1395-1399, Feb. 1989.

[5] R.D. Mindlin, "Electromagnetic radiation from a vibrating quartz plate," *Int. J. Solids Struct.* **9**, pp. 697-702, 1973.

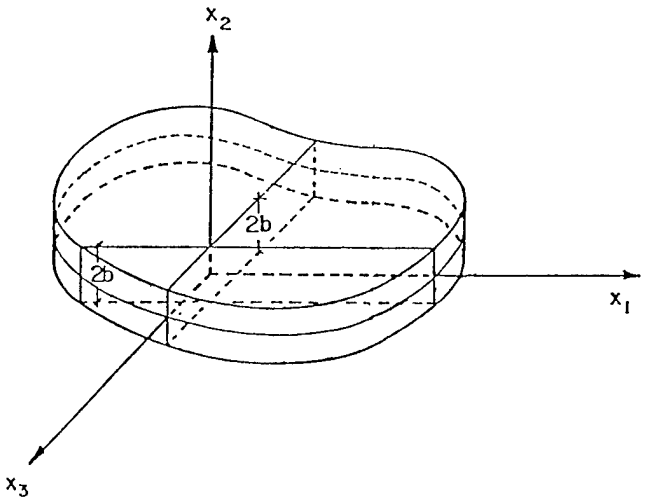
[6] P.C.Y. Lee, Y.G. Kim, and J.H. Prevost, " Electromagnetic radiation from doubly rotated piezoelectric crystal plates vibrating at thickness frequencies," *J. Appl. Phys.* **67**(11), pp. 6633-6642, June 1990.

[7] P.C.Y. Lee and J.D. Yu, "Guided EM Waves in Anisotropic Dielectric Plates", accepted for publication in *Proc. 1992 IEEE Ultrasonics Symp.*, 1992.

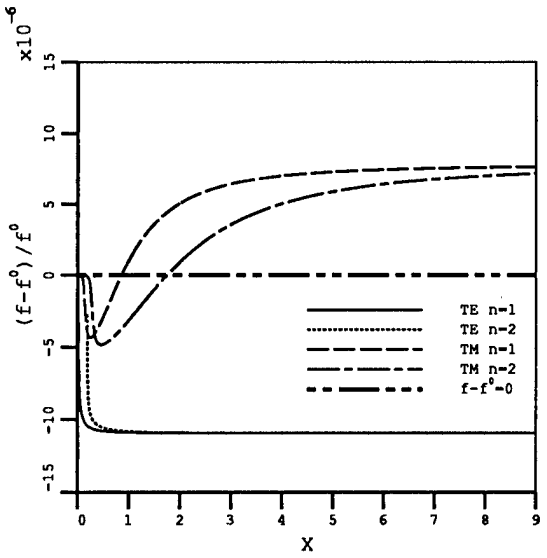
[8] H. Landolt and R. Börnstein, "Landolt-Börnstein New Series," Group III, **1**, K.H. Hellwege, ed., Springer-Verlag, Berlin, 1966.

[9] A.D. Berk, "Variational principles for electromagnetic resonators and waveguides," *IRE Trans. on Antennas and Propagation*, vol. AP-4, pp. 104-111, Apr. 1956.

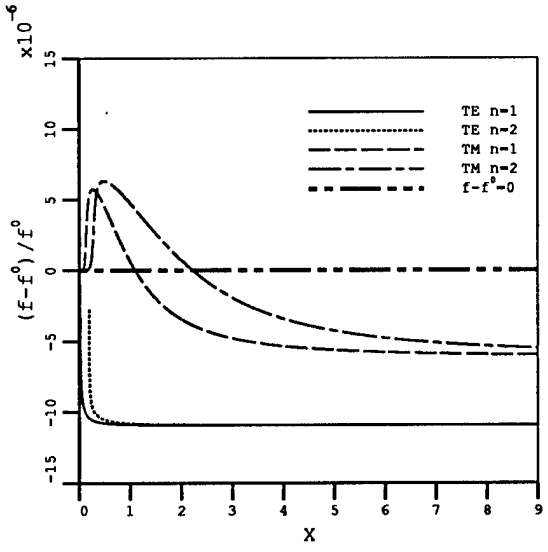
[10] S. Yamamoto, Y. Koyamada, and T. Makimoto, "Normal-mode analysis of anisotropic and gyrotropic thin-film waveguides for integrated Optics," *J. Appl. Phys.*, **43**(12), pp. 5090-5097, Dec. 1972.



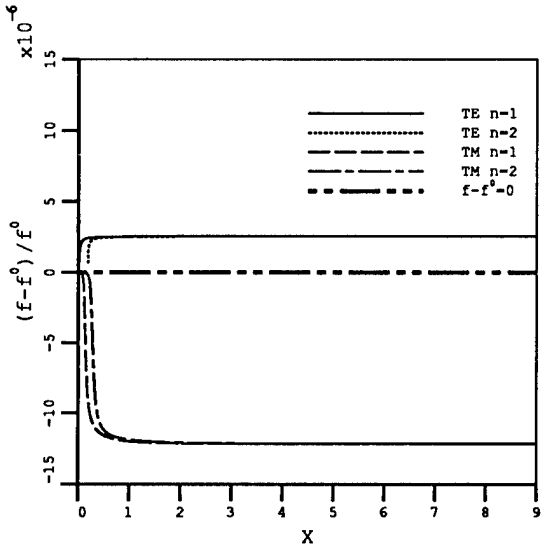
1. A dielectric plate referred to the rectangular coordinate system.



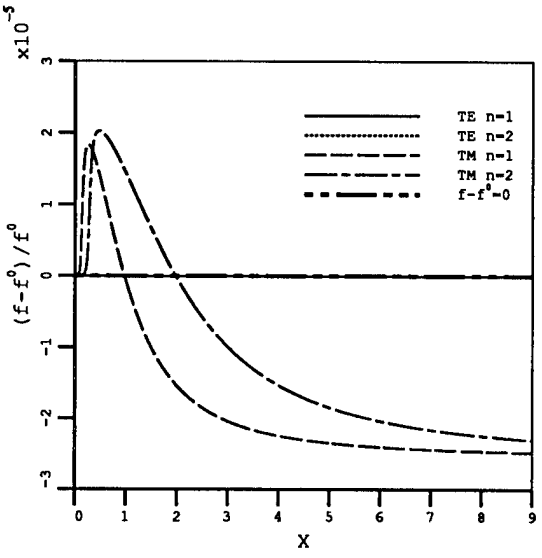
3. Relative frequency shift for $T_{22} = -1. \times 10^6 N/m^2$.



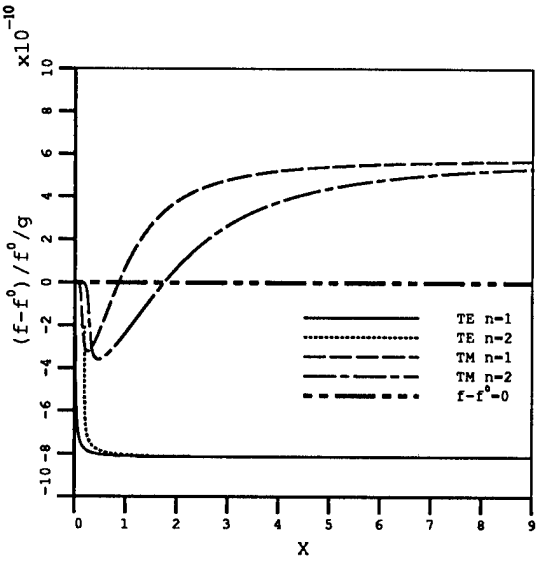
2. Relative frequency shift for $T_{11} = -1. \times 10^6 N/m^2$.



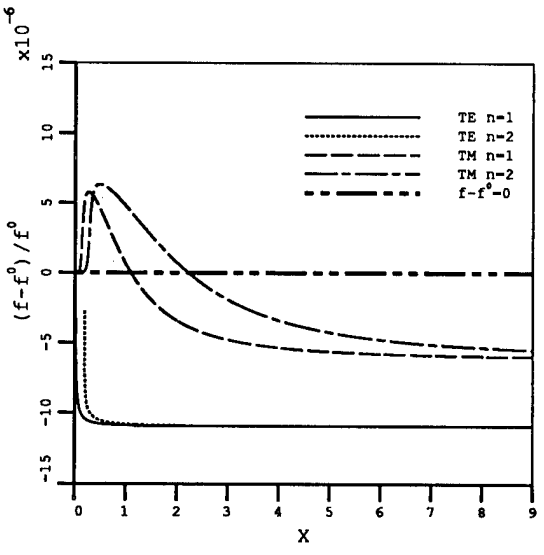
4. Relative frequency shift for $T_{33} = -1. \times 10^6 N/m^2$.



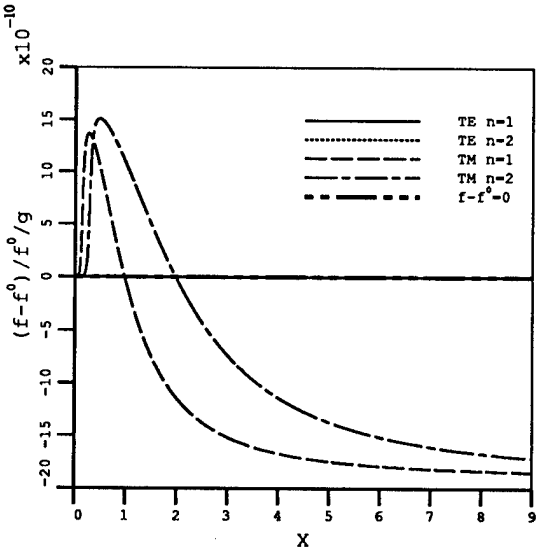
5. Relative frequency shift for $T_{23} = -1. \times 10^6 N/m^2$.



7. Relative frequency shift for vertical acceleration of $G = 1g$.



6. Relative frequency shift for $T_{11} = -1. \times 10^6 N/m^2$ from variational method with one term approximation.



8. Relative frequency shift for horizontal acceleration of $G = 1g$, $\phi = 90°$.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CALCULATION OF RADIATED ELECTROMAGNETIC POWER FROM BULK ACOUSTIC WAVE RESONATORS

C. F. CAMPBELL and R. J. WEBER

Department of Electrical Engineering and Computer Engineering
Iowa State University, Ames, Iowa 50011

ABSTRACT

In this paper, a general technique for calculating the radiated electromagnetic fields from an electrically small bulk acoustic wave resonator in the vicinity of acoustic resonance is presented. The BAW resonator is analyzed as an electrically small microstrip antenna on a piezoelectric substrate. By the application of the surface equivalence principle, the device is replaced by equivalent sources. The radiated far fields found from the equivalent sources are integrated over the upper half plane to determine the total amount of radiated EM power as a function of frequency. The result is general enough to characterize the electromagnetic radiation from arbitrarily shaped metalizations, and the antenna pattern and polarization of the radiation may also be determined. The theoretical results are experimentally confirmed by measuring the radiated EM power spectrum from lithium niobate and quartz resonators.

INTRODUCTION

With the use of piezoelectric films, the operating frequency of bulk acoustic wave resonators has been extended into the microwave frequency range [1]. As the resonant frequency of these devices continues to increase, spurious electromagnetic radiation will become more of a problem. Some prior work of note has been done on radiation from piezoelectric crystals. Mindlin and Lee have computed the radiated power from bare rotated quartz plates at thickness mode frequencies [2,3,4]. The work being proposed here differs from the prior work of Mindlin and Lee in that a bulk acoustic wave device is analyzed as a microstrip antenna with a piezoelectric substrate.

THEORY

Shown in Figure 1 is a rectangular bulk acoustic wave resonator where the piezoelectric layer and ground plane are assumed to extend to infinity. Due to the large difference between the electromagnetic and acoustic wavelengths in the piezoelectric substrate material, useful bulk acoustic wave resonators are almost always electrically small devices. The BAW resonator may be considered electrically small if the perimeter is less than 0.3 times the EM wavelength in the substrate material [5]. In an electrically small device, the conductors are considered equipotential surfaces

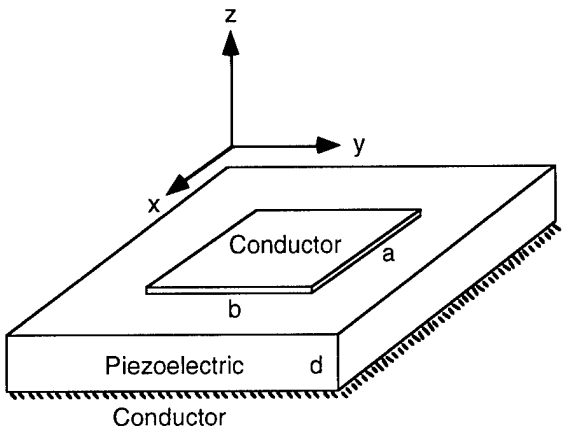


Figure 1

and the electric field is conservative. This is consistent with the quasi-static approximation for piezoelectric devices where the electric field is assumed to be irrotational and is related to the potential function by [6]

$$\mathbf{E} = -\nabla\phi \tag{1}$$

In order to estimate the radiated electromagnetic power from the device shown in Figure 1, the radiated fields must be found. One way to compute the radiated far fields is to apply the surface equivalence principle to the structure. The idea behind the surface equivalence principle is to surround the source by a closed surface S, and place equivalent sources on the closed surface. The equivalent sources will produce the same far fields outside of the closed surface as did the original source, but produce null fields inside the surface. The equivalent sources are electric and magnetic surface currents given by the following [7]:

$$\mathbf{M}_s = \mathbf{E} \times \hat{n} \tag{2}$$

$$\mathbf{J}_s = \hat{n} \times \mathbf{H} \tag{3}$$

The fields **E** and **H** are the actual electric and magnetic source fields that exist within the resonator. Since the electric field is assumed to be irrotational, the magnetic field of the device is equal to zero and only Eq. (2) needs to be considered. Summarized in Figure 2 is the

0-7803-0905-7/93 $3.00 © 1993 IEEE

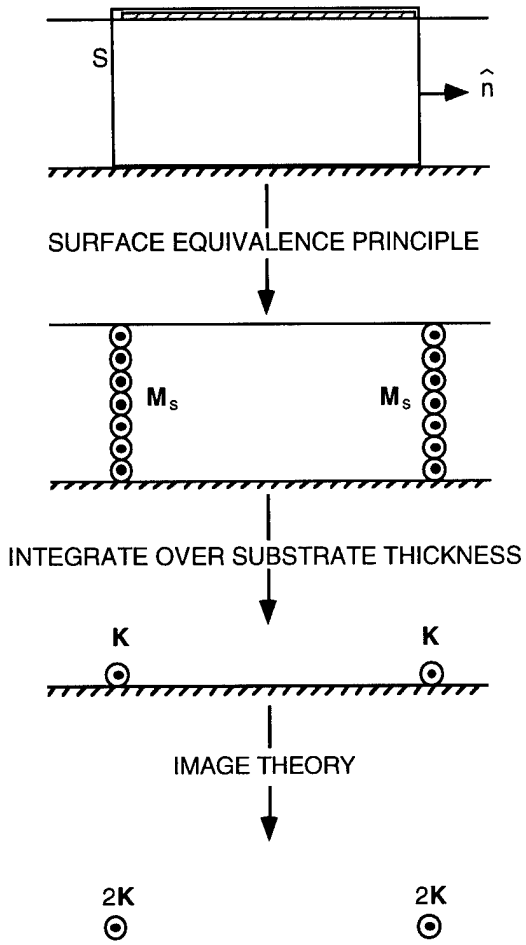


Figure 2

evolution of the electromagnetic radiation model for the BAW resonator.

First, the surface equivalence principle is applied by surrounding the structure with a closed surface S. If the conductors can be assumed to be perfect, the tangential electric field at the conductor surfaces must be zero and the magnetic surface current placed on the top and bottom of the closed surface S is also zero. The magnetic surface currents on the sides of the surface S are not zero and are the main source of the radiated fields as shown in the second part of Figure 2. Note, that in the radiation model, a ribbon of magnetic current density placed at the periphery of the device. Since the thickness of the substrate is on the order of an acoustic wavelength, it is electrically very thin, and the effect of the substrate on the radiated far fields is negligible [8]. The magnetic current density can then be integrated over the substrate thickness and condensed into a filamentary magnetic current loop **K** over an infinite ground plane as shown in the third illustration of Figure 2. The ground plane can then be removed with image theory, and the radiation model valid for the upper half plane is shown in the last part of Figure 2.

From Eq. (2), the magnetic current at the sides of closed surface S for an arbitrary electric field is equal to

$$\mathbf{K} = \hat{x}\,K_x + \hat{y}\,K_y = \int_0^d \mathbf{M}_s\,dz = \int_0^d \mathbf{E} \times \hat{n}\,dz$$

$$= (\hat{z} \times \hat{n}) \int_0^d E_z\,dz = -V(\hat{z} \times \hat{n}) \qquad (4)$$

The radiation model for the BAW resonator, valid for the upper half plane, reduces to a loop of magnetic current placed at the periphery of the device as in Figure 3.

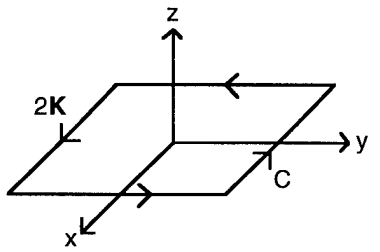


Figure 3

From Eq. (4), the magnitude of the magnetic current is equal to the voltage applied to the resonator. Since the conductors are assumed to be equipotential surfaces and the electric field conservative, this result is also valid for a three dimensional bulk acoustic wave resonator. A three dimensional device is a multiple mode structure and its electric field will be some complicated function of x,y and z. However, since the electric field is conservative, the line integral from one conductor to the other over any arbitrary path is equal to the voltage applied to the device. Assuming that the device is driven by a voltage source V_s with source impedance Z_o, from elementary circuit analysis the voltage across the device is equal to

$$V = V_s \frac{Z_{in}}{Z_{in} + Z_o} \qquad (5)$$

where Z_{in} is the impedance of the resonator.

The magnetic current in the radiation model can be computed with Eq. (4) once the input impedance of the device is known. Thus, if the impedance of the resonator is available via the Mason model, finite element computation, or measurement, Eq. (4) is valid as long as the device can be considered electrically small. The radiated far fields from an arbitrary loop of magnetic current impressed on the ground plane are [7]

$$E_\theta(r,\theta,\varphi) = \frac{j\omega\mu_o e^{-jk_o r}}{2\pi\eta_o r}\left[\tilde{K}_x(u,v)\sin\varphi - \tilde{K}_y(u,v)\cos\varphi\right] \qquad (6)$$

$$E_\varphi(r,\theta,\varphi) = \frac{j\omega\mu_o e^{-jk_o r}}{2\pi\eta_o r}\cos\theta\left[\tilde{K}_x(u,v)\cos\varphi + \tilde{K}_y(u,v)\sin\varphi\right] \qquad (7)$$

The Fourier transforms of the source are defined as

$$\tilde{K}_x(u,v) = \int_C K_x e^{j(ux+vy)} \hat{x} \cdot d\mathbf{l} \tag{8}$$

$$\tilde{K}_y(u,v) = \int_C K_y e^{j(ux+vy)} \hat{y} \cdot d\mathbf{l} \tag{9}$$

where

$$u = k_o \sin\theta \cos\varphi \qquad v = k_o \sin\theta \sin\varphi$$

$$k_o = \omega\sqrt{\varepsilon_o \mu_o} \qquad \eta_o = \sqrt{\frac{\mu_o}{\varepsilon_o}}$$

$$\varepsilon_o = 8.854 \times 10^{-12} \text{ F/m} \qquad \mu_o = 4\pi \times 10^{-7} \text{ H/m}$$

The total amount of electromagnetic power that is radiated by the bulk acoustic wave resonator into the upper half plane is equal to

$$P_{rad} = \frac{1}{\eta_o}\int_0^{2\pi}\int_0^{\pi/2} \left|\hat{\theta} E_\theta + \hat{\varphi} E_\varphi\right|^2 r^2 \sin\theta \, d\theta \, d\varphi \tag{10}$$

RESULTS

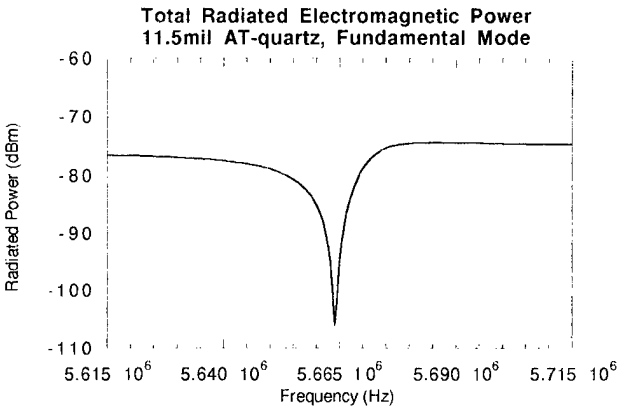
Consider the electrically small resonator shown in Figure 1. From Eqs. (8,9) the Fourier transforms are

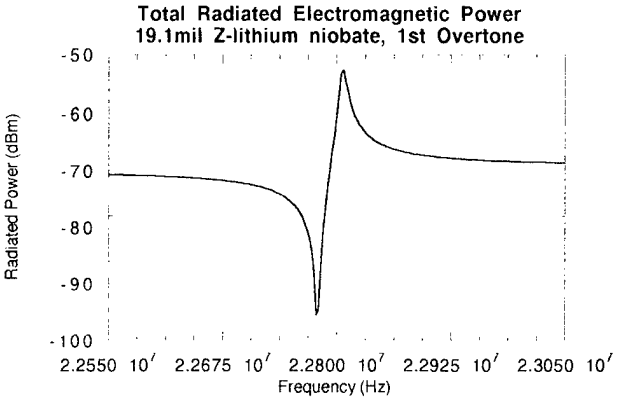
$$\tilde{K}_x(u,v) = \frac{V}{ju}(1 - e^{jvb})(1 - e^{jua}) \tag{11}$$

$$\tilde{K}_y(u,v) = -\frac{V}{jv}(1 - e^{jvb})(1 - e^{jua}) \tag{12}$$

For the case of a rectangular resonator, the radiated far fields may be written in closed form as a function of the impedance of the device. Examining Eq. (5), one would expect a drop in the radiated power at series resonance and an increase in the radiated power at parallel resonance.

To verify this, 5cm by 5cm bulk acoustic wave resonators were fabricated on 3 inch AT-cut quartz and Z-cut lithium niobate wafers. The quartz and lithium niobate wafers are 11.5mil and 19.1mil thick respectively, and the metalization is 0.9 microns of aluminum. The back side of the wafers are completely metalized, and the resonators are patterned on the front side. To apply the theory, the resonators were considered one dimensional and the input impedance was estimated with the Mason model [6]. The predicted total radiated power from the resonators for a 10V-50ohm source is plotted in Figures 4 and 5. The analysis was done for the fundamental thickness mode resonance of the quartz device and the first overtone thickness mode resonance of the lithium niobate resonator. Note that for both cases, the theory predicts a decrease in the





Figures 4-5

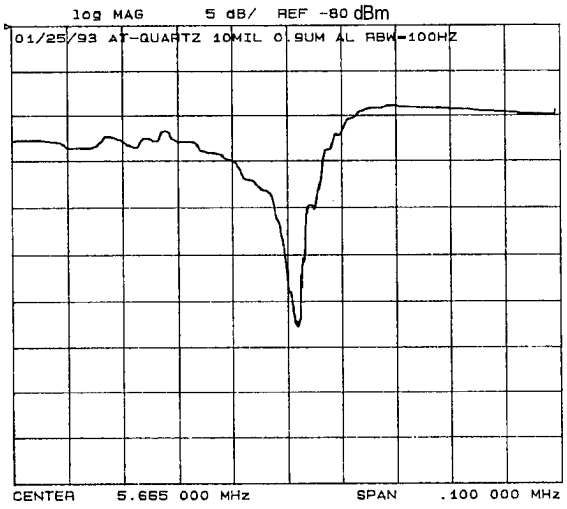
radiated power at series resonance and an increase in the radiated EM power at parallel resonance. Viscous damping loss was included in the calculation by the use of a complex elastic constant

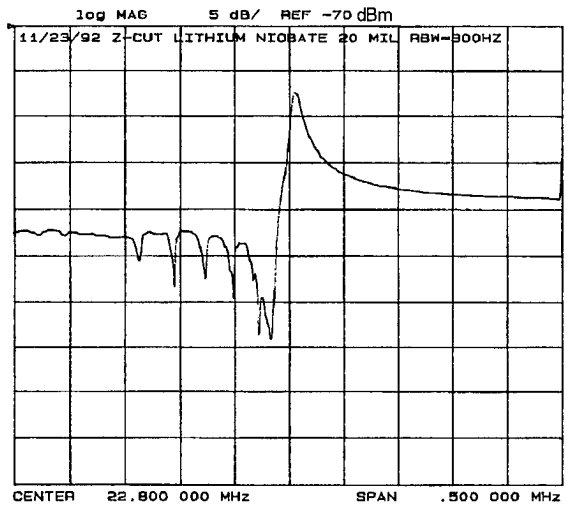
$$\hat{c} = c + j\omega\eta$$

where an estimate of the viscosity is obtained from the reported attenuation constants for quartz and lithium niobate [6,9].

To experimentally verify the theoretical results, the devices were driven by an HP8753A network analyzer, and the radiated power was picked up by an AM loopstick antenna. The AM antenna is connected to an HP3585A spectrum analyzer that is set in a peak hold mode so that the detected power may be displayed as a function of frequency. The goal of this experiment was to verify the shape of the radiation spectrum predicted by theory and not to attempt to measure the total power radiated by the resonator. What is predicted by theory is the total power radiated by the device when the Poynting vector has been integrated over the entire upper half space. For the experimental results, just the power picked up by the AM antenna is plotted. Thus, the absolute magnitude of the measured power spectrums cannot be compared with what is predicted by the theoretical calculation.

Shown in Figure 6 is the measured radiated EM power for the quartz resonator near the fundamental thickness mode. The radiated power from the lithium niobate device near the first overtone of the thickness mode is plotted in Figure 7. In both cases the radiated power drops off at series resonance and increases at parallel resonance. For the measurements, the output power of the network analyzer was set to 23dBm in order to get the detected signal level well above the noise floor of the spectrum analyzer. The measured radiated power spectrum is in excellent agreement with the theoretical computation with the exception of the spurious lateral modes which will not be predicted by the Mason model. Also consistent with the theoretical results is the relative magnitude of the radiated power increase at parallel resonance which is large for the lithium niobate device and small for the quartz device.





Figures 6-7

CONCLUSIONS

A technique for the calculation of the radiated electromagnetic power from electrically small BAW resonators has been presented. The method is valid for general BAW resonators as long as the impedance of the device is available. The theoretical results for the radiation spectrum were experimentally verified for AT-cut quartz and Z-cut lithium niobate resonators, and good agreement was obtained between experiment and theory for the shape of the radiation spectrums.

ACKNOWLEDGEMENTS

This work was supported by a U. S. Dept. of Commerce grant through the Center for Advanced Technology Development, and the Department of Electrical Engineering and Computer Engineering, Iowa State University. The graduate student was supported by a U. S. Dept. of Education Fellowship. The authors would also like to thank Crystal Technology,Inc. and P.R. Hoffman Materials Processing Corp. for supplying the lithium niobate and quartz wafers.

REFERENCES

[1] K. M. Lakin, "Modeling of Thin Film Resonators and Filters." *1992 IEEE MTT-S International Microwave Symposium*, pp. 149-152.

[2] P. C. Y. Lee, "Electromagnetic Radiation from an AT-Cut Quartz Plate Under Lateral Field Excitation." *IEEE 1988 Ultrasonics Symposium*, pp. 407-411.

[3] P. C. Y. Lee, Y. G. Kim, and J. H. Prevost, "Electromagnetic Radiation from Doubly Rotated Piezoelectric Crystal Plates Vibrating at Thickness Frequencies." *IEEE 1989 Ultrasonics Symposium*, pp. 423-428.

[4] R. D. Mindlin, "Electromagnetic Radiation From a Vibrating Quartz Plate." *Int. J. Solids Structures*, 9, pp. 697-702, 1972.

[5] W. L. Stutzman, and G. A. Thiele, *Antenna Theory and Design*. New York: John Wiley & Sons, 1981.

[6] J. F. Rosenbaum, *Bulk Acoustic Wave Theory and Devices*. Boston: Artech House, 1988.

[7] C. A. Balanis, *Advanced Engineering Electromagnetics*. New York: John Wiley & Sons, 1989.

[8] Y. T. Lo, and S. W. Lee, *Antenna Handbook*. New York: Van Nostrand Reinhold Co., 1988.

[9] B. A. Auld, *Acoustic Fields and Waves in Solids*. Vol. I - Vol. II. Malabar, Florida: Robert E. Krieger Publishing Co., 1990.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EXACT ANALYSIS OF THE PROPAGATION OF ACOUSTIC WAVES IN MULTILAYERED ANISOTROPIC PIEZOELECTRIC PLATES

By James T. Stewart and Yook-Kong Yong

Rutgers University, Dept. of Civil and Environmental Engineering, P.O. Box 909, Piscataway, N.J., 08855-0909

Abstract

Exact analysis of the propagation of acoustic waves in multilayered piezoelectric plates is performed using the transfer matrix method. A general technique for analyzing layered piezoelectric resonators under thickness and lateral field excitation is presented and is applied to the study of zinc oxide on silicon thin film resonators. Both one and two dimensional analysis with general material anisotropy is performed, and a simplified method for incorporating thin conducting electrodes on the plate's free surfaces is presented. The general methodology described is summarized into efficient algorithms to aid in the implementation of the procedures and some computational aspects are discussed. Results are presented for cutoff behavior as well as general dispersion characteristics for two and three layered plates.

Introduction

Exact analysis of the propagation of acoustic waves in multilayered piezoelectric plates is performed using a general and powerful transfer matrix approach. The transfer matrix method is a very simple technique for analyzing wave phenomena in layered media. This technique, as applied to piezoelectric crystals, has its roots in the one dimensional transmission line equivalent circuit approach originally used by electrical engineers to analyze stacked crystal filters. The most general analysis of this type was given by Ballato, Bertoni, and Tamir [2]. The transfer matrix method is simply the physical equivalent of the transmission line analysis, relying on the principles of mechanics, rather than on electric circuit analysis. Adler [3] has proposed a general method for constructing transfer matrices for both bulk and surface wave problems which is based on linear systems analysis. Nowotny, Benes, and Schmid [4] developed a general one dimensional transfer matrix method for calculating the admittance matrix for stacked piezoelectric layers with arbitrary electrode placement, and Nayfeh [5] used a transfer matrix approach to solve two dimensional problems of wave propagation at arbitrary angles in multilayered non-piezoelectric plates with monoclinic (or higher) symmetry. The present study represents a general one and two dimensional analysis of layered piezoelectric plates with general anisotropy. The one dimensional analysis presented is similar to the discussion of ref [4] with extensions to include lateral field excitation as well as a simplified technique for handling the effects of thin conducting electrodes on the plate's free surfaces. The two dimensional analysis presented represents, in principle, an extension of ref [5] to the more general case of piezoelectric layers. The two dimensional analysis is carried out for the cases of both open and short circuit conditions on the plate's free surfaces.

By definition, a transfer matrix is a linear transformation that maps a specified vector of field quantities from one point to another in a given material. When applied to a layered plate, this method greatly simplifies the analysis by guaranteeing continuity of field quantities across layer interfaces while reducing the problem to one involving these quantities evaluated at the plate's free surfaces only. More specifically, a transfer matrix for each layer is developed which maps displacements, stress tractions, electric potential, and electric charge from the layer's lower surface to it's upper surface. Satisfaction of inter-layer continuity of these field quantities leads to a simple matrix multiplication rule, from which follows a global plate transfer matrix. The plate transfer matrix equation that results is represented in terms of the field quantities evaluated on the plate's free surfaces only. From this, many problems involving thickness field excitation of layered piezoelectric resonators may be handled quite readily. It is shown that, with some modifications, this method can handle problems involving lateral field excitation as well. A simple technique for incorporating the mechanical effects of thin conducting electrodes on the plate's top and bottom surfaces is also presented

The present study involves the application of the transfer matrix method to the exact analysis of simple thickness modes and general dispersion behavior in Zinc Oxide on Silicon thin film resonators. Fundamental reso-

0-7803-0905-7/93 $3.00 © 1993 IEEE

nant and anti-resonant cutoff frequencies through both thickness and lateral field excitation are calculated for two and three layered plates. Along with these frequencies, thickness mode shapes for displacements, stress tractions, electric potential, and electric displacement are shown. Exact dispersion relations for propagating straight crested waves are also presented for these plates. Both real and imaginary branches of these dispersion relations are shown and these calculations are carried out for both open circuit and short circuit conditions. Two different crystal tensor rotations of the ZnO layers are considered, giving rise to different cases of electro-mechanical coupling.

General Problem

Consider a plate of thickness 2b constructed of N anisotropic piezoelectric layers stacked normal to the x_2 axis. The layers are rigidly bonded at each interface and are assumed to be, in general, triclinic in their material properties. The top and bottom of the plate ($x_2 = \pm b$) are optionally coated with very thin conducting electrodes with thicknesses of $2b_u$ and $2b_l$, and densities of ρ_u and ρ_l, for top and bottom, respectively. These electrodes may carry an alternating forcing potential. To simplify calculations, the quantities of interest for a given layer are referred to a layer-wise local coordinate system, $(x_1, x_2^{[r]}, x_3)$, $r = 1, \ldots, N$, whose origin corresponds to the lower face of the layer. The thickness of the r^{th} layer is denoted as $h^{[r]}$. Figure (1) shows the orientation of the N-layered plate with local coordinate systems.

Equations of Motion

To solve the wave propagation problem for the layered plate, the equations of motion must be satisfied separately in each layer. For the r^{th} layer, these equations are given as:

$$c_{ijkl}^{[r]} u_{k,li}^{[r]} + e_{kij}^{[r]} \phi_{,ki}^{[r]} = \rho^{[r]} \ddot{u}_j^{[r]}, \quad r=1,\ldots,N, \quad j=1,2,3 \qquad (1)$$

$$e_{ijk}^{[r]} u_{j,ki}^{[r]} - \varepsilon_{ij}^{[r]} \phi_{,ij}^{[r]} = 0, \qquad r=1,\ldots,N \qquad (2)$$

Where $c_{ijkl}^{[r]}$, $e_{kij}^{[r]}$, and $\varepsilon_{ij}^{[r]}$ are, respectively, the elastic, piezoelectric, and dielectric constants for the r^{th} layer, $\rho^{[r]}$ is the mass density for the r^{th} layer, and $u_j^{[r]}$, and $\phi^{[r]}$ are the displacements and electric potential for layer r.

Interfacial and Boundary Conditions

A set of solutions to equations (1) and (2) is a solution to the overall problem if the proper interfacial and boundary conditions are satisfied. The interfacial conditions are necessary to insure continuity of displacement, electric potential, traction, and electric charge across each layer interface. The interfacial conditions are given as follows:

1.) The displacements are continuous across each layer interface:

$$u_j^{[r]}(h^{[r]}) = u_j^{[r+1]}(0), \quad j=1,\ldots,3, \quad r=1,\ldots,N-1 \qquad (3)$$

2.) The electric potential is continuous across each layer interface:

$$\phi^{[r]}(h^{[r]}) = \phi^{[r+1]}(0), \quad r=1,\ldots,N-1 \qquad (4)$$

3.) The tractions are continuous across each layer interface:

$$\sigma_{2j}^{[r]}(h^{[r]}) = \sigma_{2j}^{[r+1]}(0), \quad j=1,\ldots,3, \quad r=1,\ldots,N-1 \qquad (5)$$

4.) The electric charge is continuous across each layer interface:

$$D_2^{[r]}(h^{[r]}) = D_2^{[r+1]}(0), \quad r=1,\ldots,N-1 \qquad (6)$$

The boundary conditions at $x_2 = \pm b$ are necessary to insure that the solutions take on the proper values at the two free surfaces. The electrode platings on the upper and lower surfaces are assumed to be very thin as compared to the thickness of the plate, thus allowing the stiffness of the plating to be neglected. Therefore, the mechanical effect of the electrodes is treated, to a good degree of approximation, as completely inertial. The electrical boundary conditions at $x_2 = \pm b$ take on two forms. The free surfaces may be either charge free (open circuit) or driven by an alternating potential (or as a special case, shorted). The mechanical boundary conditions are given as follows:

1.) The traction at the top surface must balance the inertial effect of the electrode plating:

$$\sigma_{2j}^{[N]}(h^{[N]}) = -2\rho_u b_u \ddot{u}_j^{[N]}(h^{[N]}), \quad j=1,\ldots,3 \qquad (7)$$

2.) The traction at the bottom surface must balance the inertial effect of the electrode plating:

$$\sigma_{2j}^{[1]}(0) = 2\rho_l b_l \ddot{u}_j^{[1]}(0), \quad j=1,\ldots,3 \qquad (8)$$

When no electrodes are present, the right hand sides of equations (7) and (8) are zero (traction free). The electrical boundary conditions are given as follows:

3.) For an open circuit condition, the top and bottom surfaces are charge free:

$$D_2^{[N]}(h^{[N]}) = 0 \qquad (9)$$

$$D_2^{[1]}(0) = 0 \tag{10}$$

4.) For a short circuit condition, the top and bottom surfaces are held at the same potential, say zero:

$$\phi^{[N]}(h^{[N]}) = 0 \tag{11}$$

$$\phi^{[1]}(0) = 0 \tag{12}$$

5.) A time harmonic forcing potential may be applied across the electrodes:

$$\phi^{[N]}(h^{[N]}) = 2\phi_0 e^{i\omega t} \tag{13}$$

$$\phi^{[1]}(0) = 0 \tag{14}$$

In equations (12) and (14), the bottom electrode is assumed to be grounded. This convention is desirable when applying the transfer matrix approach.

Simple Thickness Modes

Before studying the general dispersion characteristics of a layered plate, the cutoff behavior will be reviewed. In an infinite plate, the cutoff frequencies for straight crested waves propagating in a given direction occur when wave numbers along this direction vanish, or alternatively, wave lengths along this direction approach infinity. In this case, the components of displacement are only functionally dependant on the thickness coordinate of the plate, x_2. The vibration modes obtained for this case are known as simple thickness modes.

Formulation of The Transfer Matrix Method for Simple Thickness Modes

When the response of the plate is independent of x_1 and x_3 the equations of motion (1) and (2) simplify to:

$$c_{2jk2}^{[r]} u_{k,22}^{[r]} + e_{22j}^{[r]} \phi_{,22}^{[r]} = \rho^{[r]} \ddot{u}_j^{[r]}, \qquad r=1,...,N \tag{15}$$

$$e_{2k2}^{[r]} u_{k,22}^{[r]} - \varepsilon_{22}^{[r]} \phi_{,22}^{[r]} = 0, \qquad r=1,...,N \tag{16}$$

The electric potential can be eliminated from equations (15) and (16) by integrating equation (16) twice and solving for ϕ to give (dropping the [r] superscript):

$$\phi = \frac{e_{2k2}}{\varepsilon_{22}} u_k + B_7 x_2 + B_8 \tag{17}$$

Where B_7 and B_8 are arbitrary constants of integration. In equation (17) the superscript [r] denoting the dependance of all quantities on a given layer has been dropped to simplify the notation. It is understood that, unless stated otherwise, all quantities listed are referred to an arbitrary layer and are, in general, different for different layers. Substituting equation (17) into (15) gives an expression in terms of displacements only:

$$\tilde{c}_{2jk2} u_{k,22} = \rho \ddot{u}_j \tag{18}$$

Where $\tilde{c}_{2jk2}$ are so called piezoelectrically stiffened elastic constants, which are given as:

$$\tilde{c}_{2jk2} = c_{2jk2} + \frac{e_{22j} e_{2k2}}{\varepsilon_{22}} \tag{19}$$

To solve the wave equation (18) a general solution of the form:

$$u_j = A_j e^{i(\eta x_2 - \omega t)} \tag{20}$$

is assumed, where ω is the angular frequency and η a wave number along the x_2 direction. Substituting equation (20) into equation (18) gives the following:

$$\tilde{\Gamma}_{jk} A_k = 0 \tag{21}$$

Where $\tilde{\Gamma}_{jk}$ are the components of the Christoffel matrix for a given layer with only non-zero wave numbers along x_2 and are given by:

$$\tilde{\Gamma}_{11} = y^2 \tilde{c}_{66} - \rho v_1^2 \Omega^2 \tag{22.a}$$

$$\tilde{\Gamma}_{12} = y^2 \tilde{c}_{26} \tag{22.b}$$

$$\tilde{\Gamma}_{13} = y^2 \tilde{c}_{46} \tag{22.c}$$

$$\tilde{\Gamma}_{22} = y^2 \tilde{c}_{22} - \rho v_1^2 \Omega^2 \tag{22.d}$$

$$\tilde{\Gamma}_{23} = y^2 \tilde{c}_{24} \tag{22.e}$$

$$\tilde{\Gamma}_{33} = y^2 \tilde{c}_{44} - \rho v_1^2 \Omega^2 \tag{22.f}$$

Where $y = \eta / \frac{\pi}{2b}$ is a non-dimensionalized wave number, $\Omega = \omega / \bar{\omega}$ is a non-dimensionalized frequency with $\bar{\omega}$ the normalizing frequency (usually taken as that of the fundamental thickness-shear mode), and $v_1 = \frac{2b}{\pi} \bar{\omega}$.

A non-trivial solution to equation (21) exists

when the determinant of the coefficients $\tilde{\Gamma}_{jk}$ vanishes:

$$\left|\underset{\sim}{\tilde{\Gamma}}\right| = 0 \tag{23}$$

where $\underset{\sim}{\tilde{\Gamma}}$ is the matrix of components $\tilde{\Gamma}_{jk}$. Expanding equation (23) results in a sixth order polynomial in even powers of y with real coefficients. This polynomial will admit six roots, y_q, q = 1,...,6, and for each one of these roots, a corresponding amplitude ratio, α_{jq}, q = 1,...,6, j = 1,2,3. The amplitude ratio corresponding to a given root is simply the null space of the matrix $\underset{\sim}{\tilde{\Gamma}}$ when the value of the root is substituted into equations (22.a) through (22.f). With the six wave numbers and corresponding amplitude ratios, the displacement solution is constructed as a linear combination of six partial wave solutions:

$$u_j = \sum_{q=1}^{6} \alpha_{jq} B_q e^{i(\eta_q x_2 - \omega t)} \tag{24}$$

Where B_q are as yet undetermined coefficients. The electric potential can be reconstructed by substituting equation (24) into equation (17) to give:

$$\phi = \sum_{q=1}^{6} \alpha_{4q} B_q e^{i(\eta_q x_2 - \omega t)} + B_7 x_2 + B_8 \tag{25}$$

Where:

$$\alpha_{4q} = \frac{e_{2k2}}{\varepsilon_{22}} \alpha_{kq} \tag{26}$$

To apply the interfacial and boundary conditions, the components of traction and electric displacement normal to the plane of the plate, σ_{2j} and D_2 must be formulated. The components of σ_{2j} are given as:

$$\sigma_{2j} = c_{2jk2} u_{k,2} + e_{22j} \phi_{,2} \tag{27}$$

Substituting equations (24) and (25) into equation (27) gives:

$$\sigma_{2j} = \sum_{q=1}^{6} d_{jq} B_q e^{i(\eta_q x_2 - \omega t)} + e_{22j} B_7 \tag{28}$$

Where:

$$d_{jq} = i \frac{\pi}{2b} y_q \tilde{c}_{2jk2} \alpha_{kq} \text{ (no sum on q)} \tag{29}$$

Similarly,

$$D_2 = e_{2k2} u_{k,2} - \varepsilon_{22} \phi_{,2} \tag{30}$$

Substituting equations (24) and (25) into equation (30) gives:

$$D_2 = -\varepsilon_{22} B_7 \tag{31}$$

Equations (24), (25), (28), and (31) can be organized into a single matrix equation as:

(This equation is listed at the end) (32)

In equation (32) the extra terms, d_{j7} are given as:

$$d_{j7} = e_{22j} \qquad j=1,2,3 \tag{33}$$

and:

$$d_{47} = -\varepsilon_{22} \tag{34}$$

Equation (32) can be rewritten symbolically for the r^{th} layer as:

$$\underset{\sim}{P}^{[r]} = \underset{\sim}{X}^{[r]}(x_2^{[r]})\, \underset{\sim}{W}^{[r]}(x_2^{[r]})\, \underset{\sim}{B}^{[r]}, \qquad r=1,...,N \tag{35}$$

Where $\underset{\sim}{P}^{[r]}$ is the left hand side column vector of equation (32) for the r^{th} layer, $\underset{\sim}{X}^{[r]}$ is the matrix of equation (32) for the r^{th} layer, the matrix $\underset{\sim}{W}^{[r]}$ is a diagonal matrix given as:

(This equation is listed at the end) (36)

and $\underset{\sim}{B}^{[r]}$ is a column vector whose entries are $B_q^{[r]}$, q=1,...,8.

The vector $\underset{\sim}{P}^{[r]}$ represents all the pertinent field quantities for the r^{th} layer which are, in general, a function of x_2. Let $P^{[r]^-}$ denote these quantities evaluated at $x_2^{[r]} = 0$ (the lower face of layer r), and let $P^{[r]^+}$ denote these same quantities evaluated at $x_2^{[r]} = h^{[r]}$ (the upper face of layer r). This gives:

$$\underset{\sim}{P}^{[r]^-} = \underset{\sim}{X}^{[r]}(0)\, \underset{\sim}{B}^{[r]} \tag{37}$$

In equation (37), use is made of the fact that $\underset{\sim}{W}^{[r]}(0)$ is simply the identity matrix. Equation (37) can be solved for $\underset{\sim}{B}^{[r]}$ to give:

$$\underset{\sim}{B}^{[r]} = \underset{\sim}{X}^{[r]^{-1}}(0)\, \underset{\sim}{P}^{[r]^-} \tag{38}$$

Substituting equation (38) into equation (35) and evaluating at $x_2^{[r]} = h^{[r]}$ gives:

$$\underset{\sim}{P}^{[r]^+} = \underset{\sim}{X}^{[r]}(h^{[r]})\, \underset{\sim}{W}^{[r]}(h^{[r]})\, \underset{\sim}{X}^{[r]^{-1}}(0)\, \underset{\sim}{P}^{[r]^-} \tag{39}$$

or:

$$\underset{\sim}{P}^{[r]^+} = \underset{\sim}{\Psi}^{[r]}\underset{\sim}{P}^{[r]^-} \tag{40}$$

The matrix $\underset{\sim}{\Psi}^{[r]}$ in equation (40) is called the layer transfer matrix for layer r since it maps the values of the field quantities $\underset{\sim}{P}^{[r]}$ from their values at $x_2^{[r]} = 0$ to their values at $x_2^{[r]} = h^{[r]}$.

The interfacial conditions represented by equations (3) through (6) can be stated in vector form very simply as:

$$\underset{\sim}{P}^{[r+1]^-} = \underset{\sim}{P}^{[r]^+}, \quad r=1,...,N-1 \tag{41}$$

Using equation (40) in repeated applications of equation (41), results in a relation which maps the field quantities on the top surface of the plate to those on the plate's bottom surface. This relation is given as:

$$P^+ = \Psi P^- \tag{42}$$

Where $\underset{\sim}{P}^+ = \underset{\sim}{P}^{[N]^+}$ represents the field quantities evaluated at $x_2 = b$ and $\underset{\sim}{P}^- = \underset{\sim}{P}^{[1]^-}$ represents the field quantities evaluated at $x_2 = -b$. The matrix $\underset{\sim}{\Psi}$ is the plate transfer matrix which is given as:

$$\underset{\sim}{\Psi} = \underset{\sim}{\Psi}^{[N]}\underset{\sim}{\Psi}^{[N-1]}...\underset{\sim}{\Psi}^{[2]}\underset{\sim}{\Psi}^{[1]} \tag{43}$$

The plate transfer matrix is constructed by multiplying in reverse order the layer transfer matrices for each layer. When this is done, all the interfacial continuity conditions are automatically satisfied and the problem is reduced to one in terms of the field quantities evaluated at the plates two free surfaces.

Although the interfacial conditions are satisfied in equation (42), the vectors $\underset{\sim}{P}^+$ and $\underset{\sim}{P}^-$ must satisfy the boundary conditions at $x_2 = \pm b$ for a valid solution. The free vibration solution will be considered first for both open and short circuit conditions. Equations (7) and (8) require that the tractions at $x_2 = \pm b$ balance the effects of the electrode plating. That is, when electrodes are present, the tractions on the top and bottom surfaces take on non-zero values. The surface tractions can be made zero if the stress quantities in $\underset{\sim}{\Psi}^{[1]}$ and $\underset{\sim}{\Psi}^{[N]}$ are replaced by *effective* stresses which take into account the effects of the electrode plating. The traction boundary conditions are repeated here as:

$$\sigma_{2j}\Big|_{x_2 = b} = -2\rho_u b_u \ddot{u}_j^{[N]}\Big|_{x_2^{[N]} = h^{[N]}} \tag{44}$$

and:

$$\sigma_{2j}\Big|_{x_2 = -b} = 2\rho_l b_l \ddot{u}_j^{[1]}\Big|_{x_2^{[1]} = 0} \tag{45}$$

Using equation (24) in equation (45) gives:

$$\sigma_{2j}\Big|_{x_2 = -b} = -2\rho_l b_l \omega^2 \sum_{q=1}^{6} B_q^{[1]}\alpha_{jq}^{[1]} \tag{46}$$

And upon using equation (28), equation (46) becomes:

$$\sum_{q=1}^{6} d_{jq}^{[1]} B_q^{[1]} + d_{j7}^{[1]} B_7^{[1]} = -2\rho_l b_l \omega^2 \sum_{q=1}^{6} B_q^{[1]}\alpha_{jq}^{[1]} \tag{47}$$

Equation (47) can be rearranged to give:

$$\sum_{q=1}^{6} \bar{a}_{jq}^{[1]} B_q^{[1]} + d_{j7}^{[1]} B_7^{[1]} = 0 \tag{48}$$

Where $\bar{a}_{jq}^{[1]}$ represent the amplitude ratios of effective tractions for layer 1 which are given as:

$$\bar{a}_{jq}^{[1]} = d_{jq}^{[1]} + 2\rho_l b_l \omega^2 \alpha_{jq}^{[1]} \tag{49}$$

It is easily seen that if the electrode plating thickness, $2b_l$, is set to zero, $\bar{a}_{jq}^{[1]}$ reduces to $d_{jq}^{[1]}$. By inspecting rows 5, 6, and 7 of the matrix in equation (32) and comparing them with equation (48), it is seen that if the quantities $d_{jq}^{[1]}$ in the matrix $\underset{\sim}{X}^{[1]}(0)$ are replaced by $\bar{a}_{jq}^{[1]}$ when the first layer's transfer matrix is constructed, the traction quantities in $\underset{\sim}{P}^-$ of equation (42) will be zero, even when electrodes are present. By similar reasoning, effective tractions $\bar{a}_{jq}^{[N]}$ can be derived for the N^{th} layer as:

$$\bar{a}_{jq}^{[N]} = d_{jq}^{[N]} - 2\rho_u b_u \omega^2 \alpha_{jq}^{[N]} \tag{50}$$

It is likewise seen that if $d_{jq}^{[N]}$ in the matrix $\underset{\sim}{X}^{[N]}(h^{[N]})$ is replaced by $\bar{a}_{jq}^{[N]}$ when the N^{th} layer's transfer matrix is constructed, the top surface of the plate may be treated as traction free as well.

To apply the boundary conditions, it is useful to partition and/or reorganize the rows and columns of equation (42) as follows:

$$\begin{bmatrix} \underset{\sim}{P}_u^+ \\ \underset{\sim}{P}_\sigma^+ \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\Psi}_{uu} & \underset{\sim}{\Psi}_{u\sigma} \\ \underset{\sim}{\Psi}_{\sigma u} & \underset{\sim}{\Psi}_{\sigma\sigma} \end{bmatrix} \begin{bmatrix} \underset{\sim}{P}_u^- \\ \underset{\sim}{P}_\sigma^- \end{bmatrix} \tag{51}$$

The partitioning scheme, as will be seen, is dependant on the electrical boundary conditions.

For an open circuit condition, equations (9) and (10) must be satisfied. Equation (31) shows that D_2 is constant in each layer and, by continuity, is therefore constant throughout the plate's thickness. This means that if D_2 is set to zero on the plate's free surfaces, it must be zero throughout. This gives:

$$B_7^{[r]} = 0, \quad r=1,...,N \tag{52}$$

and since D_2 is now known, the 8th row and column of $\underset{\sim}{\Psi}$ can be ignored. In this case the partitioning in equation (51) takes the following form:

$$\underset{\sim}{P}_u^+ = \begin{bmatrix} u_1^+ & u_2^+ & u_3^+ & \phi^+ \end{bmatrix}^T \tag{53}$$

$$\underset{\sim}{P}_u^- = \begin{bmatrix} u_1^- & u_2^- & u_3^- & \phi^- \end{bmatrix}^T \tag{54}$$

$$\underset{\sim}{P}_\sigma^+ = \begin{bmatrix} \hat{\sigma}_6^+ & \hat{\sigma}_2^+ & \hat{\sigma}_4^+ \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T \tag{55}$$

$$\underset{\sim}{P}_\sigma^- = \begin{bmatrix} \hat{\sigma}_6^- & \hat{\sigma}_2^- & \hat{\sigma}_4^- \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T \tag{56}$$

$$\underset{\sim}{\Psi}_{uu} = \begin{bmatrix} \Psi_{11} & \Psi_{12} & \Psi_{13} & \Psi_{14} \\ \Psi_{21} & \Psi_{22} & \Psi_{23} & \Psi_{24} \\ \Psi_{31} & \Psi_{32} & \Psi_{33} & \Psi_{34} \\ \Psi_{41} & \Psi_{42} & \Psi_{43} & \Psi_{44} \end{bmatrix} \tag{57}$$

$$\underset{\sim}{\Psi}_{u\sigma} = \begin{bmatrix} \Psi_{15} & \Psi_{16} & \Psi_{17} \\ \Psi_{25} & \Psi_{26} & \Psi_{27} \\ \Psi_{35} & \Psi_{36} & \Psi_{37} \\ \Psi_{45} & \Psi_{46} & \Psi_{47} \end{bmatrix} \tag{58}$$

$$\underset{\sim}{\Psi}_{\sigma u} = \begin{bmatrix} \Psi_{51} & \Psi_{52} & \Psi_{53} & \Psi_{54} \\ \Psi_{61} & \Psi_{62} & \Psi_{63} & \Psi_{64} \\ \Psi_{71} & \Psi_{72} & \Psi_{73} & \Psi_{74} \end{bmatrix} \tag{59}$$

$$\underset{\sim}{\Psi}_{\sigma\sigma} = \begin{bmatrix} \Psi_{55} & \Psi_{56} & \Psi_{57} \\ \Psi_{65} & \Psi_{66} & \Psi_{67} \\ \Psi_{75} & \Psi_{76} & \Psi_{77} \end{bmatrix} \tag{60}$$

In equations (55) and (56), the terms $\underset{\sim}{\hat{\sigma}}_p$, p=6,2,4 represent effective stress quantities as given by equations (49) and (50) which are always zero on the two free surfaces. Using equations (55) and (56), equation (51) can be expanded to give:

$$\underset{\sim}{P}_u^+ = \underset{\sim}{\Psi}_{uu} \underset{\sim}{P}_u^- \tag{61}$$

$$\underset{\sim}{O} = \underset{\sim}{\Psi}_{\sigma u} \underset{\sim}{P}_u^- \tag{62}$$

To simplify the formulation, the potential at the bottom electrode is set to zero (grounded). With this, equation (62) becomes, in expanded form:

$$\begin{bmatrix} \Psi_{51} & \Psi_{52} & \Psi_{53} \\ \Psi_{61} & \Psi_{62} & \Psi_{63} \\ \Psi_{71} & \Psi_{72} & \Psi_{73} \end{bmatrix} \begin{bmatrix} u_1^- \\ u_2^- \\ u_3^- \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{63}$$

A non-trivial solution exists if the determinant of the reduced $\underset{\sim}{\Psi}_{\sigma u}$ matrix in equation (63) vanishes, i.e.:

$$\begin{vmatrix} \Psi_{51} & \Psi_{52} & \Psi_{53} \\ \Psi_{61} & \Psi_{62} & \Psi_{63} \\ \Psi_{71} & \Psi_{72} & \Psi_{73} \end{vmatrix} = 0 \tag{64}$$

For a short circuit condition, both free surfaces are held at zero potential, as given by equations (11) and (12). In this case the partitioning in equation (51) takes the following form:

$$\underset{\sim}{P}_u^+ = \begin{bmatrix} u_1^+ & u_2^+ & u_2^+ & D_2^+ \end{bmatrix}^T \tag{65}$$

$$\underset{\sim}{P}_u^- = \begin{bmatrix} u_1^- & u_2^- & u_3^- & D_2^- \end{bmatrix}^T \tag{66}$$

$$\underset{\sim}{P}_\sigma^+ = \begin{bmatrix} \phi^+ & \hat{\sigma}_6^+ & \hat{\sigma}_2^+ & \hat{\sigma}_4^+ \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T \tag{67}$$

$$\underset{\sim}{P}_\sigma^- = \begin{bmatrix} \phi^- & \hat{\sigma}_6^- & \hat{\sigma}_2^- & \hat{\sigma}_4^- \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T \tag{68}$$

$$\underset{\sim}{\Psi}_{uu} = \begin{bmatrix} \Psi_{11} & \Psi_{12} & \Psi_{13} & \Psi_{18} \\ \Psi_{21} & \Psi_{22} & \Psi_{23} & \Psi_{28} \\ \Psi_{31} & \Psi_{32} & \Psi_{33} & \Psi_{38} \\ \Psi_{81} & \Psi_{82} & \Psi_{83} & \Psi_{88} \end{bmatrix} \tag{69}$$

$$\underset{\sim}{\Psi}_{u\sigma} = \begin{bmatrix} \Psi_{14} & \Psi_{15} & \Psi_{16} & \Psi_{17} \\ \Psi_{24} & \Psi_{25} & \Psi_{26} & \Psi_{27} \\ \Psi_{34} & \Psi_{35} & \Psi_{36} & \Psi_{37} \\ \Psi_{84} & \Psi_{85} & \Psi_{86} & \Psi_{87} \end{bmatrix} \tag{70}$$

$$\underset{\sim}{\Psi}_{\sigma u} = \begin{bmatrix} \Psi_{41} & \Psi_{42} & \Psi_{43} & \Psi_{48} \\ \Psi_{51} & \Psi_{52} & \Psi_{53} & \Psi_{58} \\ \Psi_{61} & \Psi_{62} & \Psi_{63} & \Psi_{68} \\ \Psi_{71} & \Psi_{72} & \Psi_{73} & \Psi_{78} \end{bmatrix} \tag{71}$$

$$\underset{\sim}{\Psi}_{\sigma\sigma} = \begin{bmatrix} \Psi_{44} & \Psi_{45} & \Psi_{46} & \Psi_{47} \\ \Psi_{54} & \Psi_{55} & \Psi_{56} & \Psi_{57} \\ \Psi_{64} & \Psi_{65} & \Psi_{66} & \Psi_{67} \\ \Psi_{74} & \Psi_{75} & \Psi_{76} & \Psi_{77} \end{bmatrix} \tag{72}$$

As was done for the open circuit case, equation (51) is expanded with the aid of equations (67) and (68) to give:

$$\underset{\sim}{P}_u^+ = \underset{\sim}{\Psi}_{uu} \underset{\sim}{P}_u^- \tag{73}$$

$$\underset{\sim}{Q} = \underset{\sim}{\Psi}_{\sigma u} \underset{\sim}{P}_u^- \tag{74}$$

For a non-trivial solution to equations (73) and (74), the determinant of the $\underset{\sim}{\Psi}_{\sigma u}$ matrix must vanish, i.e.:

$$\begin{vmatrix} \Psi_{41} & \Psi_{42} & \Psi_{43} & \Psi_{48} \\ \Psi_{51} & \Psi_{52} & \Psi_{53} & \Psi_{58} \\ \Psi_{61} & \Psi_{62} & \Psi_{63} & \Psi_{68} \\ \Psi_{71} & \Psi_{72} & \Psi_{73} & \Psi_{78} \end{vmatrix} = 0 \tag{75}$$

Equations (54) and (75) represent the frequency equations for the layered plate for open and short circuit conditions respectively. The roots of the frequency equation for a particular set of boundary conditions (i.e., values of Ω for which either equation (54) or (75) is satisfied) are the cutoff frequencies of the layered plate. The basic procedure for obtaining these frequencies is to input a test Ω, construct the $\underset{\sim}{\Psi}_{\sigma u}$ matrix, check it's determinant, and update Ω accordingly. The procedure is repeated until an acceptable accuracy is obtained. An algorithm for finding a root of equation (54) or equation (75) is given as follows:

Algorithm 1

1.) Input data:
   i.) Input number of layers: N
   ii.) Input material properties for each layer: $c_{ijkl}^{[r]}$, $e_{ijk}^{[r]}$, $\varepsilon_{ij}^{[r]}$, and $\rho^{[r]}$, r=1,...,N
   iii.) Input electrode plating data: $2b_u$, ρ_u, $2b_l$, ρ_l
   iv.) Input initial estimate for Ω: Ω_0
2.) For r = 1 to N do:
   i.) Construct the matrix $\underset{\sim}{\Gamma}^{[r]}$ whose components are given by equations (22a) through (22f).
   ii.) Compute the six roots $y_q^{[r]}$, q=1,...,6 of $\left|\underset{\sim}{\Gamma}^{[r]}\right| = 0$, and their corresponding amplitude ratios $\alpha_{jq}^{[r]}$, q=1,...,6, j=1,2,3.
   iii.) Construct the layer transfer matrix $\underset{\sim}{\Psi}^{[r]}$ using equations (32) through (40), making the proper adjustments of stress terms as given by equations (49) and (50) when r = 1 or N.
3.) Construct the plate transfer matrix Ψ from the layer transfer matrices $\underset{\sim}{\Psi}^{[r]}$, r=1,...,N, as given by equation (43).
4.) Simplify and/or partition Ψ according to equations (53) through (60) for an open circuit condition, or equations (65) through (72) for a short circuit condition.
5.) Compute $|\Psi_{\sigma u}|$.
   i.) If $|\Psi_{\sigma u}| = 0$ (or some prescribed lower limit is satisfied) then STOP. Otherwise, update the estimate of Ω and GOTO Step 2.).

Once a value of Ω is found which satisfies the frequency equation, the vibration mode shapes can be determined. This is done by determining the constants $B^{[r]}$, r=1,...,N, and using equations (24), (25), (28), and (31), construct the distribution of the field quantities in each layer and hence throughout the thickness of the plate. When equation (74) is satisfied, a vector P_u^- which satisfies this equation is obtained as the null space of the matrix Ψ, which is now singular. From this, the vector $P^{[1]-}$ can be constructed as follows:

$$P^{[1]-} = \left[u_1^- \; u_2^- \; u_3^- \; 0 \; 0 \; 0 \; 0\right]^T \tag{76}$$

for an open circuit condition, and:

$$P^{[1]-} = \left[u_1^- \; u_2^- \; u_3^- \; 0 \; 0 \; 0 \; D_2^-\right]^T \tag{77}$$

for a short circuit condition. Using equation (38) gives:

$$B^{[1]} = X^{[1]^{-1}}(0)\, P^{[1]-} \tag{78}$$

The constants for the second layer can be found by using equations (41) and (39) to give:

$$P^{[2]-} = X^{[1]}(h^{[1]})\, W^{[1]}(h^{[1]})\, B^{[1]} \tag{79}$$

and then:

$$B^{[2]} = X^{[2]^{-1}}(0)\, P^{[2]-} \tag{80}$$

Equations (79) and (80) are applied to all the remaining layers to obtain the $B^{[r]}$ vectors necessary to complete the solution. This procedure is repeated in the following algorithm:

Algorithm 2

1.) Given: Ω such that $|\Psi_{\sigma u}| = 0$

2.) Solve for P_u^- as the null space of $\Psi_{\sigma u}$ and construct $P^{[1]-}$ using equation (76) or equation (77).

3.) Determine the vector $B^{[1]}$ for the first layer from equation (78).

4.) For r = 2 to N do:

i.) Determine $P^{[r]-}$ from equation (79).

ii.) Determine $B^{[r]}$ from equation (80).

Simple Thickness Modes In A Zinc Oxide On Silicon Thin Film Resonator

As an example, the transfer matrix method is applied to an infinite zinc oxide on silicon strip. Zinc oxide is a hexagonal (6mm) material and is piezoelectric. It will be assumed that in the reference configuration the x_2 axis is the axis of hexagonal symmetry. Silicon is a cubic material (m3m) and is not piezoelectric. The material properties for these crystals were obtained from Auld [1]. In studying the behavior of layered plates, special attention is given to the piezoelectric layers. The piezoelectric layers contain the electro-mechanical coupling necessary to transform the input electrical energy into mechanical energy in the form of acoustic waves. By transforming the material tensors of these layers, different piezoelectric behavior can be obtained, and as will be seen, different orientations of the crystallographic axes results in either thickness excitable or laterally excitable cases. In the current study, two different transformations are applied to the zinc oxide layer(s) to obtain two different modes of electro-mechanical coupling. The first case requires the rotation of the material tensors about the x_1 axis through an angle of 30° (x_1 rotated ZnO). The second case requires the rotation of the material tensors through an angle of -30° about the x_3 axis (x_3 rotated ZnO).

Simple Thickness Modes In A Zinc Oxide On Silicon Bilayer Strip

First, a two layered plate (bilayer) is studied. The first layer is constructed of Silicon and the second layer is constructed of Zinc Oxide, as shown in figure (2.a). With this configuration, two cases are studied. In the first case, an x_1 rotation of the Zinc Oxide crystal through an angle of $\theta = 30°$ is considered. In the second case, an x_3 rotation of the Zinc Oxide crystal through an angle of $\theta = 30°$ is considered. No rotations are preformed on the Silicon crystal for either case. In both cases, calculations are preformed with and without electrodes. Calculations of cutoff frequencies and thickness mode shapes are carried out on a plate of thickness 2b = 12 μm, while the thicknesses of the Si and ZnO layers are varied through different values. A useful parameter for measuring the amounts of the various materials that make up the plate is the volume fraction for the r^{th} layer which is given as:

$$C_r = \frac{h^{[r]}}{2b}, \quad r=1,...,N \tag{81}$$

For the bilayer shown in figure (2.a), a value of $C_1 = 0.0$

represents a plate consisting of all ZnO, while a value of $C_1 = 1.0$ represents a plate consisting of all Si. The cutoff frequencies are calculated for a bilayer by applying algorithm (1) and varying C_1 from 0.0 through 1.0. A useful parameter for measuring the amount of electrode material present on a plate is the "R ratio" which is defined as the ratio of the total mass of all electrodes per unit area to the mass of the plate per unit area. This quantity is given as:

$$R = \frac{2(\rho_u b_u + \rho_l b_l)}{N \sum_{r=1} \rho^{[r]} h^{[r]}} \tag{82}$$

In the present study a symmetric electrode configuration is used ($b_u = b_l$) for R values of 0.02 and 0.04, as well as the unelectroded case (R=0.00). Figure (3) shows the variation of cutoff frequencies with C_1 for the bilayer with an open circuit condition. In this figure, the solid curves represent essentially u_3 modes for x_1 rotated ZnO and essentially u_1 modes for x_3 rotated ZnO. The dashed curves represent essentially u_2 modes for both cases. This is so because the x_3 rotated ZnO can be obtained from the x_1 rotated ZnO by a simple 90° rotation in the plane of the plate. Thickness mode shapes are calculated for the bilayer by applying algorithm (2). Figures (4) through (6) show mode shapes for this plate for an open circuit condition with $C_1 = 0.4$.

Simple Thickness Modes In A 3-Layered Strip

To further study the behavior of layered crystal plates, a three layered plate will now be considered. Rather than studying a three layered plate with another non-piezoelectric layer, the more interesting problem of a plate with an additional piezoelectric phase will be considered. In this example, two differently oriented Zinc Oxide layers will be placed on a single Silicon layer as shown in figure (2.b). Once again, the thickness will be held constant at $2b = 12\mu m$. The proportion of Si will be maintained at $C_1 = 0.40$ while varying the volume fractions of the two ZnO layers between $C_2 = 0.0$ and $C_2 = 0.6$ while noting that $C_3 = 0.6 - C_2$. Figure (7) shows the variation of cutoff frequencies with C_2 for the 3-layered plate with a short circuit condition. Thickness mode shapes corresponding to $C_2 = 0.24$ with a short circuit condition are shown in figures (8) through (10).

Forced Vibrations

Piezoelectric resonators operate by transforming electrical input energy into mechanical energy. The resulting vibrations are driven by an electrical forcing function which usually takes the form of an alternating potential applied across at least one pair of electrodes. The electrodes may be constructed parallel to the plane of the plate (thickness excitation) or perpendicular to the plane of the plate (lateral excitation). To study the performance of actual resonators, it is important to compute both resonant and antiresonant frequencies under forced vibration by either thickness or lateral excitation. It will be seen that most of these frequencies have already been calculated in the previous analysis. To determine these frequencies of forced vibration for simple thickness modes, a new quantity, the electrical admittance per unit area of the plate, is introduced. The plate admittance per unit area will be denoted by Y and will take on different forms for thickness or lateral excitation. In general this quantity is given as:

$$Y = \frac{\text{Current Density}}{\text{Applied Voltage}} \tag{83}$$

Resonance occurs when the admittance approaches infinity and antiresonance occurs when the admittance approaches zero.

Thickness Excitation

Forced vibrations under thickness excitation are achieved by applying electrode plating to the top and bottom surfaces of the plate across which a time harmonic potential is applied. For this case the electrical boundary conditions take on the form as given by equations (13) and (14). The plate transfer matrix is constructed exactly as it is for the free vibration case and is partitioned according to equations (65) through (72). With a non-zero potential on the plate's upper surface, the vector $\underset{\sim}{P}^{+}_{\sigma}$ is now not equal to the zero vector. Equation (62) thus becomes:

$$\underset{\sim}{P}^{+}_{\sigma} = \underset{\sim}{\Psi}_{\sigma u} \underset{\sim}{P}^{-}_{u} \tag{84}$$

where:

$$\underset{\sim}{P}^{+}_{\sigma} = \left[2\phi_0 \; 0 \; 0 \; 0\right]^T \tag{85}$$

Solving for $\underset{\sim}{P}^{-}_{u}$ gives:

$$\underset{\sim}{P}^{-}_{u} = \underset{\sim}{\Psi}^{-1}_{\sigma u} \underset{\sim}{P}^{+}_{\sigma} \tag{86}$$

With $\underset{\sim}{P}_u^-$ known, D_2 is obtained directly because, as was shown previously, it is constant through out the thickness of the plate, and this value is taken from the last element in $\underset{\sim}{P}_u^-$ as given by equation (66). The current density, J, can now be determined as:

$$J = \dot{D}_2 = i\omega D_2 \tag{87}$$

The admittance is thus obtained as:

$$Y = \frac{i\omega D_2}{2\phi_0} \tag{88}$$

For resonance, the admittance must approach infinity. This is achieved when the matrix $\underset{\sim}{\Psi}_{\sigma u}$ is made singular, or:

$$\left|\underset{\sim}{\Psi}_{\sigma u}\right| = 0 \tag{89}$$

Equation (89) is identical to equation (75). This shows that the resonant frequencies for the thickness excitable case are the same as those for the short circuited free vibration condition. The antiresonant frequencies are obtained as the admittance approaches zero, or:

$$Y = 0 \tag{90}$$

from which it follows that:

$$D_2 = 0 \tag{91}$$

Equation (91) is satisfied only when the plate's free surfaces are charge free. Therefore, the antiresonant frequencies under thickness excitation are those obtained previously for the open circuit condition. In the case of x_1 rotated ZnO u_3 and u_2 modes are thickness excitable. Likewise, u_1 and u_2 modes are thickness excitable in x_3 rotated ZnO. In the 3-layered plate all three modes of vibration, u_1, u_2, and u_3, are thickness excitable. Figure (11) shows resonant and antiresonant frequencies as functions of C_1 for the ZnO on Si bilayer. Figure (12) shows a similar plot for essentially u_1, u_2, and u_3 modes in the 3-layered plate as functions of C_2, while holding C_1 constant at 0.40.

Lateral Excitation

For lateral excitation, the forcing potential is applied across a direction which is in the plane of the plate. In this case the plate's top and bottom surfaces remain free of electrodes and free of electric charge (traction free, open circuited). The electrical input is supplied to the plate through a non-zero constant electric field component in one direction, say x_1. As will be seen, this case is more complicated and will require modifications in the construction of the layer transfer matrices. For simple thickness modes with a non-zero E_1, the electric potential as given by equation (17) becomes:

$$\phi = \frac{e_{2k2}}{\varepsilon_{22}} u_k + B_7 x_2 + B_8 - E_1 x_1 \tag{92}$$

The equations of motion (15) and (16) remain unchanged, and hence so does the general solution given by equation (24) which is repeated here as:

$$u_j = \sum_{q=1}^{6} \alpha_{jq} B_q e^{i(\eta_q x_2 - \omega t)} \tag{93}$$

Using equations (92) and (93), an expression similar to equation (25) is obtained as:

$$\phi = \sum_{q=1}^{6} \alpha_{4q} B_q e^{i(\eta_q x_2 - \omega t)} + B_7 x_2 + B_8 - E_1 x_1 \tag{94}$$

The pertinent stress quantities thus follow as:

$$\sigma_{2j} = \sum_{q=1}^{6} d_{jq} B_q e^{i(\eta_q x_2 - \omega t)} + e_{22j} B_7 - e_{12j} E_1 \tag{95}$$

The component of electric displacement in the x_2 direction is given as:

$$D_2 = -\varepsilon_{22} B_7 + \varepsilon_{12} E_1 \tag{96}$$

Previously, it was shown that for an open circuit condition between the plate's free surfaces, $D_2 = 0$ through out the thickness. This allows equation (96) to be solved for B_7 to give:

$$B_7 = \frac{\varepsilon_{12}}{\varepsilon_{22}} E_1 \tag{97}$$

This allows B_7 and D_2 to be eliminated from the formulation. Using equation (97) in equations (94) and (95) gives:

$$\phi = \sum_{q=1}^{6} \alpha_{4q} B_q e^{i(\eta_q x_2 - \omega t)} + B_8 + \left[\frac{\varepsilon_{12}}{\varepsilon_{22}} x_2 - x_1\right] E_1 \tag{98}$$

$$\sigma_{2j} = \sum_{q=1}^{6} d_{jq} B_q e^{i(\eta_q x_2 - \omega t)} + \left[\frac{\varepsilon_{12}}{\varepsilon_{22}} e_{22j} - e_{12j}\right] E_1 \tag{99}$$

Equations (93), (98), and (99) can be written in matrix form to obtain an expression similar to equation (32) as:

(This equation is listed at the end) (100)

In equation (98) the electric potential is arbitrarily evaluated at $x_1 = 0$. In addition to the previously mentioned equations, an 8^{th} equation has been added. This equation is necessary and has the effect of simply mapping a constant E_1 from one point to the next.

The layer transfer matrix is constructed as it is for the free vibration case, except the matrix $\underset{\sim}{X}^{[r]}(h^{[r]})$ in equation (39) is replaced by the matrix of equation (100) and the quantity D_2 in $\underset{\sim}{P}^{[r]}$ and $\underset{\sim}{P}^{[r]^-}$ is replaced by E_1. The partitioning of the plate transfer matrix equation as given by equation (51) now takes on the following form:

$$\underset{\sim}{P}_u^+ = \left[u_1^+ \; u_2^+ \; u_3^+ \; \phi^+\right]^T \tag{101}$$

$$\underset{\sim}{P}_u^- = \left[u_1^- \; u_2^- \; u_3^- \; \phi^-\right]^T = \left[u_1^- \; u_2^- \; u_3^- \; 0\right]^T \tag{102}$$

$$\underset{\sim}{P}_\sigma^+ = \left[\sigma_6^+ \; \sigma_2^+ \; \sigma_4^+ \; E_1\right]^T = \left[0 \; 0 \; 0 \; E_1\right]^T \tag{103}$$

$$\underset{\sim}{P}_\sigma^- = \left[\sigma_6^- \; \sigma_2^- \; \sigma_4^- \; E_1\right]^T = \left[0 \; 0 \; 0 \; E_1\right]^T \tag{104}$$

$$\underset{\sim}{\Psi}_{uu} = \begin{bmatrix} \Psi_{11} & \Psi_{12} & \Psi_{13} & \Psi_{14} \\ \Psi_{21} & \Psi_{22} & \Psi_{23} & \Psi_{24} \\ \Psi_{31} & \Psi_{32} & \Psi_{33} & \Psi_{34} \\ \Psi_{41} & \Psi_{42} & \Psi_{43} & \Psi_{44} \end{bmatrix} \tag{105}$$

$$\underset{\sim}{\Psi}_{u\sigma} = \begin{bmatrix} \Psi_{15} & \Psi_{16} & \Psi_{17} & \Psi_{18} \\ \Psi_{25} & \Psi_{26} & \Psi_{27} & \Psi_{28} \\ \Psi_{35} & \Psi_{36} & \Psi_{37} & \Psi_{38} \\ \Psi_{45} & \Psi_{46} & \Psi_{47} & \Psi_{48} \end{bmatrix} \tag{106}$$

$$\underset{\sim}{\Psi}_{\sigma u} = \begin{bmatrix} \Psi_{51} & \Psi_{52} & \Psi_{53} & \Psi_{54} \\ \Psi_{61} & \Psi_{62} & \Psi_{63} & \Psi_{64} \\ \Psi_{71} & \Psi_{72} & \Psi_{73} & \Psi_{74} \\ \Psi_{81} & \Psi_{82} & \Psi_{83} & \Psi_{84} \end{bmatrix} \tag{107}$$

$$\underset{\sim}{\Psi}_{\sigma\sigma} = \begin{bmatrix} \Psi_{55} & \Psi_{56} & \Psi_{57} & \Psi_{58} \\ \Psi_{65} & \Psi_{66} & \Psi_{67} & \Psi_{68} \\ \Psi_{75} & \Psi_{76} & \Psi_{77} & \Psi_{78} \\ \Psi_{85} & \Psi_{86} & \Psi_{87} & \Psi_{88} \end{bmatrix} \tag{108}$$

If equation (51) is expanded and use is made of equations (101) through (104), the system can be simplified into two matrix equations:

$$\begin{bmatrix} u_1^+ \\ u_2^+ \\ u_3^+ \\ \phi^+ \\ E_1 \end{bmatrix} = \begin{bmatrix} \Psi_{11} & \Psi_{12} & \Psi_{13} & \Psi_{18} \\ \Psi_{21} & \Psi_{22} & \Psi_{23} & \Psi_{28} \\ \Psi_{31} & \Psi_{32} & \Psi_{33} & \Psi_{38} \\ \Psi_{41} & \Psi_{42} & \Psi_{43} & \Psi_{48} \\ \Psi_{81} & \Psi_{82} & \Psi_{83} & \Psi_{88} \end{bmatrix} \begin{bmatrix} u_1^- \\ u_2^- \\ u_3^- \\ E_1 \end{bmatrix} \tag{109}$$

and:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \Psi_{51} & \Psi_{52} & \Psi_{53} & \Psi_{58} \\ \Psi_{61} & \Psi_{62} & \Psi_{63} & \Psi_{68} \\ \Psi_{71} & \Psi_{72} & \Psi_{73} & \Psi_{78} \end{bmatrix} \begin{bmatrix} u_1^- \\ u_2^- \\ u_3^- \\ E_1 \end{bmatrix} \tag{110}$$

Rearranging equation (110) gives:

$$\begin{bmatrix} -\Psi_{58} \\ -\Psi_{68} \\ -\Psi_{78} \end{bmatrix} E_1 = \begin{bmatrix} \Psi_{51} & \Psi_{52} & \Psi_{53} \\ \Psi_{61} & \Psi_{62} & \Psi_{63} \\ \Psi_{71} & \Psi_{72} & \Psi_{73} \end{bmatrix} \begin{bmatrix} u_1^- \\ u_2^- \\ u_3^- \end{bmatrix} \tag{111}$$

or, symbolically:

$$\underset{\sim}{\hat{P}}_\sigma^+ = \underset{\sim}{\hat{\Psi}}_{\sigma u} \underset{\sim}{\hat{P}}_u^- \tag{112}$$

The displacements on the plates lower surface can be solved as:

$$\underset{\sim}{\hat{P}}_u^- = \underset{\sim}{\hat{\Psi}}_{\sigma u}^{-1} \underset{\sim}{\hat{P}}_\sigma^+ \tag{113}$$

From which the full vector $\underset{\sim}{P}^{[1]^-}$ is obtained as:

$$\underset{\sim}{P}^{[1]^-} = \left[u_1^- \; u_2^- \; u_3^- \; 0 \; 0 \; 0 \; 0 \; E_1\right]^T \tag{114}$$

With $\underset{\sim}{P}^{[1]^-}$ now known, the remaining interfacial values of the field quantities can be obtained from repeated applications of equation (40).

The plate admittance for the laterally excited case is given as:

$$Y = \frac{i\omega D}{E_1} \tag{115}$$

where:

$$D = \int_{-b}^{b} D_1(x_2)\,dx_2 = \sum_{r=1}^{N} D^{[r]} \tag{116}$$

and:

$$D^{[r]} = \int_{0}^{h^{[r]}} D_1^{[r]}\,dx_2^{[r]} \tag{117}$$

Using equation (92), the component of electric displacement in the x_1 direction for the r^{th} layer is given as:

$$D_1^{[r]} = \left[e_{1k2}^{[r]} - \frac{\varepsilon_{12}^{[r]}}{\varepsilon_{22}^{[r]}} e_{2k2}^{[r]}\right] u_{k,2}^{[r]} - \varepsilon_{12}^{[r]} B_7^{[r]} + \varepsilon_{11}^{[r]} E_1 \tag{118}$$

Using equation (97) and integrating between the prescribed limits, an expression for $D^{[r]}$ is obtained as:

$$D^{[r]} = \hat{e}_k^{[r]}\left[u_k^{[r]^+} - u_k^{[r]^-}\right] - h^{[r]}\left[\frac{\varepsilon_{12}^{[r]^2}}{\varepsilon_{22}^{[r]}} - \varepsilon_{11}^{[r]}\right] E_1 \tag{119}$$

where:

$$\hat{e}_k^{[r]} = e_{1k2}^{[r]} - \frac{\varepsilon_{12}^{[r]}}{\varepsilon_{22}^{[r]}} e_{2k2}^{[r]} \tag{120}$$

Therefore, once the interfacial values of the displacements are obtained, $D^{[r]}$ for each layer, r, can be computed and the results used in equation (116) to obtain the quantity D, and finally equation (115) is used to compute the admittance.

For resonance, the admittance must be made to approach infinity. This is achieved by requiring the $\hat{\underset{\sim}{\Psi}}_{\sigma u}$ matrix in equation (110) to be singular, or:

$$\left|\hat{\underset{\sim}{\Psi}}_{\sigma u}\right| = 0 \tag{121}$$

Comparison of equation (121) with equation (64) shows that the resonant frequencies for the laterally excited case are the same as those for the free vibration case with charge free and traction free surfaces. The antiresonant frequencies are obtained when the admittance approaches zero, or:

$$Y = 0 \tag{122}$$

Since the admittance is dependent on D_1 for each layer, the antiresonant frequencies cannot be accounted for by the previous analysis. Furthermore, these frequencies are only meaningful when electro-mechanical coupling through the terms $\hat{e}_k^{[r]}$ is present. Considering case of lateral excitation in the x_1 direction, predominantly u_1 modes are directly excitable in x_1 rotated ZnO and, to a lesser extent, are excitable in x_3 rotated ZnO. Predominantly u_2 modes are laterally excitable in x_3 rotated ZnO. In both x_1 and x_3 rotated ZnO, u_3 modes are not laterally excitable under E_1. The antiresonant frequencies are easily determined using algorithm (1) with step 5.) modified to calculate Y instead of $|\Psi_{\sigma u}|$. Figure (13) shows the variation of antiresonant frequencies with C_2 for the 3-layered plate under E_1 lateral field excitation.

Dispersion of Straight Crested Waves

The dispersion of straight crested waves in a layered plate will now be reviewed. For straight crested waves, propagation in the plane of the plate is assumed, while the in plane gradient of response perpendicular to the propagation direction is zero. This is analogous to the plane strain case in 2-D elastic analysis. To simplify the formulation, the propagation direction is taken as x_1. Other directions can be deduced by proper transformation of the material property tensors. For straight crested waves propagating in the x_1 direction, along with the non-zero x_2 wave numbers, wave numbers along x_1 will also be assumed non-zero. The general form of the displacements and electric potential for this case are given as:

$$u_j = A_j e^{i(\xi x_1 + \eta x_2 - \omega t)} \tag{123}$$

$$\phi = A_4 e^{i(\xi x_1 + \eta x_2 - \omega t)} \tag{124}$$

where ξ is a wave number along x_1. Substitution of equations (123) and (124) into equations (15) and (16) yields the Christoffel equations for this problem:

$$\sum_{k=1}^{4} \Gamma_{jk} A_k = 0, \quad j=1,\ldots,4 \tag{125}$$

where:

$$\Gamma_{11} = z^2 c_{11} + 2yzc_{16} + y^2 c_{66} - \rho v_1^2 \Omega^2 \qquad (126.a)$$

$$\Gamma_{12} = z^2 c_{16} + yz(c_{12} + c_{66}) + y^2 c_{26} \qquad (126.b)$$

$$\Gamma_{13} = z^2 c_{15} + yz(c_{14} + c_{56}) + y^2 c_{46} \qquad (126.c)$$

$$\Gamma_{14} = z^2 e_{11} + yz(e_{16} + e_{21}) + y^2 e_{26} \qquad (126.d)$$

$$\Gamma_{22} = z^2 c_{66} + 2yzc_{26} + y^2 c_{22} - \rho v_1^2 \Omega^2 \qquad (126.e)$$

$$\Gamma_{23} = z^2 c_{56} + yz(c_{46} + c_{25}) + y^2 c_{24} \qquad (126.f)$$

$$\Gamma_{24} = z^2 e_{16} + yz(e_{12} + e_{26}) + y^2 e_{22} \qquad (126.g)$$

$$\Gamma_{33} = z^2 c_{55} + 2yzc_{45} + y^2 c_{44} - \rho v_1^2 \Omega^2 \qquad (126.h)$$

$$\Gamma_{34} = z^2 e_{15} + yz(e_{14} + e_{25}) + y^2 e_{24} \qquad (126.i)$$

$$\Gamma_{44} = -z^2 \varepsilon_{11} - 2yz\varepsilon_{12} - y^2 \varepsilon_{22} \qquad (126.j)$$

where $z = \xi / \frac{\pi}{2b}$ is a non-dimensionalized wave number along x_1. Equations (125) and (126a) through (126j) are comparable to equations (21) and (22a) through (22f), respectively, except now an x_1 dependency has been added.

A non-trivial solution to equation (125) exists when the determinant of the coefficients Γ_{jk} vanishes:

$$|\underline{\Gamma}| = 0 \qquad (127)$$

where $\underline{\Gamma}$ is the 4x4 matrix of components Γ_{jk}. For given values of Ω and z, equation (127) can be expanded into a general 8th order polynomial in powers of y with real coefficients. This polynomial will admit eight roots, y_q, q=1,...,8, and eight corresponding amplitude ratios, α_{jq}, q=1,...,8, j=1,...,4. With this, the general solution to equation (127) can be written as:

$$u_j = \sum_{q=1}^{8} B_q \alpha_{jq} e^{i(\xi x_1 + \eta_q x_2 - \omega t)} \qquad (128)$$

$$\phi = \sum_{q=1}^{8} B_q \alpha_{4q} e^{i(\xi x_1 + \eta_q x_2 - \omega t)} \qquad (129)$$

Using equations (128) and (129), expressions analogous to equations (28) and (31) for the pertinent stress and electric displacement terms are derived as:

$$\sigma_{2j} = \sum_{q=1}^{8} d_{jq} B_q e^{i(\xi x_1 + \eta_q x_2 - \omega t)}, \quad j=1,...,3 \qquad (130)$$

and:

$$D_2 = \sum_{q=1}^{8} d_{4q} B_q e^{i(\xi x_1 + \eta_q x_2 - \omega t)} \qquad (131)$$

The terms d_{jq} are given in expanded form as:

(This equation is listed at the end) (132.a)

(This equation is listed at the end) (132.b)

(This equation is listed at the end) (132.c)

(This equation is listed at the end) (132.d)

Equations (128) through (131) are written in matrix form (omitting the $e^{i(\xi x_1 - \omega t)}$ term) as:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \phi \\ \sigma_6 \\ \sigma_2 \\ \sigma_4 \\ D_2 \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} & \alpha_{15} & \alpha_{16} & \alpha_{17} & \alpha_{18} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} & \alpha_{25} & \alpha_{26} & \alpha_{27} & \alpha_{28} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} & \alpha_{35} & \alpha_{36} & \alpha_{37} & \alpha_{38} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} & \alpha_{45} & \alpha_{46} & \alpha_{47} & \alpha_{48} \\ d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} & d_{18} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} & d_{28} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} & d_{38} \\ d_{41} & d_{42} & d_{43} & d_{44} & d_{45} & d_{46} & d_{47} & d_{48} \end{bmatrix} \begin{bmatrix} B_1 e^{i\eta_1 x_2} \\ B_2 e^{i\eta_2 x_2} \\ B_3 e^{i\eta_3 x_2} \\ B_4 e^{i\eta_4 x_2} \\ B_5 e^{i\eta_5 x_2} \\ B_6 e^{i\eta_6 x_2} \\ B_7 e^{i\eta_7 x_2} \\ B_8 e^{i\eta_8 x_2} \end{bmatrix} \qquad (133)$$

Equation (133) is written symbolically for the r[th] layer as:

$$\underline{P}^{[r]} = \underline{X}^{[r]} \underline{W}^{[r]}(x_2^{[r]}) \underline{B}^{[r]}, \quad r=1,...,N \qquad (134)$$

This equation is analogous to equation (35) except the matrix $\underline{X}^{[r]}$ is no longer dependant on x_2. In equation (134) the matrix $\underline{W}^{[r]}(x_2^{[r]})$ is now given as:

(This equation is listed at the end) (135)

Applying the same reasoning that was used in the previous development, an expression for the layer transfer matrix analogous to equations (39) and (40) is obtained as:

$$\underset{\sim}{P}^{[r]^+} = \underset{\sim}{\Psi}^{[r]}\underset{\sim}{P}^{[r]^-} \tag{136}$$

where:

$$\underset{\sim}{\Psi}^{[r]} = \underset{\sim}{X}^{[r]}\underset{\sim}{W}^{[r]}(h^{[r]})\underset{\sim}{X}^{[r]^{-1}} \tag{137}$$

Satisfaction of the interfacial continuity conditions leads to an expression for the plate transfer matrix, identical to equations (42) and (43) which are repeated here as:

$$\underset{\sim}{P}^{+} = \underset{\sim}{\Psi}\underset{\sim}{P}^{-} \tag{138}$$

where:

$$\underset{\sim}{\Psi} = \underset{\sim}{\Psi}^{[N]}\underset{\sim}{\Psi}^{[N-1]}\cdots\underset{\sim}{\Psi}^{[2]}\underset{\sim}{\Psi}^{[1]} \tag{139}$$

The mechanical effect of electrode plating on the plate's top and bottom surfaces is handled identically for the x_1 propagation problem as it is for the case of simple thickness modes. The stresses at $x_2 = \pm b$ are replaced by effective stresses which reflect the forces introduced by the inertial effects of the electrodes. This is achieved by replacing some of the constants d_{jq}, j=1,2,3 in the first and last layer transfer matrices with new constants, $\tilde{d}_{jq}$, j=1,2,3, as given by equations (49) and (50) which are repeated here as:

$$\tilde{d}_{jq}^{[1]} = d_{jq}^{[1]} + 2\rho_l b_l \omega^2 \alpha_{jq}^{[1]},\ j=1,2,3,\ q=1,\ldots,8 \tag{140}$$

$$\tilde{d}_{jq}^{[N]} = d_{jq}^{[N]} - 2\rho_u b_u \omega^2 \alpha_{jq}^{[N]},\ j=1,2,3,\ q=1,\ldots,8 \tag{141}$$

Referring to equation (137), equation (140) is applied to the $\underset{\sim}{X}^{[1]^{-1}}$ term in $\underset{\sim}{\Psi}^{[1]}$ and equation (141) is applied to the $\underset{\sim}{X}^{[N]}$ term in $\underset{\sim}{\Psi}^{[N]}$.

The dispersion equation for open and short circuit conditions is obtained by partitioning and/or reorganizing the rows and columns of equation (138) as given by equation (51) and computing the determinant of the proper submatrix. Symbolically, the frequency equation is given as:

$$\left|\underset{\sim}{\Psi}_{\sigma u}\right| = 0 \tag{142}$$

For the open circuit case, the free surfaces of the plate are charge free as before, except D_2 is no longer constant throughout the plate's thickness. Therefore, this quantity may not be eliminated from the formulation. With this, the partitioning in equation (51) becomes:

$$\underset{\sim}{P}_u^+ = \left[u_1^+\ u_2^+\ u_3^+\ \phi^+\right]^T \tag{143}$$

$$\underset{\sim}{P}_u^- = \left[u_1^-\ u_2^-\ u_3^-\ \phi^-\right]^T \tag{144}$$

$$\underset{\sim}{P}_\sigma^+ = \left[\hat{\sigma}_6^+\ \hat{\sigma}_2^+\ \hat{\sigma}_4^+\ D_2^+\right]^T = \left[0\ 0\ 0\ 0\right]^T \tag{145}$$

$$\underset{\sim}{P}_\sigma^- = \left[\hat{\sigma}_6^-\ \hat{\sigma}_2^-\ \hat{\sigma}_4^-\ D_2^-\right]^T = \left[0\ 0\ 0\ 0\right]^T \tag{146}$$

$$\underset{\sim}{\Psi}_{uu} = \begin{bmatrix} \Psi_{11} & \Psi_{12} & \Psi_{13} & \Psi_{14} \\ \Psi_{21} & \Psi_{22} & \Psi_{23} & \Psi_{24} \\ \Psi_{31} & \Psi_{32} & \Psi_{33} & \Psi_{34} \\ \Psi_{41} & \Psi_{42} & \Psi_{43} & \Psi_{44} \end{bmatrix} \tag{147}$$

$$\underset{\sim}{\Psi}_{u\sigma} = \begin{bmatrix} \Psi_{15} & \Psi_{16} & \Psi_{17} & \Psi_{18} \\ \Psi_{25} & \Psi_{26} & \Psi_{27} & \Psi_{28} \\ \Psi_{35} & \Psi_{36} & \Psi_{37} & \Psi_{38} \\ \Psi_{45} & \Psi_{46} & \Psi_{47} & \Psi_{48} \end{bmatrix} \tag{148}$$

$$\underset{\sim}{\Psi}_{\sigma u} = \begin{bmatrix} \Psi_{51} & \Psi_{52} & \Psi_{53} & \Psi_{54} \\ \Psi_{61} & \Psi_{62} & \Psi_{63} & \Psi_{64} \\ \Psi_{71} & \Psi_{72} & \Psi_{73} & \Psi_{74} \\ \Psi_{81} & \Psi_{82} & \Psi_{83} & \Psi_{84} \end{bmatrix} \tag{149}$$

$$\underset{\sim}{\Psi}_{\sigma\sigma} = \begin{bmatrix} \Psi_{55} & \Psi_{56} & \Psi_{57} & \Psi_{58} \\ \Psi_{65} & \Psi_{66} & \Psi_{67} & \Psi_{68} \\ \Psi_{75} & \Psi_{76} & \Psi_{77} & \Psi_{78} \\ \Psi_{85} & \Psi_{86} & \Psi_{87} & \Psi_{88} \end{bmatrix} \tag{150}$$

With this partitioning, equation (142) can be written explicitly as:

$$\begin{vmatrix} \Psi_{51} & \Psi_{52} & \Psi_{53} & \Psi_{54} \\ \Psi_{61} & \Psi_{62} & \Psi_{63} & \Psi_{64} \\ \Psi_{71} & \Psi_{72} & \Psi_{73} & \Psi_{74} \\ \Psi_{81} & \Psi_{82} & \Psi_{83} & \Psi_{84} \end{vmatrix} = 0 \tag{151}$$

For the short circuit condition the partitioning given by equations (65) through (72) is used and equation (142) becomes:

$$\begin{vmatrix} \Psi_{41} & \Psi_{42} & \Psi_{43} & \Psi_{48} \\ \Psi_{51} & \Psi_{52} & \Psi_{53} & \Psi_{58} \\ \Psi_{61} & \Psi_{62} & \Psi_{63} & \Psi_{68} \\ \Psi_{71} & \Psi_{72} & \Psi_{73} & \Psi_{78} \end{vmatrix} = 0 \qquad (152)$$

To study the dispersion of straight crested waves in a layered plate, wave numbers along x_1 (values of z) must be found which satisfy equation (59) for a given frequency. This is done by first specifying a frequency, and then finding values of z which satisfy equation (142). In general, there will be an infinite number of these wave numbers which can be found for a given frequency. Once a (z,Ω) pair is determined, algorithm (2) can be applied to determine the distribution of the field quantities through out the plate's thickness. The procedure for finding a single (z,Ω) pair is summarized in the following algorithm.

Algorithm 3

1.) Input data:
   - i.) Input number of layers: N
   - ii.) Input material properties for each layer: $c_{ijkl}^{[r]}$, $e_{ijk}^{[r]}$, $\varepsilon_{ij}^{[r]}$, and $\rho^{[r]}$, r=1,...,N
   - iii.) Input electrode plating data: $2b_u$, ρ_u, $2b_l$, ρ_l
   - iv.) Input Ω
   - iv.) Input initial estimate for z: z_0

2.) For r = 1 to N do:
   - i.) Construct the matrix $\underset{\sim}{\Gamma}^{[r]}$ whose components are given by equations (126.a) through (126.j).
   - ii.) Compute the eight roots $y_q^{[r]}$, q=1,...,8 of $\left|\underset{\sim}{\Gamma}^{[r]}\right| = 0$, and their corresponding amplitude ratios $\alpha_{jq}^{[r]}$, q=1,...,8, j=1,2,3,4.
   - iii.) Construct the layer transfer matrix $\underset{\sim}{\Psi}^{[r]}$ using equations (133) through (138), making the proper adjustments of stress terms as given by equations (140) and (141) when r = 1 or N.

3.) Construct the plate transfer matrix Ψ from the layer transfer matrices $\underset{\sim}{\Psi}^{[r]}$, r=1,...,N, as given by equation (139).

4.) Partition Ψ according to equations (143) through (150) for an open circuit condition, or equations (65) through (72) for a short circuit condition.

5.) Compute $|\Psi_{\sigma u}|$.
   - i.) If $|\Psi_{\sigma u}| = 0$ (or some prescribed lower limit is satisfied) then STOP. Otherwise, update the estimate of z and GOTO Step 2.).

For a specified range of frequencies, algorithm (3) is used to compute values of z, also falling in a specified range, which are plotted to obtain a dispersion relation. The dispersion relation is a plot of Ω vs. z over a certain range with like modes following regular curves. This plot is very important since it gives a graphical picture of the solution of the wave equation for a particular plate.

Figures (14) through (17) show some dispersion relations for the various plates studied thus far. Calculations were carried out for $\Omega \in [0, 1.5]$ and $Im\{z\}, Re\{z\} \in [0, 2]$. In each case the frequencies were normalized by appropriate values for ω_1. In these plots, purely real wave numbers are plotted to the right of the origin and purely imaginary wave numbers are plotted to the left. Figures (14) and (15) show dispersion relations corresponding to an x_1 rotated ZnO on Si bilayer for C_1 = 0.6. Figures (16) and (17) represent dispersion relations for the 3-layered plate for open and short circuit conditions. The calculations for this plate were carried out for C_2 = 0.24, while holding C_1 constant at 0.40.

Computational Aspects

To obtain the results contained in this study, an efficient computer program was written in FORTRAN implementing algorithms (1) through (3). Most of the calculations involving the construction and manipulation of transfer matrices are straight forward and follow simple matrix algebra. The particular linear algebra routines used in this study were obtained from the LINPACK subroutine library. At the core of the computational process are two important tasks which should be discussed.

The first basic step in constructing a layer transfer matrix is determining the x_2 wave numbers for that

layer as the roots of a polynomial, $p(y) = \sum_{i}^{n} a_i y^i$, which is itself obtained by expanding the determinant of the Christoffel matrix given by equation (23) or (127). In practice, explicit representation of this polynomial is prohibitively complicated. Some researchers have used a numerical procedure to determine the coefficients a_i in p(y) for a given frequency, and then to use these numbers in an algorithm such as Laguerre's method to obtain the roots. This procedure involves the evaluation of p(y) at a series of n sample points which is a simple matter of evaluating $|\tilde{\Gamma}|$ for particular values of y at the desired frequency, Ω. This leads to a linear system of equations, the solution of which are the constants a_i. This approach is straight forward but is computationally expensive, and in this author's opinion, is potentially inaccurate. The matrix which results from this procedure is the well known Vandermonde matrix [10] which is extremely ill conditioned and may not always lead to an accurate solution of the coefficients. The method employed in the current study was originally introduced by Müller [13]. Müller's method does not require the explicit polynomial to be supplied and does not require function derivatives to be evaluated. This method is very accurate and does not seem to require refining of the roots (at least not in this case) that other methods may need. Müllers method is also very efficient computationally, converging quadratically in the vicinity of a root. The actual implementation of Müller's method used in the current study is taken from ref. [11].

Once a root of p(y) has been found, the second step in this process is to find its corresponding amplitude ratio. It was stated earlier that the amplitude ratio corresponding to a given root is the null space of $\tilde{\Gamma}$ when the value of the root is substituted for y. The basic problem encountered is to find a non-trivial vector $\tilde{x}$ satisfying the equation $\tilde{A}\tilde{x} = \tilde{O}$ when the matrix $\tilde{A}$ is singular. The most straight forward way of doing this is to select one component of $\tilde{x}$, say x_p, set its value to 1, and solve the reduced system. This technique has been widely used by many researchers over the years to solve this type of problem. There is one serious drawback to this approach. When the term x_p in the actual solution vector is zero then the solution cannot be obtained since x_p was originally assumed to be non-zero. Situations like this arise when a pure mode exists in one of the coordinate dimensions, such as the u_1 mode in x_1 rotated ZnO or, equivalently, the u_3 mode in x_3 rotated ZnO. When working with materials with this property, the selection of x_p is not arbitrary and in general is difficult to do in a practical way. Another problem which arises in crystals with high degrees of symmetry is that of a repeated root in p(y). This occurs, for example, in Si where u_1 and u_3 modes are represented by a single degenerate solution. In this case, automatic determination of the amplitude ratios using the technique discussed above can be even more difficult. Typically, programmers get around these problems by adopting different solution schemes for each class of crystals that exhibit these properties. In the present study, a technique called a singular value decomposition (SVD) is employed which is a general and rigorous method for finding the solution space of a singular matrix without encountering the problems discussed above. The result is a general purpose program that can handle any crystal class without programming special cases. The implementation of the SVD algorithm used in the current study is G. W. Stewart's ZSVDC routine which is part of the LINPACK subroutine library.

Conclusion

The transfer matrix method developed in this chapter successfully solves the problems of simple thickness modes and propagation of straight crested waves in layered piezoelectric media. It has been found that this method is very easy to implement for numerical calculation on a computer. The systematic and straight forward methodology described allows for the easy creation of a very powerful and general procedure which can solve a wide variety of problems. The boundary conditions studied in the current analysis are the most important ones for the study of crystal resonators, but these are certainly not the only conditions that can be realized in practice. By manipulating the partitioning scheme of the plate transfer matrix, many other boundary conditions can be accommodated. It is not difficult to see how a fully general algorithm can be constructed to handle any physically realizable set of boundary conditions. Closed form solutions for layered piezoelectric resonators can, in general, not be obtained. The method presented here represents the only reliable technique for solving these problems for an arbitrary number of layers. The accuracy of these calculations can be seen by inspection of the thickness mode shapes. The thickness mode shapes presented satisfy all interfacial and boundary conditions with very good accuracy. A complete catalogue of mode shapes and dispersion relations as well as numerical results can be found in ref.[6].

*Support from the U.S. Army Research Office grant no. DAAL03-91-G-0018, and the National Research Council Associateship Program is gratefully acknowledged.

References

1.) B. A. Auld, Acoustic Fields And Waves In Solids, Vol. 1, New York, John Wiley and Sons, 1973.

2.) A. Ballato, H. Bertoni, and Theodor Tamir, "Systematic Design of Stacked-Crystal Filters by Microwave Network Methods" IEEE Trans. on Microwave Theory and Techniques, Vol. MTT-22, No. 1, Jan. 1974.

3.) E. L. Adler, "Matrix Methods Applied to Acoustic

Waves in Multilayers", IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 37, No. 6. Nov. 1990.

4.) H. Nowotny, E. Benes, and M. Schmid, "Layered Piezoelectric Resonators With An Arbitrary Number of Electrodes (General One-Dimensional Treatment)", J. Acoust. Soc. Am., 90 (3), Sept. 1991.

5.) A. H. Nayfeh, "The General Problem of Elastic Wave Propagation In Multilayered Anisotropic Media", J. Acoust. Soc. Am. 89(4), Pt. 1, April 1991.

6.) J. T. Stewart, Exact and Approximate Analysis of The Propagation of Acoustic Waves in Layered Piezoelectric Plates, Ph.D. Dissertation, Rutgers University, Oct., 1993.

7.) K. M. Lakin, G. R. Kline, R. S. Ketcham, A. R. Landin, W. A. Burkland, K. T. McCarron, S. D. Brazmen, and S. G. Burns, "Thin Film Resonator Technology", 41st Annual Frequency Control Symposium-1987, pp. 371-381.

8.) J. S. Wang, K. M. Lakin, and A. R. Landon, "Sputtered C-Axis Inclined Piezoelectric Films and Shear Wave Resonators", 1983 Ultrasonics Symposium Proceedings, pp. 144-150.

9.) Y. Mizasaka, S. Hoshino, S. Takahashi, "Advances In Structures and Fabrication Process for Thin Film Acoustic Resonators", 1987 Ultrasonics Symposium Proceedings, pp. 385-393.

10.) K. E. Atkinson, An Introduction to Numerical Analysis, New York, John Wiley & Sons, 2nd ed., 1978.

11.) S. D. Conte and C. de Boor, Elementary Numerical Analysis, An Algorithmic Approach, New York, McGraw-Hill Book Co., 3rd ed., 1980.

12.) W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Numerical Recipes The Art of Scientific Computing, Cambridge, Cambridge University Press, 1986.

13.) D. E. Muller, "A Method of Solving Algebraic Equations Using an Automatic Computer", Mathematical Tables and Other Aids To Computation (MTAC), Vol. 10, 1956, pp. 208-215.

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \phi \\ \sigma_6 \\ \sigma_2 \\ \sigma_4 \\ D_2 \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} & \alpha_{15} & \alpha_{16} & 0 & 0 \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} & \alpha_{25} & \alpha_{26} & 0 & 0 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} & \alpha_{35} & \alpha_{36} & 0 & 0 \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} & \alpha_{45} & \alpha_{46} & x_2 & 1 \\ d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} & 0 \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} & 0 \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & d_{47} & 0 \end{bmatrix} \begin{bmatrix} B_1 e^{i\eta_1 x_2} \\ B_2 e^{i\eta_2 x_2} \\ B_3 e^{i\eta_3 x_2} \\ B_4 e^{i\eta_4 x_2} \\ B_5 e^{i\eta_5 x_2} \\ B_6 e^{i\eta_6 x_2} \\ B_7 \\ B_8 \end{bmatrix} \quad (32)$$

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \phi \\ \sigma_6 \\ \sigma_2 \\ \sigma_4 \\ E_1 \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} & \alpha_{15} & \alpha_{16} & 0 & 0 \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} & \alpha_{25} & \alpha_{26} & 0 & 0 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} & \alpha_{35} & \alpha_{36} & 0 & 0 \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} & \alpha_{45} & \alpha_{46} & 1 & \frac{\varepsilon_{12}}{\varepsilon_{22}} x_2 \\ d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & 0 & \frac{\varepsilon_{12}}{\varepsilon_{22}} e_{26} - e_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & 0 & \frac{\varepsilon_{12}}{\varepsilon_{22}} e_{22} - e_{12} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & 0 & \frac{\varepsilon_{12}}{\varepsilon_{22}} e_{24} - e_{14} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} B_1 e^{i\eta_1 x_2} \\ B_2 e^{i\eta_2 x_2} \\ B_3 e^{i\eta_3 x_2} \\ B_4 e^{i\eta_4 x_2} \\ B_5 e^{i\eta_5 x_2} \\ B_6 e^{i\eta_6 x_2} \\ B_8 \\ E_1 \end{bmatrix} \quad (36)$$

$$\underset{\sim}{W}^{[r]} = \begin{bmatrix} e^{i\eta_1^{[r]} x_2^{[r]}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{i\eta_2^{[r]} x_2^{[r]}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{i\eta_3^{[r]} x_2^{[r]}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{i\eta_4^{[r]} x_2^{[r]}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{i\eta_5^{[r]} x_2^{[r]}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{i\eta_6^{[r]} x_2^{[r]}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (100)$$

$$d_{1q} = i\frac{\pi}{2b}\left[(zc_{16}+y_q c_{66})\alpha_{1q}+(zc_{66}+y_q c_{26})\alpha_{2q}+(zc_{56}+y_q c_{46})\alpha_{3q}+(ze_{16}+y_q e_{26})\alpha_{4q}\right] \tag{132.a}$$

$$d_{2q} = i\frac{\pi}{2b}\left[(zc_{12}+y_q c_{26})\alpha_{1q}+(zc_{26}+y_q c_{22})\alpha_{2q}+(zc_{25}+y_q c_{24})\alpha_{3q}+(ze_{12}+y_q e_{22})\alpha_{4q}\right] \tag{132.b}$$

$$d_{3q} = i\frac{\pi}{2b}\left[(zc_{14}+y_q c_{46})\alpha_{1q}+(zc_{46}+y_q c_{24})\alpha_{2q}+(zc_{45}+y_q c_{44})\alpha_{3q}+(ze_{14}+y_q e_{24})\alpha_{4q}\right] \tag{132.c}$$

$$d_{4q} = i\frac{\pi}{2b}\left[(ze_{21}+y_q e_{26})\alpha_{1q}+(ze_{26}+y_q e_{22})\alpha_{2q}+(ze_{25}+y_q e_{24})\alpha_{3q}-(z\varepsilon_{12}+y_q \varepsilon_{22})\alpha_{4q}\right] \tag{132.d}$$

$$\underset{\sim}{W}^{[r]} = \begin{bmatrix} e^{i\eta_1^{[r]}x_2^{[r]}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{i\eta_2^{[r]}x_2^{[r]}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{i\eta_3^{[r]}x_2^{[r]}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{i\eta_4^{[r]}x_2^{[r]}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{i\eta_5^{[r]}x_2^{[r]}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{i\eta_6^{[r]}x_2^{[r]}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & e^{i\eta_7^{[r]}x_2^{[r]}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & e^{i\eta_8^{[r]}x_2^{[r]}} \end{bmatrix} \tag{135}$$

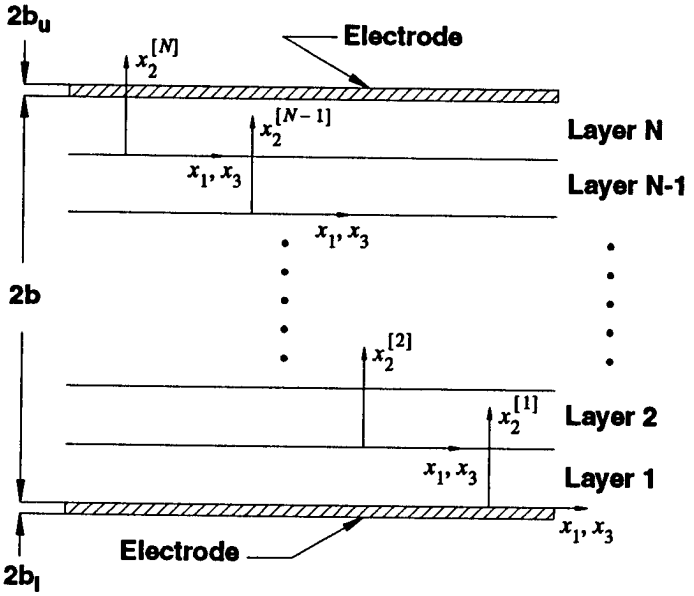
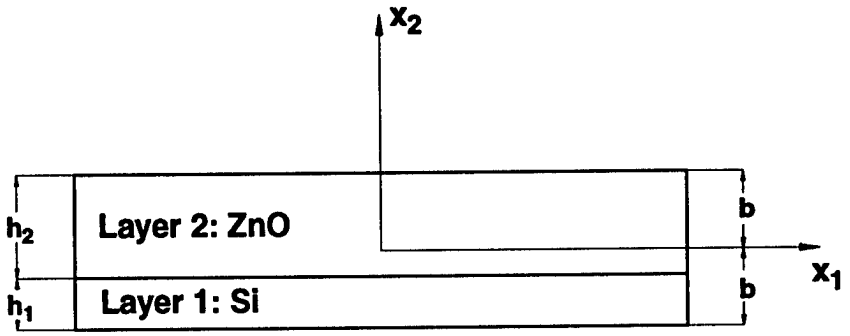
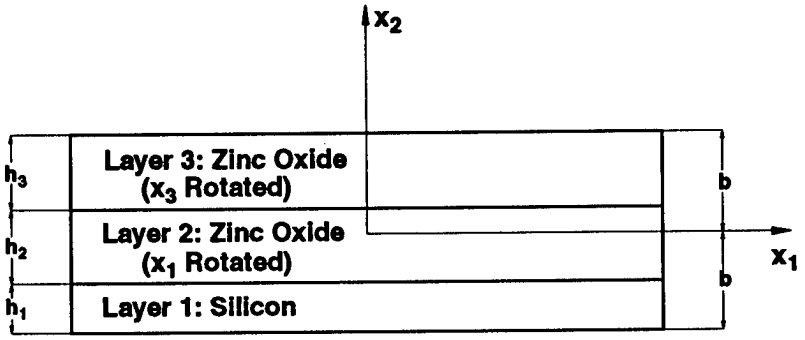


Figure 1 Orientation of N-layered plate with local coordinate systems



(a.)



(b.)

Figure 2 Configuration of Two and Three Layered Plates.

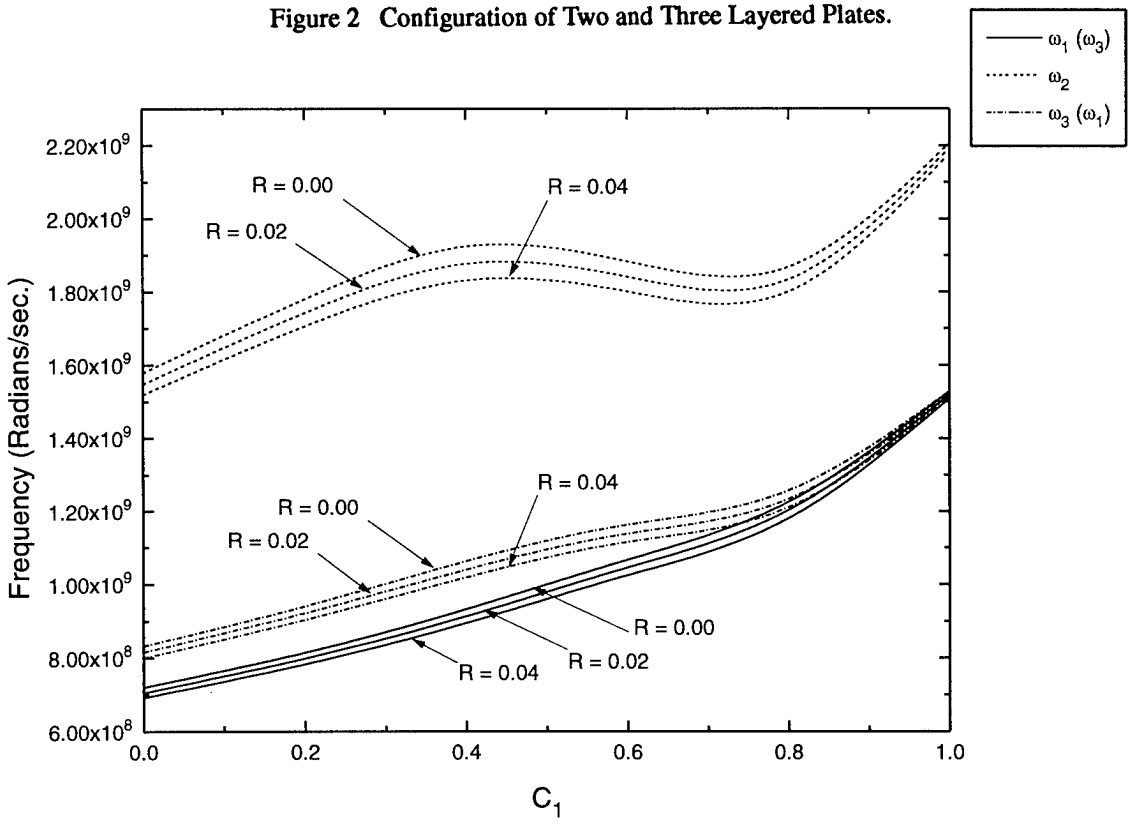


Figure 3 x_1 (x_3) Rotated ZnO on Si Bilayer. Variation of Cutoff Frequencies With C_1. Open Circuit

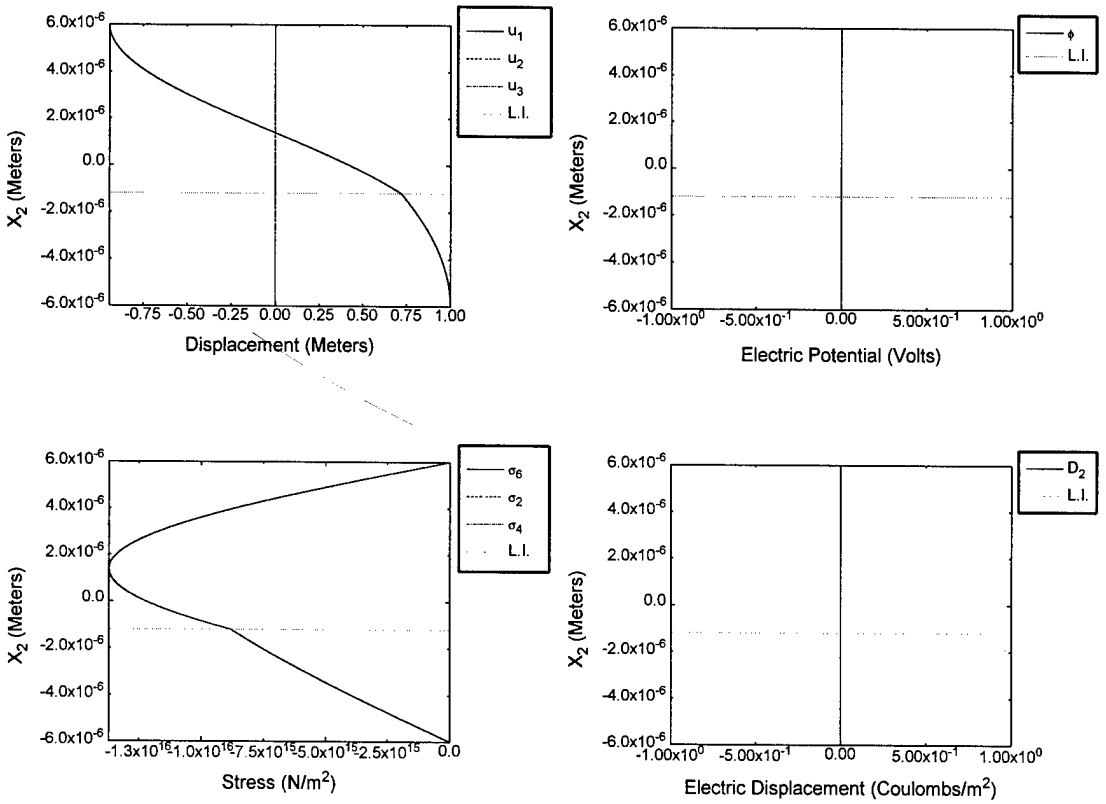


Figure 4 x_1 (x_3) Rotated ZnO on Si, Open Circuit, $C_1 = 0.4$. u_1 (u_3) Mode.

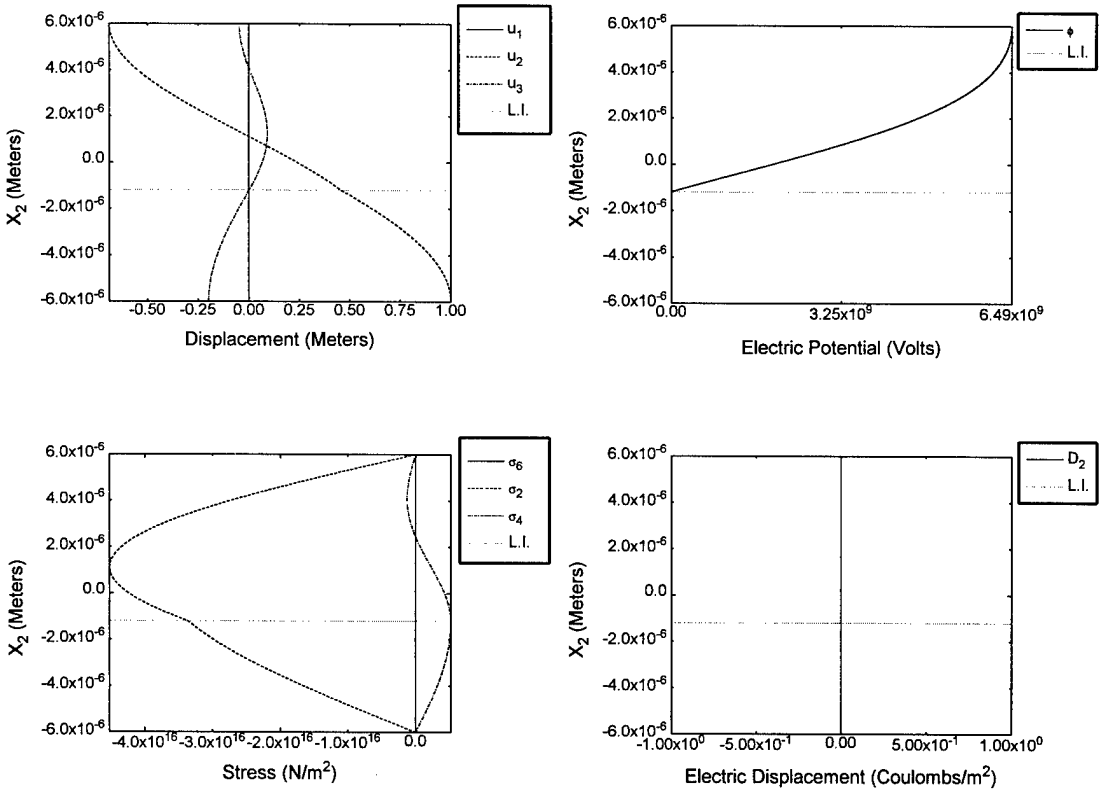


Figure 5 x_1 (x_3) Rotated ZnO on Si, Open Circuit, $C_1 = 0.4$. u_2 Mode.

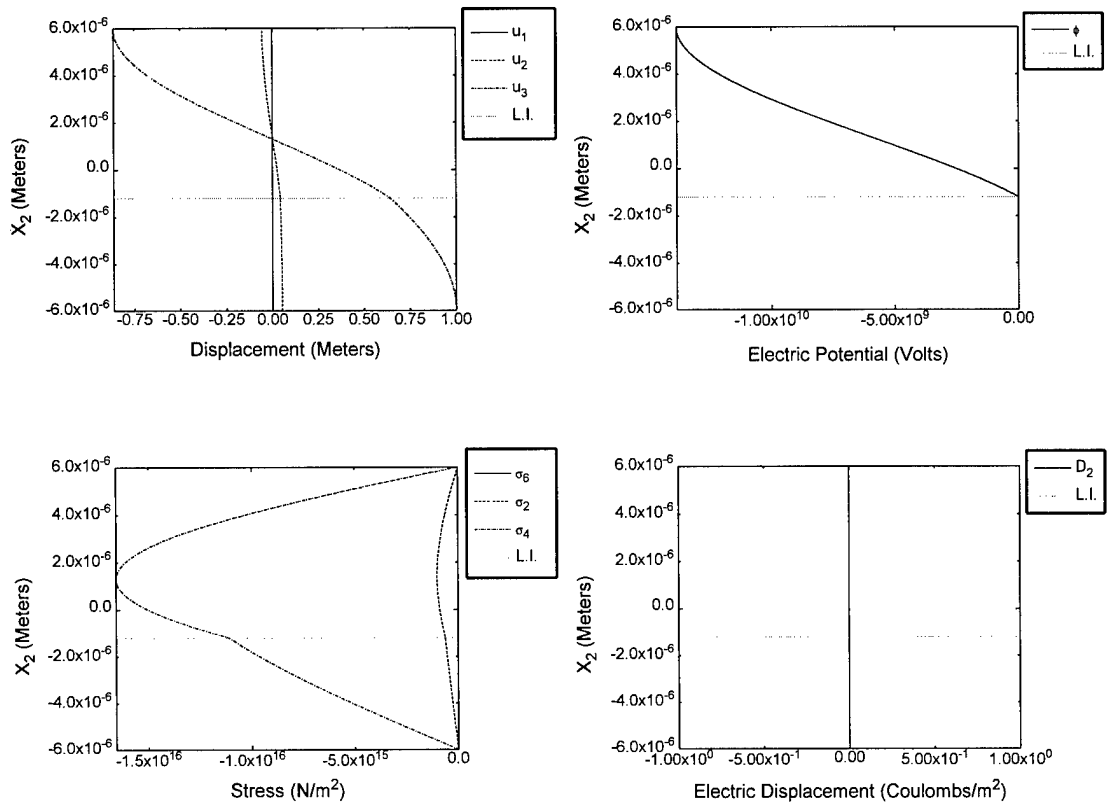


Figure 6 x_1 (x_3) Rotated ZnO on Si, Open Circuit, $C_1 = 0.4$. u_3 (u_1) Mode.

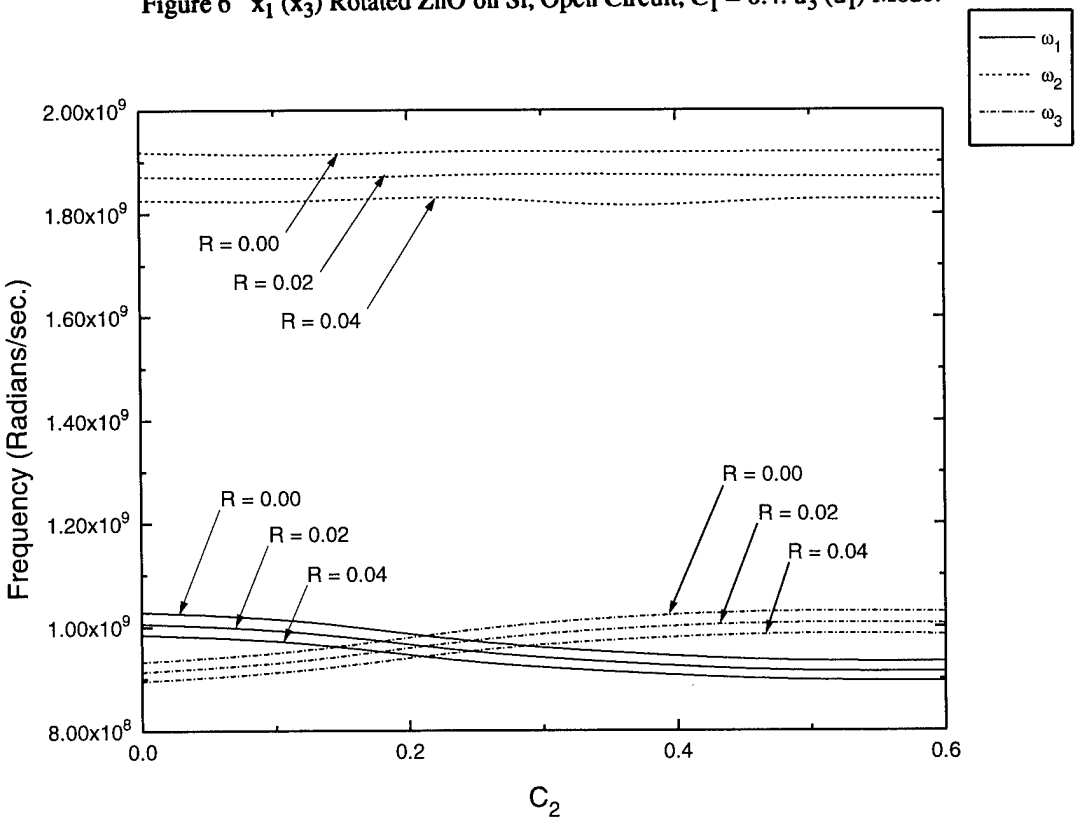


Figure 7 3-Layered Plate, $C_1 = 0.40$. Variation of Cutoff Frequencies With C_2, Open Circuit.

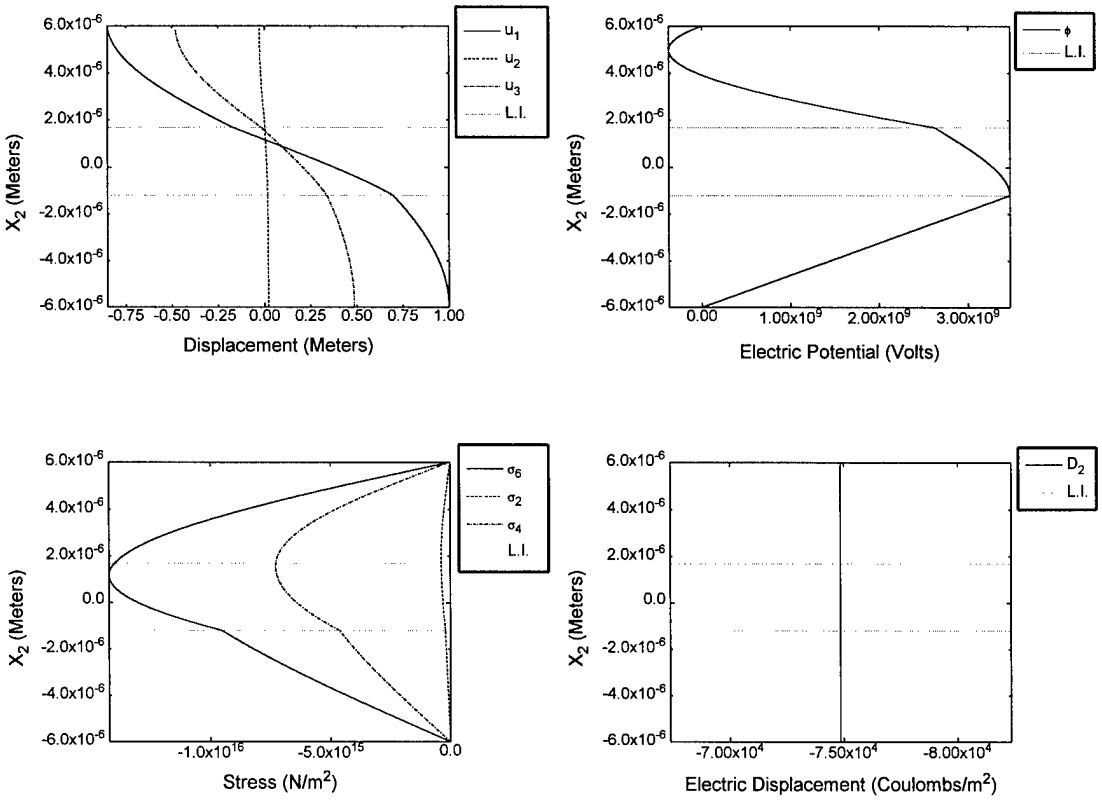


Figure 8 3-Layered Plate, $C_1 = 0.40$, $C_2 = 0.24$. Short Circuit, u_1 Mode.

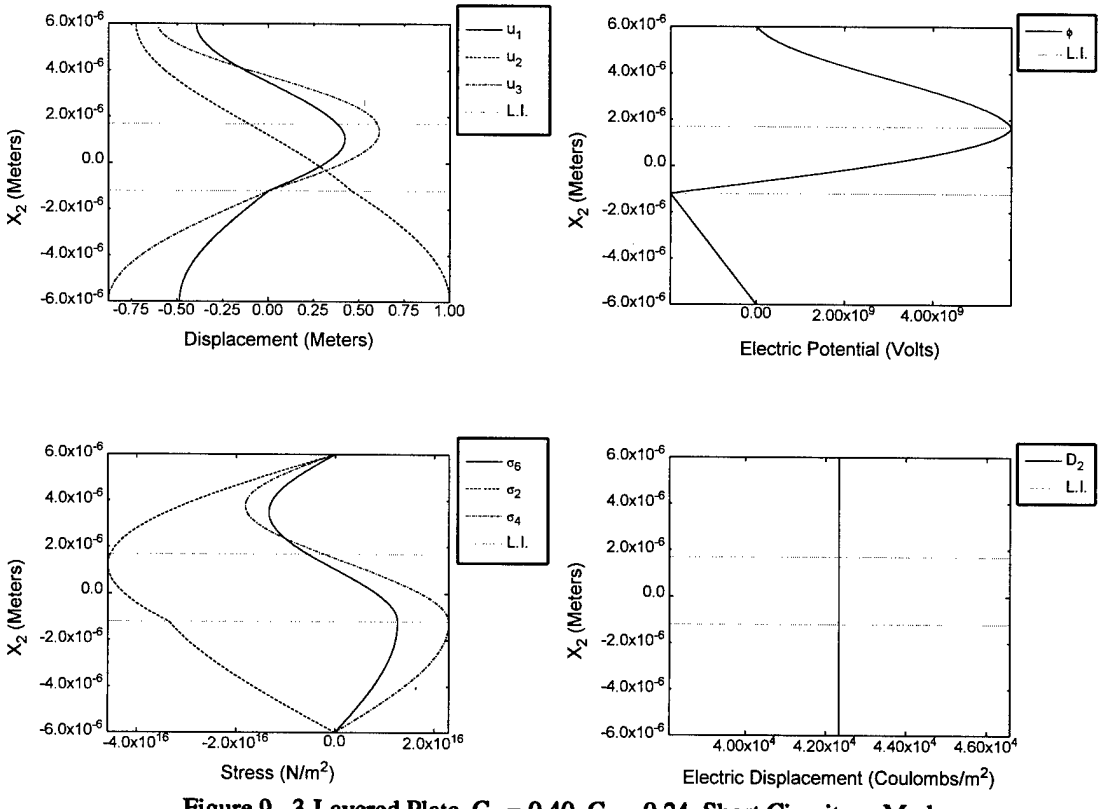


Figure 9 3-Layered Plate, $C_1 = 0.40$, $C_2 = 0.24$. Short Circuit, u_2 Mode.

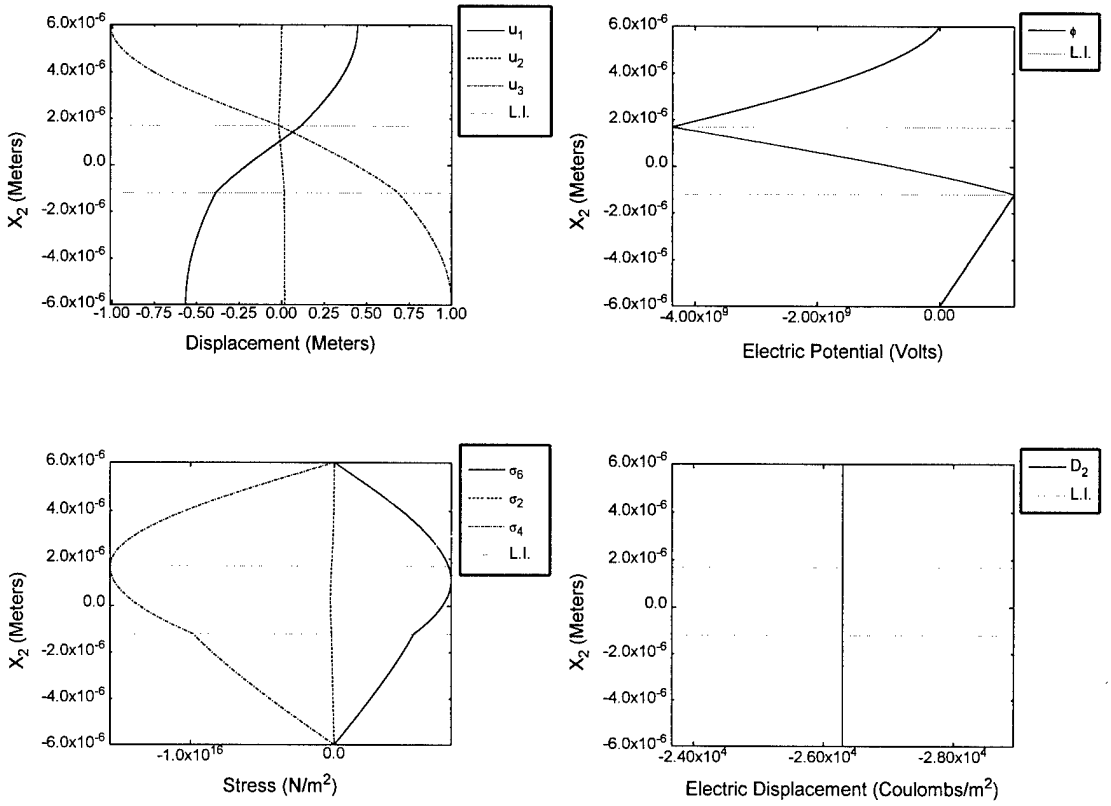


Figure 10 3-Layered Plate, $C_1 = 0.40$, $C_2 = 0.24$. Short Circuit, u_3 Mode.

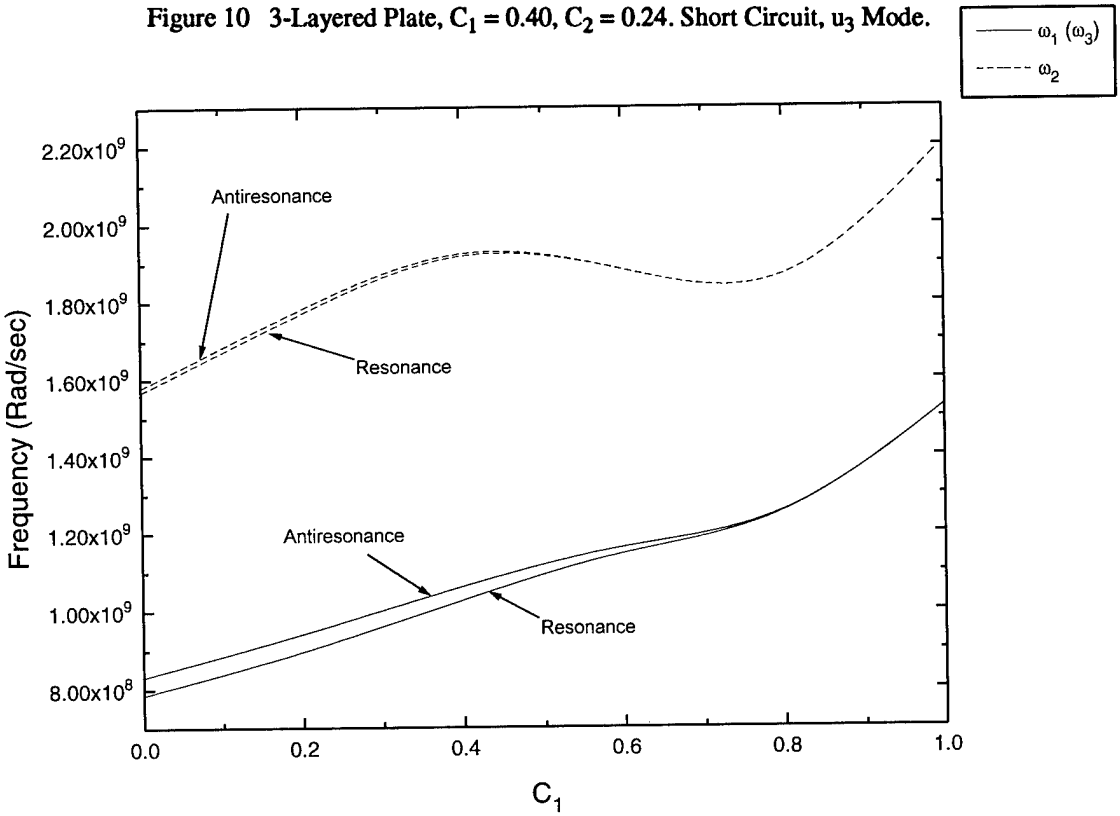


Figure 11 Resonant and Antiresonant Frequencies for x_1 (x_3) Rotated ZnO on Si Bilayer Under Thickness Field Excitation.

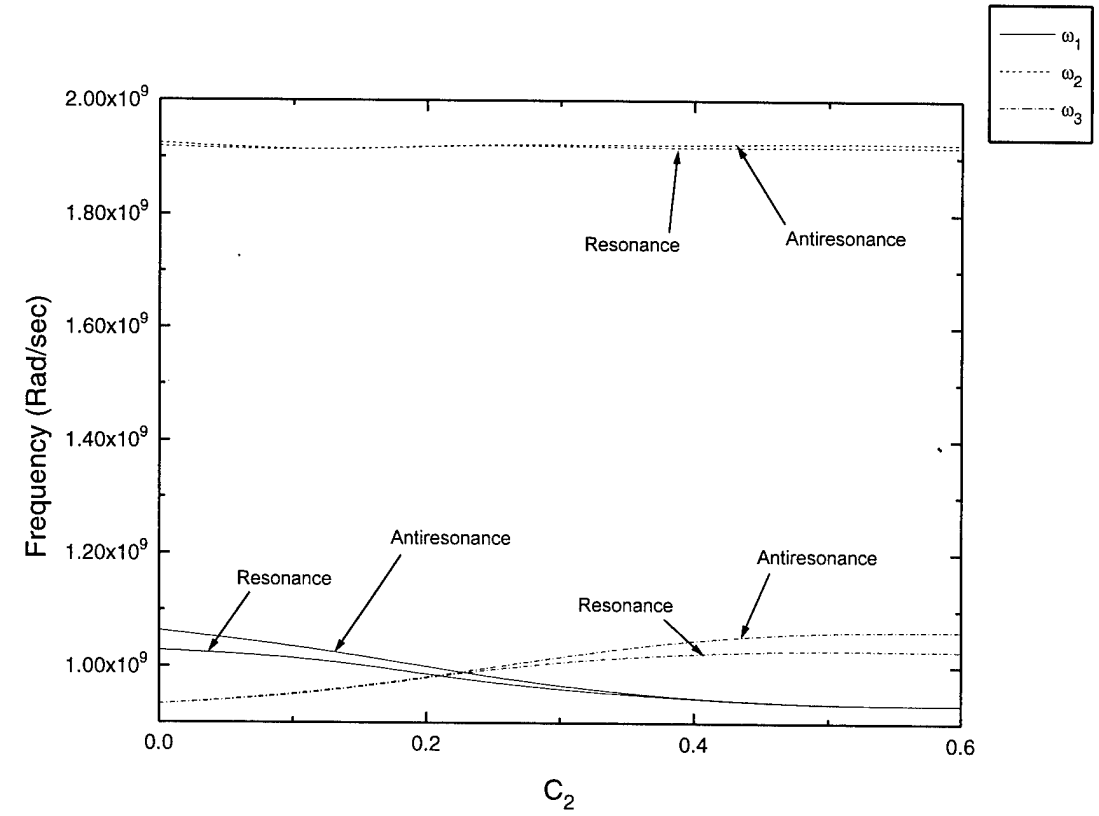


Figure 12 Resonant and Antiresonant Frequencies for 3-Layered Plate Under Thickness Field Excitation.

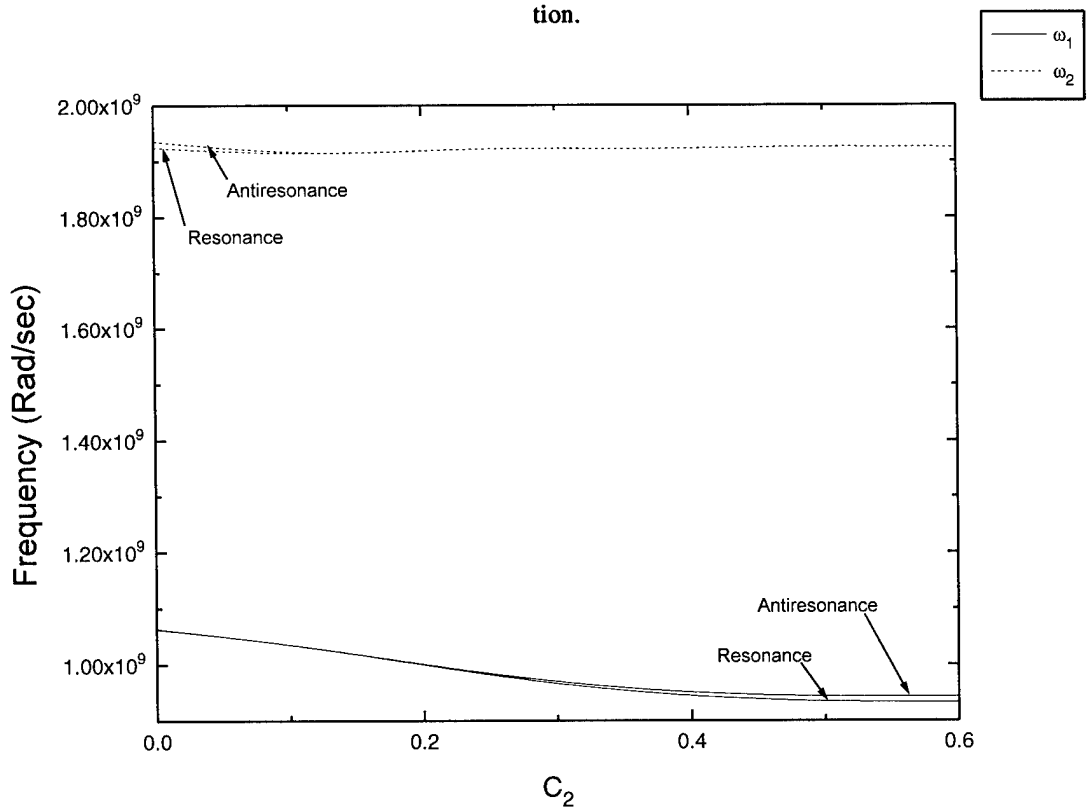


Figure 13 Resonant and Antiresonant Frequencies for 3-Layered Plate Under Lateral Field Excitation.

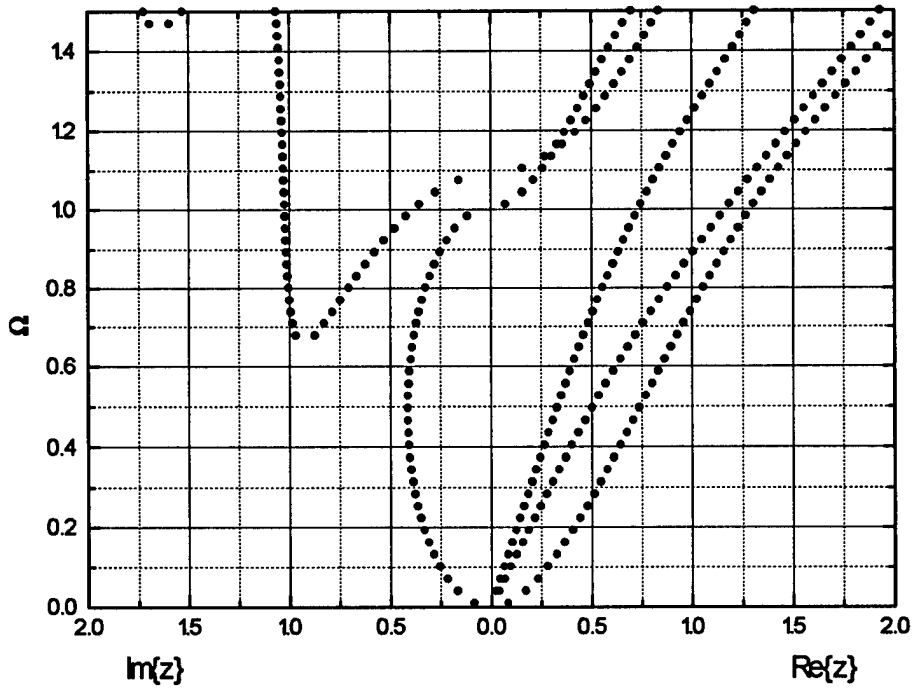


Figure 14 Dispersion relation for x_1 rotated Zno on Si Bilayer. Open Circuit, $C_1 = 0.6$

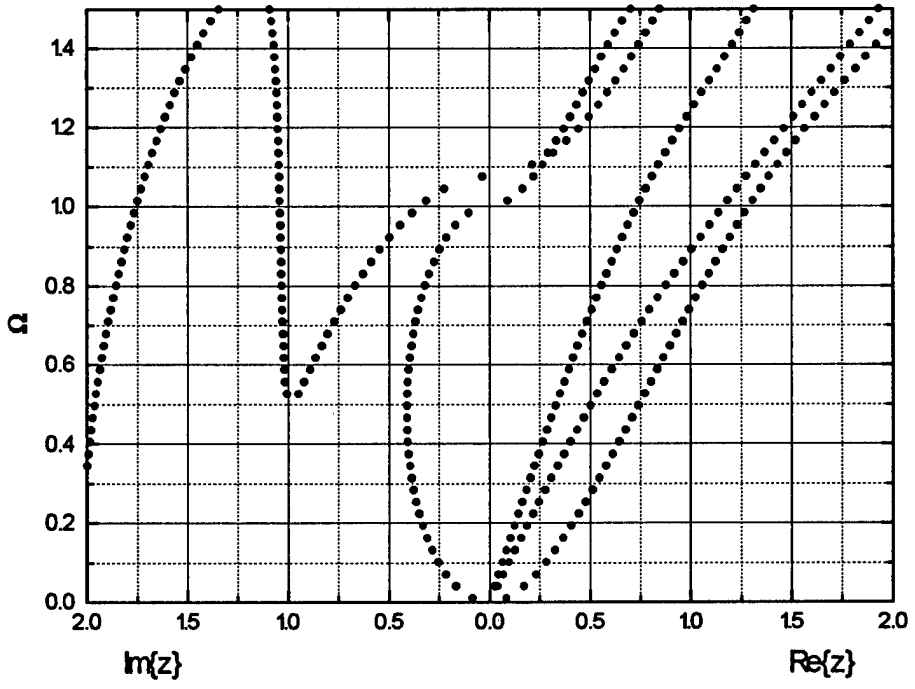


Figure 15 Dispersion relation for x_1 rotated Zno on Si Bilayer. Short Circuit, $C_1 = 0.6$

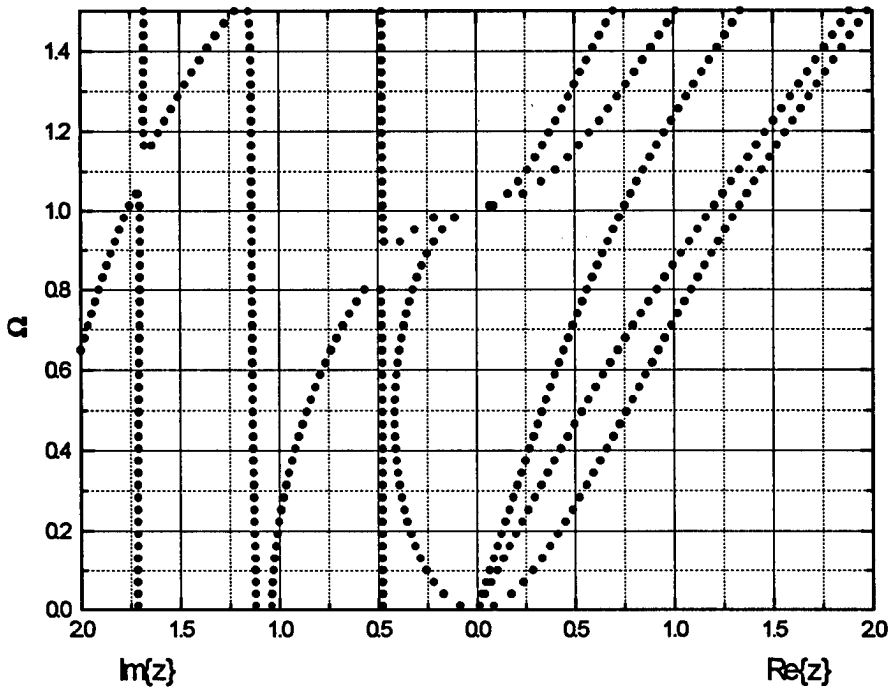


Figure 16 Dispersion relation for 3-Layered Plate. Open Circuit, $C_1 = 0.4$, $C_2 = 0.24$.

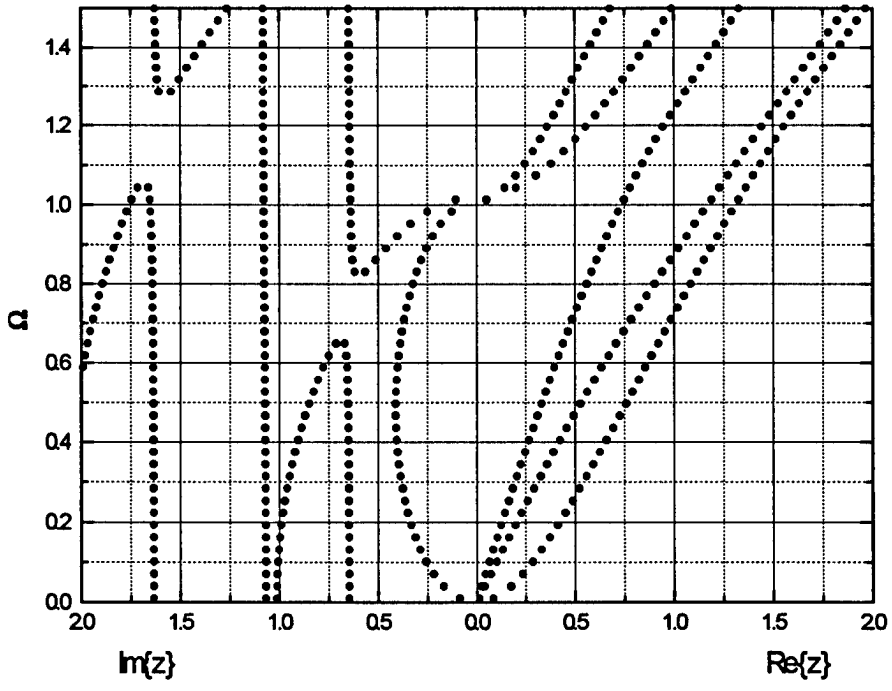


Figure 17 Dispersion relation for 3-Layered Plate. Short Circuit, $C_1 = 0.4$, $C_2 = 0.24$.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

NUMERICAL ANALYSIS OF TWO DIMENSIONAL THIN FILM RESONATORS

K.M. Lakin

TFR Technologies, Inc.
701 SE Salmon
Redmond, OR 97756

ABSTRACT

In order to gain a better understanding of thin film resonators fabricated at microwave frequencies a two dimensional numerical analysis formalism was derived and implemented. The formalism allows for arbitrary crystal anisotropy, rectangular material discontinuities, and two dimensional variations in the thickness and width dimensions. Resonators are assumed to be of infinite extent in one dimension. The formalism has been extended to include additional electrodes for two-port devices such as monolithic filters. Computed results are given for aluminum nitride resonators and filters having finite thickness metallization.

I INTRODUCTION

Thin film resonators fabricated at microwave frequencies are substantially different than quartz crystal resonators used at low frequencies. The microwave devices are different in that the metal electrodes can be a substantial fraction of the total thickness and of a large lateral extent compared to the resonator thickness. The large area results in part from a requirement that filters synthesized with thin film resonators may be designed to operate in 50 ohm impedance systems [1-2]. The area of a resonator in a filter is scaled by the source and load impedance. This feature is true of both ladder and monolithic filters. In contrast, low frequency resonators are usually designed to achieve a particular characteristic resonance such as the absence of spurs, favorable temperature coefficient, or reduced finite strain effects that determine resonator dimensions independent of source or load impedance.

Fabrication of microwave frequency resonators can give rise to step height discontinuities and edge boundaries that can significantly generate or cause plate waves to reflect. Significant plate waves can give rise to undesirable responses in finite sized resonators and filters.

Whereas extensive analytical techniques [3-6] have been applied to problems in quartz crystals of relatively simple geometry it was felt that these techniques lacked the necessary flexibility to handle potentially complex geometry and possible material inhomogeniety associated with thin film microwave acoustic resonators.

Numerical techniques have involved finite element [7,8] and finite difference methods [9]. The latter approach is used here principally because it directly solves the differential equations, places the method on a recognizable physical basis, and yields detailed solutions at each point within the device structure.

The finite difference method involves modeling the material region as a finite mesh of cells, or sub-regions, over which the equations of elasticity are applied, Fig. 1. The size of the mesh is implicitly defined as having regions sufficiently small that the mechanical displacements and electric potential can be approximated accurately by a second order function over a cell thus allowing derivatives in the equilibrium equations to be described by simple difference equations.

The mesh layout and the equilibrium region must be chosen with some care. It is computationally convenient to have all mesh points of the same type and each mesh point must generate a set of equations equal to the number of unknowns at that point. Placing mesh points on cell boundaries assures that displacements are continuous across the boundary as required. Of many possible layouts, the one shown, with mesh points at the corners and central to an equilibrium region of half the cell dimensions, seems the most useful. As will be described later, the positioning of the equilibrium region

0-7803-0905-7/93 $3.00 © 1993 IEEE

is determined by properties of derivatives in the second order finite difference approximation.

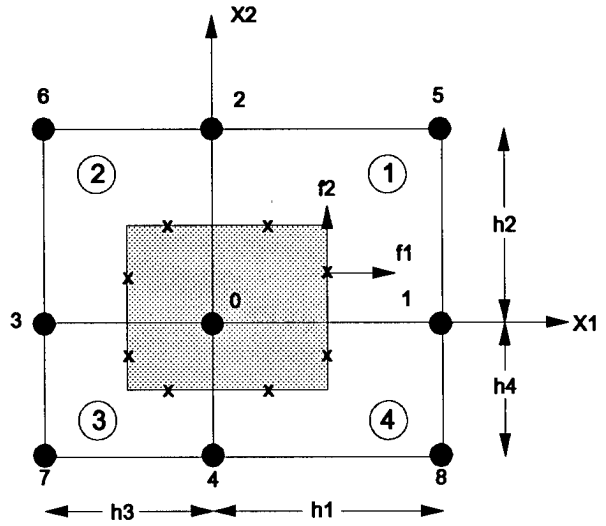


Fig 1. Mesh representation of a material region. Mesh points around the equilibrium region are labeled in sequence from central point through nearest neighbors to next nearest neighbors. Four possibly distinct material regions are contained within the nine point or four mesh area. The equilibrium region (shaded) is where surface traction forces are matched to the inertial force on the central mesh point material mass.

The overall nature of the numerical problem is illustrated in Fig. 2. Each mesh point generates up to four equations involving four variables. Crystal symmetry or other conditions can reduce the number of variables and equations as required. If the equations are "written" in a matrix form, starting from the upper left mesh point and proceeding across the shortest material dimension, the result is set of coupled equations in a sparse matrix format. The ability to store and hence numerically invert the matrix depends on the overall size of the non zero coefficients in the matrix. As illustrated in Fig. 2, the numerical bandwidth (NBW) is determined by the position of the upper left variable and last of the lower right variables in the set of local mesh equations generated by the 9 local mesh points (shaded). It is important that the formalism is efficient in generating as small NBW as possible in order to keep the store matrix as small as possible. The nine point mesh system is a good compromise between small NBW and adequate approximations for the derivatives in the finite difference equations.

In simple symmetry systems where the in-plane derivatives, (f_2 in Fig. 1) are not required the nine point system can be replaced by a five point system with somewhat lower NBW and slightly decreased accuracy. For higher levels of approximation, using more points, the NBW is significantly increased.

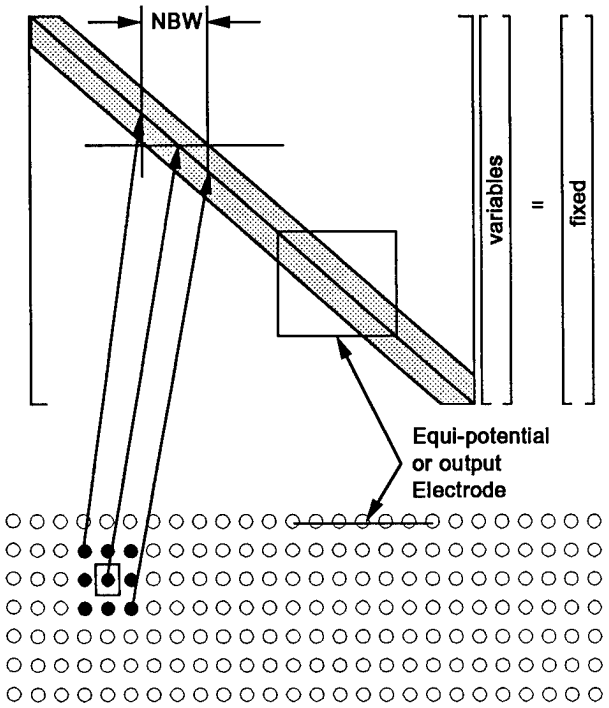


Fig. 2 Schematic of the linear equations generated by the mesh equations and a portion of the mesh. The solid dots in the mesh represent those active in generating equations and coefficients for the region of the matrix between the arrows. The box indicates an expansion of the sparse matrix dimensions caused by electrodes of unfixed potential such as floating electrodes or output electrodes in monolithic filters.

The NBW is given approximately by $NBW=2RN_v$ where R is the number of mesh rows and Nv is the number of variables per mesh point. The number of rows in the overall matrix is given by RCNv where C is the number of mesh columns. Therefore, the matrix store requirement goes as the square of the number of mesh rows.

Consequently, it is advantageous from a computational standpoint to keep the number of mesh rows at a minimum. The number of mesh points is determined by

the degree of approximation desired for the waveform being sampled. Good results have been obtained with 15 mesh points per half wavelength. The work reported here has been restricted to fundamental mode thickness resonances in order to keep the number of mesh rows to 25 or less depending upon the mode complexity.

Output or other electrodes whose potential is not fixed may require somewhat wider numerical bandwidth at local regions in the matrix as suggested by the box in Fig. 2. At a minimum a floating electrode requires that the total current sums to zero and an output electrode requires that the output current times the load impedance equals the potential of the electrode.

With the above discussion as background the following section discusses the detailed theory to be followed by computed cases as examples.

II THEORY

The detailed theory of the finite difference method used in this work starts from the point form equations for linear elasticity consisting of Newton's second law and the conservation of charge relation;

$$\frac{\partial T_{ij}}{\partial x_i} = -\rho\omega^2 u_j \qquad \frac{\partial D_i}{\partial x_i} = 0 \qquad (1)$$

Gausss' Law is applied to (1), on the small equilibrium volume in Fig. 1, to analytically reduce the order of integration by one rather than using second order finite difference equations. This approach leads to a clearer physical understanding of the process and the limits of the parabolic approximation. The result for the volume element is,

$$\int \frac{\partial T_{ij}}{\partial x_i} dv = \int T_{ij} dA_i = -\omega^2 \int \rho u_j dv \qquad (2)$$

In the limit of small volume elements the area integral in (2) is replaced by a summation and the inertial term is integrated assuming the displacement can be averaged over the volume element. The summation becomes a sum of traction forces acting on the faces of the volume element. Because the equilibrium cell has four distinct material regions it is necessary to sum the traction forces on eight surfaces defined by unit vectors in the four material regions. The forces take the form,

$$F_{1j} = T_{1j}\Delta A_1 \quad \text{and} \quad F_{2j} = T_{2j}\Delta A_2 \qquad (3)$$

where the j subscript denotes the three traction forces acting on each surface element and the left subscript index refers to the area element unit-normal index. Using the material constitutive relations in (3),

$$T_{ij} = c^E_{ijkl}\frac{\partial u_k}{\partial x_l} + e_{ijk}\frac{\partial \varphi}{\partial x_k} \qquad D_i = e_{ikl}\frac{\partial u_k}{\partial x_l} - \varepsilon^S_{ik}\frac{\partial \varphi}{\partial x_k}$$

the generalized forces become,

$$F_{1j} = \left(M_{1jk1}\frac{\Delta u_k}{\Delta x1} + N_{1jk2}\frac{\Delta u_k}{\Delta x_2}\right)\Delta A_1$$

$$F_{2j} = \left(M_{2jk2}\frac{\Delta u_k}{\Delta x_2} + N_{2jk1}\frac{\Delta u_k}{\Delta x_1}\right)\Delta A_2$$

where

$$M_{ijkl} = c^E_{ijkl} \qquad N_{ijkl} = c^E_{ijkl} \qquad j,k = 1,2,3$$

$$M_{ij4l} = e_{ijl} \qquad N_{ij4l} = e_{ijl} \qquad j = 1,2,3$$

$$M_{i4kl} = e_{ikl} \qquad N_{i4kl} = e_{ikl} \qquad k = 1,2,3$$

$$M_{i44l} = -\varepsilon^S_{il} \qquad N_{i44l} = -\varepsilon^S_{il}$$

Dropping the j and k subscripts the force equations simplify to;

$$F_1 = (M_1\frac{\Delta u}{\Delta x_1} + N_1\frac{\Delta u}{\Delta x_2})\Delta A_1$$
$$F_2 = (M_2\frac{\Delta u}{\Delta x_2} + N_2\frac{\Delta u}{\Delta x_1})\Delta A_2$$

The area elements ΔA are half the cell dimension (e.g. h2/2) times the depth dimension. The depth dimension is not carried in the equations since it will divide out of both sides of the equilibrium equation (2).

Within the parabolic approximation, derivatives formed by differences of adjacent points are valid at the midpoint between the two points. Further, derivatives perpendicular to a face vary linearly along the face. For example, in Fig. 1, $\Delta u/\Delta x_1 = (u^1 - u^0)/h1$ at $(h_1/2, 0)$ but $\Delta u/\Delta x_1 = (u^5 - u^2)/h1$ at $x_2 = h2$ so that a weighted average is used at $x_2 = h2/4$ at the midpoint on the area element. The superscripts on the displacements refer to the cell index and the subscripts have been dropped.

The in-plane derivative, f_2 in Fig. 1, cannot be determined at the area element midpoint without using mesh points beyond those already shown. Consequently the in-plane derivative must be evaluated at $x_2 = h2/2$ corresponding to the corner of the equilibrium cell.

The net force on the x_1 faces is given by,

$$F_1 = \left(F_1^1 + F_1^4\right) - \left(F_1^2 + F_1^3\right)$$

where

$$F_1^1 = \tfrac{1}{8}M_1^1 h21(u^5 - u^2 + 3u^1 - 3u^0)$$
$$+\tfrac{1}{4}N_1^1(u^5 - u^1 + u^2 - u^0)$$

$$F_1^4 = \tfrac{1}{8}M_1^4 h41(u^8 - u^4 + 3u^1 - 3u^0)$$
$$+\tfrac{1}{4}N_1^4(u^1 - u^8 + u^0 - u^4)$$

$$F_1^2 = \tfrac{1}{8}M_1^2 h23(u^2 - u^6 + 3u^0 - 3u^3)$$
$$+\tfrac{1}{4}N_1^2(u^2 - u^0 + u^6 - u^3)$$

$$F_1^3 = \tfrac{1}{8}M_1^3 h43(u^4 - u^7 + 3u^0 - 3u^3)$$
$$+\tfrac{1}{4}N_1^3(u^0 - u^4 + u^3 - u^7)$$

and the net force on the x_2 faces is given by;

$$F_2 = \left(F_2^1 + F_2^2\right) - \left(F_2^3 + F_2^4\right)$$

where

$$F_2^1 = \tfrac{1}{8}M_2^1 h12(u^5 - u^1 + 3u^2 - 3u^0)$$
$$+\tfrac{1}{4}N_2^1(u^5 - u^2 + u^1 - u^0)$$

$$F_2^2 = \tfrac{1}{8}M_2^2 h32(u^6 - u^3 + 3u^2 - 3u^0)$$
$$+\tfrac{1}{4}N_2^2(u^2 - u^6 + u^0 - u^3)$$

$$F_2^3 = \tfrac{1}{8}M_2^3 h34(u^3 - u^7 + 3u^0 - 3u^4)$$
$$+\tfrac{1}{4}N_2^3(u^4 - u^7 + u^0 - u^3)$$

$$F_2^4 = \tfrac{1}{8}M_2^4 h14(u^1 - u^8 + 3u^0 - 3u^4)$$
$$+\tfrac{1}{4}N_2^4(u^8 - u^4 + u^1 - u^0)$$

The superscripts on forces and material constants refer to material regions and those on displacements refer to mesh points as labeled in Fig. 1.

From (3) the total traction forces can now be equated to the inertial term to obtain the equilibrium equation;

$$F_T = F_1 + F_2 = -\bar{\rho}\omega^2 u^0$$

where

$$\bar{\rho} = \tfrac{1}{4}(\rho_1 h1h2 + \rho_2 h2h3 + \rho_3 h3h4 + \rho_4 h4h1)$$

By collecting all the terms associated with each mesh point a set of linear equations can be obtained describing the equilibrium at the center mesh point;

$$(\alpha_{jk}^0 - \delta_{jk}\omega^2\bar{\rho})u_k^0 + \sum_{n=1}^{8}\alpha_{jk}^n u_k^n = 0$$

where

$$\alpha^1 = \tfrac{1}{8}(-3M_1^1 h21 - 3M_1^4 h41 + M_2^1 h12 + M_2^4 h14)$$
$$+\tfrac{1}{4}(N_1^1 - N_1^4 - N_2^1 + N_2^4)$$

$$\alpha^2 = \tfrac{1}{8}(M_1^1 h21 + M_1^2 h23 - 3M_2^1 h12 - 3M_2^2 h32)$$
$$+\tfrac{1}{4}(-N_1^1 + N_1^2 + N_2^1 - N_2^2)$$

$$\alpha^3 = \tfrac{1}{8}(-3M_1^2 h23 - 3M_1^3 h43 + M_2^2 h32 + M_2^3 h34)$$
$$+\tfrac{1}{4}(-N_1^2 + N_1^3 + N_2^2 - N_2^3)$$

$$\alpha^4 = \tfrac{1}{8}(M_1^4 h41 + M_1^3 h43 - 3M_2^3 h34 - 3M_2^4 h14)$$
$$+\tfrac{1}{4}(N_1^4 - N_1^3 + N_2^3 - N_2^4)$$

$$\alpha^5 = -\tfrac{1}{8}(M_1^1 h21 + M_2^1 h12) - \tfrac{1}{4}(N_1^1 + N_2^1)$$

$$\alpha^6 = -\tfrac{1}{8}(M_1^2 h23 + M_2^2 h32) + \tfrac{1}{4}(N_1^2 + N_2^2)$$

$$\alpha^7 = -\tfrac{1}{8}(M_1^3 h43 + M_2^3 h34) - \tfrac{1}{4}(N_1^3 + N_2^3)$$

$$\alpha^8 = -\tfrac{1}{8}(M_1^4 h41 + M_2^4 h14) + \tfrac{1}{4}(N_1^4 + N_2^4)$$

The α are 4x4 matrices, u are three components of mechanical displacement plus the electric potential , and the sum over k is implied. The Kronecker delta is 1 for the first three diagonal elements and zero elsewhere. If the sum over n is moved to the right hand side (4) can be viewed as the case of a central mesh point driven by its neighbors.

In cases where the potential is fixed externally, the fourth row is dropped and the fourth column is moved to the right hand side making the overall set of equations inhomogeneous (for non-zero potentials). For floating or output electrodes the displacement current relation at each cell is changed and the fourth row must be replaced by an appropriate current equation.

III EXAMPLE SOLUTIONS

Three cases were studied as examples of the solution technique. The first was a 16 x 651 mesh having an aluminum nitride piezoelectric 12 units thick with aluminum electrodes 2 units thick. The top electrode extended from column 50 to 650 and the bottom electrode extended across the entire bottom of the plate. The edges of the plate were abruptly terminated by vacuum interfaces in order to allow plate wave reflections and transverse standing waves.

The matrix formed by this case had a NBW of 101 and consisted of approximately 30000 equations. The solution procedure was checked for numerical round-off by perturbing the last variable's solution by 10 % and calculating the effect on the first variable. This resulted in a 1% change in the first variable's solution indicating that the matrix inversion was well within the numerical range of the computation.

Impedance was calculated as a function of frequency and is shown in Fig. 3. The magnitude and phase of impedance show the effects of plate wave resonances in the form of weak spurious resonances.

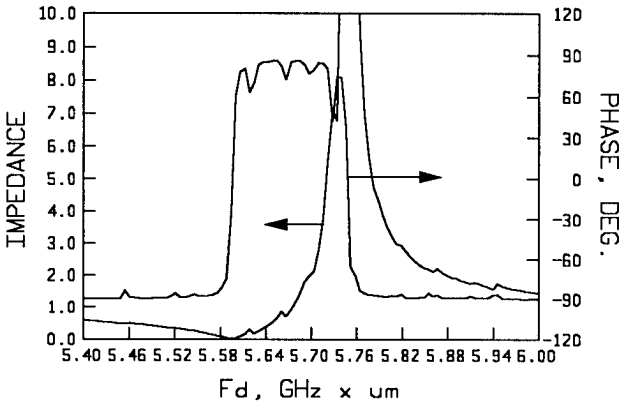


Fig. 3 Impedance plot of an AlN c-axis perpendicular resonator of width to thickness ratio approximately 40:1. The electrodes are 25% of the total plate thickness. Plate waves give rise to spurious resonances.

The vertical displacement u_2 along the top row of the piezoelectric plate, and just under the top electrode, is shown in Fig. 4 at the series resonant frequency. It has a large average value with a ripple due to resonance of plate waves in the transverse direction. The other component of displacement (u_1), not show in Fig. 4, is less than one tenth the peak value of (u_2) due to the relatively large electrode width to plate thickness ratio used in this example.

An expanded view of u_2 and voltage near the electrode edge is shown in Fig. 5. The displacement makes a smooth transition from the electrode to the unelectroded region. The potential is 1 volt at the electrode (fixed) and rises outside the electrode. Both solutions show that energy is not significantly trapped by the electrode region.

The potential variation in the depth direction, at the center of the top electrode, is shown in Fig. 6 again at the series resonant frequency. The result shows that the sampling density is adequate in the thickness dimension.

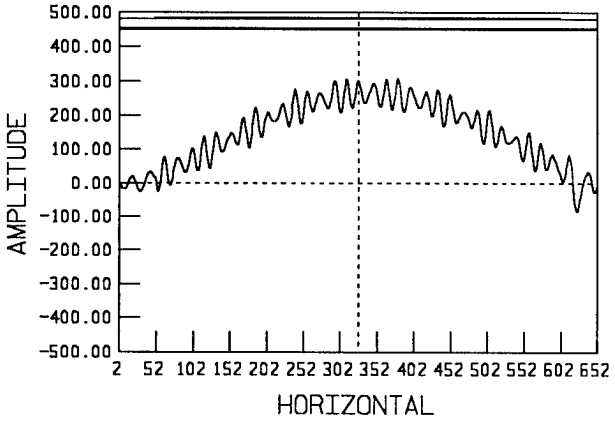


Fig. 4. Vertical displacement along the top row of the piezoelectric and under the top electrode. The ripple is due to plate waves.

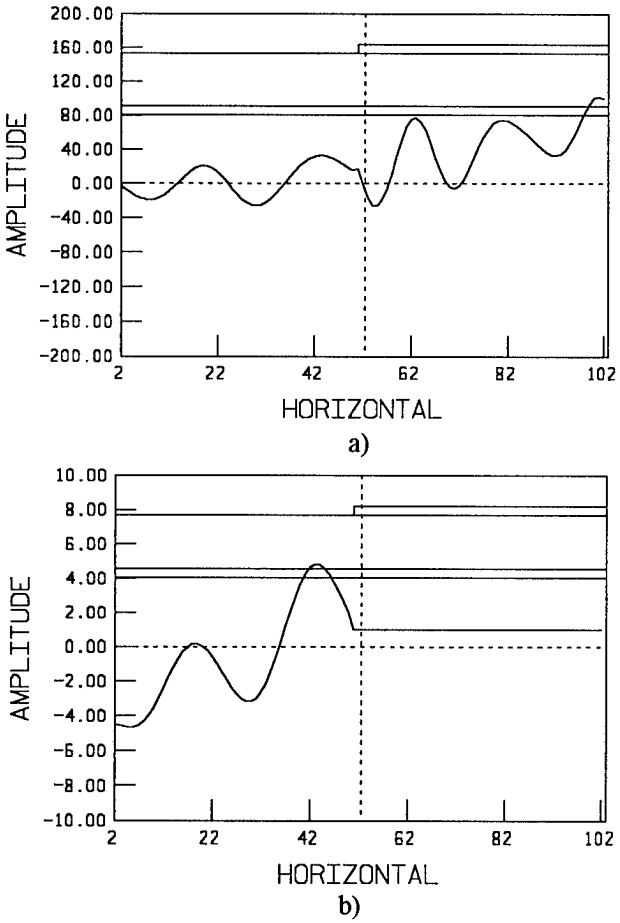


Fig. 5. Fields near the edge of the top electrode. a) Vertical displacement. b) Electrical potential. The fields show a smooth transition through the material discontinuity.

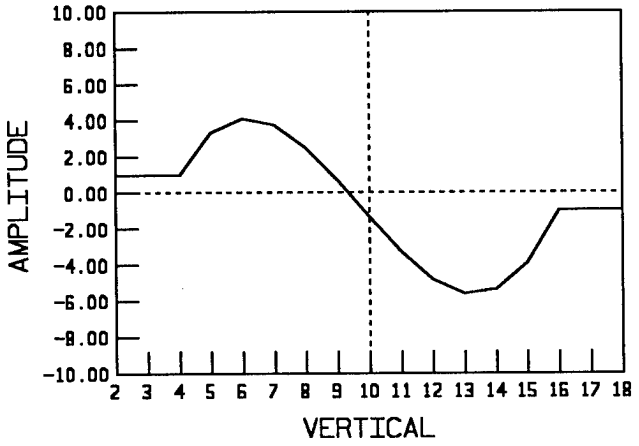
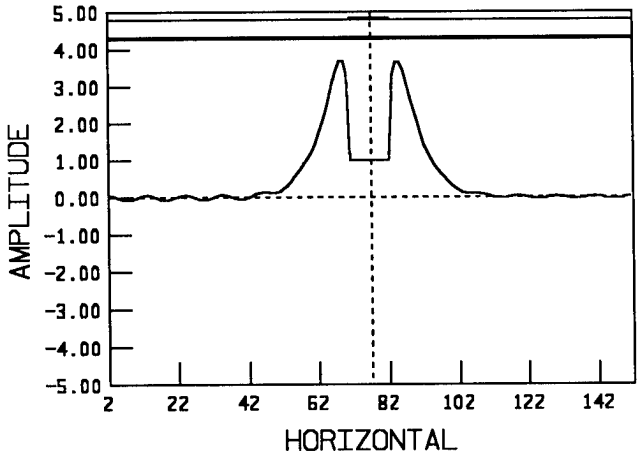
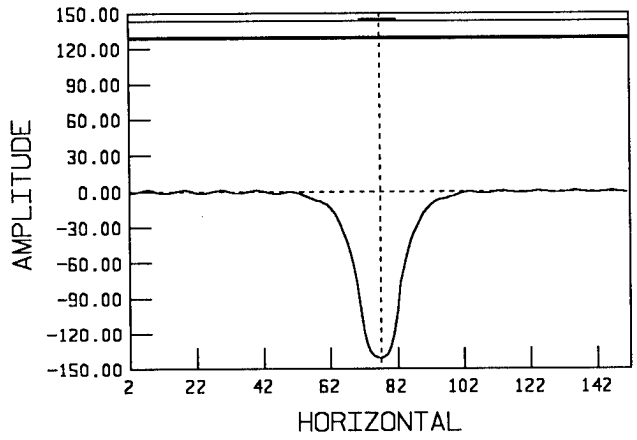


Fig. 6. Electric potential as a function of depth at series resonance.



a)



b)

Fig. 7. Field distributions near an electrode showing energy trapping. a) Electric potential, b) Electric potential as a function of depth at series resonance.

Displacement current perpendicular to the electrode is calculated using three point derivatives for improved accuracy.

Trapping effects are illustrated in Fig 7 for the case of aluminum nitride c-axis perpendicular to the page for the generation of pure shear wave displacement. The mesh size is 25 x 151. In Fig. 7b the displacement is shown well trapped by the electrode of two units thickness. The electric potential, shown in Fig. 7a, is well confined near the electrode region being fixed at the electrode to one volt.

The impedance of the structure of Fig. 7, shown in Fig. 8, is very clean without significant spurious responses. The fields in Fig. 7 were calculated at the series resonant frequency.

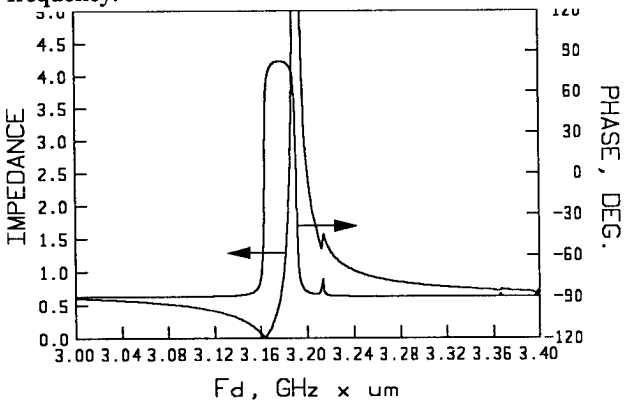


Fig. 8 Impedance of a trapped energy resonator.

The monolithic filter is a more complex structure having an input driving electrode, possible intermediate passive resonators, and an output electrode. The example used here is a four pole resonator having two grounded coupling resonators.

The displacement field in the filter near series resonance is shown in Fig. 9 along the top of the piezoelectric. The resonators are clearly coupled by an overlap of their otherwise trapped fields.

The transfer function and input impedance for the filter are shown in Fig. 10. The impedance shows the results of multiple resonator coupling and the transfer function illustrates the multi-pole response. The filter was not optimized for any particular pass-band response.

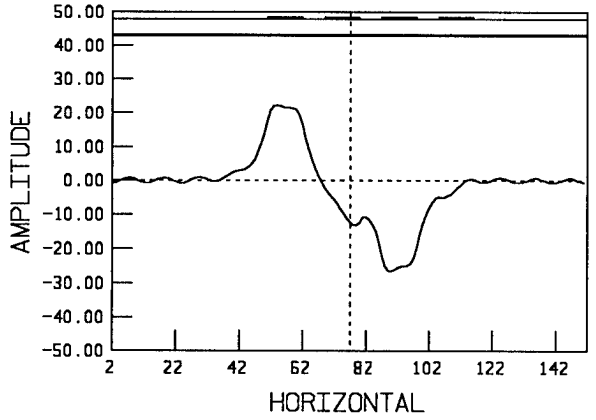
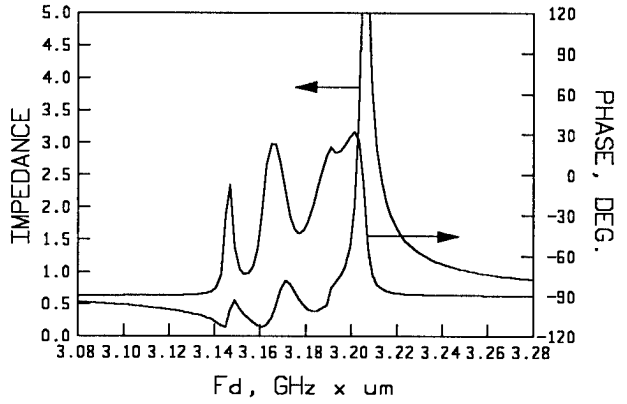
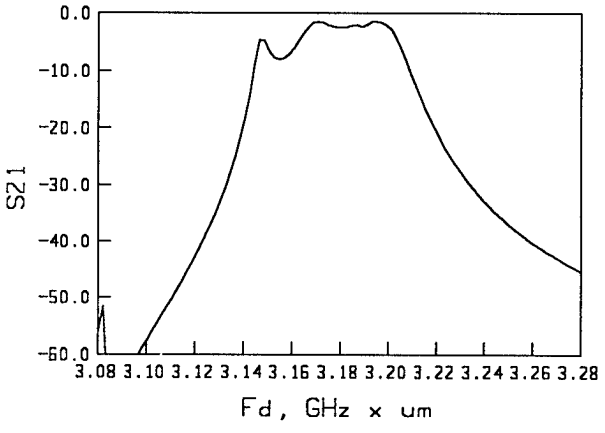


Fig. 9 Displacement distribution under electrodes in a monolithic filter.



a)



b)

Fig. 10 Frequency response of monolithic filter. a) Impedance showing effects of multi-resonator coupling. b) S21 transfer function.

IV SUMMARY AND ACKNOWLEDGMENT

A finite difference formulation was derived for the case of two dimensional resonators and applied to three examples. The formulation uses a nine point system to describe the equilibrium conditions for a central mesh point. The material regions can have arbitrary anisotropy and inhomogeniety to the cell level. A sampling density of 13 points per half wavelength was found to give adequate accuracy for resonator problems. The technique correctly predicts resonant frequencies, energy trapping, and two-port monolithic filter characteristics.

The support of the United States Government is acknowledged.

REFERENCES

1. C. Vale, J. Rosenbaum, S. Horwitz, S Krishnaswamy, and R. Moore, "FBAR Filters at GHz Frequencies", 45 th Annual Symp. of Freq. Cont. Proc. pp.332-336

2. K. Lakin, G.Kline and K. Mc Carron, "Thin Film Bulk Acoustic Filters for GPS", 1992 Ultrasonics Symp. Proc., pp471-476.

3. H. Tiersten and D. Stevens, "A Analysis of Thickness-extensional Trapped Energy Resonant Device Structures With Rectangular Electrodes In the Piezoelectric Thin Film On Silicon Configuration", J. Appl. Phys., Vol. 54, No. 10, Oct. 1983 pp 5893-5910

4. H. Tiersten and Y. Zhou, " An Analysis of Transversely Varying Thickness Modes in Quartz Resonators with Beveled Cylindrical Edges", 1993 IEEE Int. Freq. Cont. Symposium Proc.

5. J. Stewart and Y. Yong, "Exact Analysis of the Propagation of Acoustic Waves in Multilayered Piezoelectric Plates", 1993 IEEE Int. Freq. Cont. Symposium Proc.

6. P.C. Y. Lee and J. Wang, "Vibrations of AT-Cut Quartz Strips of Narrow Width and Finite Length", 1992 IEEE Freq. Cont. Symp. Proc, pp. 537-552

7. L. Clayton and E. Eernisse, "Four-point SC-cut Crystal Unit Design Calculations for Reduced Acceleration Sensitivity", 1992 IEEE Freq. Cont. Symp. Proc, pp. 582-596

8. R Lerch, "Simulation of Piezoelectric Devices by Two- and Three Dimensional Finite Elements", IEEE Trans. Ultrason., Ferro., and Freq. Cont. , Vol. 37, No. 3, p. 233-247

9. P. Lloyd and M. Redwood, J. Acoust. Soc. Am., Vol. 39, pp. 346-361.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

SELF-SUSTAINED ACOUSTOELECTRIC AND PHOTOELECTRIC OSCILLATIONS IN SYSTEMS WITH RELAXATION

Vladimir A.Vyun

Institute of Semiconductor Physics, Rus. Acad. of Sci.
Academic Lavrentjev Avenue 13, Novosibirsk 630090, RUSSIA

Investigation of new acoustoelectric and photoelectric self-sustained oscillation and self-modulation phenomena are described in systems with relaxation. Layered metal-semiconductor-dielectric or piezoelectric structures with external feedback are theoretically described and experimented. The theory presented explains the experimental results and gives the criterion for existence of the phenomena. Quasi-harmonic regime of the self-sustained oscillations is described by the Van-der-Pol equation, which can be very useful for practical applications because of it expresses relaxation properties of the structure in terms of the oscillation frequency. Fundamental possible applications of the self-sustained oscillations to new physical and chemical sensors and methods of investigation of semiconductors are suggested. Dynamic characteristics of the self- sustained oscillations under different external conditions, as temperature and illumination, are discussed.

Introduction

It is well known that acoustoelectric and photoelectric voltages arise in layered structures with a semiconductor if a propagating surface acoustic wave (SAW) and incident light interact with the semiconductor surface, respectively.

When a semiconductor sample is placed in the proximity of the surface of a piezoelectric crystal, the radio frequency (RF) electric field associated with the SAW penetrates the semiconductor and interacts with the carriers near the semiconductor surface. Nonlinearity of the acoustoelectric interaction is the source of the acoustoelectric voltage in a layered piezoelectric-semiconductor structure [1], [2]. The different types of the nonlinear acoustoelectric voltages, such as convolution and memory signals, longitudinal and transverse acoustoelectric voltages, have applications for the processing of radio and optical signals [3]-[7]. Another very interesting application of the nonlinear acoustoelectric phenomena is modeling processes with nonlinear waves [8]-[10] and nondestructive study of electrical properties of semiconductors [2], [10]-[18]. For the latter, the transverse acoustoelectric voltage (TAV) is the most sensitive to detection and of great interest.

A change in the semiconductor electron concentration under illumination of a layered metal-dielectric-semiconductor structure, is the cause of the photoelectric voltage (PEV) [19]-[21]. The PEV is applicable for photosensitive devices and in the investigation of semiconductors.

Inspite of the facts that signs of the TAV and PEV sometimes differ, shapes of the TAV and PEV are similar [22]. In the layered structures the main feature of the TAV and PEV is that they have a transient behavior after the SAW and incident light have been switched on or off. This is due to the fact that after the switching on or off of the SAW and illumination power in such a system in the semiconductor, there are electron transient processes, such as charge carrier generation, recombination, and trapping, that cause a relaxation of the TAV and PEV to their stationary values. Experimental examples of the signal waveforms of the TAV and PEV are shown in Fig. 1 and Fig. 2, obtained under impulse mode of the SAW and illumination, respectively. In Fig. 1 and Fig. 2 one can see the transient behavior of the TAV and the PEV after the beginning and end of the SAW and illumination impulses.

Another somewhat significant outcome of the transient electronic processes is bistability and multistability of acoustoelectric phenomena in layered piezoelectric-semiconductor structures, that were reviewed in [23] and in the references of this paper. In the layered structures optical bistability is also well known [24]. It is typical of the bistability and multistability that the structure can be in several static states when a fixed static external action is applied and maintained.

The structure can, on principle, be in an oscillation state when such a fixed static external action is applied and maintained. The self-sustained oscillations of the TAV were noted in [25] but have not been sufficiently studied.

The present paper presents some experimental and theoretical results on self-sustained oscillations of the TAV and PEV in the layered structure with external feedback and demonstrates their fundamental application for sensors and investigation of physical properties of semiconductors.

Experiment

Self-sustained oscillations acoustoelectric systems

We consider a layered semiconductor-piezoelectric structure of YZ-$LiNbO_3$ and n-Si with external feedback, as shown in Fig. 3. The gap between the piezoelectric and

0-7803-0905-7/93 $3.00 © 1993 IEEE

semiconductor is approximately equal to 0.1-0.2 μm. The resistivity of the semiconductor is equal to 200 Ω·cm. The frequency excitation of the SAW is equal to 110 MHz.

In this structure the surface acoustic waves, after excitation by an interdigital transducer with applied radio frequency (RF) voltage V_{in} at the frequency 110 MHz, propagate and interact with the adjacent semiconductor. The transverse acoustoelectric voltage V_{ae} in the form of a low frequency signal or less than high frequency impulse is detected by metal electrodes. The RF voltage V_o at the scheme is modulated by the amplitude modulator. Thus, the voltages V_{in} and V_o are connected by the relation

$$V_{in} = V_o F_m(V_m), \qquad (1)$$

where $F_m(V_m)$ is the modulation function of the signal V_m. The modulation depth of the input RF signal V_{in} depends on the amplitude of the low frequency (LF) signal V_m that is equal to the amplified TAV signal V_{ae}: $V_m=KV_{ae}$, where K is the amplifying coefficient K of the LF amplifier.

The scheme presented in Fig. 3 shows external feedback that makes it unstable. The self-sustained oscillations of the TAV and self-sustained modulations of the input V_{in} and output V_{out} RF signals are observed for some values of V_o and K. Self-sustained oscillations of the TAV are not observed for small amplitude of the input signal V_{in} and small amplifying coefficient K. If the above mentioned values V_{in} K are sufficiently large, low frequency self-sustained oscillations of the TAV and self- modulations of the input signal V_{in} SAW power occur in kilohertz frequency range.

The experiments were made at room temperature and some experimental results are shown in Fig. 4, where one can see the time dependencies of the amplified TAV signal V_m (solid curves) and the RF signal amplitude V_{in} dashed curves) under and without illumination and different input amplitudes V_o and coefficients K. The curves (a)-(c) in Fig. 4 refer to the different amplitudes V_o. The amplitude V_o increases from the case (a) to the case (c), as one can see from the dashed lines in Fig. 4. The periodic quasi-harmonic regime of the oscillation is transformed into a strong nonlinear mode and, finally, into a stochastic regime for a sufficiently large input signal V_o. The curves (d) in Fig. 4 refer to the case where the structure is slightly illuminated. The curves (e) in Fig. 4 refer to the case where amplification coefficient of the amplifier is more than 30 dB, as in the cases (a)-(c).

The dependencies of the TAV amplitude V_{ae} and the period of self-sustained oscillation T of the normalized input amplitude V_o/V_{oo} shown in Fig. 5, where V_{oo} is the input RF amplitude V_o when the self-sustained oscillations are initiated.

Self-sustained oscillations in photoelectric systems

The photoelectric self-sustained oscillations are investigated experimentally as shown in the Fig. 6 principle scheme. In this scheme a metal-dielectric-semiconductor structure with the external feedback was used. This layered structure performs like the one shown in Fig. 3, but, here, the substrate material need not be a piezoelectric. The structure is illuminated by a light diode. A photoelectric voltage signal V_{pe} arises at the metal electrodes of the structure under its impulse illumination by the light diode. The DC bias voltage V_g, that is applied to the light diode, is chosen to make necessary static illumination of the structure and working regime of the diode. In the feedback of the scheme the signal V_m applied to the diode is amplified by the amplifier with the amplification coefficient K and the photoelectric voltage signal V_{pe}: $V_m=KV_{pe}$. Existence of such an external feedback in the scheme makes it unstable, and in this scheme self-sustained oscillations of the photoelectric voltage V_{pe} are observed under DC bias voltage V_g and amplifying coefficient K of the LF amplifier.

The experiments are also made at room temperature. An example of the photoelectric voltage signal V_{pe} versus time t is shown in Fig. 7. The dependencies of the PEV amplitude V_{pe} and the period of the self-sustained oscillation T against the bias voltage V_g are shown in Fig. 8.

Theory

The explanation of the experimentally obtained results is based on the relaxation properties of the acoustoelectric and photoelectric voltages in layered structures and on the relaxation charge observed in the external scheme.

First, we will take into consideration the origin of the PEV and the nonlinearity of the TAV based on charge carrier generation, recombination, and trapping of a layered structure in the semiconductor. This means that the of nonlinearity, among the considered effects, is due to the carrier generation, recombination, and trapping. The mechanism is not due to nonlinear charge current.

The kinetics of the TAV and PEV may be described by the following differential equation:

$$\tau_s \frac{dV_s}{dt} + V_s = V_{so}, \qquad (2)$$

where V_S, V_{SO}, and τ_S are the semiconductor surface electric potential, its quasi-static value, and relaxation time, respectively. The equation (2) is in agreement with the experimental results shown in Fig. 1 and Fig. 2, obtained in the case of high input resistance of a receiver of the TAV and PEV. The equation (2) is also in agreement with the equations representing charge carrier trapping kinetics [26] that describe the transient electronic processes taking place at the semiconductor surface.

The relaxation time τ_S in (2) depends on the electronic parameters of the semiconductor surface such as free and coupled charge concentrations and surface electric potentials. The quasi-static value V_{SO} also depends on the above mentioned parameters of the surface and on the acoustic or light power P_{in}. Thus, the value V_{so} is a specific function of the parameter P_{in}. For example, in the case of the TAV the V_{ae} is proportional to the P_{in} , when the P_{in} is low power. In turn, the value P_{in} is another function of the V_{in} or V_m in the schemes in Fig. 3 and Fig. 6, respectively Thus finally, the value V_{SO} is a specific function F_S of the input voltage V_{in}:

$$V_{so} = F_s(V_{in}). \qquad (3)$$

Here, for the case of the PEV instead of the relation (1) we must consider $V_{in}=V_m$ or in (1) $F_m(V_m)=V_m/V_o$.

Second, the voltage V_s appears to be a voltage source in the external circuit, for which the equivalent circuit is shown in Fig. 9. The external circuit, without high input resistance of the receiver, is described by the following differential equation:

$$\frac{dV_s}{dt} = \frac{C}{C_s}\left(\frac{dV_e}{dt} + \frac{V_e}{\tau_r}\right), \qquad (4)$$

where V_e is the voltage on the metal electrodes of the structure and is equal to the V_{ae} or V_{pe}, $C=C_s+C_{in}$, $\tau_r=R_{in}\cdot C$; R_{in} and C_{in} are the input resistance and capacitance of the amplifier receiver, respectively, and C_s is the capacitance of the structure.

Finally, combining equations (1)-(4), one can obtain the following equation describing nonlinear and self-sustained oscillations:

$$\frac{d^2V_e}{dt^2} - Q(V_e)\frac{dV_e}{dt} + \Omega^2 V_e = 0, \qquad (5)$$

where

$$\Omega = (\tau_s\cdot\tau_r)^{1/2}, \qquad (6)$$

$$Q(V_e) = \frac{1}{\tau_s}\frac{C_s}{C}V_o K F_s' F_m' - \frac{1}{\tau_s} - \frac{1}{\tau_r}, \qquad (7)$$

$F_m'=dF_m(V_m)/dV_m$ and $F_s'=dF_s'(V_{in})/dV_{in}$ are the derivatives of the functions F_m and F_s with respect to their independent variables $V_m=KV_e$ and $V_{in}=V_oF_m(KV_e)$ indicated in (1) and (3) and finally expressed relative to the dynamic value V_e and the other static parameters of the system.

It is clear from the equation of nonlinear oscillation (5) that damped oscillations of the value $V_e(t)$ for $Q(0)<0$ are transformed into soft mode self-sustained oscillations of the $V_e(t)$ [27]. The condition $Q(0)>0$ is a criterion for the self-sustained oscillations of soft mode excitation. As one can see from (7), this is possible when coefficient K, the power P_{in}, and the slopes of functions F_m and F_s are sufficiently high.

The solution of the equation (5) is complicated in common form. In the small signal approximation, however, a simple quasi-harmonic solution can be obtained. This case corresponds to the Van-der-Pol equation, where it is sufficient in the expression for $Q(V_e)$ to be expanded into a series until the quadratic term. The Van-der-Pol equation thereby obtained defines the quasi-harmonic oscillation process with the cyclic frequency Ω, given by (6).

Discussion and Application

The low frequency self-sustained oscillations presented differ from various oscillations in semiconductor structures [29]. The oscillations are due to the semiconductor relaxation property, not the amplification processes or negative differential conductivity of semiconductors.

It is of practical interest to note that because the Ω is expressed in terms of the values τ_s and τ_r, the mode of self-sustained oscillations enables τ_s to be determined directly. The relaxation time τ_s depends on physical and chemical conditions of the semiconductor surface and in the environment. Moreover, it is important to note that the value τ_s is more sensitive to external action than the value V_{so} and the frequency of oscillations may be measured experimentally with high precision. It is also important to note that there is no precise method for measuring the relaxation time τ_s.

It should be especially emphasized that when τ_s is changed, for example under illumination or temperature, Ω varies according to the equation (6) given above. This effect can be used effectively in illumination or temperature acoustic and semiconductor sensor with frequency outputs. For example, an incident light on the structure reduces τ_s and causes as oscillation period (see Fig. 4d). In our experiments the oscillation frequency of 2 kHz increased by 30% with 10 K increase in temperature.

As a result of all of the above mentioned arguments, the proposed self-sustained oscillations are of interest due to their applications in sensitive and precise physical and chemical sensors and acoustoelectric and photoelectric methods of investigation of semiconductor properties [2], [11]-[23]. The sensors differ from typical acoustoelectric sensors based on high frequency resonance systems [29], [30]. The investigation methods supplement other familiar investigation methods.

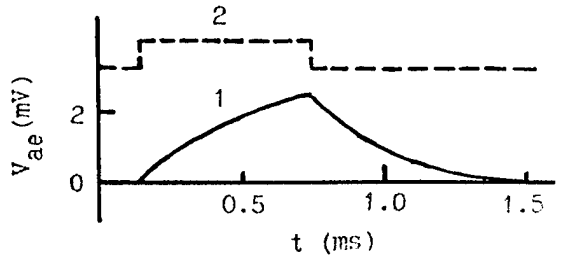


Fig. 1. Oscillogram of the TAV (solid curve 1) under SAW impulse (dashed curve 2).

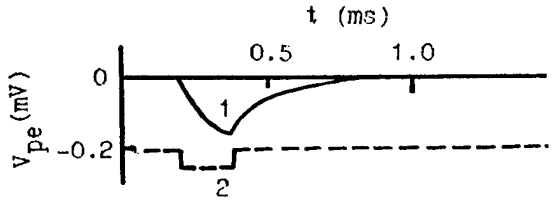


Fig. 2. Oscillogram of the PEV (solid curve 1) under illumination impulse (dashed curve 2).

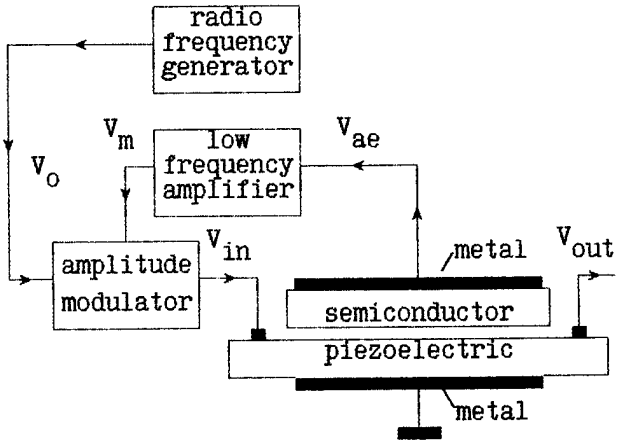


Fig. 3. Experimental setup for acoustoelectric experiments.

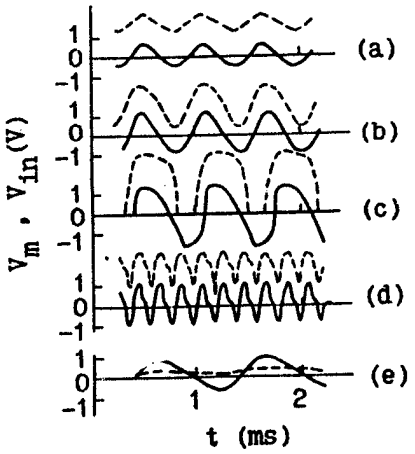


Fig. 4. Oscillograms of the amplified TV signal V_m (solid curves) and the RF input signal amplitude V_{in} (dashed curves) for K= 10 dB (a)-(d), 40 dB (e), in darkness (a)-(c), (e) and under illumination (d).

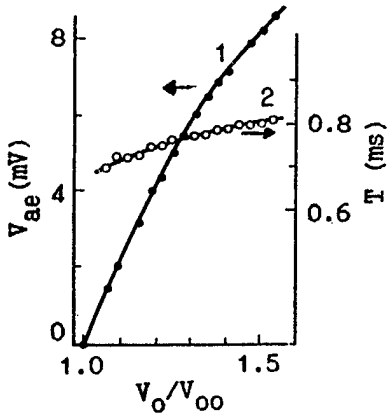


Fig. 5. The TAV amplitude V_{ae} (curve 1) and the period of self-sustained oscillation T (curve 2) versus the normalized input amplitude V_o/V_{oo}

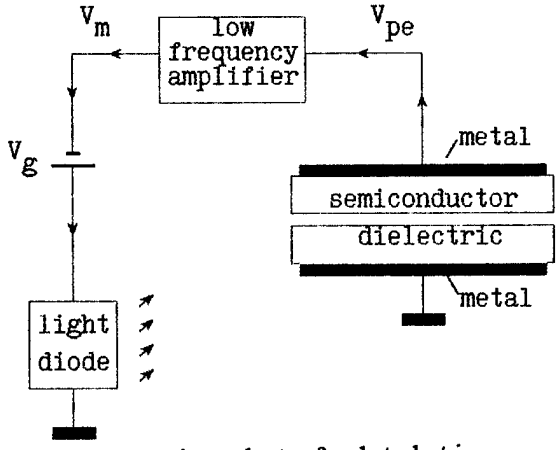


Fig. 6. Experimental setup for photoelectric experiments.

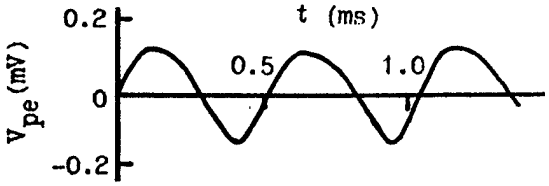


Fig. 7. Oscillogram of the photoelectric voltage.

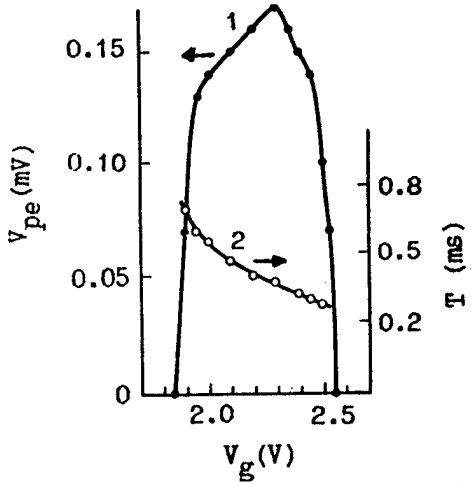


Fig. 8. The PEV amplitude V_{pe} (curve 1) and the period of self-sustained oscillation T (curve 2) versus the bias voltage V_g.

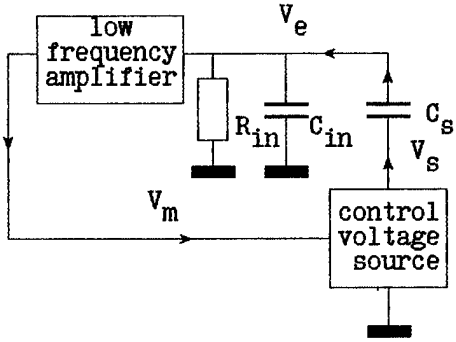


Fig. 9. Equivalent circuit.

References

[1] G.S.Kino, "Acoustoelectric interaction in acoustic-surface wave devices,"Proc.IEEE, vol. 64, No. 5, pp. 724-748, May 1976.

[2] V.A.Vyun, A.V.Rzhanov, and I.B.Yakovkin, Acoustoelectric Methods of Investigation of Semiconductor Surface. Novosibirsk: Institute of Semiconductor Physics, 1987 (in Russian)

[3] C.Campbell, Surface Acoustic Wave Devices and Their Signal Processing Applications. Boston: Academic, 1989.

[4] S.Datta, Surface Acoustic Wave Devices. Englewood Cliffs, NJ: Prentice-Hall, 1986.

[5] D.P.Morgan, Surface Wave Devices for Signal Processing. Amsterdam: Elsevier, 1985.

[6] G.S.Kino, Acoustic Waves: Devices, Imaging, and Analog Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[7] V.A.Vyun, V.V.Pnev, A.V.Tsarev, and I.B.Yakovkin, "Acoustoelectric reading of photosignals," in Proc.5th Conf. "Acoustoelectronics' 91", Varna, Bulgaria, 1991, Singapore: World Scientific Publ., 1992, pp. 146-150.

[8] M.K.Balakirev, "Nonlinear surface acoustic wave in piezocrystals and layered piezoelectric-semiconductor structure," in Nonlinear Phenomena in Solids - Modern Topics, Proc. 3th International School on Condensed Matter Physics, Varna, Bulgaria, 1984, Singapore: World Scientific Publ., 1985, pp. 191-192.

[9] V.A.Vyun, Ghiner A.V., Surdutovich G.I., and I.B.Yakovkin, "Pulse bistability on surface waves structures," in Proc. Intern. Symp. Surface Waves in Solids and Layered Structures (ISSWAS), Novosibirsk, USSR, 1986, vol. 2, pp. 173-177.

[10] V.A.Vyun, Ghiner A.V., Surdutovich G.I., and I.B.Yakovkin, "Quasi-soliton mode in bistable acoustoelectric systems," in Proc. Third Conf. "Acoustoelecrtonics'87", Varna, Bulgaria, 1987, vol. 1, pp. 344-349.

[11] I.B.Yakovkin and V.A.Vyun, "Use of acoustoelectric interaction in studies of semiconductor surface," in Proc. Intern. Symp. Surface Waves in Solids and Layered Structures (ISSWAS), Novosibirsk, USSR, 1986, vol. 1, pp. 183-200.

[12] V.A.Vyun, Yu.O.Kanter, S.M.Kikkarin, V.V.Pnev, A.A.Fedorov, and I.B.Yakovkin, "Acoustoelectric interaction of surface acoustic wave in GaAs-InGaAs superlattice," Sol. St. Commun., vol. 78, No. 9, pp. 823-825, September 1991.

[13] P.Das, M.E.Motamedi, and R.T.Webster, "Determination of semiconductor surface properties using surface acoustic waves," Appl. Phys. Lett., vol. 27, No. 3, pp. 120-122, August 1975.

[14] H.Gilboa and P.Das, "Transverse acoustoelectric voltage inversion and its application to semiconductor study: CdS," Surface Sci., vol. 62, No. 2, pp. 536-550, February 1977.

[15] H.Estrada-Vazquez, R.T.Webster, and P.Das, "Transverse acoustoelectric voltage (TAV) of high resistivity GaAs," J. Appl. Phys., vol. 50, No. 7, pp. 4942-495, June 1979.

[16] K.Varaharamyan, R.T.Webster, P.Das, and R.Bharat, "Contactless monitoring of impurity activation in ion-implantated silicon by acoustic wave techniques," J. Appl. Phys., vol. 51, No. 2, pp. 1234-1237, February 1980.

[17] B.Davari and P.K.Das, "Transient behavior of transverse acoustoelectric voltage and nondestructive characterization of semiconductor surfaces," IEEE Trans. Sonics Ultrason., vol. SU-32. No. 5, pp. 778-790, September 1985.

[18] B.Davari, M.Tabib-Azar, T.Liu, and P.Das, "Nondestructive evaluation of the semiconductor interface states' density using the transverse acoustoelectric voltage," Sol. St. Elecrton., vol. 29, No. 1, pp. 75-81, January 1986.

[19] E.O.Jonson, "Measurement of minority carrier lifetimes with the surface photovoltage," J. Appl. Phys., vol. 28, No. 11, pp. 1349-1353, November 1957.

[20] V.M.Buimistrov, A.P.Gorban, and V.G.Litovchenko, "Photovoltage induced by capture of photo-carriers by surface traps," Surface Sci., vol. 3, No. 5, pp. 445-460, May 1965.

[21] D.R.Frankl and E.A.Ulmer, "Theory of the small signal photovoltage at semiconductor surface," Surface Sci., vol. 6, No.1, pp. 115-122, January 1967. acoustoelectric voltage and nondestructive characterization of semiconductor surfaces," IEEE Trans. Sonics Ultrason., vol. SU-32, No. 5, pp. 778-790, September 1985.

[22] V.A.Vyun, O.Ivanov, M.Aleksejuk, and A.Kawalec, "Kinetics of the transverse acoustoelectric voltage and photo effects in layered structures," Acoustics Lett., vol. 14, No.12, pp. 238-240, December 1991.

[23] V.A.Vyun and I.B.Yakovkin, "Nonlinear acoustoelectric phenomena in layered structures," Int. J. Engng. Sci., vol. 29, No. 3, pp. 411-418, March 1991.

[24] H.M.Gibbs, Optical Bistability: Controlling Light with Light. Orlando: Academic, 1985.

[25] V.A.Vyun, "Self-sustained oscillation of acoustoelectric effects," in Proc. 1992 IEEE Ultrason. Symp., to be published.

[26] A.Many, Y.Goldstein, and N.B.Grover, Semiconductor Surface. New York: John Willey and Sons, 1965.

[27] N.Minorsky, Nonlinear Oscillations. Princeton: Van Nostrand, 1962.

[28] E.Scholl, Nonequilibrium Phase Transitions in Semiconductors: Self-Organization Induced by Generation and Recombination Processes. Berlin: Springer-Verlag, 1987.

[29] R.M.White, "A sensor classification scheme," IEEE Trans. Ultrason. Ferroel. Freq. Contr., vol. UFFC-34, pp. 124-126, March 1987.

[30] A.D'Amico and E.Verona, "SAW sensors," Sensors and Actuators, vol. 17, No.1, pp. 55-66, January 1989.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

RIGOROUS MODELING OF CORRUGATED SURFACES IN MICROACOUSTICS

A.R. BAGHAI-WADJI and A.A. MARADUDIN[+]
Vienna University of Technology, IAEE
Gußhausstr. 27-29, A-1040 Vienna, Austria
[+]Department of Physics, University of California
Irvine, CA 92717, U.S.A.

ABSTRACT

In the analysis and design of crystal oscillators and microacoustic resonators, an accurate characterisation of the wave propagation along periodically loaded surfaces is needed. It is a known fact that the nature of the surface loading can be electrical or mechanical or a combination of both. However, the majority of approaches presented so far consider only the effects of the electrical loading.

In this paper will be present a boundary-element-formulaion of the massloading effects in surface acoustic wave oscillators and resonators.

The method of analysis is based on the concept of periodic Green's functions and the method of weighted residuals.

A numerically calculated dispersion curve associated with a surface transverse wave in a periodic structure will also be presented.

I. INTRODUCTION

This paper concerns the application of wave number domain Green's functions to acoustic wave propagation along non-flat surfaces.

The theory applies to non-periodic as well as to periodic problems. As shown here, the formalism is extraordinarily efficient in treating periodic boundary value problems (the involved integrals transform into series). Nevertheless, the application of the Fourier transformation (which is an essential tool in our method) makes this theory also attractive for non-periodic problems.

The main idea is that by applying the Fourier transformation, the spatial co-ordinate dependence of the field quantities becomes exponential in nature, as shown below

$$e^{jk_1(x_1-x_1')}e^{jk_2(x_2-x_2')}e^{-\alpha|x_3-x_3'|}.$$

Thus differentiation with respect to, and integration over, the primed and unprimed co-ordinates, as they appear in the theory, result simply in multiplication and division by $\pm jk_1$, $\pm jk_2$ and $\pm\text{sign}(x_3 - x_3')\alpha$, respectively. These properties have promising consequences.

This is due to the fact that in many applications the spatial derivative of Green's functions appear in the underlying integral equations. Furthermore, the method of moments (MoM), which will be applied here, requires integration. (In the point-matching technique, a special case of MoM, the aforementioned integration results in sampling.)

Remark: The governing integral equations are singular integrals. Generally we have to operate with their *Cauchy's Principal Value*, and/or their *Hadamard's Finite Part* [1]. As is shown below, our solution technique offers an attractive alternative way to handle the arising integral equations.

II. BASIC INTEGRAL EQUATIONS

In an elastic medium, which is characterized by the elastic moduli $C_{\alpha\beta\mu\nu}$ and the mass density ρ, the following integral relation holds [2]:

$$\frac{1}{\rho}\sum_{\alpha\beta\mu\nu}\int\int_S n_\beta' C_{\alpha\beta\mu\nu}(u_\alpha(\vec{x}'|\omega)\times$$
$$\times\frac{\partial}{\partial x_\nu'}G_{\gamma\mu}(\vec{x},\vec{x}'|\omega)-$$
$$-G_{\gamma\alpha}(\vec{x},\vec{x}'|\omega)\frac{\partial}{\partial x_\nu'}u_\mu(\vec{x}'|\omega))dS' =$$
$$=\begin{cases} u_\gamma(\vec{x}|\omega) & \text{if } \vec{x}\in\Omega \\ 0 & \text{if } \vec{x}\notin\Omega. \end{cases} \quad (1)$$

The elastic medium occupies the volume Ω and is bounded by its surface S. The symbols appearing in these equations have the following meaning: $\vec{n}$ is the outward normal unit vector, the 3×1 vector $\vec{u}$ denotes the displacement, while the 3×3 matrix $\underline{\underline{G}}$ stands for the dyadic Green's function of an infinite elastic medium. The primed symbols refer to the source points, whereas the unprimed quantities stand for field points. In the following, we restrict our attention to Ω, the interior of the elastic medium. Assuming that the bounding surface S is stress free, the second term on the left-hand side vanishes and we are left with

$$\frac{1}{\rho}\sum_{\alpha\beta\mu\nu}\int\int_S n_\beta' dS' C_{\alpha\beta\mu\nu}(\frac{\partial}{\partial x_\nu'}G_{\gamma\mu}(\vec{x},\vec{x}'|\omega))\times$$

0-7803-0905-7/93 $3.00 © 1993 IEEE

$$\times u_\alpha(\vec{x}'|\omega) = u_\gamma(\vec{x}|\omega) \qquad \vec{x} \in \Omega. \tag{2}$$

This equation states that the condition of vibration everywhere within the medium is uniquely determined once the condition on the surface S is known. The geometry of the boundary is sketched below.

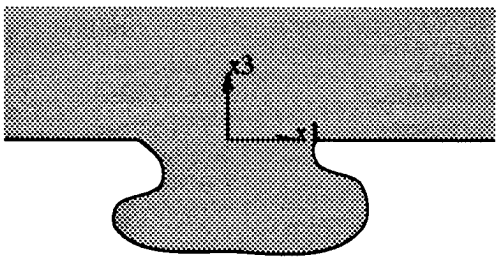


Figure 1: Geometry of interest.

As has already been shown in [3], it is much more efficient to use a matrix notation, which here transforms the above equation into the form

$$\frac{1}{\rho}\int\int_S (d\underline{\underline{S}}')^T \underline{\underline{C}}\ \underline{\underline{\nabla}}'\underline{\underline{G}}\vec{u}(\vec{x}'|\omega) = \vec{u}(\vec{x}|\omega) \qquad \vec{x} \in \Omega. \tag{3}$$

Here, the superindex T denotes the transposed form of a matrix. The matrices involved are as follows:

$$d\underline{\underline{S}}' = \begin{pmatrix} n_1' dx_2' dx_3' & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & n_3' dx_1' dx_2' \\ 0 & n_3' dx_1' dx_2' & 0 \\ n_3' dx_1' dx_2' & 0 & n_1' dx_2' dx_3' \\ 0 & n_1' dx_2' dx_3' & 0 \end{pmatrix} \tag{4}$$

$$\underline{\underline{C}} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{pmatrix} \tag{5}$$

$$\underline{\underline{\nabla}}' = \begin{pmatrix} \partial x_1' & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \partial x_3' \\ 0 & \partial x_3' & 0 \\ \partial x_3' & 0 & \partial x_1' \\ 0 & \partial x_1' & 0 \end{pmatrix} \tag{6}$$

$$\underline{\underline{G}}(\vec{x}, \vec{x}'|\omega) = \begin{pmatrix} G_{11} & 0 & G_{13} \\ 0 & G_{22} & 0 \\ G_{31} & 0 & G_{33} \end{pmatrix}. \tag{7}$$

Remarks: (i) As is clear from the structure of the elastic moduli tensor $\underline{\underline{C}}$, we restrict ourselves here to isotropic elastic media. (ii) For infinite elastic media, the Green's dyadic has the above form, which means that the vibrations polarized in the sagittal plane (x_1, x_3) and that the horizontal vibrations parallel to x_2 decouple. (iii) It should be noted that, while the matrix operator $\underline{\underline{\nabla}}$ is already known [4] it was also necessary to introduce here the matrix $d\underline{\underline{S}}$.

Using these matrices in the above integral equations and performing the necessary multiplication, we obtain

$$\begin{aligned}
&\int\int n_1' dx_2' dx_3' \frac{C_{11}}{\rho}\frac{\partial G_{11}}{\partial x_1'}u_1 + \\
&+\int\int n_1' dx_2' dx_3' \frac{C_{12}}{\rho}\frac{\partial G_{31}}{\partial x_3'}u_1 + \\
&+\int\int n_3' dx_1' dx_2' \frac{C_{44}}{\rho}\frac{\partial G_{11}}{\partial x_3'}u_1 + \\
&+\int\int n_3' dx_1' dx_2' \frac{C_{44}}{\rho}\frac{\partial G_{31}}{\partial x_1'}u_1 + \\
&+\int\int n_1' dx_2' dx_3' \frac{C_{11}}{\rho}\frac{\partial G_{13}}{\partial x_1'}u_3 + \\
&+\int\int n_1' dx_2' dx_3' \frac{C_{12}}{\rho}\frac{\partial G_{33}}{\partial x_3'}u_3 + \\
&+\int\int n_3' dx_1' dx_2' \frac{C_{44}}{\rho}\frac{\partial G_{13}}{\partial x_3'}u_3 + \\
&+\int\int n_3' dx_1' dx_2' \frac{C_{44}}{\rho}\frac{\partial G_{33}}{\partial x_1'}u_3 = u_1(x_1, x_3)
\end{aligned} \tag{8}$$

$$\begin{aligned}
&\int\int n_1' dx_2' dx_3' \frac{C_{44}}{\rho}\frac{\partial G_{22}}{\partial x_1'}u_2 + \\
&+\int\int n_3' dx_1' dx_2' \frac{C_{44}}{\rho}\frac{\partial G_{22}}{\partial x_3'}u_2 = u_2(x_1, x_3)
\end{aligned} \tag{9}$$

$$\begin{aligned}
&\int\int n_1' dx_2' dx_3' \frac{C_{44}}{\rho}\frac{\partial G_{11}}{\partial x_3'}u_1 + \\
&+\int\int n_1' dx_2' dx_3' \frac{C_{44}}{\rho}\frac{\partial G_{31}}{\partial x_1'}u_1 + \\
&+\int\int n_3' dx_1' dx_2' \frac{C_{12}}{\rho}\frac{\partial G_{11}}{\partial x_1'}u_1 + \\
&+\int\int n_3' dx_1' dx_2' \frac{C_{11}}{\rho}\frac{\partial G_{31}}{\partial x_3'}u_1 + \\
&+\int\int n_1' dx_2' dx_3' \frac{C_{44}}{\rho}\frac{\partial G_{13}}{\partial x_3'}u_3 + \\
&+\int\int n_1' dx_2' dx_3' \frac{C_{44}}{\rho}\frac{\partial G_{33}}{\partial x_1'}u_3 +
\end{aligned}$$

$$+\int\int n_3' dx_1' dx_2' \frac{C_{12}}{\rho} \frac{\partial G_{13}}{\partial x_1'} u_3 +$$

$$+\int\int n_3' dx_1' dx_2' \frac{C_{11}}{\rho} \frac{\partial G_{33}}{\partial x_3'} u_3 = u_3(x_1, x_3). \tag{10}$$

III. Shear Horizontal Waves

For the purpose of concentrating on the essential ideas in the underlying solution technique, we restrict ourselves to horizontally polarized vibrations.

Further, with regard to the fact that we are dealing here with a two-dimensional problem ($\partial/\partial x_2 \equiv 0$), it is useful to rewrite the Eq.(9) in the following form:

$$\int_\Gamma n_1(x_1', x_3') dx_3' \frac{C_{44}}{\rho} \int_{-\infty}^{\infty} dx_2' \frac{\partial G_{22}}{\partial x_1'}) u_2(x_1', x_3' +$$

$$+\int_\Gamma n_3(x_1', x_3') dx_1' \frac{C_{44}}{\rho} \int_{-\infty}^{\infty} dx_2' \frac{\partial G_{22}}{\partial x_3'} u_2(x_1', x_3') =$$

$$= u_2(x_1, x_3); \quad \vec{x} \in \Omega, \quad \vec{x}' \in \Gamma. \tag{11}$$

Before we proceed further, it is useful to perform some preparatory calculations [2] which will be used later.

IV. Preparatory Calculations

Considering a two-dimensional Fourier representation for $G_{\alpha\beta}$,

$$G_{\alpha\beta}(\vec{x}, \vec{x}'|\omega) =$$

$$= \int\int \frac{dq_1}{2\pi} \frac{dq_2}{2\pi} e^{jq_1(x_1 - x_1')} e^{jq_2(x_2 - x_2')} \times$$

$$\times G_{\alpha\beta}(q_1, q_2, \omega|x_3, x_3'), \tag{12}$$

we can write

$$\int_{-\infty}^{\infty} dx_2' \frac{\partial}{\partial x_1'} G_{\alpha\beta}(\vec{x}, \vec{x}'|\omega) =$$

$$= \int_{-\infty}^{\infty} \frac{dq}{2\pi} e^{iq(x_1 - x_1')} (-jq) g_{\alpha\beta}(q, \omega|x_3, x_3') \tag{13}$$

where we have used the notation

$$g_{\alpha\beta}(q, \omega|x_3, x_3') = G_{\alpha\beta}(q_1, 0, \omega|x_3, x_3'). \tag{14}$$

Additionally we have

$$\int_{-\infty}^{\infty} dx_2' \frac{\partial}{\partial x_3'} G_{\alpha\beta}(\vec{x}, \vec{x}', \omega) =$$

$$= \int_{-\infty}^{\infty} \frac{dq}{2\pi} \frac{\partial g_{\alpha\beta}(q, \omega|x_3, x_3')}{\partial x_3'} e^{iq(x_1 - x_1')}. \tag{15}$$

We now assume that Γ is built up from segments with finite length and parallel either to the x_1-axis or to the x_3-axis. It is convenient to parameterize the co-ordinates x_1 and x_3 on the boundary Γ, Fig.2.

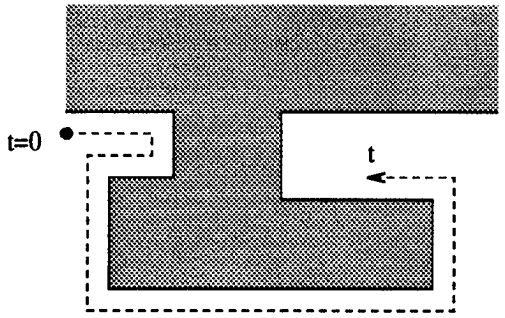


Figure 2: Boundary Γ consists of straight line seqments.

V. Parametrization of the co-ordinates on the boundary

We assume the parameter t to be the length of the intervals parallel to the x_1-axis or to the x_3-axis on the boundary.

Unit normal vector: The two-dimensionality of our problem implies that the unit normal $\vec{n}$ has two components. Thus we can write

$$\vec{n} = \frac{(n_1, 0, n_3)}{\sqrt{n_1^2 + n_3^2}}. \tag{16}$$

However, based on our assumption of dealing with subintervals parallel to the co-ordinate axes, the unit normal vector appears in our considerations in one of the following forms $(n_1, 0, 0)$ or $(0, 0, n_3)$. These are by defintion normalized vectors. Let us now consider in more detail, subintervals which are parallel to the co-ordinate axes.

Subintervalls parallel to the x_1-axis: Here we have $\vec{n}(t) = (0, 0, n_3)$. Two cases must be considered, $n_3 = -1$ and $n_3 = 1$, Fig.3.

The reader can see that, irrespective of the sign of $n_3(t)$, we have

$$x_1(t) = x_1(t_0) - n_3(t - t_0). \tag{17}$$

By differentiation we obtain

$$\frac{dx_1(t)}{dt} = -n_3 \Rightarrow dx_1 = -n_3 dt. \tag{18}$$

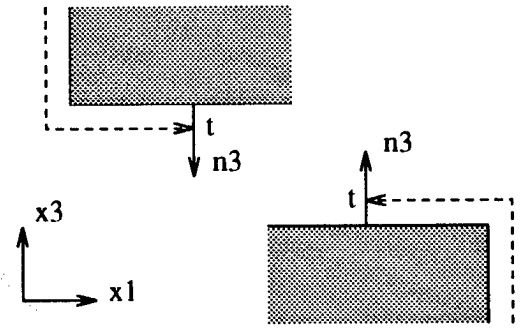


Figure 3: Subintervals parallel to the x_1-axis.

Multiplying this equation by n_3, we obtain

$$n_3 dx_1 = -n_3^2 dt. \tag{19}$$

Subintervalls parallel to the x_3-axis: In this case we have $\vec{n}(t) = (n_1, 0, 0)$. In a similar way we can show that, irrespective of the sign of $n_1(t)$, we have the following form for $x_3(t)$:

$$x_3(t) = x_3(t_0) + n_1(t - t_0), \tag{20}$$

from which we obtain

$$\frac{dx_3(t)}{dt} = n_1 \Rightarrow dx_3 = n_1 dt, \tag{21}$$

and

$$n_1 dx_3 = n_1^2 dt. \tag{22}$$

With these results we can write

$$\begin{aligned}
&\int_{-\infty}^{\infty} n_1^2(t')dt' g_a(x_1(t), x_3(t)|x_1(t'), x_3(t')) \times \\
&\times u_2(x_1(t'), x_3(t')) - \\
&- \int_{-\infty}^{\infty} n_3^2(t')dt' g_c(x_1(t), x_3(t)|x_1(t'), x_3(t')) \times \\
&\times u_2(x_1(t'), x_3(t')) = u_2(x_1(t), x_3(t))
\end{aligned} \tag{23}$$

or, in an abbreviated form,

$$\begin{aligned}
&\int_{-\infty}^{\infty} n_1^2(t')dt' g_a(t|t')u_2(t') - \\
&- \int_{-\infty}^{\infty} n_3^2(t')dt' g_c(t|t')u_2(t') = u_2(t)
\end{aligned} \tag{24}$$

where $g_a(t|t')$ and $g_c(t|t')$ are defined as follows:

$$\begin{aligned}
&\frac{C_{44}}{\rho} \int_{-\infty}^{\infty} dx_2' \frac{\partial}{\partial x_1'} G_{22}(x_1, x_3|x_1', x_3'|\omega, q) = \\
&= g_a(x_1(t), x_3(t)|x_1(t'), x_3(t')|\omega, q) = \\
&= g_a(t|t')
\end{aligned} \tag{25a}$$

$$\begin{aligned}
&\frac{C_{44}}{\rho} \int_{-\infty}^{\infty} dx_2' \frac{\partial}{\partial x_3'} G_{22}(x_1, x_3|x_1', x_3'|\omega, q) = \\
&= g_c(x_1(t), x_3(t)|x_1(t'), x_3(t')|\omega, q) = \\
&= g_c(t|t')
\end{aligned} \tag{25b}$$

with

$$\begin{aligned}
&G_{22}(x_1, x_3|x_1', x_3'|\omega, q) = \\
&= -\frac{1}{2\alpha_t c_t^2} e^{iq(x_1 - x_1')} e^{-\alpha_t |x_3 - x_3'|}
\end{aligned} \tag{26}$$

and

$$\alpha_t = \begin{cases} \sqrt{q^2 - (\omega/c_t)^2} & \text{if } q^2 > (\omega/c_t)^2 \\ -j\sqrt{(\omega/c_t)^2 - q^2} & \text{if } q^2 < (\omega/c_t)^2 \end{cases} \tag{27}$$

$$c_t^2 = \frac{C_{44}}{\rho}. \tag{28}$$

Using the above equations, we obtain

$$\begin{aligned}
&g_a(x_1, x_3|x_1', x_3'|\omega, q) = \frac{1}{2}\frac{iq}{\alpha_t} \times \\
&\times e^{iq(x_1 - x_1')} e^{-\alpha_t |x_3 - x_3'|},
\end{aligned} \tag{29a}$$

$$\begin{aligned}
&g_c(x_1, x_3|x_1', x_3'|\omega, q) = \\
&= -\frac{1}{2}\text{sigbem}(x_3 - x_3') \times \\
&\times e^{iq(x_1 - x_1')} e^{-\alpha_t |x_3 - x_3'|}.
\end{aligned} \tag{29b}$$

In contrast to the function sign(x) defined as

$$\text{sign}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0, \end{cases} \tag{30}$$

the function sigbem(x) is introduced here by the following definition:

$$\text{sigbem}(x) = \begin{cases} -1 & x < 0 \text{ or } x \to 0^- \\ 1 & x > 0 \text{ or } x \to 0^+ \end{cases}. \tag{31}$$

As will soon be clear, the introduction of sigbem(x) is essential for an adequate application of the *boundary element method* (BEM). This point deserves particular attention. BEM relies on Green's second integral theorem. The latter provides the field solution in a point (x_1, x_3) inside the domain Ω in terms of the field values on the boundary Γ. In order to generate equations for the unknowns entering into our problem, we have to investigate the behavior of the system under investigation at the limit where the field point

$(x_1, x_3) \in \Omega$ approaches the boundary points (x_1', x_3') on Γ. Thus, we will be concerned with the behavior of the system resulting from the limiting process $\{(x_1, x_3) \to (x_1', x_3')\}$ and not with a system which results from setting (x_1, x_3) equal to (x_1', x_3') $\{(x_1, x_3) = (x_1', x_3')\}$. The function sigbem$(x)$, by defintion, accounts for this minute and significant detail.

The theory established so far applies to non-periodic, as well as periodic boundary value problems. Because of the relevance of a number of theoretical features, we will concentrate on periodic problems in the following discussion.

VI. Periodic Boundary Value Problems

Let us assume that the function $f(t)$ is a T-periodic function. We can then write

$$\int_{-\infty}^{\infty} dt f(t) = \sum_{n=-\infty}^{\infty} \int_{-T/2}^{T/2} dt f(x + nT). \tag{32}$$

Up to this point no assumptions has been made regarding the form of the boundary Γ, except that it is made up of straight lines parallel to the co-ordinate axes. Consider now the case where Γ is an open periodic profile, as sketched in the figure given below. As is illustrated in the figure, the profile function may or may not be a single-valued function.

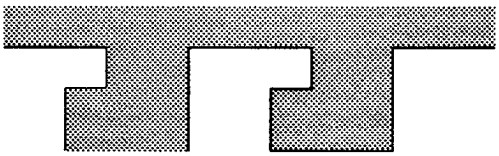


Figure 4: A muti-valued profile function.

Obviously, we can write

$$n_i^2(t + nT) = n_i^2(t); \qquad i = 1, 3, \tag{33}$$

and

$$u_2(t + nT) = u_2(t) e^{jk_0 nP}. \tag{34}$$

Remark: The value of k_0 is variable, but is restricted to the interval between $-\pi/P$ and π/P,

$$-\pi/P \geq k_0 \geq \pi/P. \tag{35}$$

Thus our integral equation results in the following equation:

$$\sum_{n=-\infty}^{\infty} \int_{-T/2}^{T/2} dt' n_1^2(t') u_2(t') e^{ik_0 nP} \times$$
$$\times g_a(t|t' + nT) -$$
$$- \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} dt' n_3^2(t') u_2(t') e^{ik_0 nP} \times$$
$$\times g_c(t|t' + nT) = u_2(t). \tag{36}$$

Further, we can write

$$g_\beta(t|t' + nT) = g_\beta(t|t') e^{-iqnP}; \quad \beta = a, c. \tag{37}$$

Inserting these results into our integral equation and changing the order of integration and summation, and rewriting, we arrive at

$$\int_{-T/2}^{T/2} dt' n_1^2(t') u_2(t') g_a(t|t') \times$$
$$\times \sum_{n=-\infty}^{\infty} e^{i(k_0 - q)nP} -$$
$$- \int_{-T/2}^{T/2} dt' n_3^2(t') u_2(t') g_c(t|t') \times$$
$$\times \sum_{n=-\infty}^{\infty} e^{i(k_0 - q)nP} = u_2(t). \tag{38}$$

Using the relation

$$\sum_{n=-\infty}^{\infty} e^{i(k_0 - q)nP} = \frac{2\pi}{P} \sum_{n=-\infty}^{\infty} \delta(q - k_0 - \frac{2\pi n}{P}) \tag{39}$$

and the defintion $k_n = k_0 + \frac{2\pi n}{P}$, we obtain

$$\int_{-T/2}^{T/2} dt' n_1^2(t') u_2(t') g_a(t|t') \frac{2\pi}{P} \times$$
$$\times \sum_{n=-\infty}^{\infty} \delta(q - k_n) -$$
$$- \int_{-T/2}^{T/2} dt' n_3^2(t') u_2(t') g_c(t|t') \frac{2\pi}{P} \times$$
$$\times \sum_{n=-\infty}^{\infty} \delta(q - k_n) = u_2(t). \tag{40}$$

Let us now take the following representations for $g_a(t|t')$ and $g_c(t|t')$:

$$g_\beta(t|t') = \int_{-\infty}^{\infty} \frac{dq}{2\pi} g_\beta(t|t'|q); \quad \beta = a, c. \tag{41}$$

Inserting the above expressions into our integral equation, and interchanging the order of the integration over

q and the summation, we then proceed to

$$\frac{1}{P}\int_{-T/2}^{T/2} dt' n_1^2(t') u_2(t') \sum_{n=-\infty}^{\infty} g_a(t|t'|k_n) - \\ -\frac{1}{P}\int_{-T/2}^{T/2} dt' n_3^2(t') u_2(t') \sum_{n=-\infty}^{\infty} g_c(t|t'|k_n) = u_2(t). \tag{42}$$

As is obvious from the above equation, we can produce the definitions

$$g_\beta^{\rm per}(t|t') = \sum_{n=-\infty}^{\infty} g_\beta(t|t'|k_n);\ \beta = a, c \tag{43}$$

which, in the following, will be refered to as the relevant periodic Green's functions for our problem.

Thus we arrive at

$$\frac{1}{P}\int_{-T/2}^{T/2} dt' n_1^2(t') u_2(t') g_a^{\rm per}(t|t') - \\ -\frac{1}{P}\int_{-T/2}^{T/2} dt' n_3^2(t') u_2(t') g_c^{\rm per}(t|t') = u_2(t). \tag{44}$$

The first term on the left-hand side of this equation subsumes the contributions from the vertical intervals to $u_2(t)$, while the second term results from horizontal ones.

We now assume a certain expansion for $u_2(t)$ in such a way that $u_2(t)$ takes a constant value $U_2^{(s)}$ on the $s^{\rm th}$ interval on the boundary Γ, i.e.

$$u_2(x_1(t), x_3(t)) = u_2(t) = \sum_{s=1}^{S} U_2^{(s)} b^{(s)}(t). \tag{45}$$

The above form implicitly assumes that the interval $(-T/2, T/2)$ has been discretized into S subintervals. The basic functions $b^{(s)}(t)$, with s varying between 1 and S, are chosen to be step functions. They take the value one on their definition range ($s^{\rm th}$ interval) and are zero elsewhere.

VII. Point-Matching Technique

For easy reference, the integral equation for the displacement $u_2(t)$ and the formulae for the relevent Green's functions are given below:

$$u_2(t) = \sum_{s=1}^{S} U_2^{(s)} \hat{n}_3^{(s)} \frac{1}{P} \int_{x_{1,b}^{(s)}}^{x_{1,e}^{(s)}} dx_1' \times \\ \times g_c^{\rm per}(x_1(t), x_3(t)|x_1', Z^{(s)}) + \\ + \sum_{s=1}^{S} U_2^{(s)} \hat{n}_1^{(s)} \frac{1}{P} \int_{x_{3,b}^{(s)}}^{x_{3,e}^{(s)}} dx_3' \times \\ \times g_a^{\rm per}(x_1(t), x_3(t)|X^{(s)}, x_3'), \tag{46}$$

with

$$g_a^{\rm per}(x_1(t), x_3(t)|X^{(s)}, x_3') = \sum_{n=-\infty}^{\infty} \frac{1}{2}\frac{jk_n}{\alpha_t^{(n)}} \times \\ \times e^{jk_n(x_1 - X^{(s)})} e^{-\alpha_t^{(n)}|x_3 - x_3'|} \tag{47a}$$

$$g_c^{\rm per}(x_1(t), x_3(t)|x_1', Z^{(s)}) = \sum_{n=-\infty}^{\infty} -\frac{1}{2} \times \\ \times {\rm sigbem}(x_3 - Z^{(s)}) e^{jk_n(x_1 - x_1')} e^{-\alpha_t^{(n)}|x_3 - Z^{(s)}|}. \tag{47b}$$

An interval which is parallel to the x_1-axis, say the $s^{\rm th}$ interval, is characterized by the co-ordinates of the beginning and end point of the interval. However, because the interval is assumed to be parallel to the x_1-axis, a single value $Z^{(s)}$ determines the x_3-co-ordinate of the interval, i.e., we have $(x_{1,b}^{(s)}, Z^{(s)})$ and $(x_{1,e}^{(s)}, Z^{(s)})$, and thus only a set of three quantities

$$(x_{1,b}^{(s)}, x_{1,e}^{(s)}, Z^{(s)})$$

uniquely specifies the $s^{\rm th}$ interval, Fig.5. Similarly, it is obvious that $s^{\rm th}$ interval, when parallel to the x_3-axis, Fig.5, can be characterized by the data

$$(X^{(s)}, x_{3,b}^{(s)}, x_{3,e}^{(s)}).$$

By inserting $g_a^{\rm per}$ and $g_c^{\rm per}$ from (47) into (46), we obtain the following equation, which is our starting equation for numerical calculations:

$$u_2(t) = \sum_{s=1}^{S} U_2^{(s)} \hat{n}_3^{(s)} \frac{1}{P} \int_{x_{1,b}^{(s)}}^{x_{1,e}^{(s)}} dx_1' \sum_{n=-\infty}^{\infty} -\frac{1}{2} \times \\ \times {\rm sigbem}(x_3 - Z^{(s)}) e^{jk_n(x_1 - x_1')} e^{-\alpha_t^{(n)}|x_3 - Z^{(s)}|} + \\ + \sum_{s=1}^{S} U_2^{(s)} \hat{n}_1^{(s)} \frac{1}{P} \int_{x_{3,b}^{(s)}}^{x_{3,e}^{(s)}} dx_3' \sum_{n=-\infty}^{\infty} \frac{1}{2}\frac{jk_n}{\alpha_t^{(n)}} \times$$

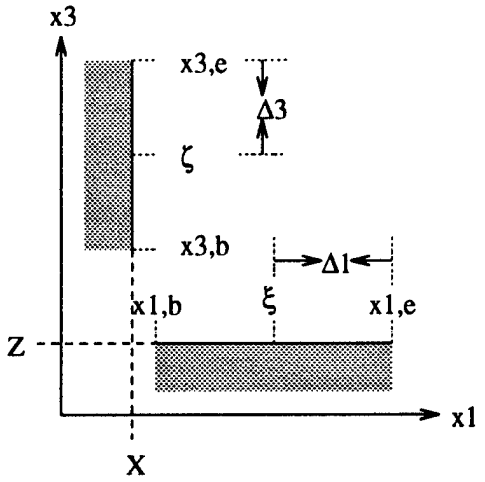


Figure 5: Parameters characterizing intervals parallel to the co-ordinate axes.

$$\times e^{jk_n(x_1-X^{(s)})}e^{-\alpha_i^{(n)}|x_3-x_3'|}. \tag{48}$$

VIII. Calculation of the Elements of the Interaction Matrix

In the present analysis, for simplicity's sake and also to concentrate on the essentials, we will restrict ourselves to single-valued profile functions. A simple example is illustrated below. Thus, while $\hat{n}_1(t)$ can have the values plus or minus one, $\hat{n}_3(t)$ is always equal to minus one.

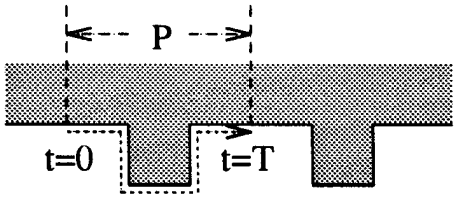


Figure 6: A single-valued profile function.

CALCULATION OF $A_{rs}^{\leftrightarrow -}$

$A_{rs}^{\leftrightarrow -}$ represents the interaction of two elements which are located parallel to the x_1-axis. The action of a source, which is distributed over the s^{th} *interval* in a constant fashion, is measured at *a point* localized at the midpoint of the r^{th} subinterval. This convention manifests itself in the symbol $A_{rs}^{\leftrightarrow -}$.

To calculate this measure, the appropriate form of $u_2(t)$ for our present calculations ($\hat{n}_1^{(s)} \equiv 0$) is the following:

$$u_2(t) = \sum_{s=1}^{S} U_2^{(s)}\hat{n}_3^{(s)}\frac{1}{P}\int_{x_{1,b}^{(s)}}^{x_{1,e}^{(s)}} dx_1' \times$$

$$\times \sum_{n=-\infty}^{\infty} -\frac{1}{2}\text{sigbem}(x_3 - Z^{(s)}) \times$$

$$\times e^{jk_n(x_1-x_1')}e^{-\alpha_i^{(n)}|x_3-Z^{(s)}|}. \tag{49}$$

Two cases must be distinguished.

CASE I: $x_3 \neq Z^{(s)}$

By changing the order of the integration and the second summation, and by performing the integration with respect to x_1', and with regard to the fact that $\hat{n}_3^{(s)}$ equals minus one, we obtain

$$u_2(t) = \sum_{s=1}^{S} U_2^{(s)}\hat{\Delta}_1^{(s)} \sum_{n=-N}^{N} \text{sign}(x_3 - Z^{(s)}) \times$$

$$\times e^{-\alpha_i^{(n)}|x_3-Z^{(s)}|}e^{jk_n(x_1-\xi^{(s)})}\text{sinc}(k_n\Delta_1^{(s)}). \tag{50}$$

In the above and in the following, the symbol $\hat{}$ on the top of the co-ordinate variables denotes non-dimensional quantities, which are scaled to P. Here, P denotes the periodicity length of the grating under consideration.

Consequently, it is useful to define the new variable $U^{(s)}$ thus:

$$U^{(s)} = U_2^{(s)}\Delta_1^{(s)}. \tag{51}$$

Next, by using Dirac's δ-functions $\delta(x_1 - \xi^{(r)})$; $r = 1, ..., S$, the process of point-matching gives us

$$A_{rs}^{\leftrightarrow -}|_{\hat{Z}^{(r)}\neq\hat{Z}^{(s)}} = \text{sign}(\hat{Z}^{(r)} - \hat{Z}^{(s)}) \times$$

$$\times \sum_{n=-N}^{N} \text{sinc}(\pi\hat{k}_n\hat{\Delta}_1^{(s)})e^{j\pi\hat{k}_n(\hat{\xi}^{(r)}-\hat{\xi}^{(s)})} \times$$

$$\times e^{-\pi\hat{\alpha}_i^{(n)}|\hat{Z}^{(r)}-\hat{Z}^{(s)}|}. \tag{52}$$

CASE II: $x_3 = Z^{(s)}$

Let us now investigate the case where x_3 approaches $Z^{(s)}$ on an interval parallel to the x_1-axis ($\hat{n}_1 \equiv 0$), as indicated in Fig.7.

Then with

$$e^{-\alpha_i^{(n)}|x_3-Z^{(s)}|} \to 1 \tag{53}$$

and

$$\hat{n}_3^{(s)}\text{sigbem}(x_3 - Z^{(s)}) = -1 \tag{54}$$

we obtain from (49)

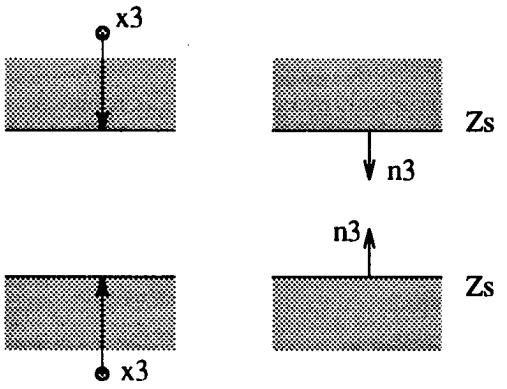


Figure 7: On the calculation of self-action.

$$u_2(t) = \sum_{s=1}^{S} U_2^{(s)} \frac{1}{2} \frac{1}{P} \int_{x_{1,b}^{(s)}}^{x_{1,e}^{(s)}} dx_1' \sum_{n=-\infty}^{\infty} e^{jk_n(x_1 - x_1')}. \quad (55)$$

Now concentrating on the summation over n, we can write

$$\sum_{n=-\infty}^{\infty} e^{jk_n(x_1-x_1')} = e^{jk_0(x_1-x_1')} \sum_{n=-\infty}^{\infty} e^{j\frac{2\pi n}{P}(x_1-x_1')}$$

$$= e^{jk_0(x_1-x_1')} \times P \sum_{n=-\infty}^{\infty} \delta(x_1 - x_1' - nP), \quad (56)$$

Thus, we obtain

$$u_2(t) = \sum_{s=1}^{S} U_2^{(s)} \frac{1}{2} \int_{x_{1,b}^{(s)}}^{x_{1,e}^{(s)}} dx_1' e^{jk_0(x_1-x_1')} \times$$

$$\times \sum_{n=-\infty}^{\infty} \delta(x_1 - x_1' - nP). \quad (57)$$

Note that *only if n equals zero can the argument of the δ-function for coinciding field and source point coordinates vanish.*

Hence the integral at the right-hand side equals δ_{rs} and we thus obtain

$$U_2^{(r)} = \sum_{s=1}^{S} U_2^{(s)} \{\frac{1}{2}\delta_{rs}\}. \quad (58)$$

By multiplying this equation by $\Delta_1^{(s)}$ and using the defintion $U^{(s)}$ as introduced above, we find that

$$A_{rs}^{\leftrightarrow -}|_{Z^{(r)} \to Z^{(s)}} = \frac{1}{2}\delta_{rs}. \quad (59)$$

CALCULATION OF $A_{rs}^{\updownarrow -}$

Calculation of $A_{rs}^{\updownarrow -}$ is quite straightforward and the application of the PMT leads directly to

$$A_{rs}^{\updownarrow -} = \text{sign}(\hat{\zeta}^{(r)} - \hat{Z}^{(s)}) \sum_{n=-N}^{N} \text{sinc}(\pi \hat{k}_n \hat{\Delta}_1^{(s)}) \times$$

$$\times e^{j\pi \hat{k}_n (\hat{X}^{(r)} - \hat{\xi}^{(s)})} e^{-\pi \hat{\alpha}_t^{(n)} |\hat{\zeta}^{(r)} - \hat{Z}^{(s)}|}. \quad (60)$$

CALCULATION OF $A_{rs}^{\leftrightarrow |}$

The symbol $A_{rs}^{\leftrightarrow |}$ denotes that the sampling point in the PMT is the point $(\hat{\xi}^{(r)}, \hat{Z}^{(r)})$, i.e., the midpoint of the r^{th} interval, which is parallel to the x_1-axis.

$$A_{rs}^{\leftrightarrow |} = \hat{n}_1^{(s)} \frac{j}{2\pi \hat{\Delta}_3^{(s)}} \sum_{n=-N}^{N} \frac{\hat{k}_n}{(\hat{\alpha}_t^{(n)})^2} e^{j\pi \hat{k}_n (\hat{\xi}^{(r)} - \hat{X}^{(s)})} \times$$

$$\times \{\text{sign}(\hat{Z}^{(r)} - \hat{x}_{3,e}^{(s)})(1 - e^{-\pi \hat{\alpha}_t^{(n)} |\hat{Z}^{(r)} - \hat{x}_{3,e}^{(s)}|}) -$$

$$-\text{sign}(\hat{Z}^{(r)} - \hat{x}_{3,b}^{(s)})(1 - e^{-\pi \hat{\alpha}_t^{(n)} |\hat{Z}^{(r)} - \hat{x}_{3,b}^{(s)}|})\}. \quad (61)$$

The function sign(x) is defined in the usual way. Thus, in cases where one of the arguments of the sign-functions involved vanishes, i.e., once we have $\hat{Z}^{(r)} - \hat{x}_{3,e}^{(s)} = 0$ or $\hat{Z}^{(r)} - \hat{x}_{3,b}^{(s)} = 0$, the corresponding terms in (61) can then be discarded. This occurs whenever we are dealing with intervals which border onto each other at the edge points. It should be emphasized that both arguments of the two sign-functions cannot vanish simultaneously.

CALCULATION OF $A_{rs}^{\updownarrow |}$

In a similar fashion, we obtain the following result for $A_{rs}^{\updownarrow |}$:

$$A_{rs}^{\updownarrow |} = \hat{n}_1^{(s)} \frac{j}{2\pi \hat{\Delta}_3^{(s)}} \times$$

$$\times \sum_{n=-N}^{N} \frac{\hat{k}_n}{(\hat{\alpha}_t^{(n)})^2} e^{j\pi \hat{k}_n (\hat{X}^{(r)} - \hat{X}^{(s)})} \times$$

$$\times \{\text{sign}(\hat{\zeta}^{(r)} - \hat{x}_{3,e}^{(s)})(1 - e^{-\pi \hat{\alpha}_t^{(n)} |\hat{\zeta}^{(r)} - \hat{x}_{3,e}^{(s)}|}) -$$

$$-\text{sign}(\hat{\zeta}^{(r)} - \hat{x}_{3,b}^{(s)})(1 - e^{-\pi \hat{\alpha}_t^{(n)} |\hat{\zeta}^{(r)} - \hat{x}_{3,b}^{(s)}|})\}. \quad (62)$$

The expression for the self-action is

$$A_{ss}^{\updownarrow |} = -\frac{j}{\pi \hat{\Delta}_3^{(s)}} \sum_{n=-N}^{N} \frac{\hat{k}_n}{(\hat{\alpha}_t^{(n)})^2} (1 - e^{-\pi \hat{\alpha}_t^{(n)} |\hat{\Delta}_3^{(s)}|}). \quad (63)$$

IX. Numerical Results

Based on the results obtained in the previous section we can write $\underline{\underline{A}}\vec{U} = \vec{U}$. The solvability condition for this equation yields the dispersion relation. Let us now consider the geometry sketched in Fig.6. The parameters of the rectangular corrugation are the following: width to period = 1:3, height to period = 1:3. Both the ridge and the substrate are made from the same elastic material which is characterized by the bulk transverse velocity $c_t(= \sqrt{C_{44}/\rho})$. Fig.8 shows the result for the dispersion curve in the Brillouin diagram. The horizontal axis is the normalized wave number $\hat{k}(= \pi k_0/P)$, while the vertical axis is the normalized frequency $\hat{\omega}(= P\omega/\pi c_t)$. In the normalized quantities, the slope of the dispaersion curve at $\hat{k} = 0$ is unity.

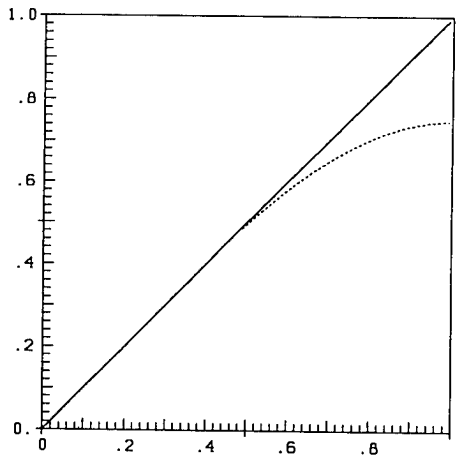


Fig.8: Dispersion realtion

X. Summary

We have outlined a solution technique with a wide applicability range for engineering problems. The important features of the method are: (1) It is applicable to arbitrarily shaped electrode geometries. (2) It is applicable to arrays consisting of electrodes with materials differing from the substrate-material. (3) Periodic problems can be tackled with a minimum amount of computer resources. (The computation of a point on the dispersion curve is of the order of a second on medium scale computers.) (4) It can be shown that the same solution procedure applies to waves with sagittal polarization.

Acknowledgements

One of the authors (ARBW) thanks the Ministry of Science and Research in Austria for a Kurt Goedel Grant at the University of California, Irvine. He is grateful for generous allocations of computer time at the Vienna University of technology.

References

[1] R.P. Kanval, *Generalized functions, Theory and Techniques*, Academic Press, Mathematics in Science and Engineering, Vol. 171, 1983.

[2] N.E. Glass, R. Loudon, and A.A. Maradudin, *Propagation of Rayleigh Surface Waves across a Large-Amplitude Grating*, Physical Review B, Vol. 24, No. 12, Dec. 1981.

[3] A.R. Baghai-Wadji and A.A. Maradudin, *Scattering of a Sagittal Acoustic Surface Wave on a Large Aplitude Ridge or on a deep Groove*, Proccedings of Nineteenth Annual Review of Progress in Qualitative Nondestructive Evaluation, University of california, San Diego, La Jolla, California, 1992.

[4] B.A. Auld, *Acoustic Fields and Waves in Solids*, Vol.I, 1973, John Wiley & Sons.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

FREQUENCY TEMPERATURE CHARACTERISTICS OF THE x-LENGTH STRIP RESONATORS OF AT-CUT QUARTZ

J. Zelenka* and P.C.Y. Lee**

*Dept. of Electrical Engineering, Technical University of Liberec,
Halkova 6, CS-461 17 Liberec, Czech Republik
**Dept. of Civil Engineering and Operations Research,
Princeton University, Princeton, NJ 08544 USA

Abstract

The equations of motion of the coupled thickness-shear, thickness-length flexure, width-shear and width-length flexure deduced by neglecting of the piezoelectric coupling deduced from a system of one-dimensional equations of motion for AT-cut quartz strip resonators (by Lee and Wang, 1992) is employed for the study of the frequency temperature characteristics of the x-length AT-cut quartz strip resonators. The computed dispersion curves, frequency spectrum and the thicknes-shear resonance frequency-temperature curves (the last two as a function of dimensions ratios) are given.

Introduction

The analysis of the vibrations of AT-cut quartz strips of narrow width and finite length has been published by Lee and Wang [1] in 1992. In the mentioned paper, one-dimensional equations for the modes of vibration in strip width and for frequencies upto and including the fundamental thickness-shear have been deduced from the two-dimensional, first-order equations for piezoelectric crystal plates, given by Lee, Syngellakis, and Hou [2], by expanding the mechanical displacements and electric potentials in series of trigonometric functions of the width coordinate. The neglecting of the piezoelectric properties and elastic stiffness c_{56} of the plate made it possible to select four groups of the modes of vibrations. They were the thickness-shear and thickness length flexure vibration and their first twist-overtone, the length-extension, width-stretch, and symmetric width-shear vibrations and the width-shear, width-length flexure and antisymmetric width-stretch vibrations.

The thickness-shear resonance is the main resonance of the strip for applications. The thickness-shear resonance frequency temperature dependence can be predicted from the frequency equation of the coupled thickness-shear and thickness-length flexure vibration of the strip given by Lee and Wang in [1] when the temperature changes of the elastic stiffnesses and thermal expansion coefficients are included. But the more precise analysis of the effect of length-to-thickness (a/b) and width-to-thickness (c/b) ratios of the strip on the resonance frequency temperature dependence requires to consider also the influence of the coupling with the other modes of vibrations.

By neglecting the coupling of the anti-symmetric width-stretch mode with width-shear and width-length flexure modes of vibrations for c/b less then 3.78 Lee deduced a set of four coupled displacement equations of motion from the one-dimensional equations for strip resonator given in [1]. These four displacement equations of motion accommodate the coupling of thickness-shear, thickness-length flexure, width shear and width-length flexure vibrations.

In the present paper, the frequency equation of the four coupled displacement equations for AT-cut quartz crystal strip is obtainedby setting piezoelectric constants $e_{i6} = 0$. The temperature dependent material properties are included in the dispersion relation and frequency equation. The resonance frequency temperature dependences as the functions of a/b and c/b ratios are computed.

Temperature Dependent Material Properties

The AT-cut quartz strip resonators shown with its coordinates and dimensions in Fig. 1 is considered in subsequent discussion. Similarly as in references [4], [5] and [6] we express the influence of the thermally biased homogenius strain by means of terms β_{ki} and D_{ijkl}.

The term β_{ki} is given by the relation [4], [5]

$$\beta_{ki} = \delta_{ki} + \alpha^{\Theta}_{ki}, \tag{1}$$

where δ_{ki} is a *Kronecker delta* and

$$\alpha^{\Theta}_{ki} = \alpha^{(1)}_{ki}\Theta + \alpha^{(2)}_{ki}\Theta^2 + \alpha^{(3)}_{ki}\Theta^3. \tag{2}$$

In (2) $\alpha^{(n)}_{ki}$ are n-th order thermal expansion coefficients (measured by Bechmann, Ballato and Lukaszek [6] and corrected by Kosinski, Gualtieri and Ballato [7]) and Θ is the temperature change, $\Theta = T - T_0$.

The tensors α^{Θ}_{ki} and β_{ki} are predominantly diagonal tensors with the off-digonal terms in the order of magnitude of 10^{-6} as compared with the diagonal terms. Therefore, by neglecting the off-diagonal terms, we have ($\beta_{ki} = 0$ for $k = i$).

$$\beta_k = \beta_{kk} = 1 + \alpha^{\Theta}_{kk}, \quad \text{(nosum)}. \tag{3}$$

The term D_{ijkl} is given by the relation

$$D_{ijkl} = C_{ijkl} + D^{(1)}_{ijkl}\Theta + D^{(2)}_{ijkl}\Theta^2 + D^{(3)}_{ijkl}\Theta^3, \tag{4}$$

0-7803-0905-7/93 $3.00 © 1993 IEEE

where

$$D_{ijkl}^{(q)} = \frac{1}{q!}C_{ijkl}^{(q)} + C_{ijklmn}\alpha_{mn}^{\Theta}, \tag{5}$$

and C_{ijkl} and C_{ijklmn} are the second and third order elastic stiffness of quartz, while $C_{ijkl}^{(1)}$, $C_{ijkl}^{(2)}$ and $C_{ijkl}^{(3)}$ are respectively the first temperature derivatives, second effective temperature derivatives and third effective temperature derivatives. Values of the temperature derivatives were calculated and reported in reference [5] and [6]. The magnitudes of D_{ijkl} where reported in reference [5].

Coupled Thickness-Shear, Thickness-Length Flexure, Width-Shear and Width-Length Flexure Vibrations

The displacement equations of motion of the coupled TSh, tlF, WSh and wlF modes of vibrations including of the thermal expansion coefficients β_i are

$$\begin{aligned}
&c_{56}(\beta_2\beta_3 u_{3,11}^{(00)} + \frac{\alpha_1}{c}\beta_1\beta_2 u_{1,1}^{(01)}) \\
&+c_{66}(\beta_2^2 u_{2,11}^{(00)} + \frac{\alpha_1}{b}\beta_1\beta_2 u_{1,1}^{10)}) = \varrho(1+R)\ddot{u}_2^{(00)}, \\
&c_{55}(\beta_3^2 u_{3,11}^{(00)} + \frac{\alpha_1}{c}\beta_1\beta_3 u_{1,1}^{(01)}) \\
&+c_{56}(\beta_2\beta_3 u_{2,11}^{(00)} + \frac{\alpha_1}{b}\beta_1\beta_3 u_{1,1}^{(10)}) = \frac{\varrho}{\alpha_2^2}(1+R)\ddot{u}_3^{(00)}, \\
&\bar{c}_{11}\beta_1^2 u_{1,11}^{(01)} - \frac{2\alpha_1}{c}c_{55}(\beta_1\beta_3 u_{3,1}^{(00)} + \frac{\alpha_1}{c}\beta_1^2 u_1^{(01)}) \\
&-\frac{2\alpha_1}{c}c_{56}(\beta_1\beta_2 u_{2,1}^{(00)} + \frac{\alpha_1}{b}\beta_1^2 u_1^{(10)}) = \varrho(1+R)\ddot{u}_1^{(01)}, \\
&\tilde{c}_{11}u_{1,11}^{(10)} - \frac{2\alpha_1}{b}c_{66}(\beta_1\beta_2 u_{2,1}^{(00)} + \frac{\alpha_1}{b}\beta_1^2 u_1^{(10)}) \\
&-\frac{2\alpha_1}{b}c_{66}(\beta_1\beta_3 u_{3,1}^{(00)} + \frac{\alpha_1}{c}\beta_1^2 u_1^{(10)}) = \varrho(1+2R)\ddot{u}_1^{(10)},
\end{aligned} \tag{6}$$

and the stress-displacement relations

$$\begin{aligned}
T_1^{(01)} &= 2\bar{c}_{11}\beta_1 u_{1,1}^{(01)}, \\
T_1^{(10)} &= 2\tilde{c}_{11}\beta_1 u_{1,1}^{(10)}, \\
T_5^{(00)} &= 4(c_{55}(\beta_3 u_{3,1}^{(00)} + \frac{\alpha_1}{c}\beta_1 u_1^{(01)}) \\
&+ c_{56}(\beta_2 u_{2,1}^{(00)} + \frac{\alpha_1}{b}\beta_1 u_1^{(10)})), \\
T_6^{(00)} &= 4(c_{56}(\beta_3 u_{3,1}^{(00)} + \frac{\alpha_1}{c}\beta_1 u_1^{(01)}) \\
&+ c_{66}(\beta_2 u_{2,1}^{(00)} + \frac{\alpha_1}{b}\beta_1 u_1^{(10)})).
\end{aligned} \tag{7}$$

We choose the modes of vibrations to have the form

$$\begin{aligned}
u_1^{(01)} &= \sum_{r=1}^{4} A_{1r}\cos\xi_r x_1 e^{i\omega t}, \\
u_1^{(10)} &= \sum_{r=1}^{4} A_{2r}\cos\xi_r x_1 e^{i\omega t}, \\
u_2^{(00)} &= \sum_{r=1}^{4} A_{3r}\sin\xi_r x_1 e^{i\omega t}, \\
u_3^{(00)} &= \sum_{r=1}^{4} A_{4r}\sin\xi_r x_1 e^{i\omega t}
\end{aligned} \tag{8}$$

which satisfy (6), provided

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} A_{1r} \\ A_{2r} \\ A_{3r} \\ A_{4r} \end{pmatrix} = 0, \tag{9}$$

where

$$\begin{aligned}
a_{11} &= \frac{\bar{c}_{11}}{c_{66}}\beta_1^2\bar{\xi}_r^2 + (\frac{b}{c})^2 c_{55}^*\beta_1^2 - \Omega^2(1+R), \\
a_{12} &= \frac{b}{c}c_{56}^*\beta_1^2, \\
a_{13} &= \sqrt{2}\frac{b}{c}c_{56}^*\beta_1\beta_2\bar{\xi}_r, \\
a_{14} &= \sqrt{2}\frac{b}{c}c_{55}^*\beta_1\beta_3\bar{\xi}_r, \\
a_{21} &= a_{12}, \\
a_{22} &= c_{11}^*\beta_1^2\bar{\xi}_r^2 + c_{66}^*\beta_1^2 - \Omega^2(1+2R), \\
a_{23} &= \sqrt{2}c_{66}^*\beta_1\beta_2\bar{\xi}_r, \\
a_{24} &= \sqrt{2}c_{66}^*\beta_1\beta_3\bar{\xi}_r, \\
a_{31} &= \frac{1}{2}a_{13}, \\
a_{32} &= \frac{1}{2}a_{23}, \\
a_{33} &= c_{66}^*\beta_2^2\bar{\xi}_r^2 - \Omega^2(1+R), \\
a_{34} &= c_{56}^*\beta_1\beta_3\bar{\xi}_r^2, \\
a_{41} &= \frac{1}{2}a_{14}, \\
a_{42} &= \frac{1}{2}a_{24}, \\
a_{43} &= a_{34}, \\
a_{44} &= c_{55}^*\beta_3^2\bar{\xi}_r^2 - \frac{\Omega^2}{\alpha_2^2}(1+R).
\end{aligned} \tag{10}$$

The normalized frequency and wave number are defined by

$$\Omega = \frac{\omega}{\frac{\pi}{2b}\sqrt{\frac{c_{66}}{\varrho}}}, \qquad \bar{\xi}_r = \frac{2b}{\pi}\xi_r. \tag{11}$$

The dispersion relation

$$[\ a_{ij}\ ] = 0 \tag{12}$$

yield four frequency branches as shown in Fig. 2.

For traction-free ends of the strip, we required, at $x_1 = \pm a$

$$T_1^{(01)} = T_1^{(10)} = T_5^{(00)} = T_6^{(00)} = 0. \tag{13}$$

Substitution of (8) into (7) and, in turn, into (13) results in

$$\sum_{r=1}^{4}\gamma_{1r}\bar{\xi}_r A_{4r}\sin\xi_r a = 0,$$

$$\sum_{r=1}^{4}\gamma_{2r}\bar{\xi}_r A_{4r}\sin\xi_r a = 0,$$

$$\sum_{r=1}^{4}(\frac{1}{\sqrt{2}}\frac{b}{c}c^*_{55}\gamma_{1r}\beta_1+\frac{1}{\sqrt{2}}c^*_{56}\gamma_{2r}\beta_1+c^*_{56}\gamma_{3r}\beta_2$$
$$+c^*_{55}\beta_3)A_{4r}\cos\xi_r a=0,$$
$$\sum_{r=1}^{4}(\frac{1}{\sqrt{2}}\frac{b}{c}c^*_{56}\gamma_{1r}\beta_1+\frac{1}{\sqrt{2}}c^*_{66}\gamma_{2r}\beta_1+c^*_{66}\gamma_{3r}\beta_2$$
$$+c^*_{56}\beta_3)A_{4r}\cos\xi_r a=0, \qquad (14)$$

where

$$\gamma_i j=\frac{A_{ij}}{A_{4j}}, \qquad j=1,2,3,4 \qquad (15)$$

are the amplitude ratios which can be computed from (14).

The vanishing of the determinant of the coefficients matrix of (14) gives the frequency equation which must be solved in conjunction with dispersion relation (12).

The elastic stiffnesses used in equations given above depend on the temperature and are defined by the relations

$$\begin{aligned}
\bar{c}_{pq} &= D_{ppqq}-\frac{D_{pp22}D_{22qq}}{D_{2222}} \quad p,q=1,3\\
\bar{c}_{p4} &= D_{pp23}-\frac{D_{pp22}D_{2223}}{D_{2222}} \quad p=1,3\\
\bar{c}_{4q} &= D_{23qq}-\frac{D_{2322}D_{22qq}}{D_{2222}} \quad q=1,3\\
\bar{c}_{44} &= D_{2323}-\frac{D_{2322}D_{2223}}{D_{2222}}\\
c_{55} &= D_{1313}\\
\hat{c}_{pq} &= \bar{c}_{pq}-\frac{\bar{c}_{p4}\bar{c}_{4q}}{\bar{c}_{44}} \quad p,q=1,3\\
\tilde{c}_{11} &= \hat{c}_{11}-\frac{\hat{c}_{13}\hat{c}_{31}}{\hat{c}_{33}},\\
c^*_{11} &= \frac{\tilde{c}_{11}}{c_{66}}, \quad c^*_{55}=\frac{c_{55}}{c_{66}},\\
c^*_{56} &= \frac{c_{56}}{c_{66}}, \quad c^*_{66}=\frac{D_{1212}}{c_{66}}, \qquad (16)
\end{aligned}$$

where

$$c_{66}=[D_{1212}]_{\Theta=0}.$$

We note that in (16) no summation over the repeated indices.

Frequency-Temperature Characteristics of Thickness-Shear Resonance

Computational result of resonance frequency as a function of the length-to-thickness ratio a/b for a fixed width-to-thickness ratio c/b = 3.78, R = 0 and AT-cut ($\theta = 35.167^0$) is shown in Fig. 3.

It can be seen from the frequency spectrum of the strip given in Fig. 3 that for predominant thickness-shear vibrations, the strip resonators must have the a/b ratios near the values 11.05, 12.91, 14.64, 16.36 and 18.05. The thickness shear resonance frequency-temperature characteristic for these a/b ratios are given in Fig. 4. The influence of the a/b ratio is greater for the small values of the a/b ratio.

The resonance-frequency temperature characteristics for a few ratios a/b near the value a/b = 11.05 are given in Fig. 5. The resonance-frequency temperature dependence changes very rapidly if the a/b ratio is far from the inflexion point of the frequency vs a/b ratio curve.

References

[1] P.C.Y. Lee and Ji. Wang. "Vibrations of AT-cut quartz strips of narrow width and finite length". *Proc. 46th Ann. Freq. Control Symposium,* 1992, pp.

[2] P.C.Y. Lee, S. Syngellakis and J.P. Hou. "A two-dimensional theory for high frequency vibrations of piezoelectric crystal plates with or without electrodes". *J. Appl. Phys.*, 61(4), pp. 1249-1262, 1987.

[3] P.C.Y. Lee and Y.K. Yong. "Temperature derivatives of elastic stiffness derived from the frequency-temperature behavior of quartz plates". *J. Appl. Phys.*, 56, pp. 1514-1521, 1984.

[4] Y.K. Yong and P.C.Y. Lee. "Frequency temperature behavior of flexural and thickness-shear vibrations of rectangular rotated Y-cut quartz plates". *Proc. 39th Ann. Freq. Control Symposium,* 1985, pp. 415-426.

[5] P.C.Y. Lee and Y.K. Yong. "Frequency-temperature behavior of thickness vibration of doubly rotated quartz plates affected by plate dimensions and orientations". *J. Appl. Phys.*, 60, pp. 2327-2342.

[6] R. Bechmann, A.D. Ballato and T.J. Lukaszek. "High-order temperature coefficients of the elastic stiffnesses and compliances of alpha-quartz". *Proc.IRE,* 50, pp. 1812-1816, 1962.

[7] J.A. Kosinski, J.G. Gualtieri and A. Ballato. "Thermoelastic coefficients of alpha quartz". *IEEE Trans. Son. Ultrason. Freq. Control,* 39, pp. 502-507, 1992.

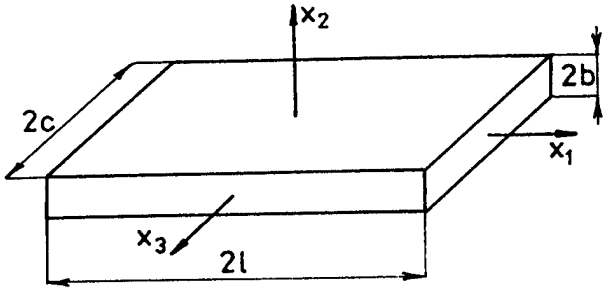


Fig. 1. An x-length AT-cut quartz strip resonator.

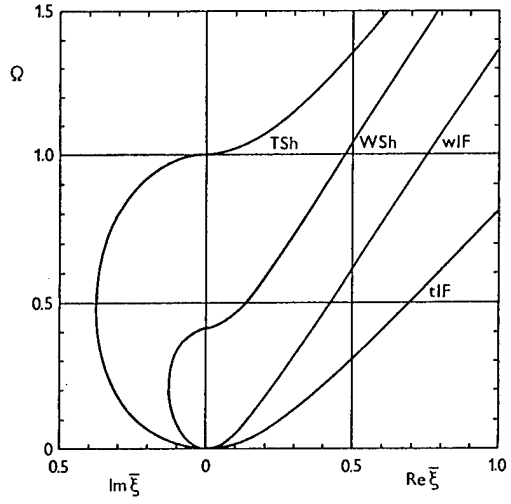


Fig. 2. Dispersion curves of coupled thicness-shear (TSh), thickness-length flexure (tlF), width-shear (WSh) and width-length flexure (wlF) vibrations of an AT-cut strip with $c/b = 3.78$.

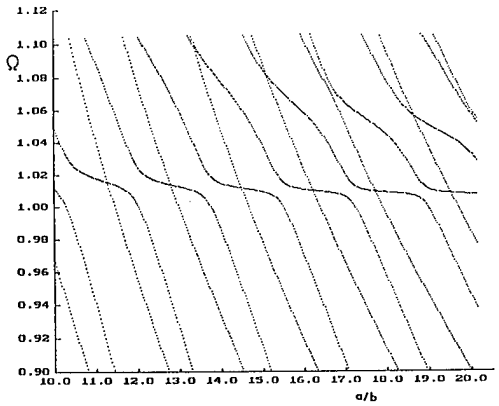


Fig. 3. Ω vs. a/b of coupled thicness-shear, thickness-length flexure, width-shear and width-length flexure vibrations in an AT-cut quartz strip with $c/b = 3.78$.

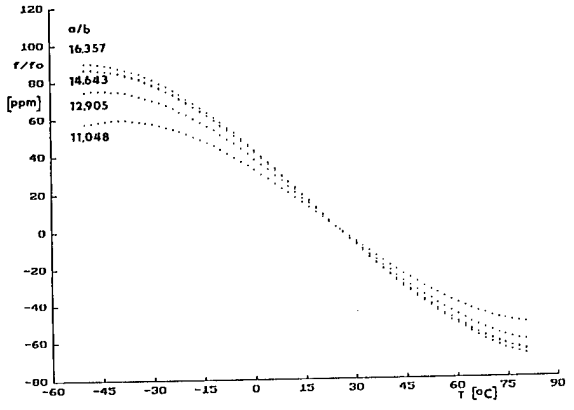


Fig. 4. Predict thickness-shear resonance frequency temperature curves for AT-cut ($\theta = 35.167^0$) quartz strip for a/b ratio variable and $c/b = 3.78$.

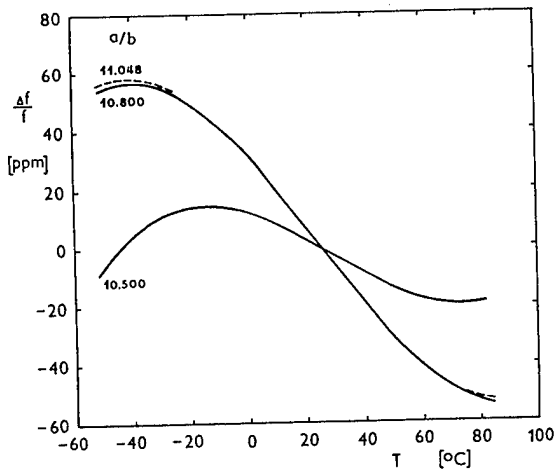


Fig. 5. Predict thickness-shear resonance frequency temperature curves for AT-cut ($\theta = 35.167^0$) quartz strip for three values of the ratio a/b near $a/b = 11.05$ and $c/b = 3.78$.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Forced Vibrations of KT-Cut Width-Extensional Mode Quartz Crystal Resonators

Hirofumi KAWASHIMA

Seiko Electronic Components Co., Ltd.
Tochigi-shi Japan, 328

Abstract

This paper describes an analysis of forced vibrations for KT-cut width-extensional mode quartz crystal resonators, which has a zero temperature coefficient at room temperature and also an absolute value of its second order temperature coefficient β is smaller than that of the DT-cut. The objectives of this paper are to clarify the influence on frequency temperature behavior due to a piezoelectric term and errors of temperature coefficients of elastic stiffness constants and the piezoelectric term that are used in the calculation.

First, a frequency equation and a transcendental equation are derived including a piezoelectric term, i.e., the frequency equation is given as one dimensional-approximate equation.

Next, from the equation the frequency constant and each temperature coefficient of α, β and γ are calculated theoretically, so that the piezoelectric term causes the increase of approximately 0.4 percent for the frequency constant. On the other hand, with respect to the temperature coefficients, it influences α slightly and β, γ markedly. In particular, the calculated values of β and γ are in very good agreement with the experimental ones, when the errors of elastic stiffness constants and the piezoelectric terms are taken into account.

Finally, resonant frequencies of fs=13.6, 16.3 and 30.7 MHz are taken as an example. The difference of a cut angle for the calculated and measured values is found to be 0.1°, 1.1° and 1.3° for each frequency.

§1 Introduction

It was shown [1]-[3] by the author that a KT-cut width-extensional mode quartz crystal resonator has a zero temperature coefficient at room temperature and also its second order temperature coefficient β has a value of $-1.25\times10^{-8}/℃^2$, an absolute value of which is smaller than that of the DT-cut with $\beta=-1.8\times10^{-8}/℃^2$, in addition, a KT-cut resonator is obtained with a small frequency deviation over a wide temperature range of +30 to +130℃ because α and β reach approximately zero at high temperature region and the third order temperature coefficient γ also has a value of about eighty percent smaller than that of an AT-cut resonator, further since resonant frequency for a width-extensional mode is inversely proportional to width dimension, a KT-cut resonator of approximately 5 MHz to 33MHz can be realized at fundamental.

However, the difference of $\Delta\theta=2.5°$ at a cut angle θ takes place between the calculated and measured values of the first order temperature coefficient α. As one of this reason, it is conceivable that a piezoelectric term was neglected in the previous report.

The objective of this paper is to clarify the affection on the first, second and third order temperature coefficients α, β, γ of a KT-cut resonator due to a piezoelectric term. In addition, another objective is to clarify the influence on frequency temperature behavior due to the errors of temperature coefficients of a piezoelectric term and elastic stiffness constants that are used in the calculation.

First, stresses T_1, T_2 including a piezoelectric term are derived under the condition that thickness z_0 of the resonator is extremely smaller than width x_0 and length y_0. Further, assuming that length y_0 is much larger than width x_0, a frequency equation which is given as one dimensional-approximate equation, is derived.

Next, from this equation the influence on resonant frequency and frequency temperature behavior due to a piezoelectric term is clarified, as a consequence, the piezoelectric term is found to influence β and γ remarkably.

Finally, when an error of the temperature coefficients for the elastic stiffness constants and the piezoelectric term is taken into consideration, the frequency temperature behavior agrees well in the calculated and measured values.

0-7803-0905-7/93 $3.00 © 1993 IEEE

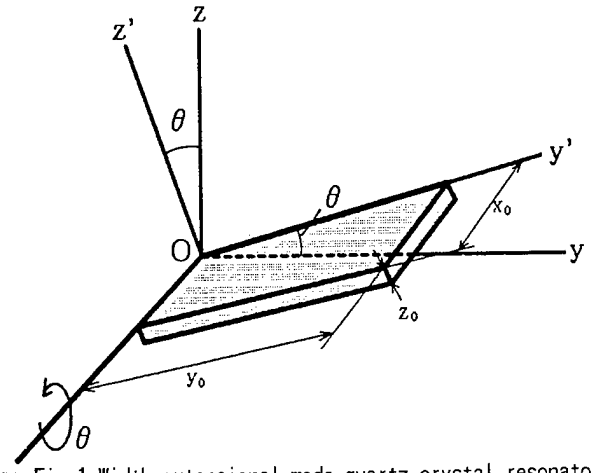


Fig. 1 Width-extensional mode quartz crystal resonator and its coordinate system.

§2 Derivative of Frequency Equation

Figure 1 shows a width-extensional mode quartz crystal resonator and its coordinate system. The resonator of a Z-plate is rotated with rotation angle θ around the x-axis. In addition, this plate is represented by (ZYw)θ in IEEE expression, and KT-cut generally has a value of a cut angle $\theta = 24^\circ$ to 28° dependent on a combination of a cut angle θ and an aspect ratio z_0/x_0.

Figure 2 shows two kinds of resonator shapes (a), (b) for width-extensional mode quartz crystal resonators that is treated in this paper and the electrode constructions (c), (d) of sectional view of the vibrational portion. As the present resonator's shapes and the electrode constructions are described in detail in the references [1], [2], a brief description will be performed here.
A resonator consists of a vibrational portion and supporting portions and the vibrational portion has a dimension of width x_0, length y_0 and thickness z_0. Also, the supporting portions are designed not so as to suppress the vibration in the width x_0 direction and simultaneously, so that vibrations of an overtone for a length-extensional mode avoid coupling to the fundamental vibration of a width-extensional mode. In addition, the excitation electrodes are deposited on both sides of the vibrational portion as shown in Fig. 2 (c), (d), so that the width-extensional mode resonator that is treated in this paper is a resonator to extend and contract largely in the width x_0 direction by an electric field of x-axis direction.

Figure 3 shows a modified shape of resonator's shapes (a), (b). Similar to the NS-GT cut resonator[4]-[6] introduced by the author and coworkers, the supporting portions for the present resonator, as described in the references [1], [2], have each mass m and can be regarded as

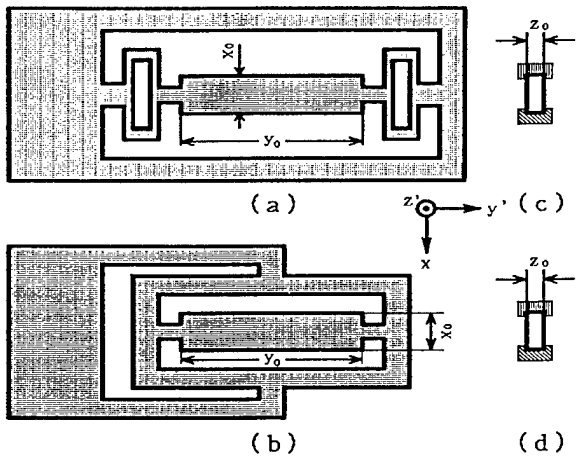


Fig. 2 Resonator shapes (a), (b) for width-extensional mode resonator and its electrode construction (c), (d) of section of vibrational portion.

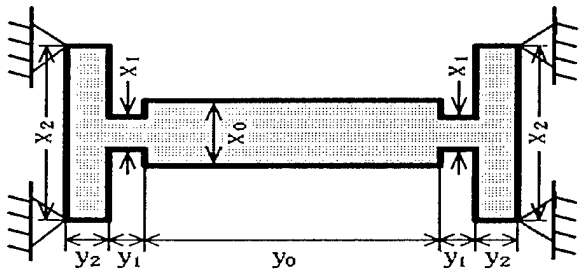


Fig. 3 Modified shape of the resonator shapes (a), (b).

resistance (work done by external force) to the vibration of the vibrational portion. Moreover, as shown in Fig. 2 (c), (d), an electric field is applied only in the x-axis direction because the resonator is fabricated from a plate rotated about the x-axis and the excitation electrodes are deposited on the surfaces perpendicular to the x-axis. A frequency equation including a piezoelectric term will be derived below.

When an elastic stiffness constant c^E_{ij}, a piezoelectric constant e_{ij} and an electric field E_1 of the x-axis direction are taken, a relationship between stress T_i and strain S_j for the present resonator that is rotated about the x-axis is given as

$$\begin{aligned}
T_1 &= c^E_{11}S_1 + c^E_{12}S_2 + c^E_{13}S_3 + c^E_{14}S_4 - e_{11}E_1 \\
T_2 &= c^E_{21}S_1 + c^E_{22}S_2 + c^E_{23}S_3 + c^E_{24}S_4 - e_{12}E_1 \\
T_3 &= c^E_{31}S_1 + c^E_{32}S_2 + c^E_{33}S_3 + c^E_{34}S_4 - e_{13}E_1 \\
T_4 &= c^E_{41}S_1 + c^E_{42}S_2 + c^E_{43}S_3 + c^E_{44}S_4 - e_{14}E_1 \\
T_5 &= c^E_{55}S_5 + c^E_{56}S_6 \\
T_6 &= c^E_{65}S_5 + c^E_{66}S_6 \qquad (1)
\end{aligned}$$

and also an electric displacement D_1 of the x-

axis direction is given as

$$D_1=e_{11}S_1+e_{12}S_2+e_{13}S_3+e_{14}S_4+\varepsilon_{11}^{S}E_1 \quad (2)$$

where ε_{11}^{S} : Dielectric constant.

Now, when width x_0 and length y_0 are much larger than thickness z_0 ($x_0, y_0 \gg z_0$), stresses T_3, T_4 and T_5 are satisfied with

$$T_3=T_4=T_5=0 \quad (3)$$

Accordingly, from eq. (3), equations (1) and (2) are transformed into

$$\begin{aligned}
T_1&=c_{11}'^{E}S_1+c_{12}'^{E}S_2-e_{11}'E_1\\
T_2&=c_{21}'^{E}S_1+c_{22}'^{E}S_2-e_{12}'E_1\\
T_6&=c_{66}'^{E}S_6\\
D_1&=e_{11}'S_1+e_{12}'S_2+\varepsilon_{11}'^{S}E_1
\end{aligned} \quad (4)$$

where

$$\begin{aligned}
c_{ij}'^{E}&=\bar{c}_{ij}^{E}-\bar{c}_{i4}^{E}\bar{c}_{4j}^{E}/\bar{c}_{44}^{E} \quad (i,j=1,2)\\
\bar{c}_{ij}^{E}&=c_{ij}^{E}-c_{i3}^{E}c_{3j}^{E}/c_{33}^{E} \quad (i,j=1,2,4)\\
c_{66}'^{E}&=c_{66}^{E}-c_{56}^{E}c_{65}^{E}/c_{55}^{E}\\
e_{1j}'&=\bar{e}_{1j}-\bar{e}_{14}\bar{c}_{j4}^{E}/\bar{c}_{44}^{E} \quad (j=1,2)\\
\bar{e}_{1j}&=e_{1j}-e_{13}c_{j3}^{E}/c_{33}^{E} \quad (j=1,2,4)\\
c_{ij}^{E}&=c_{ji}^{E}\\
\varepsilon_{11}'^{S}&=\bar{\varepsilon}_{11}^{S}+\bar{e}_{14}^{2}/\bar{c}_{44}^{E}\\
\bar{\varepsilon}_{11}^{S}&=\varepsilon_{11}^{S}+e_{13}^{2}/c_{33}^{E}
\end{aligned}$$

In addition, when length y_0 is much larger than width x_0 ($y_0 \gg x_0$), we can take $\partial/\partial y=0$. Also, when displacement u_1 of x-axis direction is taken and stresses T_1, T_2 which are entirely separated from stress T_6 are taken into account, from the electrical condition that $\mathrm{div}\boldsymbol{D}=\partial D_1/\partial x=0$, the equation of motion reduces to

$$\rho\frac{\partial^2 u_1}{\partial t^2}-c_{11}'^{D}\frac{\partial^2 u_1}{\partial x^2}=0 \quad (5)$$

where

$$c_{11}'^{D}=c_{11}'^{E}+\frac{e_{11}'^{2}}{\varepsilon_{11}'^{S}}$$

and ρ is the density of quartz and t is the time. While the boundary condition is given as

$$c_{11}'^{D}\frac{\partial u_1}{\partial x}-\frac{e_{11}'}{\varepsilon_{11}'^{S}}D_1=0 \quad \text{for } x=\pm x_0/2 \quad (6)$$

From eqs. (5) and (6), a resonant frequency and a transcendental equation can be derived.

Now, by employing an angular frequency ω, $u_1(x,t)$ can be taken as

$$\begin{aligned}
u_1(x,t)&=u_1(x)\cos\omega t\\
u_1(x)&=A\sin n'x+B\cos n'x
\end{aligned} \quad (7)$$

Substituting eq. (7) into eq. (5) and from eq. (6),

$$A=\frac{e_{11}'}{c_{11}'^{D}\varepsilon_{11}'^{S}}\frac{1}{n'}\frac{D_1}{\cos n'}, \quad B=0$$

$$u_1=\frac{e_{11}'}{c_{11}'^{D}\varepsilon_{11}'^{S}}\frac{1}{n'}\frac{D_1}{\cos n'}\sin n'x$$

$$n'=\omega/v, \quad v=\sqrt{c_{11}'^{D}/\rho}, \quad n'=nx_0/2 \quad (8)$$

Furthermore, an electric potential ϕ is calculated from a relation of $E_1=-\partial\phi/\partial x$ and when a voltage that is applied between both electrode terminals is taken as V, an electric displacement D_1 of eq. (2) is given as

$$D_1=e_{11}'\frac{2A}{x_0}\sin n'-\varepsilon_{11}'^{S}\frac{V}{x_0} \quad (9)$$

Therefore, displacement u_1 reduces to

$$u_1=\frac{e_{11}'}{c_{11}'^{D}}\frac{(-V)}{nx_0}\frac{1}{\cos n'}\frac{\sin n'x}{\left(1-\dfrac{e_{11}'^{2}}{c_{11}'^{D}\varepsilon_{11}'^{S}}\dfrac{\tan n'}{n'}\right)} \quad (10)$$

Since an electric displacement D_1 was given as eq. (9), input admittance Y_f is easily calculated from the following relation that is given as a ratio of electric current I between electric terminals and voltage V, namely,

$$\begin{aligned}
Y_f&=\frac{I}{-V}=\frac{j\omega\iint D_1 dydz}{-V}\\
&=j\omega\left(\frac{\varepsilon_{11}'^{s}y_0z_0}{x_0}\right)+j\frac{e_{11}'^{2}y_0z_0}{c_{11}'^{D}}\\
&\quad\times\frac{2\omega}{nx_0^2}\frac{1}{\cot n'-\dfrac{k_{11}'^{2}}{n'}}
\end{aligned} \quad (11)$$

where $k_{11}'^{2}=\dfrac{e_{11}'^{2}}{c_{11}'^{D}\varepsilon_{11}'^{S}}$

By replacing n' by $m\pi/2$ (m:constant) and from the resonance condition that $Y_f=\infty$, a resonant frequency f and a transcendental equation are given as

$$\begin{aligned}
f&=\frac{m}{2x_0}\sqrt{\frac{c_{11}'^{D}}{\rho}}\\
\frac{\tan(m\pi/2)}{m\pi/2}&=\frac{1}{k_{11}'^{2}}
\end{aligned} \quad (12)$$

From eq. (12) the influence on the resonant frequency f and the frequency temperature behavior due to a piezoelectric term can be clarified. Also, it is easily understood that, when k_{11}' is zero approximately, m is odd numbers for symmetry mode and even numbers for asymmetry mode.

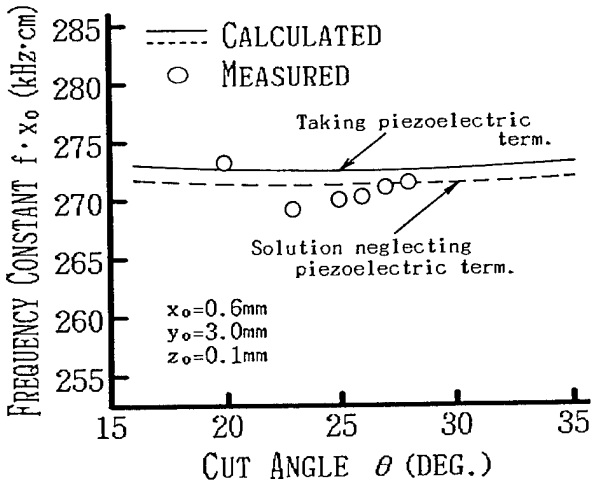


Fig. 4 Relationship between cut angle θ and frequency constant ($f \cdot x_0$) when piezoelectric term is taken into account.

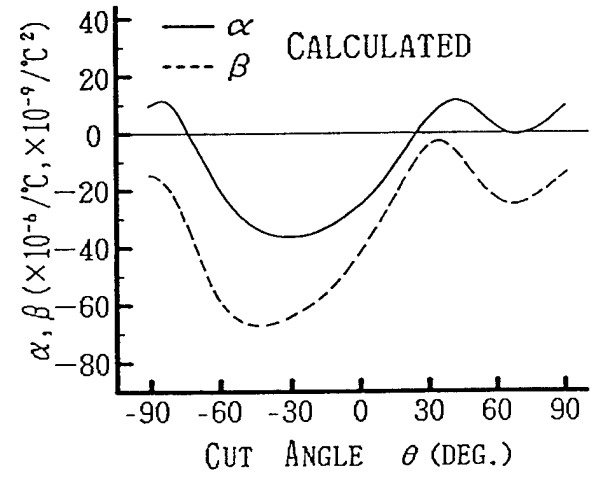


Fig. 5 Relationship of the first- and second-order temperature coefficients versus cut angle θ when piezoelectric term is taken into account. (Calculated values)

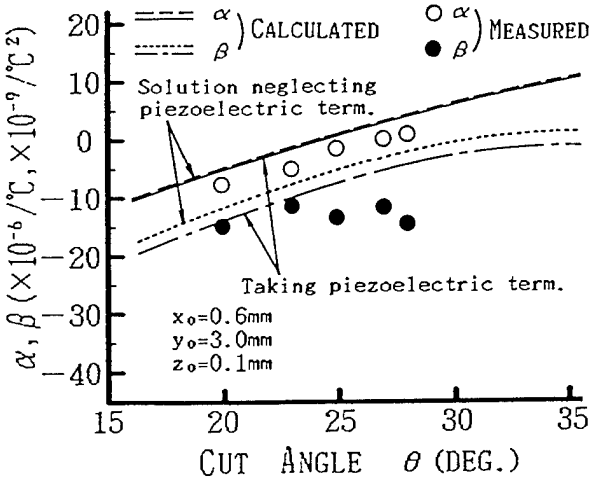


Fig. 6 Theoretical and measured values of the first- and second-order temperature coefficients α, β versus cut angle θ when piezoelectric term is taken into account.

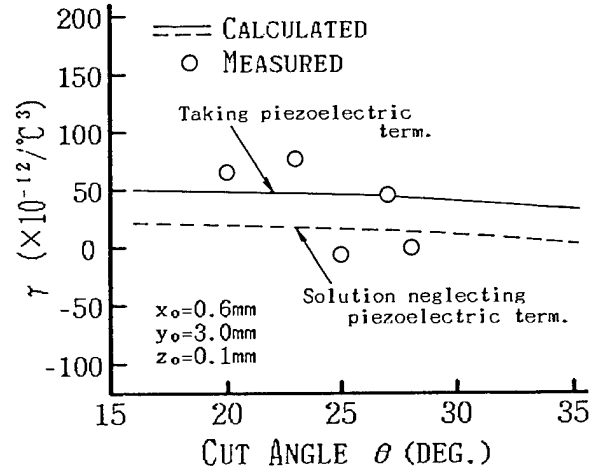


Fig. 7 Theoretical and measured values of the third order temperature coefficient γ versus cut angle θ when piezoelectric term is taken into account.

§3 Results and Discussion

According to the following order, the results are discussed. First, a variation of frequency constant ($f \cdot x_0$) versus a cut angle θ is shown when a piezoelectric term is taken into account and neglected. Next, a relationship between a cut angle θ and the first, second and third order temperature coefficients α, β, γ, especially, the influence on α, β, γ is clarified due to a piezoelectric term and an error of temperature coefficients for elastic stiffness constants and the piezoelectric term. Finally, the calculated results of frequency temperature behavior are compared with the measured data, so that both results are found to agree well with the error of a cut angle $\Delta\theta = 0.1°$ to $1.3°$. In this calculation, Aruga's constants[7] are used. In addition, Taylor's series is expanded at temperature of 20℃ and also, various characteristics that are shown below are those for fundamental vibration.

3.1 Frequency constant

Figure 4 shows a relationship between a cut angle θ and frequency constant ($f \cdot x_0$) of a width-extensional mode quartz crystal resonator (hereafter, WE resonator) with width $x_0 = 0.6$mm($f \fallingdotseq 4.5$MHz), length $y_0 = 3.0$mm, thickness $z_0 = 100\,\mu$m, $x_1 = 80\,\mu$m, $y_1 = 200\,\mu$m, $x_2 = 0.5$mm and $y_2 = 80\,\mu$m of the vibrational portion. As an electromechanical transformation efficiency for a quartz crystal resonator is low, it is predicted beforehand that the influence on frequency constant due to a piezoelectric term is slight, as is shown in Fig. 4,

Table 1 Temperature coefficients[7] of elastic stiffness constants and piezoelectric terms used in the calculation.

($\dot{c}^E_{ij}/c^E_{ij}$, $\dot{\bar{A}}/\bar{A}$, $\dot{\bar{B}}/\bar{B}$: 10^{-4}/℃, $\ddot{c}^E_{ij}/c^E_{ij}$, $\ddot{\bar{A}}/\bar{A}$, $\ddot{\bar{B}}/\bar{B}$: 10^{-8}/℃², $\dddot{c}^E_{ij}/c^E_{ij}$, $\dddot{\bar{A}}/\bar{A}$, $\dddot{\bar{B}}/\bar{B}$: 10^{-12}/℃³)

	$\dot{c}^E_{ij}/c^E_{ij}$, $\dot{\bar{A}}/\bar{A}$, $\dot{\bar{B}}/\bar{B}$			$\ddot{c}^E_{ij}/c^E_{ij}$, $\ddot{\bar{A}}/\bar{A}$, $\ddot{\bar{B}}/\bar{B}$			$\dddot{c}^E_{ij}/c^E_{ij}$, $\dddot{\bar{A}}/\bar{A}$, $\dddot{\bar{B}}/\bar{B}$		
	Mi (α)	M (α)	Ma (α)	Mi (β)	M (β)	Ma (β)	Mi (γ)	M (γ)	Ma (γ)
c^E_{11}	-0.4614	-0.4566	-0.4518	-17.69	-16.47	-15.25	-2477	-2094	-1711
c^E_{12}	-25.2679	-25.4726	-25.6773	-497.12	-549.25	-601.38	-10602	-26949	-43296
c^E_{13}	-6.0841	-6.1248	-6.1655	-140.22	-150.59	-160.96	-19048	-15795	-12542
c^E_{14}	0.9896	0.9798	0.9700	-4.07	-6.56	-9.05	-2314	-3095	-3876
c^E_{33}	-1.8812	-1.8741	-1.8670	-34.48	-32.67	-30.86	-157	-726	-1295
c^E_{44}	-1.7144	-1.7113	-1.7082	-46.78	-45.98	-45.18	-2654	-2404	-2154
$\frac{e_{11}}{\sqrt{\varepsilon^S_{11}}}$	-1.177	-0.802	-0.427	-354.6	-259.0	-163.4	64250	94230	124210
$\frac{e_{14}}{\sqrt{\varepsilon^S_{11}}}$	-4.24	-2.54	-0.84	-891	-458	-25	95500	231200	366900

$\bar{A}=e_{11}/\sqrt{\varepsilon^S_{11}}$, $\bar{B}=e_{14}/\sqrt{\varepsilon^S_{11}}$, $\bullet\equiv\partial/\partial T$ (T : Temperature)

the piezoelectric term increases approximately 0.4 percent for the frequency constant ($f \cdot x_0$). In addition, the calculated and measured values agree well with an error of 1.3 percent at maximum at a cut angle of $\theta=20°$ to $28°$.

3.2 Frequency temperature coefficients

Figure 5 shows a relationship between a cut angle θ and the first and second order temperature coefficients α, β of a WE resonator calculated by eq.(12). α reaches zero at cut angles $\theta=-72.8°$, $24.7°$, $63.7°$ and $75.9°$, while an absolute value of β represents the minimum value at $\theta=34.7°$, the value of which is $\beta=-0.29\times10^{-8}/℃^2$. In comparison to the reference[1] reported previously, the piezoelectric term has remarkable influence on β more than α. For the largeness where the piezoelectric term has an influence on α, β, a detailed description will be performed in comparison with the measured data in the vicinity of KT-cut.

Figure 6 shows a relationship between a cut angle θ and the first and second order temperature coefficients α, β of a WE resonator with the same dimension and shape as the WE resonator in Fig. 4, when a piezoelectric term is taken into account. A cut angle θ where gives $\alpha=0$ is $\theta=24.7°$ in the calculation and $\theta=27°$ in the experiments, an error of $2.3°$ takes place between them. This error is improved by $0.2°$ as compared with the error of $2.5°$ when the piezoelectric term was neglected, but, the degree of the improvement is extremely low. On the other hand, for β when $\alpha=0$, $\beta=-0.91\times10^{-8}/℃^2$ in the calculation and $\beta=-1.25\times10^{-8}/℃^2$ in the experiments, β obtained by the present analysis approaches to the measured data fairly, as compared with $\beta=-0.60\times10^{-8}/℃^2$ obtained when the piezoelectric term was neglected. Thus, the piezoelectric term has outstanding influence on β more than α. As described above, the calculated values of α, β represent an inclination to approach to the measured data, but, it is still unenough. As one of this reason, it is conceivable that there is an error of temperature coefficients for elastic stiffness constants and a piezoelectric term used in the present calculation. The temperature coefficients with the errors are discussed later.

Figure 7 shows a relationship between a cut angle θ and the third order temeprature coefficient γ of a WE resonator with the same dimension and shape as the WE resonator in Fig. 4, when a piezoelectric term is taken into consideration. The solid and broken lines indicate the results, respectively, when the piezoelectric term is taken into account and neglected. It is well understood from Fig. 7 that the piezoelectric term has remarkable influence on γ. In addition, γ has a value of $44.3\times10^{-12}/℃^3$ in the calculation at a cut angle $\theta=27°$ where gives $\alpha=0$ and $47.9\times10^{-12}/℃^3$ in the experiments. Thus, both results agree well when the piezoelectric term is taken into account. The same results as that for $\theta=27°$ are obtained for a cut angle θ of $20°$ and $23°$. However, the

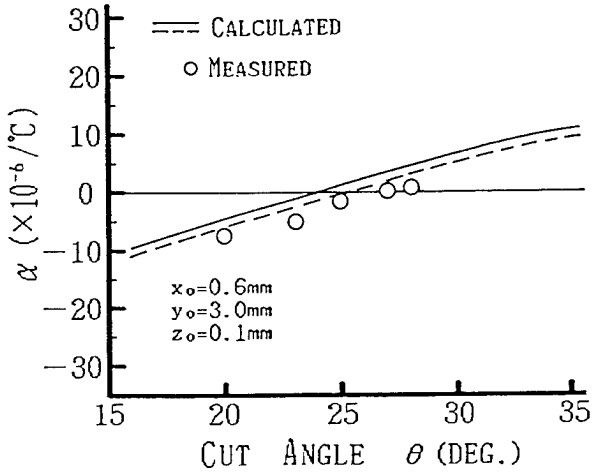


Fig. 8 Relationship between cut angle θ and the first order temperature coefficient α when errors of temperature coefficients of elastic stiffness constants and piezoelectric term are taken into account.

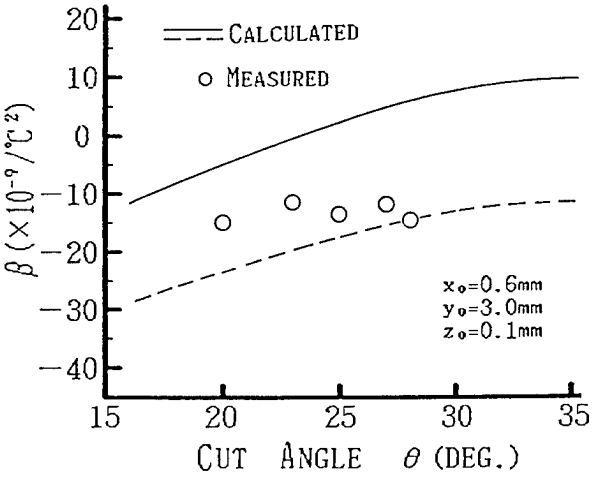


Fig. 9 Relationship between cut angle θ and the second order temperature coefficient β when errors of temperature coefficients of elastic stiffness constants and piezoelectric term are taken into account.

measured values for a cut angle θ of 25° and 28° agrees well rather with the calculated values obtained by neglecting the piezoelectric term than those obtained by taking the piezoelectric term into account. This reason is unclear, the behavior of γ will be further clarified in the experiments of resonators with another dimensions.

Table 1 shows the first, second and third order temperature coefficients of elastic stiffness constants c_{ij}^{E} and piezoelectric terms of $\overline{A}=e_{11}/\sqrt{\varepsilon_{11}^{S}}$ and $\overline{B}=e_{14}/\sqrt{\varepsilon_{11}^{S}}$ introduced by Aruga[7] that are used in the calculation. Since Aruga's constants are treated in the theory of error, the temperature coefficients of c_{ij}^{E}, $\overline{A}$ and $\overline{B}$ are represented by the symbols of Mi and Ma which indicate the minimum and maximum values of α, β, γ respectively.

Figure 8 shows a relationship between a cut angle θ and the first order temperature coefficient α of a WE resonator due to an error of the first order temperature coefficient for the elastic stiffness constants and the piezoelectric term. The solid and broken lines indicate the range of α when the error was taken into account, it is understood from Fig. 8 that the result of the broken line approaches closely to the measured values. In this case, $\alpha=0$ at $\theta=25.5°$, i.e., the cut angle θ is improved in $\Delta\theta=0.8°$ as compared with a cut angle $\theta=24.7°$ where gives $\alpha=0$ when the errors of the temperature coefficients were neglected. However, this value has still the difference of $\Delta\theta=1.5°$ compared with $\theta=27°$ where gives $\alpha=0$ in the experiments. As another reason, it is, therefore, conceivable because the lateral motion of the thickness direction that is caused by Poission's ratio was neglected.

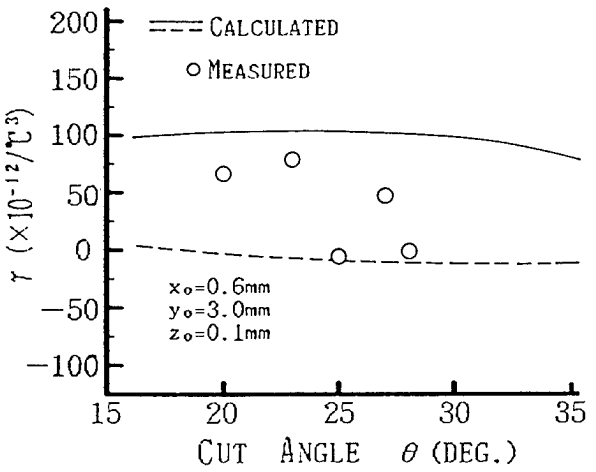


Fig. 10 Relationship between cut angle θ and the third order temperature coefficient γ when errors of temperature coefficients of elastic stiffness constants and piezoelectric term are taken into account.

Figure 9 shows a relationship between a cut angle θ and the second order temperature coefficient β of a WE resonator that is caused by an error of the second order temperature coefficient for the elastic stiffness constants and the piezoelectric term. Similar to α in Fig. 8, the solid and broken lines indicate the range of β. As is apparent from Fig. 9, the measured values of β in case of $\theta=20°$ to 28° are in the range of the calculated values when the error of the temperature coefficients is taken into consideration.

Figure 10 shows a relationship between a cut angle θ and the third order temperature coefficient γ of a WE resonator that is caused by an error of the third order temperature

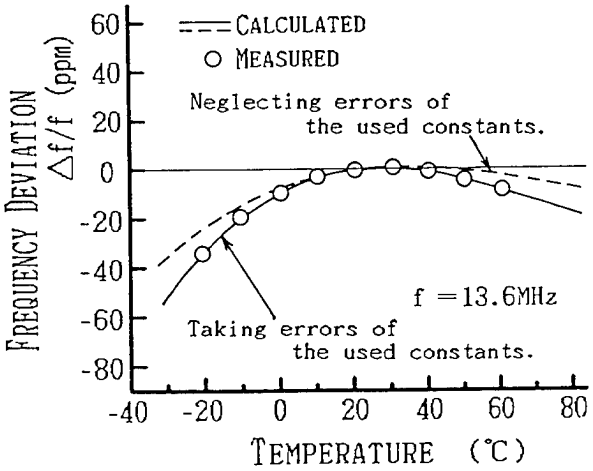


Fig. 11 Calculated frequency temperature behavior when errors of temperature coefficients of elastic stiffness constants and piezoelectric term are taken into account, and the measured data.

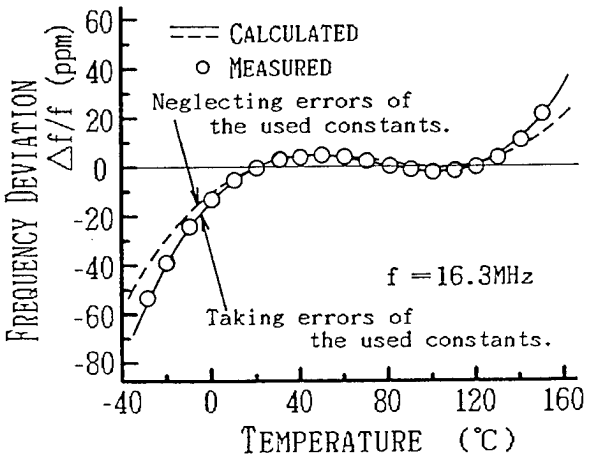


Fig. 12 Another example of calculated frequency temperature behavior when errors of temperature coefficients of elastic stiffness constants and piezoelectric term are taken into account, and the measured data.

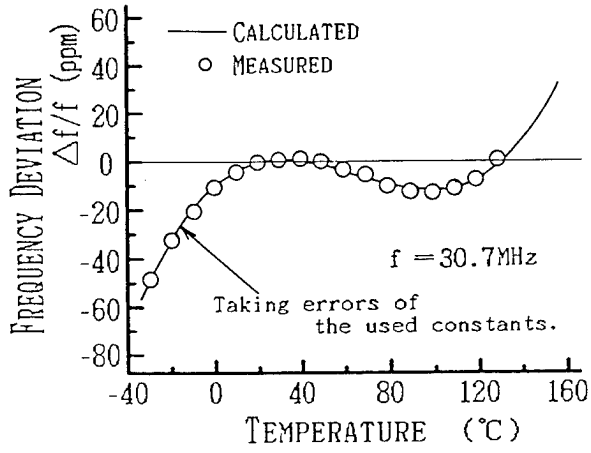


Fig. 13 Further example of calculated frequency temperature behavior when errors of temperature coefficients of elastic stiffness constants and piezoelectric term are taken into account, and the measured data.

coefficient for the elastic stiffness constants and the piezoelectric term. The measured data are entirely in the range of the calculated values obtained by taking the errors as well as β shown in Fig. 9. Thus, the measured values for β, γ are in good agreement with the calculated ones when the errors of the temperature coefficients are taken into account.

3. 3 Frequency temperature behavior

Until now, an examination was performed on the temperature coefficients of α, β, γ. From the results of the examination, it is discussed how the calculated and measured values for frequency temperature behavior cause the difference.

Figure 11 shows the frequency temperature behavior of a WE resonator with a frequency f= 13. 6MHz. This is in case that a turnover temperature point is set up in the vicinity of room temperature. The circles are the measured values and the resonator with width $x_0 = 0.2$mm, length $y_0 = 3.0$mm, thickness $z_0 = 100\,\mu$m, $x_1 = 80\,\mu$m, $y_1 = 200\,\mu$m, $x_2 = 0.5$mm and $y_2 = 80\,\mu$m (Resonator's shape(a)) is formed from a plate with a cut angle $\theta = 24°$. While the broken line is in case of $\theta = 24.8°$ and the turnover temperature point agrees approximately with that of the measured values. In this case, the difference of 0. 8° occurs in the measured and calculated values. On the contrary, the solid line indicates the calculated result when an error of temperature coefficients for the elastic stiffness constants and the piezoelectric term is taken into account. The calculated value of frequency temperature behavior for $\theta = 24.1°$ agrees well with the measured one, namely, the difference of a cut angle in the calculated and measured values is 0. 1° .

Figure 12 shows an example of frequency temperature behavior for a WE resonator(Resonator's shape (b), $x_0 = 160\,\mu$m, $y_0 = 4.3$mm, $z_0 = 100\,\mu$m, $x_1 = 50\,\mu$m, $y_1 = 200\,\mu$m, $x_2 = 0.5$mm and $y_2 = 80\,\mu$m) which is so designed that the first order temperature coefficient α has a slightly positive value in the vicinity of room temperature. The resonator has a frequency of f=16. 3MHz, the circles are the measured values for $\theta = 27°$ and the broken lines are the calculated ones for $\theta = 25.0°$. Both results agree sufficiently with an error of a cut angle θ of 2. 0° , while the solid line indicates the calculated value when an error of the temperature coefficients for the elastic stiffness constants and the piezoelectric term is taken into account. The calculated value is in very good

agreement with the measured one for $\theta = 25.9°$. In this case, an error of a cut angle θ for the measured and calculated values is 1.1° and the error of both results is further decreased.

Figure 13 shows an example of frequency temperature behavior for a WE resonator with a frequency f=30.7MHz ($x_0 = 82\mu m$, $y_0 = 4.3mm$, $z_0 = 50\mu m$, $x_1 = 50\mu m$, $y_1 = 200\mu m$, $x_2 = 0.6mm$ and $y_2 = 80\mu m$). The measured values for $\theta = 27°$ are in good agreement with the calculated value at $\theta = 25.7°$, when an error of the temperature coefficients for the elastic stiffness constants and the piezoelectric term is taken into account. In this case, an error of both results is 1.3°.

Thus, the calculated and measured values of frequency temperature behavior for the present width-extensional mode quartz crystal resonator agree well by taking into account an error of the temperature coefficients for the elastic stiffness constants and the piezoelectric term that are used in the calculation.

§4. Conclusions

In this paper, a study was performed with a view to clarifying the influence on frequency temperature behavior of a KT-cut width-extensional mode quartz crystal resonator due to an error of temperature coefficients for the piezoelectric term and the elastic stiffness constants.

First, a frequency equation and a transcendental equation are derived from input admittance which is calculated from the partial differential equation of one dimensional-approximate equation and the boundary condition derived under the condition that thickness for a WE resonator is much smaller than width and length and also, the length is much larger than the width. That is to say, the frequency equation is given as one dimensional-approximate equation.

Second, from the obtained frequency equation and the transcendental equation, one clarified the influence on frequency constant and frequency temperature coefficients due to a piezoelectric term, so that the piezoelectric term has very slight influence on the frequency constant and the first order temperature coefficient, rather, noticeable influence on the second and third order temperature coefficients. In addition, since it is conceivable that the difference of the calculated and measured values for the first, second and third order temperature coefficients α, β, γ is due to an error of temperature coefficients for the elastic stiffness constants that are used in this calculation, the influence on α, β, γ due to the error was examind. As a consequence, it was shown that the error has remarkable influence on β, γ and also, that the measured values of β, γ are in the region of the calculated ones when the error of the temperature coefficients for the elastic stiffness constants and the piezoelectric term was taken into account.

Finally, the calculated values of frequency temperature behavior were compared with the measured ones for a WE resonator with a frequency of fs=13.6, 16.3 and 30.7MHz, when an error of temperature coefficients for the elastic stiffness constants and the piezoelectric term that are used in this calculation was taken into account. The difference of both results is 0.1°, 1.1° and 1.3°, respectively, the calculated and measured values were found to be in sufficiently good agreement by taking the error of the temperature coefficients into account.

A coming subject is to clarify the influence on frequency temperature behavior due to lateral motion, especially, in the thickness direction.

Acknowledgements

The author wishes to thank Prof. T. Ikeda of Kanagawa University for his kind advise on an analysis method of this study and K. Sunaga for his help in the numerical calculation by computer.

References

[1] H. Kawashima: "New Cuts for Width-Extensional Mode Quartz Crystal Resonators", Proc. the 1992 IEEE Freq. Control Symposium, pp. 525-531 (1992).

[2] H. Kawashima: "A New Cut of Width-Extensional Mode Quartz Crystal Resonators", Trans. IEICE (A), Japan, J76-A, 4, pp. 572-578(1993).

[3] H. Kawashima: "Frequency Temperature Behavior and Electrical Characteristics of New Cuts for Width-Extensional Mode Quartz Crystal Resonators", IEICE, Japan, US92-23, pp. 69-75(1992).

[4] H. Kawashima: "A Variational Analysis of GT Cut Quartz Crystal Resonators with the Supporting Portions at Both Ends", Trans. IEICE (A), Japan, J68-A, 8, pp. 771-778(1985).

[5] H. Kawashima: "Variational Analysis of GT Cut Quartz Crystal Resonators with the Supporting Portions at the Ends", Proc. the 40th Annual Symp. on Frequency Control, pp. 193-200(1986).

[6] H. Kawashima: "The Influence on Frequency Temperature Behavior due to the Supporting Portions of GT Cut Quartz Crystal Resonators Formed by an Etching Process", Trans. IEICE(A), Japan, J73-A, 6, pp. 1031-1038(1990).

[7] M. Aruga: "Adiabatic Elastic Stiffness Constants and Their Temperature Characteristics for Quartz Crystal Resonators", Bulletin of Tokyo Institute of Technology, A-2, pp. 88-182 (1956).

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

APPLICATION OF TC-CUT QUARTZ RESONATORS EXCITED BY LATERAL ELECTRICAL FIELD

I.V.Abramson, A.N.Dikidzhi, A.I.Kutalev, S.V.Plachotin

Omsk Research Instrumentation Institute, Omsk 644063, Russia

Abstract

Subject of this paper is designing TC-cut crystals for S-TCXOs that don't contain special means to reduce the temperature instability. With this object in view single, two-crystal and three-crystal TC-cut units have been developed and studied. The units show in temperature range (-10 +55)°C frequency instability from several ppm for single crystal down to $\pm 0.3 \cdot 10^{-6}$ for three-crystal units.

Moreover XOs using the developed resonators are described. Owing to excellent characteristics of the units the oscillators have short-term stability and phase noise level close to those of miniature OCXO, while their size, consumpted power and warm-up time are identical with the parameters of TCXOs.

Introduction

TC-cut crystals (YXbl/13°54'/-32°) have properties close to that of RT-cut crystals (YXbl/15°/-34°): smooth temperature-frequency dependence and very high Q-factor. However they are more convenient in X-ray orientation owing to presence of atomic plane 13.3 with XX' angle 13°54'.

In spite of the fact that TC- and RT-cuts are known for many years (RT-cut have been suggested by Bechman in 1962, and TC-cut in 1975 [1]) they are not used till now. The reason is unsatisfactory spectrum containing besides useful C-mode spurious B-mode being much more active than the first one and arranged close to it in frequency (about 7% higher). Suppression of the unwanted resonance by circuit means results in complexity in XO tunning as well as degradation in its stability.

New prospects for usage of the TC-cut crystals appear with using the lateral excitation. As indicated in [2] properly designed TC-cut crystals with lateral excitation operating on the fifth overtone and the fundamental mode have good electrical parameters and sufficiently suppressed B-mode.

As it follows from studies of the TC-cut resonator performances [2] they can be used successfully as in high stability OCXOs (when excited on the fifth overtone) as in simple XOs without circuit means to reduce the temperature instability (when excited on the fundamental mode).

It has been known that the temperature stability of XOs can be considerably improved by using two-crystal stabilization where the interaction between couple of the crystals having appropriate frequency vs. temperature characteristics is exploited. The two-crystal stabilization technique was first suggested by Koga in 1936 [3] and then subsequently studied by other scientists [4], [5]. However the method has not found practical application because of absence of the quartz crystals that possess necessary properties - smooth temperature-frequency characteristic, which could be easy moved along temperature axis by changing the plate's orientation angle and would be repeatable in production. Apparently the TC-cut crystals with lateral excitation that meet the requirements most fully provide new opportuneties for this technique.

Goal of this work is development and exploration of multiple resonators containing two or three coupled TC-cut crystals as well as designing XOs using the developed resonators.

0-7803-0905-7/93 $3.00 © 1993 IEEE

Self-compensated TC-cut resonators.

Single-crystal units.

TC-cut crystals have smooth f vs. T characteristics in a form of regular parabola (Fig. 1).

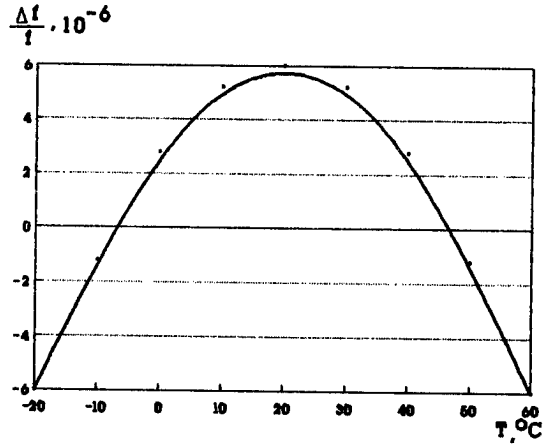


Fig. 1. Temperature-frequency characteristic of TC-cut crystal.

The turnover temperature To can be shifted along T axis by changing the angle of the plate orientation β (angle of the plate rotation arround the X'-axis). Since dependance To vs. β is not strong - about 1°C per 1' that provides a moderate dispersion of the temperature-frequency characteristics (usually within (2÷3)°C).

As one can see when a nominal frequency is tuned to be exactly between the maximum and minimum frequencies over a temperature range the resonator has minimum temperature instability. A temperature instability over any temperature range ΔT is described by expression

$$\frac{\Delta f}{f} = \pm 10^{-9}\cdot\Delta T^2 \qquad (1)$$

and is shown in Fig. 2 (curve 1). On comparing the data from Fig. 2 with those for AT-cut crystals one can argue that TC-cut crystals provide the temperature stability to be unattainable for AT-resonators taking into account their f vs. T dependance to be more susceptable to the plate orientation errors.

Electrical parameters of TC-cut resonators operating on frequencies 5 and 10

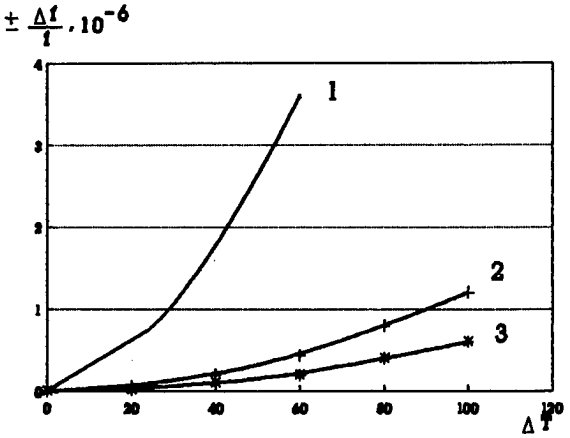


Fig. 2. The frequency instability of resonator versus operating temperature range: 1-single crystal; 2-two-crystal resonator;3-three-crystal resonator.

MHz on the fundamental mode C are given in table 1.

Table 1.

Frequency, MHz	Parameters on the C-mode			Resistance on the B-mode
	Rq, Ohms	Lq,H	Q,10^6	
5	70	6	2.7	120
10	75	1.9	1.6	140

It's evident from the table that crystals have moderably suppressed B-mode and not-too-large motional resistance on C-mode that allows to use them in simple oscillator's networks.

Two-crystals units.

Network of two-crystal unit is illustrated in Fig. 3. Two quartz crystals of TC-cut with identical electrical parameters and shifted along T - axis temperature-frequency characteristics are connected in parrallel. In series with them there is placed capacitor providing the couple between the crystals and being of greate concern in shaping the resulting temperature-frequency dependance.

The unit has two resonance frequencies that differ from partial frequencies of the separate crystals. The first of the frequencies $\omega 1$ is between series resonance frequencies of the crystals. The second one $\omega 2$ is between series resonance highest

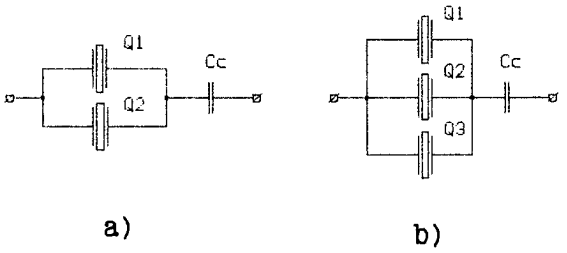


Fig. 3. Networks of TC-cut multiple resonators: a) two-crystal resonator; b) three-crystal resonator.

frequency of the separate crystals and the highest parallel resonance frequency.

To provide mutual temperature compensation of the TC-cut crystals the two-crystal unit must operate on frequency ω_2 that can be easy realized at properly adopted the capacitance value as well as by limiting an operating temperature range within resonable boundaries.

The f vs. T dependances of the double crystal unit are shown in Fig. 4 at different values of capacitance Cc. It's evident that characteristics are much more smooth than those of separate crystals. As capacitance Cc decreases a couple between the crystals increases that results in the

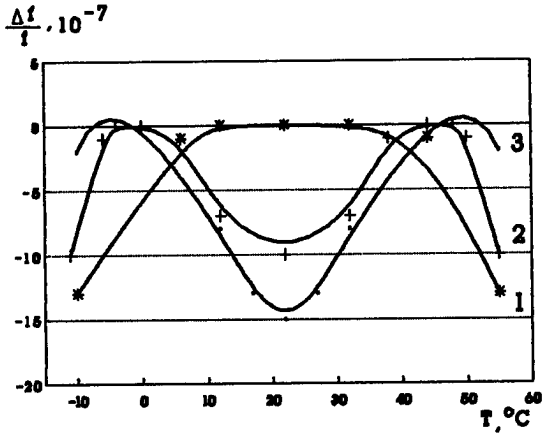


Fig. 4. The frequency versus temperature dependance of TC-cut two-crystal resonator: 1.Cc=5pF; 2.Cc=10pF; 3.Cc=18pF.

curve shape become more flat. At certain capacitance value the f vs. T curve has only one extremum (curve 1).

As it follows from Fig. 4 stability about $\pm 5 \cdot 10^{-7}$ over a temperature range (-10 +55)°C is attainble. For another temperature range an ultimate f vs. T stability of two-crystal unit can be defined as:

$$\frac{\Delta f}{f} = \pm\ 1.6 \cdot 10^{-10} \cdot \Delta T^2, \qquad (2)$$

where $\Delta T = T_{o1} - T_{o2}$, that is difference between turnover temperatures of the crystals.

Dependance (2) is shown in Fig. 2 and exibites strikly advantages of two-crystal stabilisation over single-crystal one.

The motional resistance of the TC-cut two-crystal units depends on electrical parameters of the separate crystals, their frequency separation and the coupling capacitance value. As the second parameter is function of temperature the resistance is also temperature dependent (Fig. 5). From theoretical analisis and experimental data it follows that minimum resistance value occurs in the midle of a temperature range where the crystals' frequencies coinside. The top value takes place at temperature range boundaries where the frequency distance attaines a maximum. For different values of the capacitance Cc the limits of the resistance deviation can be defined as:

$$R_{max} = R_o(1 + C_o/C_c)^2,$$
$$R_{min} = R_{max}/2\ , \qquad (3)$$

where Ro-resistance of the single crystal; Co - parallel capacity of the multiple resonators being of 1.5-2 pF.

Q-factor of the TC-cut two-crystal units measured with phase methode and vibration amplitude attenuation technique appears to be identical with that for single crystals (table 1). So, the motional inductance repeating the motional resistance variation from coupling capacitance value and the crystals' frequencies separation may be below the single crystal inductance. Tipically it's about 4 H for 5 MHz resonator and about 1 H for 10 MHz resonators.

Tree-crystal unit.

Three-crystal unit network similar to

that of two-crystal unit contains TC-cut crystals connected in parallel and a coupling capacitor placed in series with the crystals (Fig. 3).

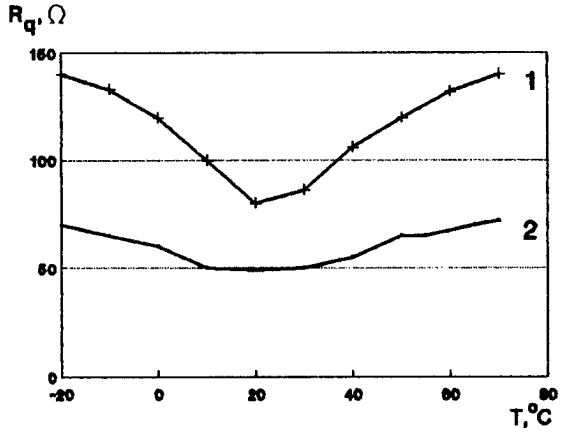


Fig. 5. The motional resistance versus temperature for TC-cut multiple resonators: 1.two-crystal resonator (Cc= 8pF); 2. three-crystal resonators (Cc=33pF).

That unit has three resonance frequencies during which the highest one arranged between highest series resonance frequency of the crystals and the highest parallel resonance frequency is useful. Like case of two-crystals compensation a steady operation on this resonance is easy obtained at not-too-large Cc value and at limited crystal frequency separation.

Resumptive f vs. T curves are displayed in Fig. 6 at different Cc values. As a temperature distance between separate curves is here far less than in case of two-crystal unit the temperature instability is reduced considerably and can be achieved at the more weak coupling.

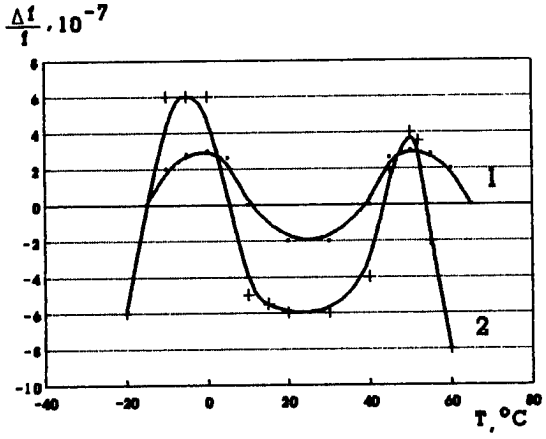


Fig. 6. The frequency versus temperature dependance of TC-cut three-crystal resonator: 1. Cc=23pF; 2. Cc=33pF.

The unit ultimate temperature instability vs. a temperature range is indicated in Fig. 2 (curve 3) to be about 2 times lower than for two-crystal units. In not too-wide temperature range (-10 +55)°C the instability is less than $\pm 3 \cdot 10^{-7}$, that corresponds to the figures of TCXOs.

The motional resistance of three-crystal units varies with temperature (Fig. 6), during which the maximum value being defined by expression (2) is always less than one for two-crystal units due to larger value of the coupling capacitance Cc at almost identical Co (Co here is about 2÷2.5 pF). The mimimum resistance in the midle of temperature range is lower than for two-crystal units owing to operation of the third crystal.

Q-factor of three-crystal resonators like two-crystal instance ramaines the value of TC-cut single crystals. Then motional inductance repeating the resistance variations is always less than one of two-crystal unit (usually about 3.5 H for 5 MHz resonator).

The multiple units construction

TC-cut resonator considered in the paper have been realized in construction containing one, two or three quartz plates placed in single vacuum glass holder. The integration of the crystals in single volume provides a minimum of temperature difference between the plates during ambient temperature changes that results in high temperature-dynamic stability of the resonator.

Proccess of the resonator production is conventional for TC-cut crystals excepting a procedur of final tuning the frequency which in the multiple resonators must meet more rigit accuracy requirements.

Designing the S-TCXOs based on the TC-cut multiple resonators.

The TC-cut units' properties considered above allow to develop XOs possessing a set of characteristics to be unattainable with other kinds of oscillators.

The major of the properties is very flat f vs. T dependance that permits to obtaine temperature stability to be tipi-

cal of TCXOs without circuit means of the compensation. That leads to essential simplification of the oscillator network and its tunning procedure.

A network of the oscillator is shown in Fig. 7. The circuit contains 4 transitors and consists of Pirce type oscillator

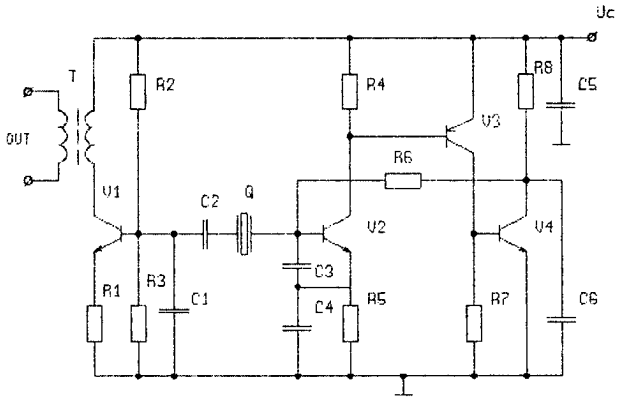


Fig. 7. Network of oscillator using the TC-cut multiple resonators.

(on V2) vibration amplitude autocontroller (on V3, V4) and amplifier (on V1). As there are not voltage sensible components in the circuit (to be inherent in TCXO) the oscillator has small f vs. input voltage instability (about $2 \cdot 10^{-9}$ 1/V) even in absence of special voltage stabilisator.

As motional inductance of the units is very large - (1÷3) H (50 - 100 time greater than one of AT-cut crystals usually used in TCXOs) the oscillator frequency is essentially inresponsive to changes in the circuit components parameters under ambient factors and aging.

Furthermore high Q-factor and large inductance of a resonator are essential prerequisite for achievement of low phase noise level of a XO.

As it has been obtained from the theoretical analisise of the oscillators' networks to be similar to one in Fig. 8. white phase noise spectral density is decribed by expression:

$$L(f) = \frac{F \cdot k \cdot T \cdot (R_q + r)}{8\pi^2 f^2 I_q^2 L_q^2} + \frac{2\,F \cdot k \cdot T}{I^2 r}, \quad (4)$$

where F - the noise factor of the active device; k - Boltzmann's constant; T - temperature (in $^{\circ}$K); f - the frequency offset from the carrier; I - current through resonator; Rq - motional resistance of resonator; Lq - motional inductance of resonator; r - real-part impedance on the amplifier input considered in the resonator loop.

The first term describing white frequency noise is defined by resonator parameters and degrades fastly with frequency.In contrast the second term - white phase noise floor - is defined only by power dissipated on a resonator.

It's evident from the equition when current through a resonator is limited by low aging requirements the noise level near the carrier can be reduced with raising motional inductance and loaded Q of resonator.The white phase noise floor declines with raising r but that leads to unwanted increasing the first term of equition (4).

Thus high Q-factor and large inductance of resonator are nessesary prerequisite to achieve spectral purity of XOs. In fact that parameters of the TC-cut units are close to those of 5 and 10 MHz third-overtone SC-cut crystals. So,the S-TCXO using the developed resonators can have the noise characteristics to be at the level of high stability OCXOs.

The typical phase noise characteristic of the developed XOs is shown in Fig. 8 to have the noise level of - 150 dBc/Hz at offset frequency 100 Hz and white phase noise floor about - 158 dBc/Hz.

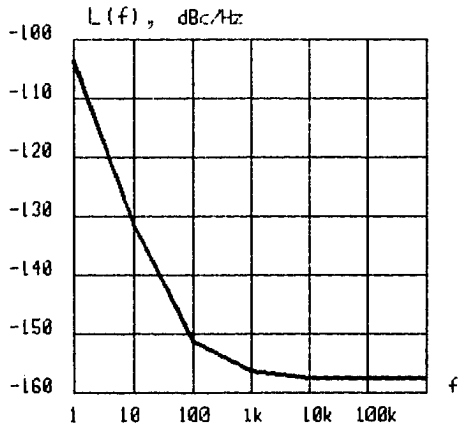


Fig. 8. Phase noise level characteristic of S-TCXO based on the multiple resonators.

Short-term instability of the developed oscillators being in inverse proportion with Q of resonator is found to be

less than $1 \cdot 10^{-12}$ per 1 s that is about the instability of high stability OCXOs.

To sum it up the most important characteristics of S-TCXOs designed on the base of the TC-cut multiple resonators are depicted in table 2.

Table 2.

Characteristics	S-TCXOs		
	two-crystal	three-crystal	two-crystal
Frequency, MHz	5	5	10
Aging/Year, 10^{-7}	2	2	3
Temp.stab.in range (-10 +55)°C, 10^{-6}	1.0	0.5	1.0
Stability, $\sigma(\tau)$ (τ = 1 s)	$1 \cdot 10^{-12}$	$1 \cdot 10^{-12}$	$1 \cdot 10^{-12}$
Size (cm^3)	10	20	10
Warmup Time (s)	1	1	1
Power (mW)	45	45	45

Comparation of data in table 2 and Fig.8 with similar data for different kinds of oscillators [6] indicates that developed devices in temperature stability, warm-up time, power consumption and sizes are identical with TCXOs while in short-term stability and phase-noise level they are close to high-stability OCXOs.

It's important to emphasize that none of the modern oscillators have the combination of properties to be inherent in the S-TCXOs. So, application of the developed devices in portable communication systems is promissing to improve their basic technical characteristics.

Conclusion.

The TC-cut multiple resonators in not-too-wide temperature range provide the f versus T stability to be identical with that of TCXOs. This fact along with comparatively low motional resistance and good spectrum of the resonators allow to design simple and miniature XOs having low consumption and momentary setting a frequency. At the same time extremely high Q-factor and large inductance of the resonators permit to obtaine noise level and short term stability close to those of high stability OCXOs.

As a main part of cost of the S-TCXOs is due to the multiple quartz resonator an important problem is improvements in technology of their manufacturing and tuning.

Acknowledgement.

The authors express a gratitude to Ms. N. Kuchatova for help in the experiments and preparing the paper.

Reference.

[1] С.Ф. Травкина, А.Е. Караульник,В.Шин, М.И.Ярославский, "Способ изготовления кварцевых резонаторов", А.С.N 474906, 1975.

[2] I.U. Abramson, "Characteristics of TC-cut crystal units excited by lateral field", in Proceedings of 7th European Frequency and Time Forum, Neuchatel, 1993.

[3] J. Koga, "Notes on Piezoelectric Quartz Crystal. "Proc. IRE", 1936, v.24, N3.

[4] В.Т. Алдабаев "Двухкварцевый генератор на резонаторах с квадратичными температурно-частотными характеристиками", Труды НИИТП, N3, 1971.

[5] D. Fewing and C. Ince, "A self-compensating Crystal Oscillator", "Marconi Review", vol. 31, N169, 1968.

[6] J.R. Vig, "Introduction to Quartz Frequency Standarts", Research and Development Technical Report SLC-TR-92-1.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THIN ROTATED Y-CUT QUARTZ RESONATORS VIBRATING IN B-MODE OVER A WIDE TEMPERATURE RANGE

MITSUO NAKAZAWA*, MAKOTO TAKEUCHI*, HIDEKI YAMAGUCHI**
AND ARTHUR BALLATO***

* SHINSHU UNIVERSITY (500 Wakasato, Nagano-shi, 380,Japan)
** MIYOTA Co.Ltd. (4107-5 Miyota, Nagano-ken, 380-02, Japan)
*** US ARMY ELECTRONICS LABORATORY
(Fort Monmouth, NJ 07703-5601, U.S.A)

Abstract

The paper describes both new theoretical analyses and experimental results in order to develop the thin rotated Y-cut quartz resonators vibrating in b-mode which could show reliable frequency-temperature characteristics under the high temperature region.

1 Introduction

It is well known that the single crystal alpha-quartz is one of the most important materials used for resonators, because it has excellent properties in elastic, piezoelectric, chemical, optical and thermal fields.

In this paper, a study is presented of the fundamental equation between the turnover temperature and the angle of cut for the thin rotated Y-cut quartz crystal resonator vibrating in b-mode over a wide temperature range. We show both theoretical and experimental results concerning the frequency-temperature characteristics of the crystal resonators, in which Taylor expansion theory is used.

The resonator vibrating in b-mode has a greater frequency constant than one of the quartz crystal resonator vibrating in c-mode. This is an important advantage at the highest frequency. Furthermore, the turning point in the frequency-temperature curves is both easily and linearly controlled by adjusting the cut angle. In particular, this paper describes the resonators having reliable frequency-temperature characteristics in the high temperature region.

2 Theory

From Ref.(1) we have the following equation for the turnover temperature T vs the doubly rotated cut angles ϕ and θ of the crystal resonators:

$$\begin{aligned}&\tfrac{1}{2}d_{kkl}(\theta-\theta_0)^2\\&+\{(d_{kl}+d_{kll}(T-T_0)+d_{hkl}(\phi-\phi)\}(\theta-\theta_0)\\&+\left\{d_{lh}+d_{llh}(T-T_0)+\tfrac{1}{2}d_{lhh}(\phi-\phi_0)\right\}(\phi-\phi_0)\\&+\left\{\alpha+\beta(T-T_0)+\tfrac{1}{2}\gamma(T-T_0)^2\right\}=0,\end{aligned} \tag{1}$$

where

$$\alpha = d_l = \left[\frac{1}{f}\frac{\partial f}{\partial T}\right]_{\phi_0,\theta_0,T_0},$$

$$\beta = d_{ll} = \left[\frac{1}{f}\frac{\partial^2 f}{\partial T^2}\right]_{\phi_0,\theta_0,T_0},$$

$$\gamma = d_{lll} = \left[\frac{1}{f}\frac{\partial^3 f}{\partial T^3}\right]_{\phi_0,\theta_0,T_0},$$

$$d_h = \left[\frac{1}{f}\frac{\partial f}{\partial \phi}\right]_{\phi_0,\theta_0,T_0},$$

$$d_{hh} = \left[\frac{1}{f}\frac{\partial^2 f}{\partial \phi^2}\right]_{\phi_0,\theta_0,T_0},$$

$$d_{hhh} = \left[\frac{1}{f}\frac{\partial^3 f}{\partial \phi^3}\right]_{\phi_0,\theta_0,T_0},$$

$$d_k = \left[\frac{1}{f}\frac{\partial f}{\partial \theta}\right]_{\phi_0,\theta_0,T_0},$$

$$d_{kk} = \left[\frac{1}{f}\frac{\partial^2 f}{\partial \theta^2}\right]_{\phi_0,\theta_0,T_0},$$

$$d_{kkk} = \left[\frac{1}{f}\frac{\partial^3 f}{\partial \theta^3}\right]_{\phi_0,\theta_0,T_0},$$

0-7803-0905-7/93 $3.00 © 1993 IEEE

$$
\begin{aligned}
d_{hk} &= \left[\frac{1}{f}\frac{\partial^2 f}{\partial\phi\partial\theta}\right]_{\phi_0,\theta_0,T_0}, \\
d_{kl} &= \left[\frac{1}{f}\frac{\partial^2 f}{\partial\theta\partial T}\right]_{\phi_0,\theta_0,T_0}, \\
d_{lh} &= \left[\frac{1}{f}\frac{\partial^2 f}{\partial T\partial\phi}\right]_{\phi_0,\theta_0,T_0}, \\
d_{hhk} &= \left[\frac{1}{f}\frac{\partial^3 f}{\partial\phi^2\partial\theta}\right]_{\phi_0,\theta_0,T_0}, \quad \cdots . \qquad (2)
\end{aligned}
$$

Now, we investigate the thin rotated Y-cuts vibrating in b-modes as shown in Fig.1, therefore, we may set as

$$
\begin{aligned}
&\phi_0 = 0, \quad \theta_0 = 0, \\
&\Delta\phi = \phi - \phi_0 = 0, \quad \Delta\theta = \theta - \theta_0 = \theta.
\end{aligned} \qquad (3)
$$

From equations (1)-(3), we have the following quadrature relations as

$$
\begin{aligned}
a\theta^2 &+ 2h\theta\Delta T \ + b\Delta T^2 + 2g\theta + 2f\Delta T + c \\
&= \sum_{i,j=0}^{2} a_{ij}x_i x_j = 0, \qquad (4)
\end{aligned}
$$

or

$$
(\Delta T + f_{11}\theta + g_{11})(\Delta T + f_{22}\theta + g_{22}) + k = 0, \quad (5)
$$

where

$$
\begin{aligned}
a &= \frac{d_{kkl}}{\gamma} \equiv a_{11}, \\
b &= 1 \equiv a_{22}, \\
c &= \frac{2\alpha}{\gamma} \equiv a_{00}, \\
2f &= \frac{2\beta}{\gamma} \equiv 2a_{02}, \\
2g &= \frac{2d_{kl}}{\gamma} \equiv 2a_{10}, \\
2h &= \frac{2d_{kll}}{\gamma} \equiv 2a_{12},
\end{aligned} \qquad (6)
$$

$$
\begin{aligned}
&a_{ij} = a_{ji}, \\
&x_0 = 1, \quad x_1 = \theta, \quad x_2 = \Delta T, \\
&f_{11} = h + \sqrt{h^2 - a}, \quad f_{22} = h - \sqrt{h^2 - a}, \\
&g_{11} = f + \sqrt{f^2 - (c - k)}, \\
&g_{22} = f - \sqrt{f^2 - (c - k)}, \\
&k = c - f^2 - \frac{(g - fh)}{(a - h^2)}^2, \quad \Delta T = T - T_0.
\end{aligned} \qquad (7)
$$

Now, from equation (4) we obtain the following results:

$$
\Delta T = -(f + h\theta) \pm \sqrt{(f + h\theta)^2 - (a\theta^2 + 2g\theta + c)}, \qquad (8)
$$

or

$$
\begin{aligned}
&\Delta\theta = \theta = \tfrac{1}{a}\{-(g + h\Delta T) \\
&\pm\sqrt{(g + h\Delta T)^2 - a(b\Delta T^2 + 2f\Delta T + c)}\}. \qquad (9)
\end{aligned}
$$

If rearranged with $\theta = 0$, equation (7) coincides with the old one. In equations (1),(4) and (5), the constants $a, b, c, f, g, h, f_{11}, g_{11}$ etc. are the ones defined by the Taylor's expansion series. According to the thickness vibration theory, the frequency-temperature coefficients α, β, and γ show the non-linear change with the angle of cut for the quartz crystal resonator varying.

From the universality of the quadrature for the coefficients a_{ij} in equation(4), at first, we can choose the most suitable values for α, β and γ from the experimental results, and after that the values a, b, g and h are decided experimentally.

3 Experiments

The typical specification of the crystal resonators are as follows:

- kinds of crystal resonator: rotated Y-cut disk,
- cut angle:
  (yxwl) ϕ/θ, $\phi = 0°$, $\theta = -40°\ 18' \pm 2'$,
- frequency constant for b-mode:
  $f \cdot t = 2.4920$ MHz,
- series resonance frequency: $f = 25.3949$ MHz,
- thickness of crystal resonator: $t = 98.13\mu$m,
- surface roughness: GC ♯3000,
- electrode: (Cr + Ag) film,
- capsule: HC-49/U.

Some features of the equivalent electric circuits for the crystal resonator are as follows at 25 ℃:

- Q value: Q=93560,
- capacitance raito: $r = C_0/C_1 = 465$,
- series resonance registance: $R_1 = 5\Omega$.

The block diagram for the observed oscillation frequency-temperature characteristics is shown in Fig.2.

Figs.3 and 4 show the frequency-temperature characteristics for Nos,3 and 5, respectively.

4 Calculation

From the observed oscllation frequency-temperature characteristics for the rotated Y-cut disk resonator vibrating in thickness-shear b-mode as shown in Figs.5 and 6, we obtain the following values of **least-squares** method at $T_0 = 30$ ℃:

$$\begin{aligned} \alpha &= 1.892 \times 10^{-5}[℃]^{-1}, \\ \beta &= -2.569 \times 10^{-8}[℃]^{-2}, \\ \gamma &= -6.582 \times 10^{-10}[℃]^{-3}. \end{aligned} \tag{10}$$

Hence, we have the following results:

$$\begin{aligned} 2a_{02} &= \frac{2\beta}{\gamma} = 78.06, \\ a_{00} &= \frac{2\alpha}{\gamma} = -57490.10. \end{aligned} \tag{11}$$

From equations (4), (6) and experimental results, we obtain the following equations:

$$\begin{pmatrix} \theta_1^2 & \theta_1\Delta T_1 & \theta_1 \\ \theta_2^2 & \theta_2\Delta T_2 & \theta_2 \\ \theta_3^2 & \theta_3\Delta T_3 & \theta_3 \end{pmatrix} \begin{pmatrix} a_{11} \\ 2a_{12} \\ 2a_{10} \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}. \tag{12}$$

Hence,

$$\begin{pmatrix} a_{11} \\ 2a_{12} \\ 2a_{10} \end{pmatrix} = \begin{pmatrix} \theta_1^2 & \theta_1\Delta T_1 & \theta_1 \\ \theta_2^2 & \theta_2\Delta T_2 & \theta_2 \\ \theta_3^2 & \theta_3\Delta T_3 & \theta_3 \end{pmatrix}^{-1} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}, \tag{13}$$

where

$$\begin{aligned} &A_i = -(\Delta T_i^2 + 2a_{02}\Delta T_i + a_{00})\Delta T_i = T_i - T_0, \\ &T_0 = 30\ ℃ \qquad for \quad i = 1, 2, 3. \end{aligned} \tag{14}$$

From Ref.(2), we have the following experimental results:

Table 1. Experimental results for the rotated Y-cut resonators

cut angle $(\theta_i)(degrees)$	turnover temp. T_i(℃)	temp. difference ΔT_i(℃)
-49.15	20.00	-10.00
-47.20	64.47	34.47
-46.10	90.00	60.00

From equation (4), (6) and experimental results, we obtain the following equations:

$$\begin{pmatrix} a_{11} \\ 2a_{12} \\ 2a_{10} \end{pmatrix} = \begin{pmatrix} -2109.89 \\ 93.59 \\ -103948.60 \end{pmatrix} \tag{15}$$

and

$$\begin{aligned} &f_{11} = 112.36, \quad f_{22} = -18.78, \\ &g_{11} = 859.54, \quad g_{22} = -781.48, \quad k = 614215.03. \end{aligned} \tag{16}$$

Furthermore, it is found that the following conditions are satisfied:

$$\Delta = \begin{vmatrix} a & h \\ h & b \end{vmatrix} < 0$$

and

$$D = \begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix} \neq 0. \tag{17}$$

Therefore, the θ-ΔT characteristics curve expressed by equation (4) is a hyperbola.

Similarly, from our experimental results and Ref.(2), we have the following Tab.2:

Table 2. Experimental results for thin rotated Y-cut resonators

cut angle θ_i(degrees)	turnover temp. T_i(℃)	temp. difference ΔT_i(℃)
-49.15	20.00	-10.00
-46.10	90.00	60.00
-40.30	232.00	202.00

Hence, we obtain the following results:

$$a_{11} = -2090.21, \quad 2a_{12} = 92.73,$$
$$2a_{10} = -102989.80,$$
$$f_{11} = 111.48, \quad f_{22} = -18.75 \quad g_{11} = 857.65,$$
$$g_{22} = -779.59, \quad k = 611130.19 \quad . \tag{18}$$

What is more, we have

$$\Delta = \begin{vmatrix} a & h \\ h & b \end{vmatrix} < 0, \tag{19}$$

and

$$D = \begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix} \neq 0. \tag{20}$$

It is found that the numerical values in equations (14) and (15) are nearly the same ones as compared with equation (17).

From equations (4)-(18), we have Figs.5 and 6. In Fig.5 we can predict the turnover temperature T in the range of $\theta \doteq -50^\circ \sim -30^\circ$, and $\theta \doteq -0.5^\circ \sim -0^\circ$.

From these pictures we can easily predict the cut angle of the reliable frequency-temperature characteristics for any high temperature range such as 200 ℃. By using equation (8), Figs.5 and 6, it is found that when $\theta \doteq -41^\circ\ 33'$ or $\theta \doteq -0^\circ\ 11'$ in thin rotated Y-cut, the quartz resonator shows the turnover temperature at T=200 ℃. Hence, these resonators will be designed with a frequency tolerance of ±3 ppm over the temperature range 195 ℃ to 205 ℃.

5 Conclusion

We have carried out new theoretical analyses in order to develop the quartz resonators which could show reliable frequency temperature characteristics under the high temperature region.

Our theory coincides well with the experimental results.

This work was supported by San-ei Denshi Kogyo Co. Ltd., Miyota Co, Ltd., Shin-etsu Kagaku Kogyo Co. Ltd. and Kinseki Co. Ltd. in Japan.

The authors wish to thank Dr.Hitohiro Fukuyo, Professor Hiroshi Takahashi of Shinshu University and Messrs. Atsushi Araki, Takao Umeda for their help and suggestions.

References

1. M. Nakazawa, M. Takeuchi, T. Moriizumi, H. Niiyama, A. Ballato and T. Lukaszek: A theoretical study of quartz crystal resonators for chemical sensors, vibrating in c-mode, 1992 IEEE Frequency Control Symposium, pp. 610-613, May. 1992.

2. I. Koga: Piezoelectricity and high frequency (in Japanese), Ohm Co. Ltd., 1937; T. Shinada: Theory and practices for quartz crystal oscillator (in Japanese) , Ohm Co. Ltd., 1955: R. Bechmann, A. Ballato and T. Lukaszek: Higher order temperature coefficients of the elastic stiffnesses and compliances of alpha-quartz, U.S. Army ERAD Laboratory, 1963; W. G. Cady: Piezoelectricity, Dover, 1964.

3. M. Nakazawa, M. Takeuchi, Xu Guanghua, T. Moriizumi and H. Niiyama: A theoretical study of the quartz crystal resonators vibrating in b-mode, and having reliable frequency-temperature characteristics in the region of higher temperature than 200 ℃, '92 H-CJS, 1992.

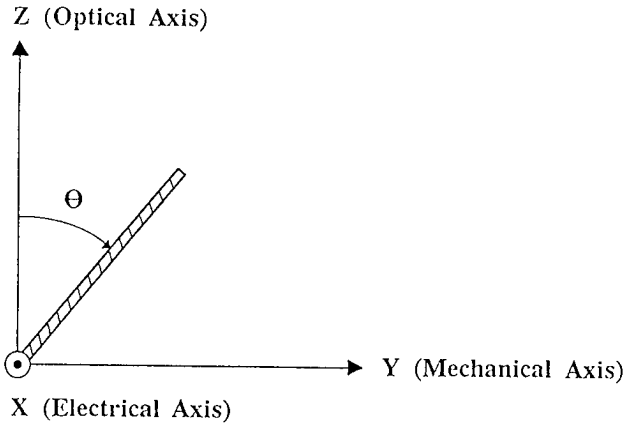


Fig.1. **Right-handed Cartesian coordinate system and rotated Y-cut quartz plate vibrating in the b-mode.**

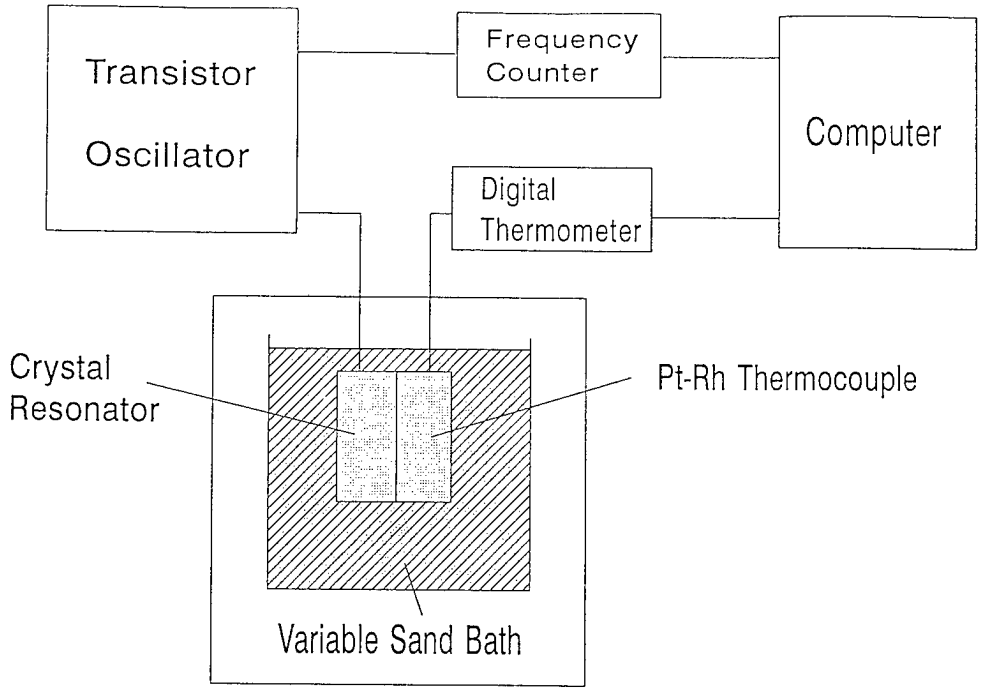


Fig.2. Circuit diagram for frequency-temperature characteristics.

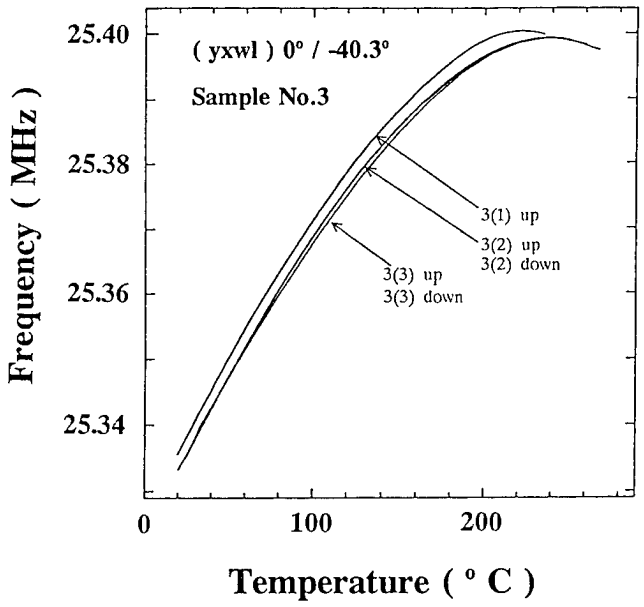


Fig.3. **Frequency change vs temperature characteristics of the quartz crystal resonator (yxwl) 0° / -40.3°.**

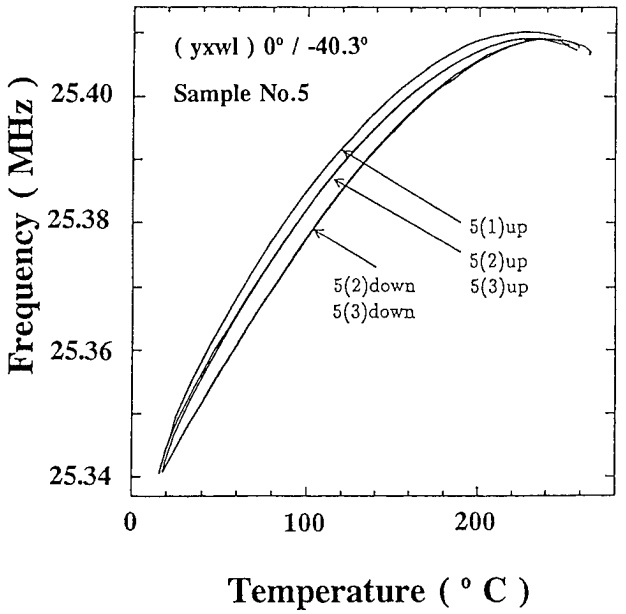


Fig.4. **Frequency change vs temperature characteristics of the quartz crystal resonator (yxwl) 0° / -40.3°.**

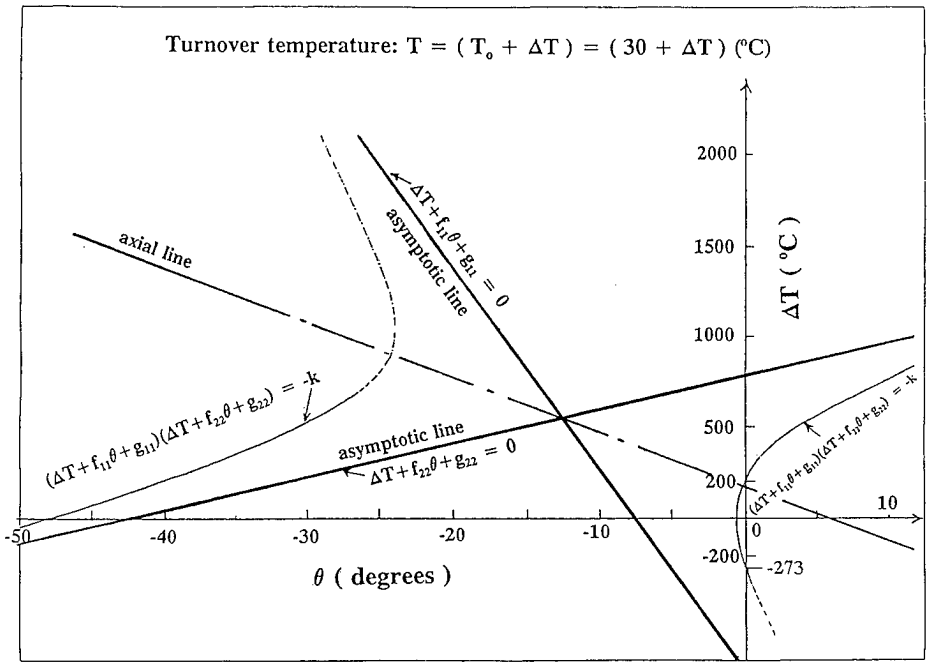


Fig.5. Turnover temperature vs angle of cut characteristics (calculated).

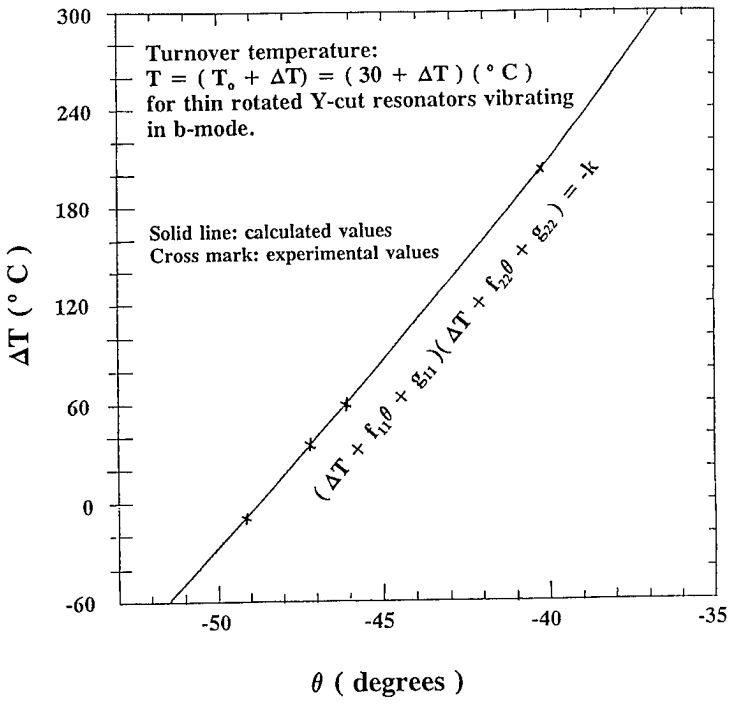


Fig.6. Turnover temperature characteristics vs angle of cut.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

RESEARCH AIMED AT DESIGNING A MINIATURE RUGGEDIZED CONFIGURATION OF BVA-TYPE CRYSTAL UNIT

A.E.Karaulnik, V.B.Grouzinenko

Scientific Research Institute "Fonon"
Moscow, Russia

Abstract

Results of investigations are presented allowing to expect the creation of highly stable crystal units at 5 Mc, 5th overtone, having a configuration characterized by special methods of mounting the piezoelectric element and of double hermetical sealing the latter.

Experimental samples have been fabricated for which all the mounting and hermetical sealing operations were carried on within a single vacuum cycle. The samples subjected to tests gave assuring results in aging and mechanical strength values. The Q-value of the samples reached 2.5×10^6, the frequency shifts were within (5.10^{-11} - 5.10^{-12}) per day. Crystal units of the configuration being reported have a volume of $3,5cm^3$ and maintain their efficiency under influence of vibrations up to 5000Hz at accelerations up to 40g.

As to the usage of an air-gap configuration for crystal units the most impressive results were first received by Smagin A.G. in 1957 [1, 2], when using AT-cut convexo-convex plates crystal units were fabricated having quality factors of the order of 20 millions. Oscillators using such type of crystal units achieved frequency stability equal to approximately 1.10^{-10} /day. Nevertheless the mentioned type of crystal units did not find a wide application for by that time crystal units at 2.5Mc, 5th overtone,had been designed [3], characterized by much simpler production processes, the quality factor

0-7803-0905-7/93 $3.00 © 1993 IEEE

approximately 6 millions and the frequency stability about 10^{-9}/month as well. For three decades up to the present time different modifications of that crystal unit configuration were most widely spread among the precision and high-stable quartz crystals. Those modifications differed in nominal frequencies, holder types, electrode materials and hermetical sealing techniques. Yet those variants were common in that 1)electrodes were realised in a thin film form and were vacuum-plated directly onto the surfaces of crystal elements; 2)high quality surface finish of the element was used; 3)crystal elements and exciting electrodes had optimum dimensions and 4)thorough cleaning of all the parts of the crystal unit construction was carried out prior to the hermetical sealing operation.

Nevertheless in spite of the advances in the modern technology the aging of the crystal units mentioned above due to the electrodes overplated just on the element surface had an order of magnitude higher value as compared to the intrinsic aging of the quartz crystal element itself. Therefore further improvements of the crystal oscillator frequency stability are conditioned with the use of crystal unit designs based on the air-gap excitement. The Renaissance of such crystal unit configuration just began about 15 years ago since R.Besson's researches resulted in the design well known at present as BVA-type [4-7]. The main difference as compared to analogous air-gap constructions being widely applied at present, is the minimum gap (of about 10 microns) between the exciting electrodes and the surfaces of crystal elements.

BVA-crystal units being mass-produced have the following main parameters: quality factor 2.5 millions, adjustment tolerance 0.6×10^{-6}, motional resistance approximately 70 Ohm and motional inductance about 60 H. The frequency drift due to the aging may equal to 5×10^{-12}/day [6].

Investigations have been carried on by the authors and co-workers which resulted in designing some experimental samples of comparatively small ruggedized crystal units at 5 Mc, 5th overtone, with an air-gap excitement of the crystal element. The crystal unit in question consists of three parts: a crystal element, a base and a cover, all of which are made of quartz and have the same crystallographic orientation. The crystal element represents a usual plano-convex quartz lens of 15 mm in diameter with a sphere radius of 60 mm. The base is a quartz disk having a recess of approximately 10 microns in depth produced by ion-plasma processing and a polished peripheral part with little mounting sites (crosspieces) for a crystal element. The cover of 18 mm in diameter has a recess with a depth slightly exceeding the crystal element thickness. The crystal element, the base and the cover were subjected to special chemical and thermal treatment. Then all three parts were placed into a vacuum chamber and on the surfaces intended for being combined a layer of indium was evaporated. Afterwards while all the parts of the crystal unit remaining under vacuum, the crystal element and the base were combined together by using manipulators and then pressed. After that the cover was placed over the base with the crystal element and the operation of combining and pressing was repeated.

Fig.1 shows the tensile strength of the junction versus the applied pressure for lapped and polished samples. The abscissa is the applied pressure, the ordinate is the tensile strength. Curve 1 is for lapped samples, curve 2 is for polished ones. High pressure values caused twinning in the quartz plate and mechanical damages of quartz parts of the crystal unit, while low pressure resulted in inadequate junction strength.

As a result of the experiments it was found that the pressure value guaranteed

the tensile strength of the junction not worse than 10×10^{6} ewton/m^2. Such pressure value is sufficient to ensure leakage values less than 1×10^{-12} W and an adequate design strength under severe mechanical effects. Thus, due to the contact of the juvenile surfaces of the thin indium layers crystal units were received in the form of discs with the diameter of 18 mm and with an air-gap excitement. One of the exciting electrodes was an indium film evaporated onto the inner surface of the base, the other one was a chrom-copper-nickel (or any other metal) layer which was plated onto the surface of the cover after the final frequency adjustment under vacuum.

The frequency adjustment of crystal units with an air-gap excitement may be fulfilled by changing the gap value between the crystal element and the electrodes. The crystal unit configuration being described allows to carry on frequency adjustment by lapping off or by etching off the cover outer surface of the crystal tablet-unit.

The accuracy of the final frequency adjustment is defined basically by the temperature-frequency change of the crystal element and by the frequency change due to the "membrana effect" of the crystal tablet-unit. The value of the temperature frequency change may be reduced by introducing a frequency correction which value may be defined from frequency versus temperature curves for individual crystal elements.

From Fig.2 where the membrana-effect acting upon tablet-units is presented, one may see that with any atmospheric pressure change equal to 150 mm of Hg the frequency change of the crystal unit makes up the value of approximately $(0.2-0.35) \times 10^{-6}$. The abscissa is the atmospheric pressure change, the ordinate is the frequency change. The real frequency shift values received after lapping off the cover of the tablet-unit are given in Fig.3. The abscissa is the change of the gap between the upper exciting electrode

and the crystal element. The ordinate is the frequency change.

After the operations of final adjustment and vacuum-plating the outer exciting electrode the tablet-unit is mounted into a cold-weld metal holder and the operation of welding is performed (holder dimensions: 13 mm in height, 24 mm in diameter).

The samples fabricated in the course of the experiments were subjected to different type tests, the results of which are presented in the Table given below. The electrical parameters evidence of their good reproducibility from sample to sample. The Table shows that the adjustment tolerance for those crystal units is 0.5×10^{-6} relative to the nominal frequency 5,000,040 Hz.

Crystal units in question possess high mechanical strength values (withstand vibrational loads up to 5000 Hz at 40 g).

Simultaneously these crystal units have low aging values and dimensions about 3 times less compared to crystal units of other air-gap configurations. Fig.4 shows the aging frequency changes of two concrete samples of the crystal units. Please, note that the frequency changes given in Fig.4a have a value measured in 10^{-11} and the changes given in Fig.4b have a value measured in 10^{-10}.

The results received during the investigation and the experiments lead to the following preliminary conclusions:

- the use of highly plastic materials (for example indium characterized by high pressure of elastic steams) as a spacer component for mounting and hermetical sealing minimizes both the mechanical and thermomechanical stresses in crystal elements and the mass transfer inside crystal units and as a result of that reveals good perspectives in acheiving high frequency stability and quality factors;

- the application of an indium laying (spacer) as a mounting part of the design allows to use a crystal element

of a traditional configuration (without any crosspieces which join the vibrating part of the crystal element with the mounting structure of the crystal unit). This makes it possible to reduce the dimensions of the design as a whole and to ensure its high mechanical strength.

Acknowledgment

The authors wish to express their appreciation and sincere thanks to B. Buckshtam for carrying out the experiments and to P.E.Kandyba for his support of the research.

References

[1]. Смагин А.Г. К вопросу о роли внутреннего трения в кристаллах пьезокварца. "ДАН СССР", 1957, 21.1, 3, стр. 425-426.

[2]. Смагин А.Г. Кварцевый резонатор с добротностью 17 Миллионов. "Измерительная техника", 1957, I-II,N 1, стр. 41-42.

[3]. A.W.Warner. "Design and Performance of Ultraprecise 2.5 Mc Quartz Crystal Units". Bell. Syst. J., 1960, 39, 5, p. 1193-1217.

[4]. R.Besson. "A New Piezoelectric Resonator design". Proc. 30th ASFC, p.78-81, 1976.

[5]. R.Resson. "A New Electrodeless Resonator design". Proc. 31st ASFC, p.147-152, 1977.

[6]. R.Resson, U.Peier."Further Advances on B.V.A. Quartz Resonators". Proc. 34th ASFC, p. 175-182, 1980.

[7]. R.Besson et al. "Performance of New Oscillators Designed for Electrodeless crystals". Proc. 34th ASFC, p. 457-462, 1980.

[8]. E.P.Graf, U.R.Peier. "BVA Quartz Crystal Resonator and Oscillator Production. A Statical Review". Proc. 37th ASFC, p. 492-500, 1983.

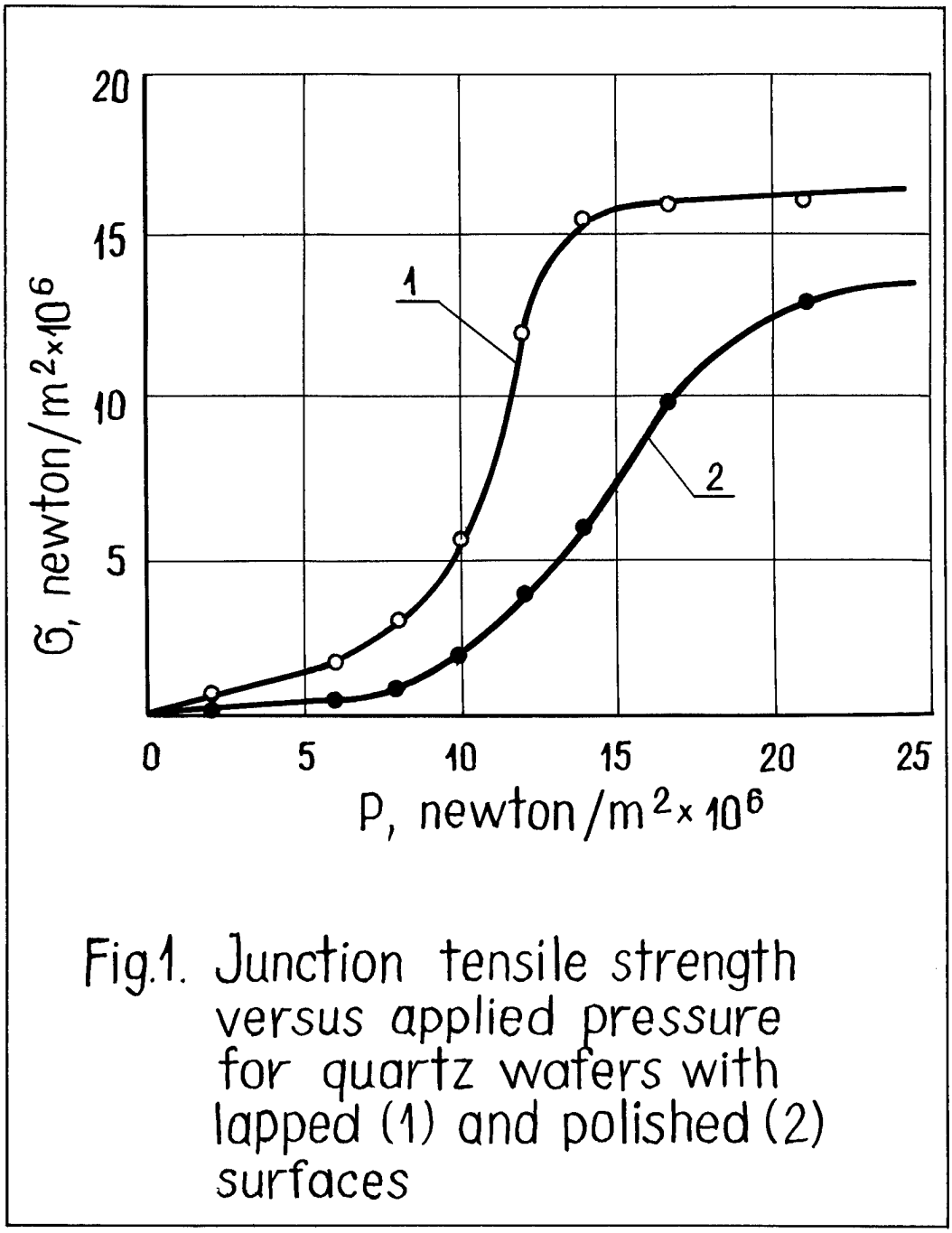


Fig.1. Junction tensile strength versus applied pressure for quartz wafers with lapped (1) and polished (2) surfaces

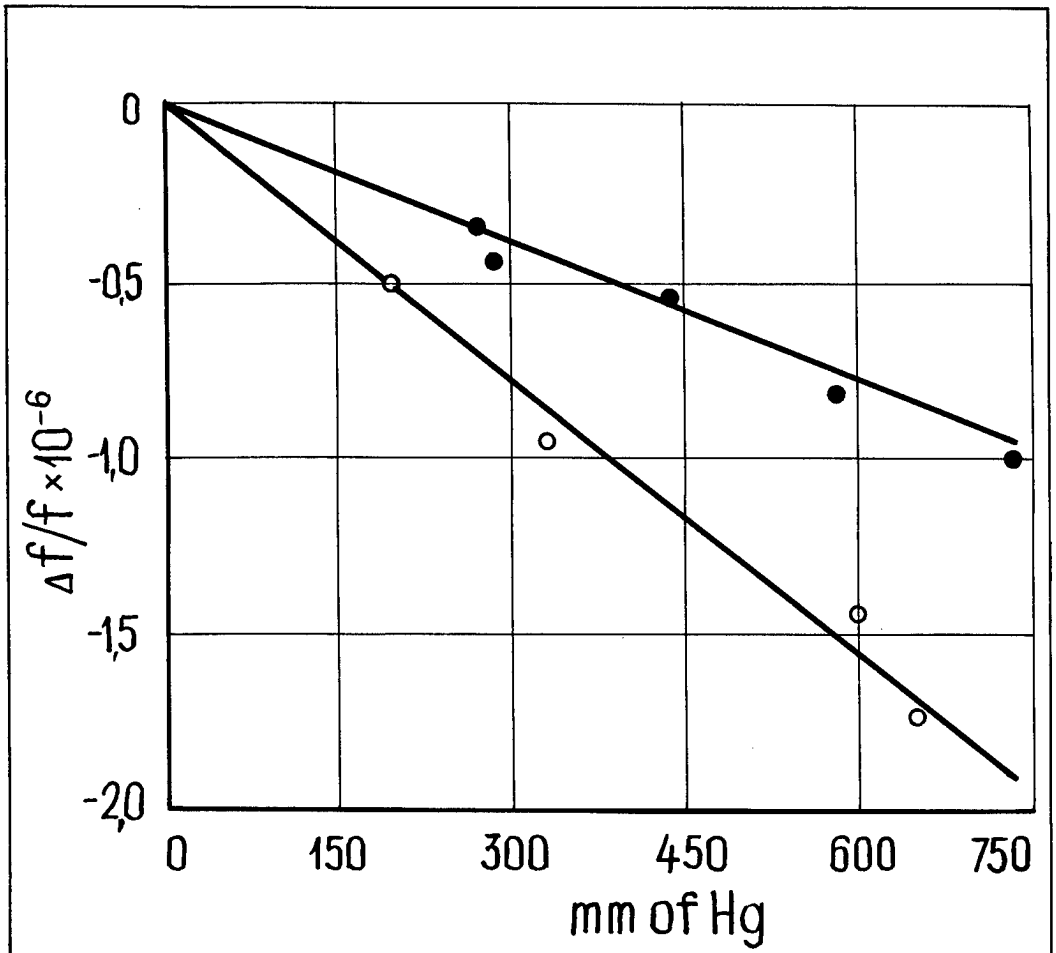


Fig.2. Quartz crystal disk-unit freguency versus applied pressure value (membrane effect)

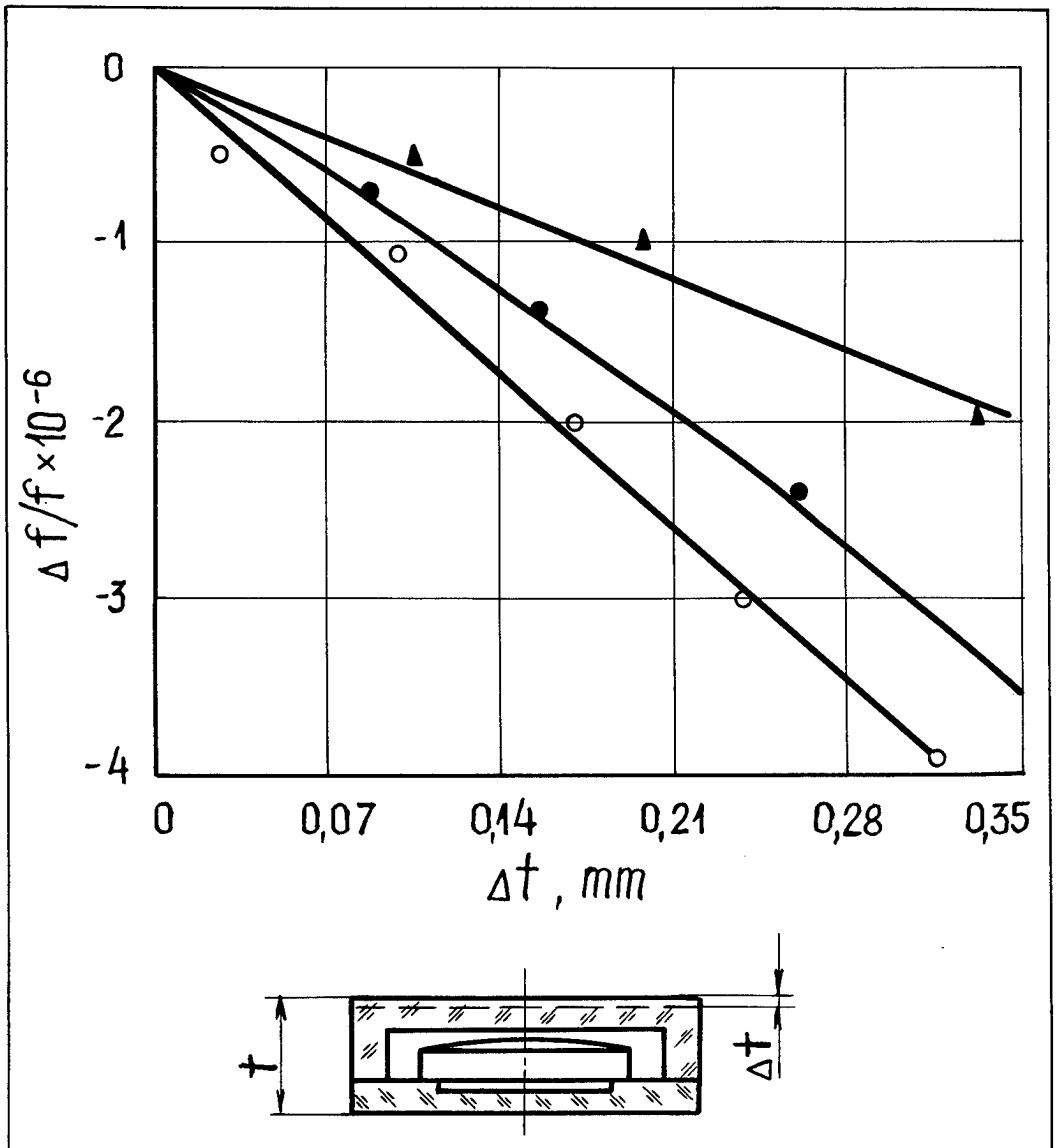


Fig.3. Quartz crystal disk-unit freguency change due to reducing the cover thickness

Table

Number of sample, N N	Freguency, Hz	Quality factor, $\times 10^6$	Motional resistance, Ohm
1	5 000 040	2,55	450
2	5 000 038	2,60	430
3	5 000 042	2,37	500
4	5 000 039	2,56	450
5	5 000 038	2,56	450
6	5 000 039	2,56	450

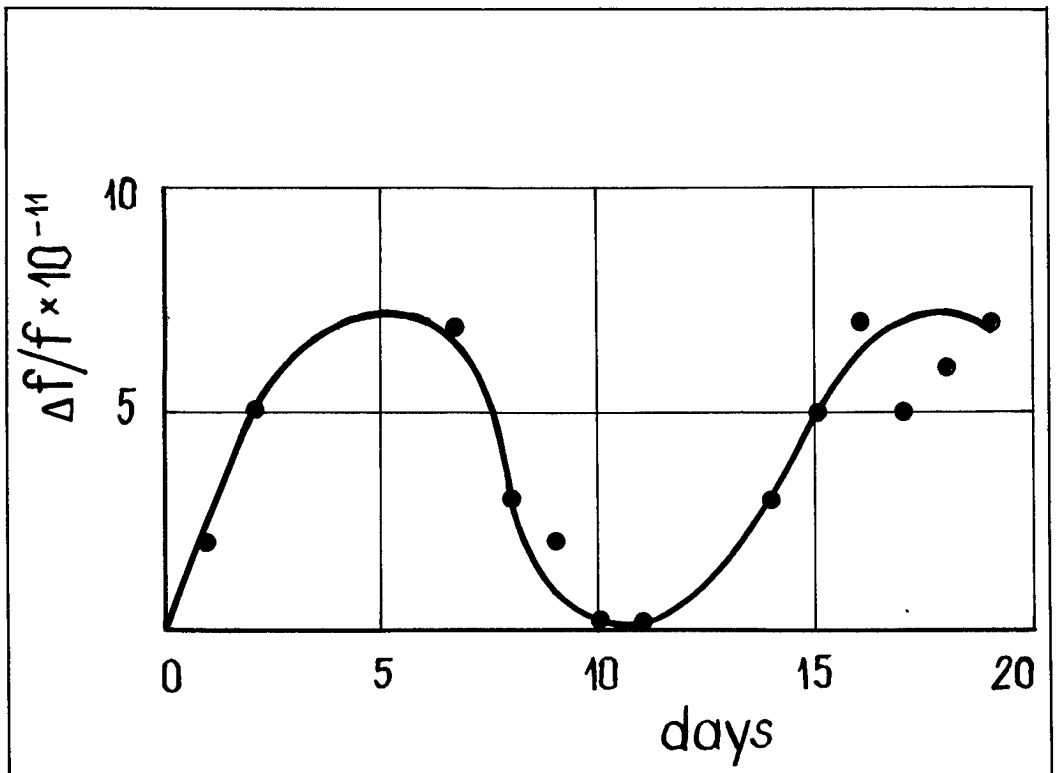


Fig. 4a. Ageing of experimental crystal units at 5 MHz, 5th overtone, after preliminary temperature runs for 60 days

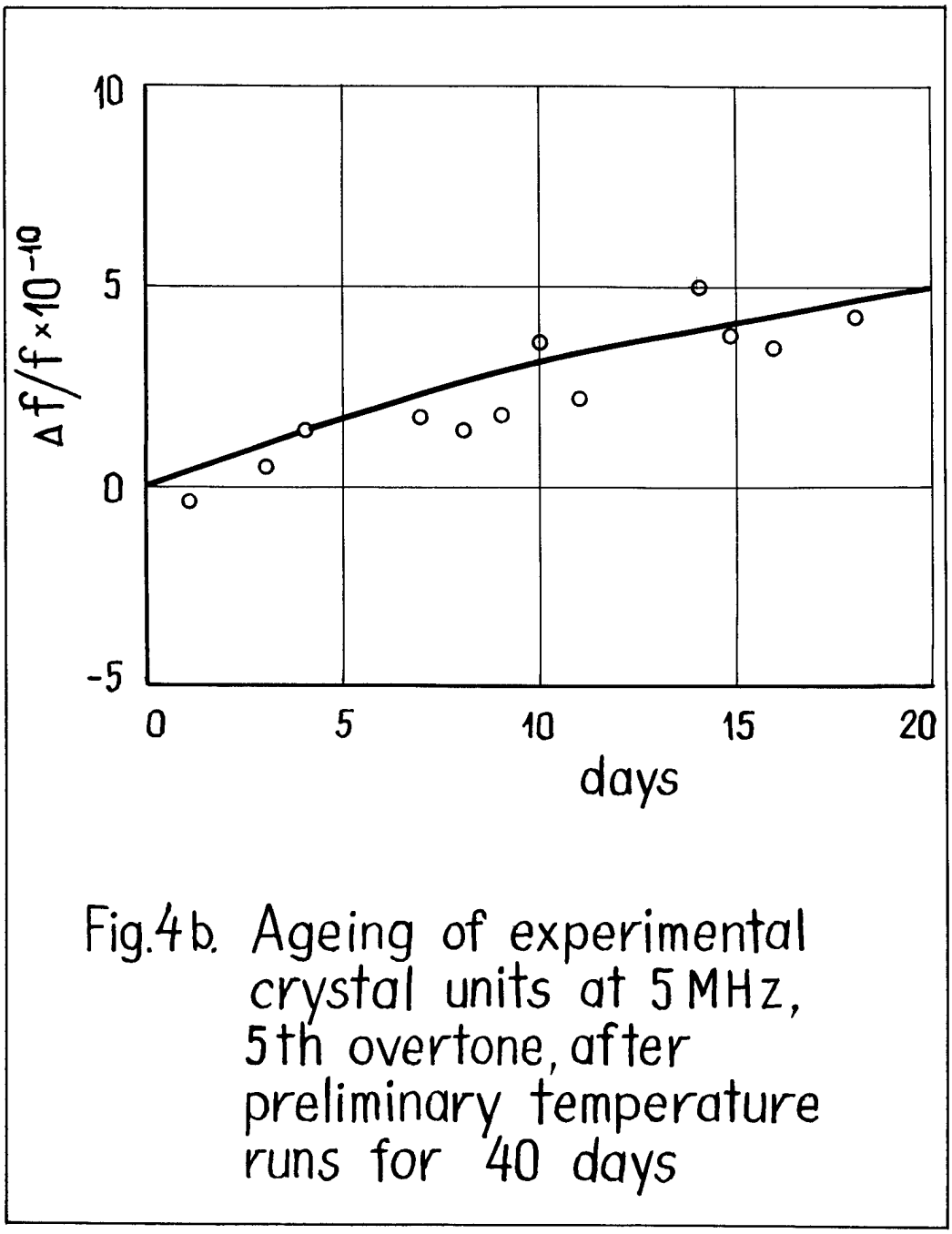


Fig.4b. Ageing of experimental crystal units at 5 MHz, 5th overtone, after preliminary temperature runs for 40 days

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE EFFECTS OF METAL DEPOSITIONS UPON ACCELERATION INDUCED FREQUENCY SHIFTS FOR QUARTZ CRYSTAL UNITS

LARRY D. CLAYTON and ERROL P. EERNISSE
Quartztronics, Inc. 1020 Atherton Dr., Bldg. C
Salt Lake City, UT 84123
(801) 266-6994 FAX (801) 266-7985

Abstract

Past acceleration sensitivity calculations focused upon development of a rugged 4-point SC-cut crystal unit with low acceleration sensitivity. A new 4-point mount called the horizontal strip mount resulted from these calculations. Horizontal strip mounts were formed and several crystal units assembled for testing. Acceleration sensitivities measured for these crystal units are much larger than predicted. Thin patches of gold were deposited off-center on the resonator surface to adjust mode shape and location and reduce acceleration sensitivity. Iterative depositions of Au had little effect upon the acceleration sensitivities of the horizontal strip mount crystal units.

To asses the effectiveness of mass loading, deposition test results for several crystal unit designs were compiled. Deposition histories are presented for crystal units with acceleration sensitivities which improved to varying degrees. This comparison includes two 4-point mount crystal unit designs with resonators similar to those found in the horizontal strip mount units. Acceleration sensitivities for these crystal units were largely unaffected by mass loading. The biggest reductions in acceleration sensitivity were achieved for asymmetric 2-point crystal units which also contain resonators similar to those found in the horizontal strip mount units.

An analysis of the influence of resonator stress distributions upon acceleration sensitivity revealed that out-of-plane shear stresses play a critical role. Defects which enhance inherent out-of-plane asymmetry increase these stresses. Reductions in acceleration sensitivity were not achieved for the 4-point crystal units because the out-of-plane shear stresses exhibit even symmetry. The asymmetric 2-point mount produces resonator stresses which may be combined by movement of the mode to partially cancel frequency shifts arising from acceleration. Out-of-plane stresses decrease and in-plane stresses dominate acceleration sensitivity as the resonator thickness is reduced. Crystal units which contain thin resonators had lower inherent acceleration sensitivities and responded more readily to mode adjustments.

Introduction

A solution technique developed to accurately calculate frequency shifts for quartz resonators arising from virtually any combination of loading and boundary conditions has been described in several recent publications [1-3]. This solution technique combines finite element analysis (FEA) with numerical integration of the frequency perturbation integral derived by H.F. Tiersten [4]. Finite element analysis is used to compute mechanical stresses induced by external actions and distributed throughout the resonator. A program has been written to retrieve element stresses and nodal displacements from solution files generated for FEA models of quartz resonators. This program computes eigenvector components and eigenvalues for fundamental modes and odd-harmonics of the thickness mode triad for singly- and doubly-rotated quartz resonators [5]. A transformation of the thickness mode displacement components to the global coordinate system for the FEA model includes off-center translations in the plane of the resonator. Components of the symmetric stress tensor and antisymmetric rigid rotation tensor from the FEA solution are contracted with the second- and third-order elastic coefficients and the second-order compliance coefficients of quartz in the model coordinate system to form effective nonlinear elastic coefficients [4, 6]. The nonlinear elastic coefficients are contracted with spatial derivatives of the mass-normalized thickness mode displacement components transformed to the model coordinate system to form the perturbation integral. Piecewise integration of the perturbation integral is performed numerically over elements comprising a significant portion of the active region of the resonator thickness mode to compute stress-induced frequency shifts.

This solution technique is general in that any combination of ideal or imperfect mount and resonator geometries may be modeled. Any crystallographic orientation may be chosen for the resonator and any load or combination of loads may be applied. Frequency shifts may be calculated for the mode located at the resonator center or offset in the plane of the resonator.

0-7803-0905-7/93 $3.00 © 1993 IEEE

Independent test results have been used to verify the accuracy of frequency shifts computed using this solution technique. Variations in force-frequency coefficients with azimuthal angle ψ calculated for AT- and SC-cut resonators agreed closely with measured results [1, 3]. The minimum acceleration sensitivity predicted for an AT-cut crystal unit in Reference 3 is in good agreement with minimums achieved by experimentally altering mode shape and location [7]. Calculations described in Reference 3 showed that acceleration sensitivity for manufactured crystal units is influenced by mode location and imperfect resonator stress distributions arising from mount and crystal unit defects.

The solution technique has been used to design a new 4-point mount which in theory produces low acceleration sensitivities for SC-cut crystal units [2]. Low acceleration sensitivities were predicted for the horizontal strip mount design. Mount dimensions and an optimum mount orientation and mode location were calculated to minimize acceleration sensitivity for the horizontal strip mount crystal units.

In the present work horizontal strip mounts were formed and several crystal units were assembled with 10MHz, 3rd, SC-cut resonators for testing. Acceleration sensitivities measured for these crystal units are significantly larger than the minimum $\vec{\Gamma}$ predicted for the ideal crystal unit in Reference 2. Larger than predicted acceleration sensitivities for the horizontal strip mount crystal units are attributed to imperfect resonator stresses and mode shapes arising from manufacturing defects.

Thin patches of gold were deposited off-center on the resonators in the horizontal strip mount crystal units. Off-center deposition of thin metal films alters mode shape and location and influences crystal unit acceleration sensitivity [8-11]. Acceleration sensitivities may be reduced for manufactured crystal units using the judicious addition of mass to the resonator.

Iterative depositions of Au to adjust mode shape and location had little effect upon acceleration sensitivities for the horizontal strip mount crystal units. Deposition histories are presented for the horizontal strip mount as well as other 2- and 4-point mount SC-cut crystal units. These show that significant reductions in acceleration sensitivity have been achieved using mass loading for SC-cut crystal units with two asymmetric mounts. Acceleration sensitivities for 4-point mount crystal units with resonators similar to those found in the horizontal strip mount crystal units have not responded to mode trimming. Limited reductions in acceleration sensitivity were achieved for SC- and AT-cut 4-point crystal units with thin high frequency resonators.

In an effort to understand the experimental results, a detailed investigation of stress distributions and frequency shifts arising from mount reactions to gravitational loading has been performed. This investigation revealed that out-of-plane shear stresses strongly influence acceleration sensitivity calculations. Increases in these stresses arise from crystal unit defects which enhance inherent out-of-plane crystal unit asymmetry. Evidence suggests that defects of this type have significantly increased the acceleration sensitivities of horizontal strip mount as well as other crystal unit designs.

Improvements in acceleration sensitivity have not been achieved for 4-point crystal units by altering mode shape and location because the dominant out-of-plane shear stresses have even symmetry with respect to the in-plane axes of the resonator. These stresses are not small at the center of the resonator and shifting the mode has little effect upon the contribution of even symmetry stresses to acceleration sensitivity. Unlike biaxial tensile or compressive stresses in the resonator plane, out-of-plane shear stresses are not inherently compensated for by the SC-cut.

Theory and experiments show that two asymmetric mounts produce resonator stresses which may be combined by movement of the mode to partially cancel frequency shifts arising from gravitational loading. Four-point mount crystal units with thin high frequency resonators have lower inherent acceleration sensitivities and are more responsive to mode trimming because out-of-plane shear stresses are reduced. Out-of-plane shear stresses will decrease and planar stresses will increasingly dominate gamma vector calculations as resonator thickness is reduced.

Horizontal Strip Mount Crystal Units

Work at Quartztronics on a project to develop analytical, empirical, and manufacturing processes to measure and reduce the acceleration sensitivity of quartz crystal units has recently been concluded. Crystal units selected for this project contain a plano-convex 10MHz, 3rd, SC-cut resonator (ϕ=22.4°, θ=34.3°) with a diameter of 0.55" and a radius of curvature of 3/8 diopter (≈55.6"). A thickness of 0.021" at the resonator center produces a fundamental C-mode frequency of approximately 3.336MHz (N_o=1811 Hz-m). The resonator is supported at four points distributed at equal angular intervals around its circumference.

Acceleration sensitivity for quartz crystal units is typically expressed as a vector quantity known as the gamma vector

$$\vec{\Gamma} = \Gamma_i \hat{i} + \Gamma_j \hat{j} + \Gamma_k \hat{k} \qquad (1)$$

Gamma vector components $\vec{\Gamma}_i$ are fractional frequency shifts produced by unit acceleration along three orthogonal directions. Fractional frequency shifts arising from acceleration in any direction are computed as the scalar product of the gamma vector with the acceleration vector

$$\frac{\Delta f}{f_o} = \vec{\Gamma} \cdot \vec{a} \qquad (2)$$

where f_o is the frequency for no acceleration. Gamma vectors for precision crystal units typically have magnitudes ranging from 10^{-9} to 10^{-10} per g.

Acceleration sensitivities were calculated in Reference 2 for 4-point mount SC-cut crystal units using the solution technique described above. The mounts in these crystal units are located asymmetrically to one side of the resonator. Prior studies [12-16] have identified asymmetry about the central plane of the resonator as an important factor contributing to increased acceleration sensitivity. Out-of-plane symmetry was sacrificed in these crystal units to achieve compact designs which are robust and readily manufactured. An even number of equally spaced mounts produces orthogonal planes of geometric symmetry normal to the resonator. Calculations produced a mount geometry and orientation and mode location which ideally compensate for the inherent asymmetry of the crystal unit design and significantly reduce acceleration sensitivity.

Acceleration sensitivities predicted in Reference 2 for the mode at the center of the resonator and mount orientations ranging from ψ=-90° to ψ=+90° were smallest for the horizontal strip mount design. Off-center mode locations to achieve minimum acceleration sensitivities were also smallest for the horizontal strip mount design. These factors lead to the selection of the horizontal strip mount for development of low acceleration sensitivity 4-point mount crystal units.

Additional calculations were performed to characterize the influence of critical dimensions upon the acceleration sensitivity of horizontal strip mount crystal units. A final mount design emerged for which minimum acceleration sensitivities were predicted and which facilitates precise repeatable fabrication. Four mounts are integrally formed from a single strip of 0.005" thick 302 stainless steel as shown in Figure 1. The mounts are intimately joined by a cupped ring which rigidly fixes each mount with consistent boundary conditions. This design maintains precise control of mount spacing and alignment but necessitates that electrical connections between header pins and the resonator electrodes be made using fine wire leads rather than through the mounts. A minimum acceleration sensitivity of $\vec{\Gamma}$=0.22x10^{-10}/g was predicted for this crystal unit design with an optimum mount orientation of ψ=+30°;-60° and the mode located off-center 0.028cm in the optimum direction. Figure 2 depicts the angular orientation of the horizontal strip mounts with respect to the crystallographic axes (X", Z") of the SC-cut resonator.

One hundred horizontal strip mounts were formed using precision metal stamping to minimize deviations from the ideal mount geometry depicted in Figure 1. Resonators and header bases were purchased for crystal units assembled with horizontal strip mounts. A fixture was designed to precisely align and position the mounts on header bases for resistance spot welding. A second fixture was designed to precisely orient the SC-cut resonators in the mounts at the optimum orientation. Adhesives with a variety of characteristics and cure schedules were selected to bond the resonators to the mounts. These adhesives were chosen to evaluate whether bond characteristics influence crystal unit acceleration sensitivity.

Nine horizontal strip mount crystal units depicted in Figure 3 were assembled for testing. Anaerobic structural adhesives, two part epoxies, a polyimide adhesive, and a compliant silicone rubber were used in seven crystal unit assemblies to attach the resonator to mounts at the optimum orientation (i.e., ψ=+30°;-60°). Two crystal units were assembled with the resonator orientation reversed (i.e., ψ=-30°;+60°). These latter crystal units were tested to insure that the crystal units had not been assembled with the resonator upside-down.

Acceleration Sensitivity Testing of the Horizontal Strip Mount Crystal Units

Acceleration sensitivities were measured with the horizontal strip mount crystal units and oscillator circuit mounted inside a vibration fixture. The vibration fixture was attached to the armature coil of a small electrodynamic shaker table. Threaded connections on each face of the fixture facilitate vibration of the crystal units along three mutually orthogonal axes in positive and negative orientations. Sinusoidal signals generated by an audio oscillator were fed to a low-distortion amplifier which drives the shaker. An accelerometer mounted to the fixture and aligned with the vibration direction was used to monitor and control vibration frequency f_V and level V_{ACC}.

Sidebands generated by vibration induced modulation of the oscillator signal were used to calculate gamma vector components for the horizontal strip mount crystal units. Sideband amplitude and frequency were measured using the system depicted in Figure 4. An HP frequency synthesizer generates a signal used to mix the

oscillator output signal down to 5.25MHz. The mixed oscillator signal, including sidebands, is phase-locked to a 5.25MHz VCXO to eliminate the carrier. A Fast Fourier Transform (FFT) spectrum analyzer is then used to measure sideband amplitude following ensemble averaging of spectral estimates for a series of measurement time histories. Phase differences between the vibration induced sideband signal and the accelerometer output were measured using the cross-spectrum function of the FFT analyzer. Phase angle measurement were used to assign direction to the gamma vector components. Gamma vector components generated by sinusoidal vibration along each axis and for each orientation are computed using

$$\vec{\Gamma}_i = \frac{V_{SB} S_{VCXO}}{f_o V_{ACC} S_{ACC}} \tag{3}$$

where V_{SB} and V_{ACC} are sideband and accelerometer signal amplitudes and f_o is the oscillator frequency. The VCXO sensitivity S_{VCXO} converts sideband amplitude to an equivalent acceleration induced frequency shift and the accelerometer sensitivity S_{ACC} converts V_{ACC} to acceleration response in g's.

Measurements of vibration induced sidebands were repeated for each crystal unit using the system depicted in Figure 5. Vibration modulated oscillator output is fed directly to a spectrum analyzer in this system. The offset function of the spectrum analyzer is used to measure sideband to carrier power ratios $\mathcal{L}_i(f_V)$ in decibels (dBc). Gamma vector components are given by

$$\vec{\Gamma}_i = \left(\frac{2 f_V}{g_i f_o}\right) 10^{\frac{\mathcal{L}_i(f_V)}{20}} \tag{4}$$

where f_V is the modulation frequency and g_i is the PEAK acceleration level in g's.

Preliminary measurements of gamma vector components for a variety of crystal units were performed to characterize measurement system capabilities. Reversing the crystal unit orientation with respect to the accelerometer orientation for vibration along each axis should produce gamma vector components of roughly the same magnitude with phase angle differences of approximately 180°. System errors which limit measurement capability are characterized by inconsistent gamma vector components and ambiguous phase angle measurements which do not reverse ≈180°.

Error sources which adversely affect measurement accuracy and limit the minimum resolvable gamma vector component were identified following a lengthy trouble-shooting process. Foremost among these is coherent noise at the modulation frequency arising from magnetic interference. Magnetic interference generated sidebands at the modulation frequency without the fixture and oscillator attached to the shaker. These sidebands varied significantly as the fixture was moved to various locations in the general vicinity of the shaker. Magnetic interference was suppressed by shielding sources of magnetic fields, eliminating multiple instrumentation grounds, and twisting conductors carrying the shaker drive signal.

Inconsistent clamping forces applied to the crystal unit and oscillator circuit upon assembly into the vibration fixture also contribute to measurement error. Clamping forces applied to the header base are coupled through the mounts to the resonator. Reversing the fixture orientation altered the clamping forces producing ambiguous component measurements. Asymmetric clamping forces which distort the oscillator circuit have a similar effect upon the gamma vector components. Systematic tests were performed to characterize fixture design features which contribute to ambiguous component measurements. The fixture design was modified to eliminate crystal unit clamping forces and to minimize forces which distort the oscillator circuit.

A less significant source of error for this measurement system is noise arising from vibration of the oscillator power and signal cables. The vibration fixture was designed to maintain the oscillator circuit and low-noise coaxial cables at a fixed position while the crystal unit was rotated independently for vibration along the in-plane axes. Component measurements for positive and negative fixture orientations and each axis were averaged to cancel cable modulation effects which remain substantially fixed [17].

Acceleration sensitivity testing began for the horizontal strip mount crystal units once the dominant sources of measurement error had been reduced. Sideband amplitudes and phase angles were measured for vibration along three mutually orthogonal axes using the PLL/FFT measurement system depicted in Figure 4. Gamma vector components were computed for positive and negative orientations of the crystal units. Components calculated for reverse orientations along each axis were averaged to eliminate cable effects. Sideband measurements were repeated using the test system depicted in Figure 5 for verification. Gamma vector components measured for positive and negative crystal unit orientations were typically in good agreement and phase angles consistently reversed approximately 180°.

Table 1 lists each horizontal strip mount crystal unit tested. Base line measurements of $\vec{\Gamma}$ are listed in Table 1 for each measurement system. Acceleration sensitivities measured for the horizontal strip mount crystal units using the PLL/FFT test system range from $\vec{\Gamma}$=11.8x10^{-10}/g to $\vec{\Gamma}$=30.6x10^{-10}/g. These acceleration

sensitivities are much larger than $\vec{\Gamma}=0.22\text{x}10^{-10}/\text{g}$ predicted for the optimum mount orientation and mode location.

Acceleration sensitivities measured using both test systems typically differed by less than 15%. Adhesive characteristics including bond flexibility, strength, and cure schedule did not consistently affect acceleration sensitivity. Reversing the mount orientation did not produce significantly different gamma vector measurements.

Deposition Histories for 2- and 4-Point Mount SC-Cut Crystal Units

Acceleration sensitivities measured for the horizontal strip mount crystal units are much larger than predicted. Defects arising from unavoidable manufacturing and assembly tolerances as well as material imperfections distort resonator mode shapes and stress distributions and contribute to increased acceleration sensitivity and $\vec{\Gamma}$ scatter. To compensate for inherent crystal unit asymmetry and manufacturing defects, experimental techniques to adjust mode shape and location may be employed to reduce acceleration sensitivity.

Experimental results presented in Reference 8 show that mode shape imperfections contribute to large $\vec{\Gamma}$ scatter and increased acceleration sensitivity for manufactured crystal units. Recent studies suggest that minimum acceleration sensitivities for crystal units which do not exhibit perfect symmetry occur for the mode offset from the resonator center [2, 3, 7-11, 17]. These studies indicate that reduced acceleration sensitivities and improved precision may be achieved for manufactured crystal units by adjusting mode shape and location.

Mode shape and location may be adjusted using an experimental technique described in References 8 through 11 which involves iterative depositions of thin metal patches off-center on the resonator active region. These patches are typically deposited along orthogonal axes in the resonator plane at positive and negative locations an equal distance from the resonator center. Adding mass to the resonator surface pulls the mode in the direction of the deposition through energy trapping. Following each deposition, the acceleration sensitivity measurement is repeated to project the location and thickness of additional depositions. Variations in acceleration sensitivity may be projected assuming each component varies linearly with increased deposition thickness. Deposition thicknesses, which in theory minimize $\vec{\Gamma}$, may be calculated following depositions along each axis.

This technique has been used successfully to influence acceleration sensitivity for a variety of 2- and 4-point mount crystal units [8]. Significant reductions in acceleration sensitivity have been achieved for a Hewlett-Packard 2-point mount crystal unit design. These crystal units contain a plano-convex 10MHz, 3rd, SC-cut resonator with a diameter of 0.59" and a radius of curvature of 1/2 diopter (≈42.0"). The resonator is mounted at two points on diametrically opposed nickel strips which are 0.005" thick and 0.05" wide. The mounts are rotated by $\psi\approx 60°$ with respect to the X" crystallographic axis of the SC-cut resonator and are of unequal length and stiffness. Figure 6 depicts characteristic changes in acceleration sensitivity with each deposition for six asymmetric 2-point mount crystal units selected from a larger test group. Acceleration sensitivities for all six 2-point mount crystal units were improved to varying degrees. The biggest improvement for this group of crystal units was from $\vec{\Gamma}=19.9\text{x}10^{-10}/\text{g}$ to $\vec{\Gamma}=2.0\text{x}10^{-10}/\text{g}$. The scatter between minimum and maximum acceleration sensitivities decreased from $12.88\text{x}10^{-10}/\text{g}$ before mode trimming to $7.32\text{x}10^{-10}/\text{g}$ after.

Efforts to reduce larger than predicted acceleration sensitivities for the horizontal strip mount crystal units began with off-center depositions of thin gold patches. Depositions were located along the X- and Z-axes depicted in Figure 2 using a shadow mask. Evaporated Au was deposited on the resonator surface through the shadow mask in a cryo-pumped vacuum system. The thickness of metal deposited on the resonator surface was determined by a standard quartz resonator thin film thickness monitor with a sensitivity of approximately 15.6 Hz/Å for Au.

Figure 7 depicts changes in the gamma vector following each deposition for the first six crystal units assembled with the horizontal strip mounts at the optimum mount orientation. Table 1 lists base line measurements of $\vec{\Gamma}$ for the horizontal strip mount crystal units along with final acceleration sensitivities measured following completion of the deposition process. Iterative depositions of Au of varying thickness to adjust mode shape and location had little effect upon the acceleration sensitivities of the horizontal strip mount crystal units. The largest in-plane component (i.e., along either X or Z) for each crystal unit remained virtually unchanged despite efforts to reduce $\vec{\Gamma}$ by changing the size, thickness, and off-center location of the deposition. Only the Y-component, which in some cases reversed direction, was significantly altered by iterative mass deposition. Changes in the Y-component had little effect

upon overall acceleration sensitivities which were dominated by much larger in-plane components for each crystal unit. Small slope coefficients for the large in-plane components resulted in linear projections of very thick depositions to achieve significant improvements in $\vec{\Gamma}$. Base line acceleration sensitivities measured for the first six crystal units ranged from $\vec{\Gamma}=15.4\text{x}10^{-10}$/g to $22.4\text{x}10^{-10}$/g. Following depositions of increasing thickness acceleration sensitivities ranged from $\vec{\Gamma}=17.3\text{x}10^{-10}$/g to $25.4\text{x}10^{-10}$/g. Similar acceleration sensitivities and deposition results were obtained for horizontal strip mount crystal units assembled with the mount orientation reversed (i.e., ψ=-30°;+60°).

Historical evidence exists for crystal units with acceleration sensitivities which cannot be significantly improved by small adjustments in mode shape and location. The authors of Reference 8 indicate that it was not possible to obtain significant improvements in acceleration sensitivities for 10MHz, 3rd, SC-cut 4-point mount crystal units obtained from two different manufacturers. These crystal units contain resonators which are very similar to those mounted in the horizontal strip mount crystal units. Acceleration sensitivities could not be significantly altered, even with thick platinum patches which eventually peeled from the resonator surface.

Deposition test results for selected crystal units of various designs have been compiled for comparison. The manufacturers of these crystal units are listed in Table 2 along with the resonator dimensions and mount configurations. A variety of mask designs and deposition metals have been used in an effort to improve acceleration sensitivities for these crystal units. Various test systems and vibration fixtures have been used to measure gamma vector components. Figure 8 depicts characteristic deposition histories for three 4-point mount crystal unit designs listed in Table 2. Figure 9 depicts deposition histories for six SC-cut, 50MHz, 5th 4-point mount crystal units.

In marked contrast to improvements achieved for 2-point crystal units with asymmetric mounts, the 4-point mount crystal units containing SC-cut, 10MHz, 3rd plano-convex resonators resisted all efforts to improve acceleration sensitivity. In half the cases depicted in Figure 8 the acceleration sensitivity increased with each deposition and additional depositions did not reverse this trend. Results for the 4-point mount 50-MHz crystal units were mixed with maximum reductions in acceleration sensitivity ranging from negligible to a factor of 1.8. In some cases minimum acceleration sensitivities were achieved and additional depositions reversed linearly projected trends towards improvement.

Apparently certain crystal unit designs are inherently amenable to mode trimming techniques and improvements in acceleration sensitivity are readily achieved while other designs resist improvement. Experimental techniques to adjust mode shape and location should be more effective for resonators with low contour number and less energy trapping. Tests have shown that the asymmetric 2-point mount crystal units consistently responded to mass loading and significant reductions in acceleration sensitivity were achieved while the horizontal strip mount crystal units resisted all efforts to reduce $\vec{\Gamma}$. The resonators in the asymmetric 2-point mount crystal units have a slightly higher contour than those mounted in the horizontal strip mounts. Therefore, resistance to mode trimming must arise from fundamental differences in resonator stress distribution rather than small differences in curvature and energy trapping.

Resonator Stress Distributions and Frequency Shifts Arising from Gravitational Loading

Resonator stress distributions produced by gravitational loading along three orthogonal axes have been computed for finite element models of 2- and 4-point mount crystal units. Frequency shifts arising from these stress distributions have been estimated using the perturbation integral simplified for normalized thickness mode displacement gradients combined with average stress components computed at the center of the resonator. This solution technique, for the sake of simplicity, neglects the effect upon frequency shift of spatial interaction between the normalized thickness mode displacement gradients and biasing stress components. A simplified approach was opted for to avoid time consuming numerical integration of the perturbation integral for each model, load condition, and mount orientation. Precise gamma vector calculations may be performed using the more rigorous solution technique described at the first of this paper once general trends have been established using the simple approach.

Finite element models of 2- and 4-point mount crystal units have been created. These include a model of the horizontal strip mount crystal unit described earlier in this paper and a model of the same crystal unit with the mounts along the Z-axis removed. A third model of a 2-point mount crystal unit with the mount at the +X location significantly stiffer has also been created. This model is similar to the 2-point crystal units with asymmetric mounts for which significant improvements in acceleration sensitivity have been achieved by adjusting mode shape and location.

Resonator stress distributions have been calculated for each mount configuration and gravitational loading along the global axes depicted in Figure 2. Stress components were output for nodes in the central region of the resonator along the X- and Z-global axes and through the resonator thickness. Nodal stresses along the X- and Z-axes and through the resonator thickness were averaged to produce effective stress components at the center of the resonator. Table 3 lists effective stress components calculated for each crystal unit model and load condition.

Fractional frequency shifts arising from gravitational loading along the crystal unit axes may be estimated using the stress components averaged at the resonator center. These are the gamma vector components which may be combined to obtain estimates of acceleration sensitivity for each crystal unit model. Fractional frequency shifts are computed for the effective stress components at the resonator center using stress coefficients. Stress coefficients are the frequency shifts per unit stress computed using the simplified approach for each stress component at the center of the resonator. Stress coefficients computed for variations in the stress orientation ψ are plotted in Figure 10 for an SC-cut resonator vibrating in a C-mode fundamental or odd-harmonic overtone.

Shear and normal stresses acting out of the resonator plane (i.e., σ_{yy}, σ_{xy}, σ_{yz}) are often neglected in calculations of stress induced resonator frequency shifts. The effect of these stress components is generally considered to be insignificant for resonators in which the thickness is small in comparison to the diameter.

Consistent with this assumption, gamma vector components were originally calculated for the symmetric 4-point or horizontal strip mount model using the planar stresses (i.e., σ_{xx}, σ_{zz}, σ_{xz}) averaged at the resonator center from Table 3. Figure 11 depicts variations in the gamma vector and its components with stress orientation ψ for the SC-cut stress coefficients and the in-plane stresses at the resonator center. Stress induced frequency shifts are generally largest for gravitational loading along the Y-axis. Gravitational loading along the Y-axis produces biaxial normal stresses in the resonator plane with the same sign and even symmetry with respect to the X- and Z-axes. Equal stress coefficients of opposite sign over all ψ angles cancel frequency shifts arising from uniform normal stresses in the resonator plane giving the stress compensated or SC-cut its name. Slight differences in the planar normal stresses at the resonator center for gravitational loading of the 4-point model along the Y-axis account for the gamma vector component in this direction

Gravitational loading in the X- and Z-directions for the 4-point mount model produces planar stresses with odd-symmetry about the X-and Z-axes. Odd symmetry stress distributions produce no net frequency shift when the center of the stress symmetry coincides with the mode center. The X- and Z-components contribute little to the acceleration sensitivity of the symmetric 4-point mount model in Figure 11 because on average the odd symmetry planar stresses cancel.

Acceleration induced frequency shifts have also been calculated for the 4-point horizontal strip mount crystal unit model using the more rigorous solution technique. Frequency shifts are calculated for element mechanical stresses and rigid rotations contracted with normalized thickness mode displacement gradients distributed throughout a portion of the resonator volume over which the numerical integration is performed. Anisotropic material behavior may be modeled as well as movement of the mode in the general vicinity of the resonator center. Figure 12 depicts variations in the gamma vector and its components for the mode located at the resonator center and changes in mount orientation ψ for the more rigorous solution technique. The Y-component is not dominant in Figure 12 and the gamma vector is larger over all mount (i.e., stress) orientations. The increase in $\vec{\Gamma}$ arises primarily from increased contributions from the X- and Z-components.

The gamma vector and its components for each solution technique in Figures 11 and 12 exhibit 90° rotational symmetry consistent with the equal angular spacing of the mounts. Inclusion of anisotropy improves the accuracy of gamma vector calculations performed using numerical integration of the perturbation integral with increased spatial resolution of resonator mode shape and stress distributions. Symmetry of the planar stresses and their contribution to the gamma vector for the 4-point mount model should be essentially the same regardless of solution technique. Therefore, increases in the X- and Z-components and to a lesser degree for the Y-component found using the rigorous solution technique must arise from the out-of-plane stresses originally ignored for the simplified approach. Gravitational loading in the X- and Z-directions produces out-of-plane shear stresses with even symmetry about the in-plane axes for the 4-point model. These out-of-plane shears exceed the other average stress components at the resonator center listed in Table 3 for the 4-point model.

Gamma vector calculations were repeated for the 4-point model using all six stress components averaged at the resonator center and the SC-cut stress coefficients. Figure 13 depicts variations in the gamma vector and its components for the full stress tensor. The X- and Z-components are significantly increased by the

inclusion of the even symmetry shear stresses normal to the resonator plane. The magnitude of the gamma vector also increases and its direction changes with the addition of larger in-plane components.

Out-of-plane shear stresses dominate the simple calculation of the in-plane gamma vector components because they are largest at the resonator center. More importantly, the stress coefficients for the out-of-plane shears are significantly larger than the planar stress coefficients for most orientations. To produce a gamma vector component with a magnitude of $2x10^{-9}$/g the planar normal stresses must have a magnitude of at least 2,150.5 Dyne/cm^2 per g of loading for the maximum stress coefficients. The minimum uniform shear stress in the resonator plane which will produce a component of this magnitude is 1,070 Dyne/cm^2 per g. If gravitational loading produces biaxial normal stresses of the same sign in the resonator plane the frequency shift will be significantly reduced or canceled for the SC-cut. If gravitational loading produces planar stresses with odd symmetry about the in-plane axes the gamma vector will also be reduced by stress cancellation at the mode center. The out-of-plane shear stress components require a magnitude of at least 310.6 Dyne/cm^2 per g of loading to produce a gamma vector component of $2x10^{-9}$/g. Clearly these stress components should not be ignored, particularly for the calculation of in-plane gamma vector components for SC-cut crystal units. The only stress component listed in Table 3 which is consistently negligible for each crystal unit model is the out-of-plane normal stress σ_{yy}. With rare exception [12], the importance of the contribution of the out-of-plane shear stresses to crystal unit acceleration sensitivity has not been widely appreciated.

In-plane gamma vector components computed using the stress coefficients are significantly larger than those computed using the more rigorous solution technique. Averaging the nodal stresses along the in-plane axes over the resonator thickness must artificially exaggerate the contribution of the out-of-plane shears to these components. Large in-plane gamma vector components arising from the out-of-plane shears for the simple calculation are, however, more consistent with measurements of these components for the horizontal strip mount crystal units. This similarity led to the realization that large gamma vector components consistently measured for the horizontal strip mount crystal units along the X- and Z-axes could arise from the out-of-plane shear stresses. These components are not readily reduced by altering the mode shape and location because the out-of-plane shears have even symmetry with respect to the in-plane axes of the crystal unit. Even symmetry combined with large stress coefficients significantly limits any improvements in $\vec{\Gamma}$ which may be achieved by mode trimming. The Y-components are smaller and more responsive to mode trimming for the horizontal strip mount crystal units because the dominant planar normal stresses for this component are inherently compensated for by the SC-cut.

Mount, resonator, and crystal unit imperfections which arise from imprecise fabrication of the horizontal strip mount crystal units must contribute to increases in the out-of-plane shear stresses. Increases in out-of-plane shear stresses will arise from defects which enhance inherent out-of-plane asymmetry for these crystal units. Mounting the resonator so that its center of mass is offset in the Y-direction with respect to the effective plane of its support points is an example of this type of defect. Increasing the contour of a plano-convex resonator will also increase this type of crystal unit asymmetry.

Gravity induced frequency shifts were calculated next for the crystal unit model with two symmetric mounts. These calculations were performed using the SC-cut stress coefficients and the complete stress tensor averaged at the resonator center for each gravitational load. Figure 14 depicts variations in the gamma vector and its components with ψ for this model. Gravitational loading aligned with the mounts along the X-axis produced normal stress σ_{xx} with odd symmetry about the Z-axis and shear stress σ_{xz} with odd symmetry about the X-axis. Odd symmetry stresses cancel at the resonator center and the small X-component in Figure 14 arises primarily from the even symmetry out-of-plane shear σ_{xy}.

Gravitational loading in the Z-direction produces normal and shear stresses with odd symmetry in the resonator plane for which there is little contribution to the frequency shift. Gravitational loading in the Z-direction produces a large out-of-plane shear σ_{yz} with even symmetry for which a sizable Z-component is calculated. Gravitational loading along the Y-axis of the symmetric 2-point model produces biaxial normal stresses in the resonator plane which are opposite in sign. These stresses have even symmetry about the in-plane axes and are not small at the resonator center. Stress coefficients which are equal but opposite in sign combine with the in-plane normal stresses, which are also opposite in sign, to produce a large Y-component. Zero crossing for the normal stress coefficients in the plane of an SC-cut resonator occur at approximately ψ=85° and ψ=175°. Orienting the mounts (stresses) at these angles produces minimums in the gamma vector for the 2-point model depicted in Figure 14. Even symmetry for the biaxial normal stresses limits the reduction in acceleration sensitivity which may be achieved for crystal

units with the mounts rotated away from the stress coefficient zero crossings.

Gamma vector calculations performed using the SC-cut stress coefficients concluded with the model of a 2-point mount crystal unit with one mount stiffer than the other. This model is similar to the asymmetric 2-point mount crystal units for which significant reductions in $\vec{\Gamma}$ were achieved using mode trimming. Figure 15 depicts variations in the gamma vector and its components with ψ for this model. Comparing the curves in Figure 15 with those in Figure 14, the X-component is increased by increasing the stiffness of the mount at the +X location. The curves for the Y- and Z-components are shifted along the ψ axis and these components are reduced somewhat in magnitude. Increases in the X-component are more than offset by reductions in the Y- and Z-components, and the acceleration sensitivity predicted for the asymmetric crystal unit is reduced relative to that for the symmetric 2-point model.

Stiffening the mount at the +X location offsets the odd symmetry stress σ_{xx} along the X-axis for gravitational loading in the X-direction. The planar shear σ_{xz} with odd symmetry about the X-axis is virtually unaffected by altering the mount stiffness. Stiffening the mount also produces planar normal stress σ_{zz} and out-of-plane shear σ_{xy} each with even symmetry about the in-plane axes. Net average values at the resonator center for these even symmetry stresses combine with the offset odd symmetry stress σ_{xx} to increase the magnitude of the X-component. Increased frequency shifts arising from the even symmetry stresses may be partially offset by altering the contribution of the odd symmetry stress σ_{xx} through movement of the mode along the X-axis.

Biaxial normal stresses in the resonator plane with even symmetry and opposite signs are produced by gravitational loading along the Y-axis for both the symmetric and asymmetric 2-point mount models. An out-of-plane shear σ_{xy} is also produced by rotation of the resonator resulting from the stiffer mount in the asymmetric 2-point model. This out-of-plane shear has odd symmetry with respect to the Z-axis and movement of the mode along the X-axis alters the average σ_{xy} computed at the mode center. This stress component may be used to partially offset frequency shifts arising from the biaxial normal stresses to reduce the Y-component.

Loading the asymmetric 2-point mount model with 1g in the Z-direction produces normal stress σ_{xx} with odd symmetry about the X-axis. Gravitational loading along Z also produces in-plane shear σ_{xz} and out-of-plane shear σ_{yz} both with odd symmetry about axes offset along X. Movement of the mode along the X-axis will reduce the contribution of the odd symmetry shears to the Z-component without increasing the contribution of σ_{xx}.

Calculations performed using the SC-cut stress coefficients and stress tensors averaged at the resonator center indicate that reductions in all three gamma vector components for the 2-point model with asymmetric mounts may be achieved by movement of the mode along the X-axis. More accurate calculations may be performed using the rigorous solution technique to confirm this conclusion. Evidently lack of symmetry in the mounts produces odd symmetry stresses which may be used through movement of the mode center to partially offset frequency shifts arising from gravitational loading. Deposition histories depicted in Figure 6 for the asymmetric 2-point mount crystal units appear to support this conclusion.

Conclusion

A brief summary describing a rigorous solution technique which may be used to accurately calculate quartz resonator frequency shifts arising from mechanical stresses has been presented. This solution technique has been used to design a new mount, called the horizontal strip mount, to achieve low acceleration sensitivities for 4-point SC-cut crystal units. Nine crystal units have been assembled with horizontal strip mounts for acceleration sensitivity testing.

Acceleration sensitivities measured for the horizontal strip mount crystal units are much larger than predicted for the ideal crystal unit design. Efforts to reduce the acceleration sensitivities of the horizontal strip mount crystal units using off-center depositions of Au to alter mode shape and location were unsuccessful. Deposition histories were presented for additional 4-point mount crystal units with different mount configurations and resonators similar to those mounted in the horizontal strip mounts. Iterative depositions of Au also had little effect upon the acceleration sensitivities for these crystal units. Deposition histories were presented for asymmetric 2-point mount crystal units. These crystal units responded to mode trimming and significant reductions in acceleration sensitivity were achieved.

Certain crystal unit designs are strongly influenced by adjustments in mode shape and location while others resist change. Resistance to change in acceleration sensitivity arises from fundamental differences in stress distributions created in the resonator by mount reactions to gravitational loading. This investigation has shown that out-of-plane shear stresses by virtue of larger stress coefficients typically dominate calculations of gamma vector components. Increases in

these stresses arise from defects which enhance out-of-plane asymmetry. Evidence suggests that these types of defects have significantly increased the acceleration sensitivity of the horizontal strip mount as well as other crystal units. Significant improvements in acceleration sensitivity have not been achieved for 4-point crystal units by altering mode shape and location because the out-of-plane shear stresses exhibit even symmetry about the resonator planar axes. Theory and experiments show that an asymmetric 2-point mount produces stresses which may be combined by movement of the mode to reduce frequency shifts arising from gravitational loading.

Limited improvements in acceleration sensitivity have been achieved for 4-point mount 50MHz, 5th, SC-cut crystal units using mode trimming. These crystal units may have been more responsive to mode trimming simply because the resonator blank is thinner with respect to its diameter. Out-of-plane shear stresses diminish and planar stresses increasingly dominate acceleration sensitivity as the resonator thickness decreases. Crystal units containing thin high frequency resonators have lower inherent acceleration sensitivities and respond more readily to mode trimming techniques.

Acknowledgments

This work has been supported by the U.S. Army LABCOM ETDL under Contract No. DAAL01-90-C-0009

References

[1] L.D. Clayton and E.P. EerNisse, "Quartz Resonator Frequency Shifts Computed Using the Finite Element method," International Journal for Numerical Methods in Engineering, **36**, 1993, pp. 385-401.

[2] L.D. Clayton and E.P. EerNisse, "Four-Point SC-cut Crystal Unit Design Calculations for Reduced Acceleration Sensitivity," in Proceedings 46th Annual Symposium on Frequency Control, 1992, pp. 582-596.

[3] L.D. Clayton and E.P. EerNisse, "Frequency Shift Calculations for Quartz Resonators," in Proceedings 45th Annual Symposium on Frequency Control, 1991, pp. 309-320.

[4] H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," Journal of the Acoustical Society of America, **64**, 1978, pp. 832-837.

[5] D.S. Stevens and H.F. Tiersten, "An Analysis of Doubly Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," Journal of the Acoustical Society of America, **79**, 1986, pp. 1811-1826.

[6] H.F. Tiersten, "Electroelastic Interactions and the Piezoelectric Equations," Journal of the Acoustical Society of America, **70**, 1981, pp. 1567-1576.

[7] R.C. Smythe and W.H. Horton, "Adjustment of Resonator G-Sensitivity by Circuit Means," in Proceedings 44th Annual Symposium on Frequency Control, 1990, pp. 437-443.

[8] E.P. EerNisse, et al., "Experimental Evidence for Mode Shape Influence on Acceleration-Induced Frequency Shifts in Quartz Resonators," IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, **37** (6), 1990, pp. 566-570.

[9] E.P. EerNisse, L.D. Clayton, and M.H. Watts, "Distortions of Thickness Shear Mode Shapes in Plano-Convex Quartz Resonators with Mass Perturbations," IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, **37** (6), 1990, pp. 566-570.

[10] E.P. EerNisse and R.W. Ward, "Crystal Resonator with Low Acceleration Sensitivity and Method of Manufacture Thereof," U.S. Patent No. 4,837,475, June 6, 1989.

[11] E.P. EerNisse, et al., "Crystal Resonator with Low Acceleration Sensitivity and Method of Manufacture Thereof," U.S. Patent No. 4,935,658, June 19, 1990.

[12] H.F. Tiersten and Y.S. Zhou, "The Increase in the In-Plane Acceleration Sensitivity of the Plano-Convex Resonator Resulting from Its Thickness Asymmetry," Journal of Applied Physics, **71** (10), 1992, pp. 4684-4692.

[13] H.F. Tiersten and Y.S. Zhou, "On the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators Supported Along Rectangular Edges," Journal of Applied Physics, **70** (9), 1991, pp. 4708-4714.

[14] R. Besson, et al., "Design of Bulk Wave Quartz Resonators Insensitive to Acceleration," in Proceedings 33rd Annual Symposium on Frequency Control, 1979, pp. 337-345.

[15] R.L. Filler, J.A. Kosinski, and J.R. Vig, "The Effect of Blank Geometry on the Acceleration Sensitivity of AT and SC-Cut Quartz Resonators," in Proceedings 36th Annual Symposium on Frequency Control, 1982, pp. 215-219.

[16] P.C.Y. Lee and M.S.H. Tang, "Acceleration Effect on the Thickness Vibrations of Doubly Rotated Crystal Resonators," in Proceedings 41st Annual Symposium on Frequency Control, 1987, pp. 277-281.

[17] M.H. Watts, et al., "Technique for Measuring the Acceleration Sensitivity of SC-Cut Quartz Resonators," in Proceedings 42nd Annual Symposium on Frequency Control, 1988, pp. 442-446.

Table 1. Acceleration Sensitivities Measured for the Horizontal Strip Mount Crystal Units

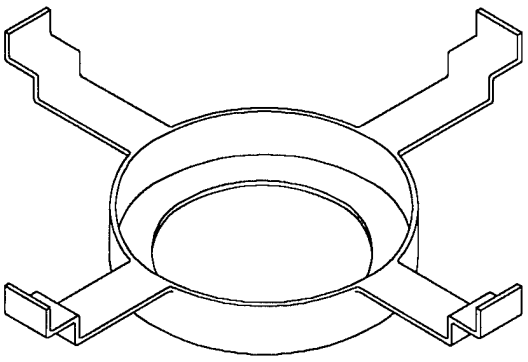
			PRE-	DEPOSITION	POST-	DEPOSITION
Crystal Unit No.	Resonator Adhesive	Mount Orientation Angle ψ	PLL/FFT System $\vec{\Gamma}$ x10^{-10}/g	Spectrum Analyzer $\vec{\Gamma}$ x10^{-10}/g	PLL/FFT System $\vec{\Gamma}$ x10^{-10}/g	Spectrum Analyzer $\vec{\Gamma}$ x10^{-10}/g
1	LOCTITE UV	+30°;-60°	17.1	NA	18.6	18.1
2	EPOTEK H72	+30°;-60°	20.3	23.4	20.6	21.9
3	EPOTEK 301-2	+30°;-60°	16.3	17.9	17.3	19.9
4	EPOTEK 353ND	+30°;-60°	15.4	NA	17.4	18.85
5	ECCOBOND 45	+30°;-60°	22.4	25.4	25.4	27.9
6	71-2LMI	+30°;-60°	21.1	NA	21	NA
7	EPOTEK 353ND	-30°;+60°	16	17.6	18	18.6
8	EPOTEK 353ND	-30°;+60°	11.8	13.25	11.6	11.1
9	RTV-SILICONE	+30°;-60°	30.6	34.95	NA	NA

Table 2. SC-Cut Crystal Units for Which Deposition Test Results Have Been Obtained

Crystal Unit Supplier	Mount			Resonator				
	No.	General Shape	Angle ψ	Freq. (MHz)	C-mode Overtone	Diameter (in.)	Thickness (in.)	Curvature (in.)
Hewlett-Packard	2	Unequal Strip	≈+60°	10	3RD	0.59	0.021	42
Piezo Crystal	4	V-Clip	0°	10	3RD	0.55	0.021	18
Piezo Technology	4	V-Clip	0°	10	3RD	0.55	0.021	21
FEI	4	Step	0°	10	3RD	0.375	0.021	8
FEI	4	Step	0°	50	5TH	0.375	0.007	21
Quartztronics	4	Horizontal Strip	+30°	10	3RD	0.55	0.021	55.6

Table 3. Stress Components Averaged at the Resonator Center for Each Crystal Unit Model

Stress	4-Point	Ideal	Mount	2-Point	Ideal	Mount	2-Point	Asymmetric	Mount
Dyne/cm^2	1-g X	1-g Y	1-g Z	1-g X	1-g Y	1-g Z	1-g X	1-g Y	1-g Z
σ_{xx}	7.4	-2546	-23.2	6.2	-6272	-9.8	-2013	-4144	-28.6
σ_{yy}	29.4	6.6	0	-4.4	6.6	-12.3	0	-15.2	-4.6
σ_{zz}	11.4	-2754	19.8	-5.2	1554	-25.6	445	1091	-17.4
σ_{yz}	-23.4	0	0	1.1	1.7	-558.2	1.5	14.6	-141
σ_{xz}	4.4	-46.6	0	-10.8	-65.8	-49.3	-26	-48.2	60.3
σ_{xy}	-488.9	0	-103	58	-5.6	8.4	146.4	-463	3.2



HORIZONTAL STRIP MOUNT DESIGN

Figure 1. Optimum Horizontal Strip Mount Designed for Low Acceleration Sensitivity SC-Cut Crystal Units

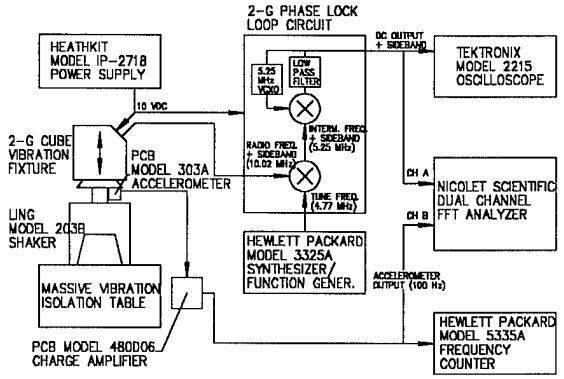


Figure 4. Measurement System with the Phase Lock Loop Circuit and FFT Spectrum Analyzer

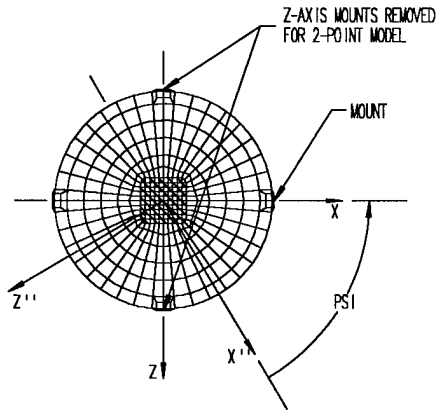


Figure 2. Mount (Stress) Orientation Relative to the Crystallographic Axes of the SC-Cut

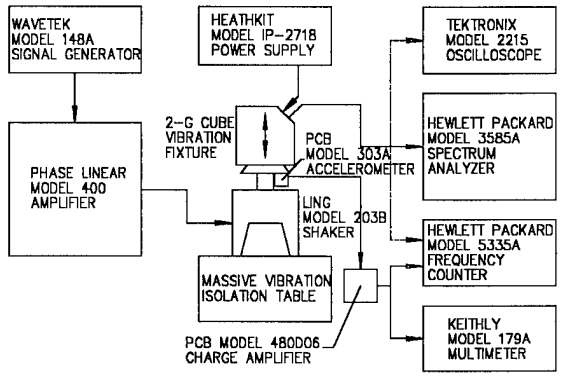


Figure 5. Measurement System with the Spectrum Analyzer

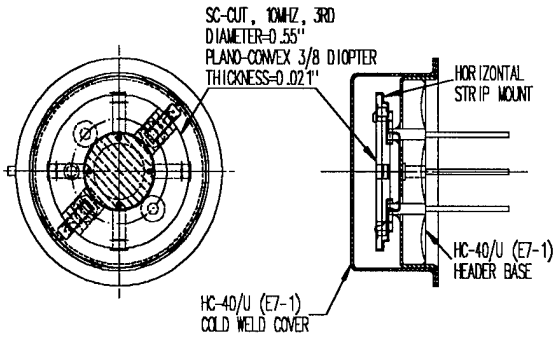


Figure 3. SC-Cut, 10MHz, 3rd 4-Point Horizontal Strip Mount Crystal Unit

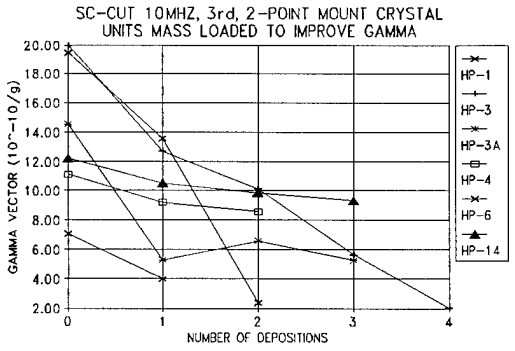


Figure 6. Deposition Histories for the SC-Cut, 10MHz, 3rd Asymmetric 2-Point Mount Crystal Units

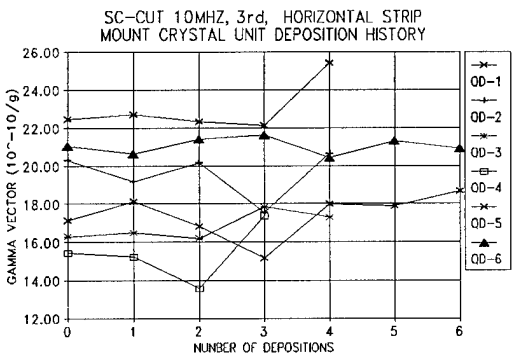


Figure 7. Deposition Histories for 4-Point Horizontal Strip Mount Crystal Units

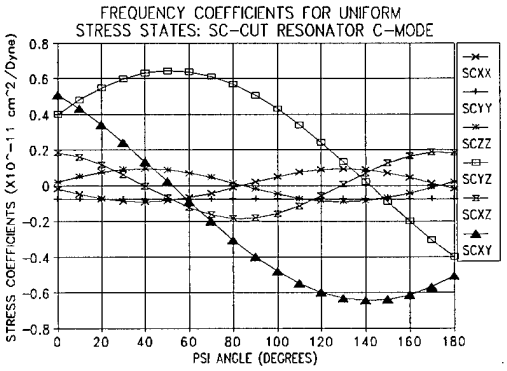


Figure 10. Stress Coefficients for SC-Cut Resonators

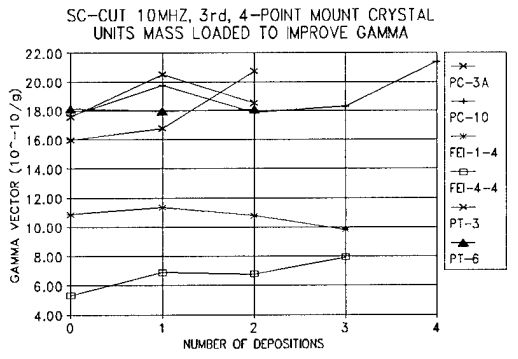


Figure 8. Deposition Histories for SC-Cut, 10MHz, 3rd 4-Point Crystal Units of Various Designs

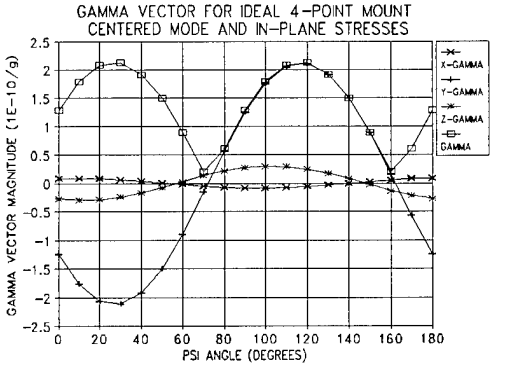


Figure 11. Gamma Vector and Its Components for the 4-Point Mount Crystal Unit and Planar Stresses

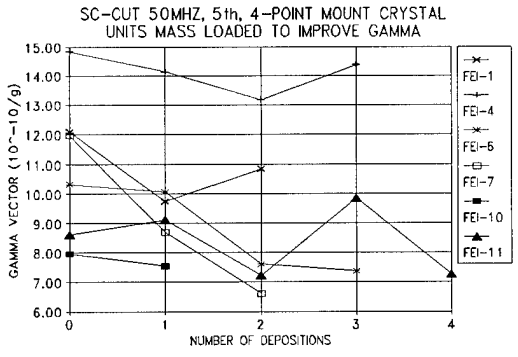


Figure 9. Deposition Histories for the SC-Cut, 50 MHz, 5th 4-Point Mount Crystal Units

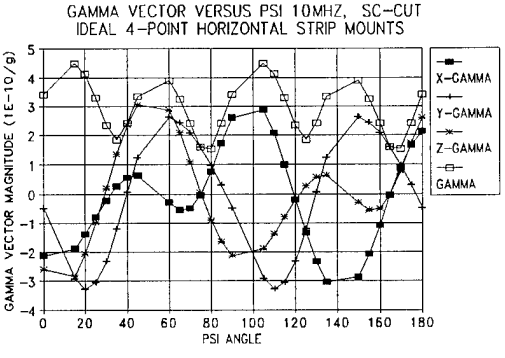


Figure 12. Gamma Vector for the 4-Point Mount Crystal Unit from the Rigorous Solution Technique

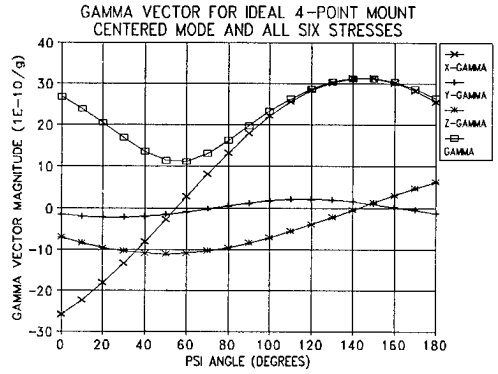


Figure 13. Gamma Vector and Its Components for the 4-Point Mount Crystal Unit and a Full Stress Tensor

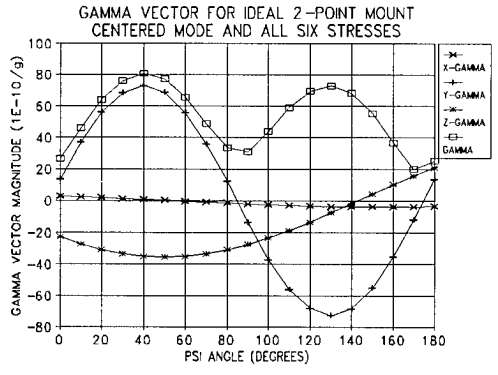


Figure 14. Gamma Vector and Its Components for the Symmetric 2-Point Mount Crystal Unit

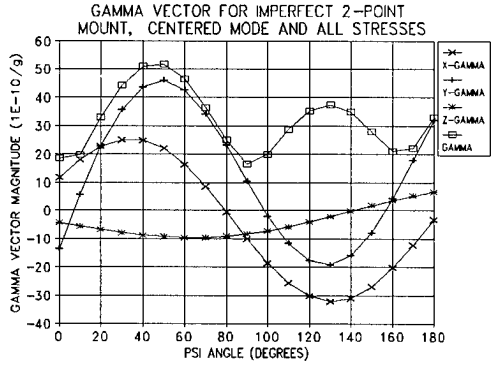


Figure 15. Gamma Vector and Its Components for the Asymmetric 2-Point Mount Crystal Unit

INVESTIGATIONS OF GOLD FILMS ON QUARTZ CRYSTALS

Mary Hendrickson[1], Sally Laffey[2] and John R. Vig[1]
[1]U.S. Army Research Laboratory, Ft. Monmouth, NJ 07703
[2]Vitronics Inc., 15 Meridian Rd., Eatontown, NJ 07724

Abstract

Changes at the interface between the metal electrode and the quartz plate have been identified as possible instability mechanisms in high precision resonators. This study was undertaken in order to investigate: the nature of the gold-quartz interface, if gold diffuses into quartz under "normal" conditions, and whether or not the adhesion of gold electrodes changes with time. Various metal adhesion tests, etching tests, heating tests, and surface analyses were employed during the investigations. Resonators fabricated in 1973 were compared with those processed nearly 20 years later.

The results showed no evidence of electrode diffusion. The electrodes of 20-year old resonators could be removed with Scotch Magic tape as easily as those from recently fabricated resonators. Even the weak adhesion of 3M Post-it Notes can remove the gold from both old and new samples. Testing for gold in the quartz with Rutherford backscattering and SIMS analysis indicated no more gold in the 1973 samples than in the recently fabricated ones. Upon removal of the gold electrodes from both old and new samples, a "ghost" image of the electrodes can be produced by lightly etching the quartz plates. These ghosts have been found to be due to differences in contamination levels between the areas under and outside the electrodes. When old and new samples were UV/ozone cleaned prior to etching, no ghost images were produced by the etching. No evidence was found for any disturbance of the quartz surface by the gold electrodes.

Gold electrodes on quartz have been found to consist of a weakly adhering central area and a strongly adhering narrow rim. The narrow rim is a <40 nm thick film that results from shadowing by the evaporation mask. The strong adhesion of this rim can be reproduced by depositing 10 nm to 40 nm thick pure gold films onto quartz plates. However, when additional gold is deposited onto such thin films in order to produce the typical gold thickness on quartz resonators, i.e., a total of ~100 nm, the adhesion again becomes weak. The thicker films (both layers) can be readily peeled off the quartz with Scotch tape. No satisfactory explanation has been found for the strong adhesion of the thin gold film.

Introduction

Changes at the interface between the pure gold electrode and the quartz plate have been identified as possible instability and failure mechanisms in high precision resonators. Studies by Belser and Hicklin [1] indicated the possibility of gold diffusion into quartz as resonators age. Gold has also been shown to diffuse into quartz at high temperatures under high electric fields [2]. Additionally, when electric fields are applied to the electrodes of resonators, rapid frequency drift can result, even at normal laboratory ambient temperatures [3], presumably due to diffusion effects. Whether or not gold diffuses into quartz under standard processing and extended storage or usage conditions, and if the adhesion of gold films on quartz resonators changes with time have not been reported in the literature. If such diffusion or adhesion changes do occur, then these could be mechanisms for instabilities such as aging [4], hysteresis [5] and noise [6,7]. Therefore, a study was undertaken to investigate if gold diffuses into quartz, the nature of the gold-quartz interface, and whether or not the adhesion of gold electrodes changes with time. The approach used in this investigation was to compare resonators fabricated in 1973 with recently processed resonators. Various metal adhesion tests, etching tests, heating tests, and surface analyses were employed during the investigations.

The "Ghost" Problem

Belser and Hicklin's neutron activation analysis and electron microprobe X-ray analysis results indicated gold diffusion into quartz [1]. Belser and Hicklin had also found that, when the electrodes were chemically removed and the quartz blanks were etched for a short time in a saturated solution of ammonium bifluoride, a shadow image of the metal electrode patterns became visible on the quartz surfaces. The electrodes seemed to have left a "ghost" image on the quartz surfaces. These images seemed to indicate that some type of change had occurred at the quartz surfaces, possibly gold diffusion. The initial step in the present investigation was to verify the observations of Belser and Hicklin.

Initially, AT-cut units fabricated in 1973 were examined. These units were the oldest precision resonators that were available. These units had been

0-7803-0905-7/93 $3.00 © 1993 IEEE

stored in a cabinet at room ambient conditions. The units had pure gold electrodes on mechanically polished surfaces and had been cold-weld sealed into HC-36 packages. The hermetically sealed packages were cut open, and the electrodes were chemically removed from the quartz plates using a cyanide based etchant. The quartz blanks were then inspected using an optical microscope (at 70X) for a ghost image of the electrode pattern. No image was visible on the samples at this point. The ghost patterns became visible upon etching the samples in a saturated solution of ammonium bifluoride at 75°C for three seconds. Figure 1 shows a typical ghost pattern generated on the surface of a quartz plate.

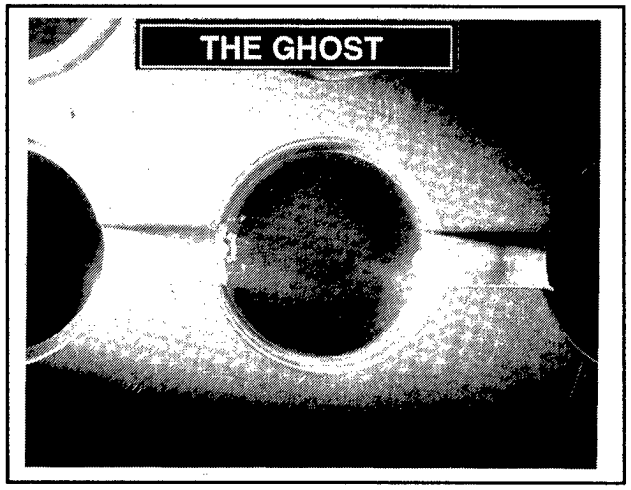


Figure 1 Ghost pattern etched in quartz blank.

Recently processed resonators were similarly examined for the presence of ghost images. These units were SC-cuts with pure gold electrodes, the quartz blanks had been finished with 3 micron abrasive, then chemically polished in a 4:1 ammonium fluoride - hydrofluoric acid solution [8]. The blank cleaning process [9] prior to metallization of the recent samples consisted of a 15 minute immersion in a 1:1:5 solution of NH_4OH : H_2O_2 : H_2O at 80°C (SC1), rinsing in DI water, a 15 minute immersion in a 1:1:5 solution of HCL : H_2O_2 : H_2O at 80°C (SC2), rinsing in DI water, drying in UHP filtered nitrogen, then five minutes UV/ozone cleaning [10] on each side. Prior to deposition, the chamber was pumped to a base pressure of 7×10^{-8} torr, then the samples were heated to 200°C for approximately one hour until the vacuum recovered to 1×10^{-7} torr. The electrodes were deposited using a CVC SC-6000 electron beam multimetal evaporator, at a rate of 0.5 nm/sec. A total of 100 nm of gold was deposited onto the quartz surfaces. After the metallization, no further cleaning was performed. (No information was available on the cleaning of the 1973 samples.)

The "recent" samples were prepared six months to one year prior to the tests and had been stored in hermetically sealed packages until the tests. The electrodes were chemically removed using a cyanide based etchant. Upon etching the quartz blanks in a saturated solution of ammonium bifluoride at 75°C for three seconds, the ghost images became visible on these samples also. This test was repeated on quartz blanks immediately after metal deposition, with the same results. Therefore, the conditions responsible for producing the "ghost" image did not appear to be time dependent but occurred during or after electrode deposition. The generation of this image also appeared to be independent of the type of cut, as the 1973 vintage resonators were AT-cuts and the recent resonators were SC-cuts. Further testing was required to verify these results, to determine if gold diffusion occurs and to identify the mechanism for the ghost pattern.

Various possibilities were considered in order to explain the ghost patterns. After chemical removal of the electrodes from both new and old samples, two surface conditions were observed on the quartz surfaces. Upon removing the quartz sample from the liquid etchant, a hydrophilic surface was observed on the quartz surface in the exact shape and location of the electrode metal, and a hydrophobic surface was observed outside the electrode areas.

One possibility considered for the existence of two different surface conditions was that the electrode deposition process changed the hydroxyl groups on the quartz surface [11,12]. It is well known that when a quartz surface is heated, surface hydroxyl groups are desorbed at temperatures above 150°C. The desorption is reversible up to about 350°C. Above 350°C, the surface begins to be irreversibly dehydrated due to the formation of siloxane (Si-O-Si) bridges. Upon heating to above 500°C, the contact angle increases rapidly. (Above 800°C, the quartz surface becomes completely hydrophobic, with a contact angle of about 40°.) It is also known that, during electrode deposition, the quartz surface is heated to well above the nominal substrate temperature [13-15]. The extra heat results from radiated heat from filaments (also from electron bombardment and ion bombardment where applicable), and from the energy released during condensation of the metal. The temperature rise can be high when the film is very thin; e.g., in one report, a temperature rise of up to 400°C was observed at a 10 nm thickness during deposition of Ag at a rate of 0.3 nm/s [13,15]. If the extra heating irreversibly dehydrated the surface, then the surface under the electrode would have become hydrophobic, which was not so.

Another possibility for explaining these hydrophilic/hydrophobic surfaces is differences in surface

contamination. The quartz surface outside the metal electrode is more likely to be contaminated than the area beneath the electrode since the outside area is exposed to subsequent processing after metallization, and it can adsorb outgassing products inside the hermetically sealed enclosure. Once the electrode is removed, a cleaner quartz surface is exposed which may have a hydrophilic surface. The etching rate of the quartz after the electrode is removed may be affected by the surface contaminants in the area outside the electrode. This may cause the ghost patterning of the quartz blank in the shape of the electrode.

To test this hypothesis, 1973 vintage resonators and recently processed resonators were opened and UV/ozone cleaned [10] before electrode removal. After chemically removing the electrodes, the blanks were etched for three seconds in saturated ammonium bifluoride and inspected for the "ghost" images of the electrodes. No images were detected on either sample. All areas of the quartz blanks exhibited hydrophilic surfaces. This test was repeated three times using one sample of each vintage per test. No ghost patterns were found. Therefore, the "ghost" images appear to be due to etching differences caused by surface contamination around the electrodes rather than to gold diffusion (or other gold-quartz interactions).

Metal Adhesion Studies

In theory, if gold gradually diffuses into quartz, then the adhesion of gold electrodes would improve as resonators age. The electrodes of resonators manufactured 20 years ago should adhere more strongly than the electrodes of recently fabricated resonators. Therefore, some simple metal adhesion tests were conducted on 1973 vintage resonators fabricated with pure gold electrodes, and on recently fabricated samples that also had pure gold electrodes. The adhesion tests revealed that the gold electrodes could be readily peeled off from both types of samples by attaching Scotch Magic Tape [16] to the electrodes and pulling on the tape in a direction that is perpendicular to the plane of the film. The electrodes of both types of samples could also be readily peeled off with the very weak adhesive of 3M Post-it-Notes [16].

Rutherford backscattering (RBS) and secondary ion mass spectroscopy (SIMS) were used to determine the gold concentration in the quartz in the area from where the electrode had been peeled. The RBS results in Figure 2 show that gold was detected on the surface of the quartz on both the 1973 sample and the recent sample, at levels of 1 - 2 x 10^{12} atoms/cm^2, and no gold was detectable below the surface. Slightly more gold was detected on the blank of the recently fabricated sample than on the 1973 sample. This could be attributed to the different

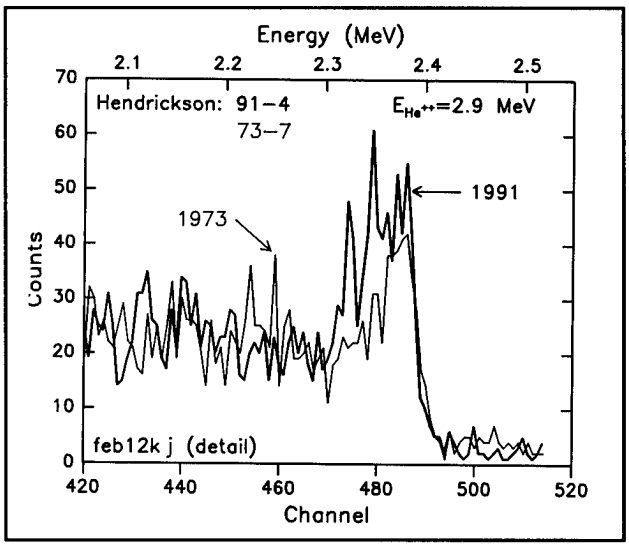


Figure 2 RBS analysis of old and recent samples.

fabrication processes used, i.e., the 1973 sample had a mechanically polished surface while the recent sample had a 3 micron lapped and chemically etched surface. If gold diffusion had occurred, the 1973 sample would have exhibited the higher gold content. The results obtained from the SIMS analysis, shown in Figures 3 and 4, also indicate the existence of gold near the surface of the quartz on both the new and the old samples. The amounts of gold detected on the 1973 sample and the recent sample are similar and do not suggest any gold diffusion.

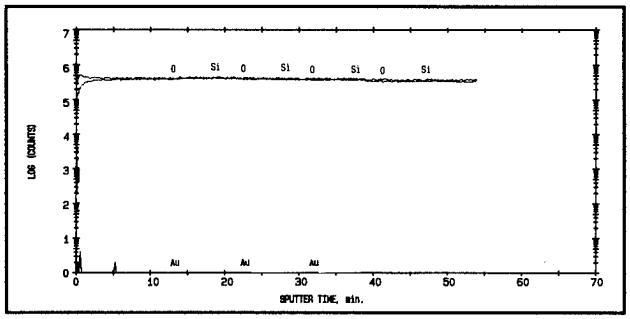


Figure 3 SIMS analysis of 1973 vintage resonator after gold electrode was peeled off.

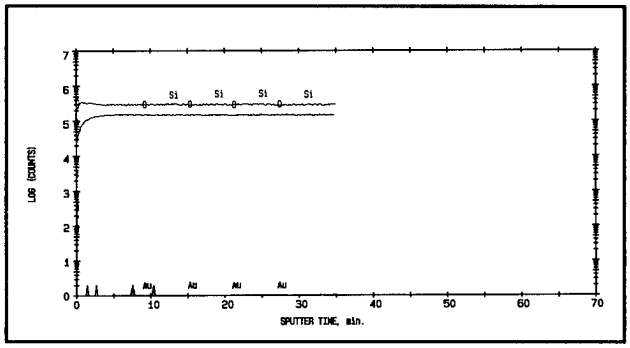


Figure 4 SIMS analysis of recent resonator after gold electrode was peeled off.

A closer inspection of these samples revealed that the adhesive on the tape had removed all of the gold except a thin narrow rim of gold outlining the electrode pattern. These metal rims could not be removed with Scotch Magic tape even after several attempts. The Scotch tape tests left the gold rims intact on both 1973 and freshly fabricated samples, and the rims are, therefore, not an artifact of aging. The rims appear as though they are formed from the shadowing of the evaporated metal through the electrode mask. The rims are visibly thinner than the bulk of the electrode metal, and the adhesion of these gold rims to the quartz is much greater than that of the thicker central portion of the electrode metal. The thickness of the central portion of the electrode was approximately 100 nm. An SEM micrograph was used to determine that the average thickness of the thin rim of gold was between 30 and 40 nm.

The entire gold rim could be readily removed with a gold etchant [17], at room temperature. Rubbing with a paper towel or scratching with a wooden stick failed to remove any of the rim, but scratching the film with the tip of a stainless steel tweezer did remove gold. Ultrasonic agitation [18] for one hour (20 min each, sequentially, in acetone, water, and isopropyl alcohol) also failed to remove the strongly adhering rim. The ultrasonic agitation did remove some flakes of gold from the inside of the rim, i.e., from the side facing the center of the electrode region, thus indicating an intermediate adhesion strength for this intermediate thickness film.

Various explanations for the strong adhesion of the rim were considered, but none proved to be satisfactory. One possible explanation was that, although the adhesion of a pure gold film to a perfect quartz surface is weak, imperfection on the quartz surface can result in tight binding sites for the gold atoms [13]. The imperfections can be, for example, defects in the bulk quartz that intersect the surface (i.e., dislocations), surface contamination, and etch features resulting from deep-etching the crystal blank.

Another possibility considered was that a very thin gold film can assume the surface crystalline structure of clean, single crystal quartz, whereas a thick gold film has a different crystalline structure, thus causing a lattice mismatch and, thereby, a weaker adhesion [19], i.e., the thinner gold forms a strained epitaxial layer on the quartz surface. As the gold thickness increases, the strains also increase. After a certain thickness, it becomes energetically more favorable for the gold at the interface to be "epitaxial" on the thick gold film. If this hypothesis is correct, then similarly deposited thin gold films should not adhere to glass or to unetched quartz.

A series of tests was conducted to determine the reasons for the strong adhesion of the thin gold films. To see if there was a thickness dependence to the adhesion, samples were metallized with the approximate gold rim thickness measured by the SEM, i.e. 30 to 40 nm. The Scotch tape test" failed to lift off gold from any area of these electrodes. The test was repeated several times and the time lapse between deposition and test varied from an hour to two weeks. Some samples were tested again after four months and the "Scotch tape test" still could not lift off any gold. Additional samples were prepared with approximately 10 nm thick gold electrodes. The same results were obtained, i.e., no gold could be peeled off with the Scotch Magic tape. On a few of these samples, the adhesive on the tape peeled off the tape backing because it adhered so strongly to the clean gold electrode.

The adhesion of thermally evaporated and electron beam evaporated gold was also compared. Quartz blanks were cleaned using the same procedure as described earlier: SC1, SC2, and UV/ozone. One side of the blank was metallized with 10 nm of thermally evaporated gold. The other side was coated with 10 nm of e-beam evaporated gold. Vacuum was not broken between depositions. The evaporation rates for both types of evaporations were between 0.2-0.3 nm/sec. Although the grain size appeared smaller in the thermal gold than in the e-beam gold, no differences in adhesion were found. The "Scotch tape test" produced no peeling on either side.

A series of tests was conducted to investigate the effect of removing the amorphous layer of SiO_2, if any, using a 50% concentration fluoboric acid in water [20]. AT-cut quartz blanks were prepared by mechanically polishing the surfaces with a one micron rare earth slurry. Two of these samples were metallized with 100 nm of gold after polishing, another two samples were etched in a saturated solution of ammonium bifluoride at 75°C before electrode deposition, and the last two samples were etched in the saturated solution and then also etched in fluoboric acid before metal deposition. In all cases, an improvement in surface quality was observed after the fluoboric acid dip. The samples were then metallized with 100 nm of gold. Electrode adhesion tests were performed on each of these samples. The adhesions were again poor which indicates that the fluoboric acid had no obvious effect on the adhesion of the gold to the quartz surface.

The adhesion of thin (10 nm) and thick (100 nm) gold films was tested on both crystalline and noncrystalline substrates. Both p- and n-type single crystal silicon wafers were coated with 10 nm and 100 nm of gold. No peeling was observed on the 10 nm samples and complete peeling was observed on the 100 nm samples. Window glass was then coated with 10 nm and

100 nm of gold. On one set of samples, the window glass was etched prior to metallization. The same results were obtained on both the etched and unetched surfaces, i.e. the "Scotch tape test" produced no peeling at the 10 nm thickness and complete peeling at the 100 nm thickness. Table 1 summarizes the adhesions tests on crystalline and noncrystalline substrates. ("Standard Polish" was the as-received surface;the polishing processes are unknown to the authors.) The results show that, independent of whether or not the substrate is crystalline or noncrystalline, etched or unetched, the adhesion of 10 nm thick gold is much greater than that of 100 nm thick gold.

Table I Gold film adhesion results

Substrate	Preparation	Etching	10nm Au	100nm Au
AT-cut Quartz	1micron polish	No	No Peeling	Peeled
AT-cut Quartz	1micron polish	Yes	No Peeling	Peeled
Silicon Wafer 100 P-Type	Standard Polish	No	No Peeling	Peeled
Silicon Wafer 111 P-Type	Standard Polish	No	No Peeling	Peeled
Window Glass	Standard Polish	No	No Peeling	Peeled
Window Glass	Standard Polish	Yes	No Peeling	Peeled

Surface cleaning was same for all: SC1 + SC2 + UV/O_3

During thin film deposition, island formation is one of the early stages in the film growth process [13,14]. During this stage, the atoms migrate to nucleation sights on the surface of the substrate, thereby possibly increasing the adhesion of the film. It was surmised that if a thin layer of gold, i.e., 10 nm, was deposited on the quartz surface and given time to migrate to tight binding points, and then a thicker layer of gold was added to the initial layer, the thicker film may maintain the strong adhesion of the 10 nm film. This turned out not to be the case. The deposition chamber (a CVC multimetal system) was pumped to 7 x 10^{-8} torr and heated to nominally 224°C for one hour. The deposition was performed after the chamber cooled to below 50°C. The first 10 nm was deposited at 0.1 nm per second. A waiting time of 10 minutes preceded the second layer of gold which was deposited at 0.5 nm per second. The resulting two-layer film peeled off easily using the Scotch tape test. There was no discernible increase in gold adhesion.

Heat Treatments and SIMS Analysis

Tests were conducted to determine if heat, without an electric field, would induce gold diffusion or change the electrode adhesion characteristics. Recently fabricated samples and samples made in 1973 were removed from their packages and subjected, in pairs, to heat cycles at 300°C and 600°C. One pair was placed in a vacuum furnace and heated to 300°C for 24 hours. The topography of the electrode metal was not visibly changed by either heat treatment. The electrodes were easily removed from these samples using Scotch Magic tape. No obvious increase in adhesion of the gold to the quartz could be observed. The thin rim of gold outlining the periphery of the electrode was observable, as discussed earlier.

SIMS analysis was conducted on these samples. Several different spots within the electrode areas on each sample were analyzed. No gold was detected on either sample. (See Figure 5.)

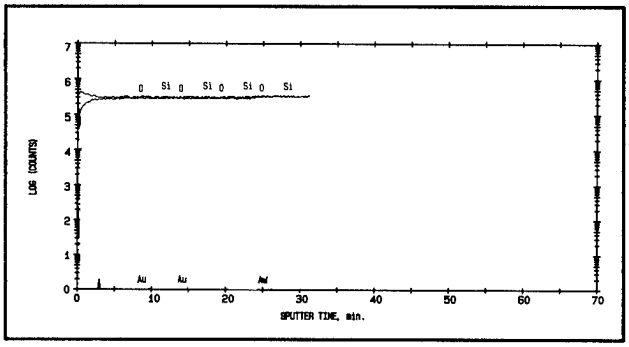


Figure 5 SIMS analysis of recent sample after 300°C heat treatment.

During one heat cycle test, the thermocouple completely malfunctioned and the temperature of two samples increased to approximately 600°C (which is above the α-β transition of quartz and is also above the gold-silicon eutectic temperature). One was a 1973 sample and the other, a newly fabricated unit. The electrodes of both samples appeared to have changed drastically. Under high magnification, small islands of gold within the electrode pattern were visible, as is shown in Figures 6 and 7. The

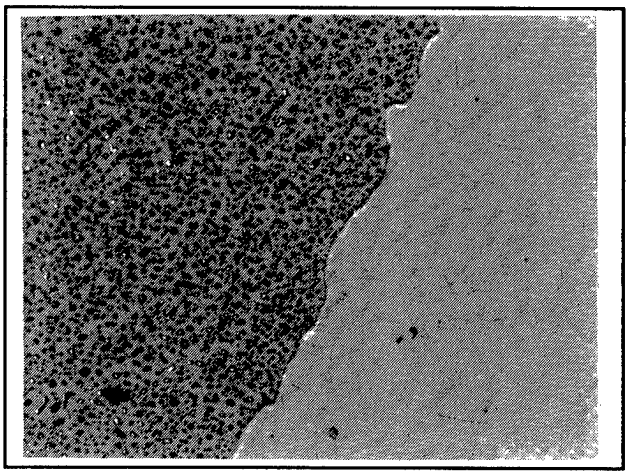


Figure 6 Photograph of 1973 sample after 600°C heat treatment.

1973 sample appeared to have larger areas of intact electrode metal, but on both samples, the gold appeared to have moved or to have evaporated. This could be attributed to the level of cleanliness of the surface or the fact that the 1973 sample had a polished surface and the 1991 sample had a 3 micron lapped chemically polished surface. Adhesion tests on these samples showed a strong adhesion. No peeling was observed upon the "Scotch tape test".

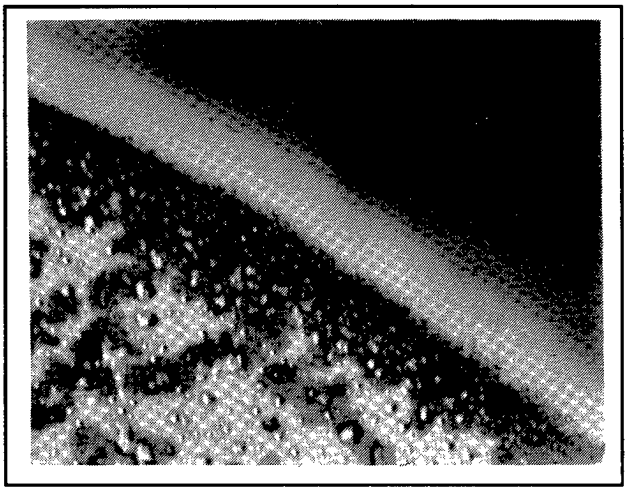


Figure 7 Photograph of recent sample after 600°C heat treatment.

Comments on the Adhesion of Gold Electrodes

When a pure gold electrode (without an adhesive layer) is deposited onto a clean quartz plate, the electrode consists of a weakly adhering central portion, a strongly adhering narrow rim, and, probably, a transition region of changing film thickness, where the adhesion changes rapidly from strong to weak. All but the narrow rim can be readily removed with even the very weak adhesion of Post-it Notes. The electrode adhesion, however, is strong enough not only for typical applications but also for severe environment applications. For example, high-shock resonators with pure gold electrodes can readily withstand shocks of 30,000 g of 12 ms duration [21]. No failures occurred due to electrodes coming loose. Resonators with pure gold electrodes can also withstand very high drive levels. For example, in Filler's studies of the drive level sensitivity of SC-cut resonators [22,29], he was able to use drive levels as high as 24 ma (corresponding to >100 mW of dissipation), without driving off the electrodes. Filler has also used ~20 ma of driving current in order to test resonators for the effect of high drive current on short term stability subsequent to the high driving. (He found that some resonators' stability improves, some do not change, and a few are destroyed by the high driving.)

If the electrodes affect the stability of resonators, it may be that changes in the rim area are responsible. The effects of varying this rim area are yet to be determined.

It is "well known" that gold films do not adhere to oxide surfaces, such as quartz [13,23,24]. "The generally accepted criterion for adhesion between a metal film and an oxide substrate is that the metal must be oxygen active so as to react chemically with the oxide surface, forming an interfacial reaction zone..." [23]. Since gold does not form a stable oxide under normal conditions (having a heat of oxide formation of +19kcal/mol), according to conventional wisdom, it is not supposed to adhere to oxide surfaces. Good adhesion has been obtained by sputtering in an oxygen atmosphere [25], by irradiation of the film subsequent to deposition [26-28], and by having the gold diffuse into the substrate material [23] (e.g., when the substrate is a high-lead glass or tin-oxide doped glass into which the gold can diffuse).

Since the gold adhesion showed no noticeable improvement 20 years after deposition, and since neither the thinner nor the thicker gold films were irradiated, the diffusion effects and the irradiation effects cannot explain the strong adhesion of the thinner films. That sputtering in an argon atmosphere produced poorly adhering gold films whereas sputtering in an oxygen-containing atmosphere resulted in strongly adhering films [13] may be relevant to our results. Sputtering in an oxygen atmosphere probably resulted in atomic oxygen which made the substrates super-clean, as does UV/ozone cleaning. However, no satisfactory explanation has been found for the strong adhesion of the thin gold film and the drastic decrease in adhesion when the film thickness exceeds about 40 nm.

Conclusions

No evidence was found of electrode diffusion or other permanent changes due to pure gold electrodes. The ghost images of the electrodes appear to be due to surface contamination because when the surfaces are properly cleaned prior to etching, no ghosts result. It is possible that the diffusion reported by Belser and Hicklin [1] was due to the resonators' having been exposed to a voltage across the electrodes in an oscillator or during testing [3].

Gold electrodes on quartz have been found to include a weakly adhering central area and a strongly adhering narrow rim. The narrow rim is a < 40 nm thick film which results from shadowing by the evaporation mask. The very strong adhesion of this rim can be reproduced by depositing 10 nm to 40 nm thick pure gold films onto quartz plates, silicon wafers or window glass. However, when additional gold is deposited onto such thin films, to

produce the typical gold thickness on quartz resonators, i.e., a total of ~100 nm, the adhesion again becomes weak. The thicker films (both layers) can be readily peeled off the quartz with Scotch Magic tape. No satisfactory explanation has been found for the strong adhesion of the thin gold film.

Gold is often the preferred electrode material for high-precision resonators, primarily because of its weak interaction with the quartz surface, its lack of a stable oxide under normal conditions, and its ease of cleaning [10]. Other electrode materials, such as chromium, aluminum, and nickel, have stronger adhesion to quartz, but also produce much higher stresses at the electrode-quartz interface [30]. The weak adhesion of gold on quartz in properly made resonators results in no reliability or ruggedness problems. The question of whether or not the electrodes contribute to instabilities is yet to be answered.

Acknowledgements

The authors would like to acknowledge the surface analysis work provided by T. Monahan, R. Lareau (SEM and SIMS), and R. Pfeffer (RBS).

References

1. R.B. Belser and W.H. Hicklin, 23rd Ann. Symp. on Frequency Control, pp. 132-142, 1969.

2. J. Gualtieri, "The Influence of Temperature and Electric Field on the Etch-Channel Density in Swept Cultured Quartz," 39th Ann. Symp. on Frequency Control, pp. 247-258, 1985.

3. R. L. Filler, et al., "Aging Studies on Quartz Crystal Resonators and Oscillators," Proc. 38th Ann. Symp. on Frequency Control, pp. 225-231, 1984, AD-A217381.

4. J. R. Vig and T. R. Meeker, "The Aging of Bulk Acoustic Wave Resonators, Filters and Oscillators," Proc. 45th Ann. Symp. on Frequency Control, pp. 77-101, 1991, IEEE Cat. No. 91CH2965-2.

5. J. A. Kusters and J. R. Vig, "Hysteresis in Quartz Resonators - A Review," IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, pp. 281-290, 1991.

6. J. J. Gagnepain, "Fundamental Noise Studies of Quartz Resonators," Proc. 30th Ann. Symp. on Frequency Control, pp. 84-91, 1976.

7. F. L. Walls, et al., "A New Model of I/f Noise in BAW Quartz Resonators," Proc. 1992 IEEE Frequency Control Symp., pp. 327-333.

8. J. R. Vig, R. Brandmayr, and R. L. Filler, "Etching Studies on Singly and Doubly Rotated Quartz Plates," Proc. of the 33rd Ann. Symp. on Frequency Control, pp. 351-358, 1979.

9. J. R. Vig, "Contamination Control Methods - A Review of Recent Progress," Proc. of the 44th Ann. Symp on Frequency Control, pp. 201-206, 1990.

10. J. R. Vig, "Ultraviolet-Ozone Cleaning of Semiconductor Surfaces," in The Handbook of Semiconductor Wafer Cleaning Technology, W. Kern, editor, Noyes Publications, 1993.

11. M. L. Hair, "The Molecular Nature of Adsorption on Silicon Surfaces," Proc. 27th Ann. Symp. on Frequency Control, pp. 73-78, 1973.

12. M. L. White, "Clean Surface Technology," Proc. 27th Ann. Symp. on Frequency Control, pp. 79-88, 1973.

13. K. L. Chopra, Thin Film Phenomena, pp. 162, 316-321, McGraw-Hill Book Co.,1969.

14. C. A. Neugebauer, "Condensation, Nucleation and Growth of Thin Films," in Handbook of Thin Films Technology, edited by L. I. Maissel and R. Glang, McGraw-Hill Book Company, 1970.

15. M. V. Belous and C. M. Wayman, J. Appl. Phys., Vol. 38, pp. 5119-5124, 1967.

16. 3M Commercial Office Supply Div., St. Paul, MN 55144-1000.

17. The etchant was Technistrip, manufactured by Technic, Inc., and contains potassium cyanide, lead oxide, and benzoic acid.

18. The ultrasonic cleaner was a 40 kHz, 340 watt unit made by Branson Cleaning Equipment Co..

19. E. Poindexter, private communication, March 1993.

20. L. Morawiec, "State of the Art VHF Quartz Blanks Processing," Proc. 13th Piezoelectric Devices Conference and Exhibition, Electronic Industries Assoc., pp. 1-5, 1991.

21. R. L. Filler, et. al., "Ceramic Flatpack Enclosed AT and SC-cut Resonators," Proc. 1980 IEEE Ultrasonics Symp., pp. 819-824, 1980.

22. R. L. Filler, "The Amplitude-Frequency Effect in SC-cut Resonators," Proc. of the 39th Ann. Symp. on

Frequency Control, pp. 311-316, 1985.

23. D. M. Mattox, "Thin Film Metallization of Oxides in Microelectronics," Thin Solid Films, vol. 18, pp. 173-186, 1973; also in Proc. 27th Ann. Symp. on Frequency Control, pp. 89-97, 1973.

24. K. E. Haq, K. H. Behrndt, and Ilse Kobin, "Adhesion Mechanism of Gold-Underlayer Film Combinations to Oxide Substrates," J. Vac. Sci. & Techn. vol. 6, pp. 148-152, 1969.

25. D. M. Mattox, "Influence of Oxygen on the Adherence of Gold Films to Oxide Substrates," J. Appl. Phys., vol. 37, pp. 3613-3615, 1966.

26. P. Benjamin and C. Weaver, "Adhesion of Metal Films to Glass," Proc. Royal Soc. London, Ser. A, vol. 254, pp. 177-183, 1960.

27. S. A. Varchenya and G. P. Upit, "Adhesive Strength of the Interface Between a Metal Film and Crystalline Quartz," Thin Solid Films, vol. 122, pp. 59-62, 1984.

28. I. V. Mitchell, et al., "Thin Film Adhesion Changes Induced by Electron Irradiation," Appl. Phys. Lett., vol. 44, pp. 193-195, 1984.

29. R. L. Filler, private communication, June 1993.

30. K. Haruta and W. J. Spencer, "Strain in Thin Metal Films on Quartz," J. Appl. Physics, vol. 37, pp. 2232-33, 1966.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PRECISION FREQUENCY TRIMMING OF SAW AND STW RESONATORS USING Xe^+ HEAVY ION BOMBARDMENT

V. Sh. Aliev[+] and I. Avramov [*]

+ Institute of Semiconductor Physics Rus. Acad. of Sci. SB,
13 Academician Lavrentyev Aven. Novosibirsk 630090 RUSSIA
* Institute of Solid State Physics Bulgarian Acad. of Sci.,
72 Tzarigradsko Chaussee Blvd.,Sofia 1784 BULGARIA

A method for precision frequency trimming of surface acoustic wave (SAW) and surface transverse wave (STW) based resonant devices using a Xe^+ heavy ion bombardment technique is described. The devices are donwtrimmed in frequency in an in-situ monitoring process by means of a Kaufmann type ion source that allows first a rough and then a fine frequency trimming with an accuracy of 1 ppm. An improvement of the device insertion loss and unloaded Q as a result of the trimming process is achieved. Single mode 776 MHz STW resonators can be downtrimmed by more than 5000 ppm without deteriorating their parameters while SAW resonators allow a much lower frequency downshift. The method is simple and can cost effectively be applied to SAW and STW device fabrication.

Introduction

Plasma etching techniques are widely used for frequency trimming of SAW resonators on piezoelectric quartz . By removing part of substrate material between electrodes it is possible to achieve a desired resonator frequency and improve device parameters. Potentials of ion reactive etching techniques for making of SAW devices with quartz crystalline substrate have been studied in the paper [1]. CHF_3 gas was used for etching of quartz and etching rates have been obtained from 3 to 30nm/min. However, the fine frequency trimming require lower rates of etching. The etching rate may be decreased by lowering the ion current density. But the fall in the current density gave rise to formation of a fluorine carbon containing polymer film on the surface and, consequently, to uncontrolled etching.Therefore, physical sputtering (ion etching) with ions of noble gas instead of ion-chemical etching with CHF_3 gas is preferred for fine frequency trimming.

Another feature required from ion etching is that the etching rate must be considerably lower for aluminum electrodes than that for the quartz substrate. The rate of ion etching is dictated by a sputtered yield which weakly depends on chemical nature of material and increases with the growth of ion mass. The relation between the etching rate of aluminum and quartz may be improved by using a blockage of sputtering of metals in the presence of chemical reaction gas [2]. By supplying oxygen into a vacuum chamber the sputtering rate of aluminum may be decreased several times while that of oxygen containing compounds is not practically changed. In our work for a partial pressure of oxygen about $2 \cdot 10^{-3}$ Pa the sputtering rate of aluminum was lower than 5Å/min (Xe^+,0.2 ma/sm), whereas that of quartz was 40 Å/min.

Experimental

Fig.1 shows schematically an experimental apparatus for precision frequency trimming process. The sample 7 of a SAW (or STW) resonator was attached to the holder 9 in the vacuum chamber 1 which was pumped with the turbomoleculer pump 2 by 850 l/s rate. The ion current density was measured with the Faraday cylinder 11. The testing system 3 consisted of a network-analyzer and frequency counter. The SAW (or STW) resonator cable connected through the holder 9 and the vacuum high frequency current lead-in to the testing system. Additional loss in the electrical line of the holder 9 was about 0.5 dB for the STW and 1 dB for the SAW resonators.

Trimming frequency process was performed in the sequence of steps:

1.The resonator substrate was inserted to the holder.The vacuum chamber was pumped down to $5 \cdot 10^{-4}$ Pa.

2.The starting parameters of the resonator were measured.

3.The ion source was switched on (the shield 8 was down). Working parameters of ion source were set up.

0-7803-0905-7/93 $3.00 © 1993 IEEE

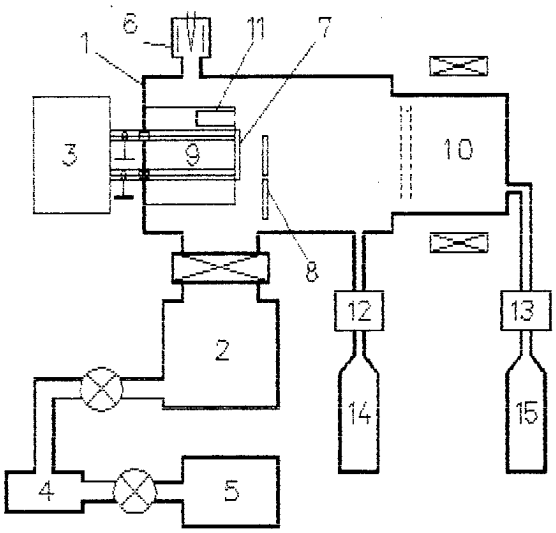


Fig.1.Schematic diagram of the experimental apparatus:(1) vacuum chamber; (2) turbomolecumer pump; (3) testing system; (4) vacuum trap; (5) rotary pump; (6) ionization vacuum gauge; (7)substrate of SAW or STW resonator; (8) shield; (9) holder; (10) ion source; (11) Faraday cylinder; (12),(13)needle valves; (14) oxygen vessel; (15) Xe vessel.

4.Oxygen gas was supplied to about $3\ 10^{-3}$ Pa.

5.The shield was up and the frequency trimming was made. A frequency response was controlled by means of the testing system.

The samples of SAW and STW resonators were made using the back photolitography technology. A standard double- electrode transducer had geometric parameters:
for the STW resonators: a width of electrodes - 1.6 mkm, an electrode/spacing widths ratio - 1, the number of the transducer electrodes -150, the number of the reflection array electrodes - 900, a transducer aperture - 0.65mm;
for the SAW resonators: a width of electrodes - 2.26 mkm, an electrode/spacing widths ratio - 1, the number of the transducer electrodes -100, the number of the reflection array electrodes - 400, a transducer aperture - 0.35mm.

Results

Starting parameters of the STW resonator, which was inserted in the vacuum holder, were following:
F0 = 774082 kHz is a central frequency,
ΔF(3dB) = 369 kHz is a 3 dB bandwidth,
SLS = 7 dB, is a sidelobe level loss,
IL = 7.1 dB is an insert loss.

A frequency response of the unetched STW resonator is shown in Fig.2. Once the ion source has been switched and the shield has been opened, the downshift of the central frequency and satellite was observed (Fig.3). Concurrent with the shift of the frequency, the insert loss

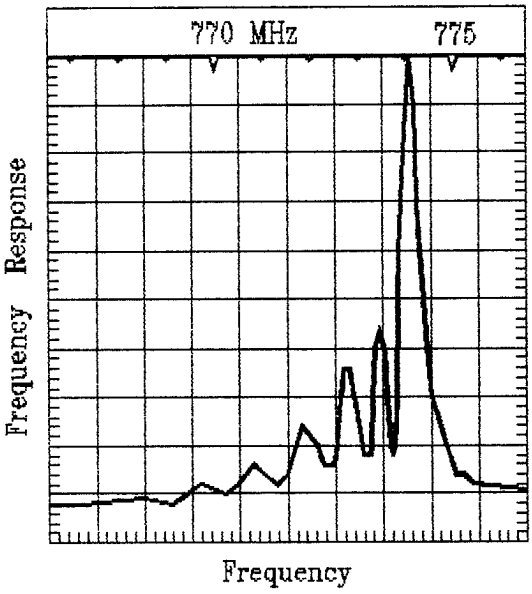


Fig.2.Frequency response of the unetched STW resonator.

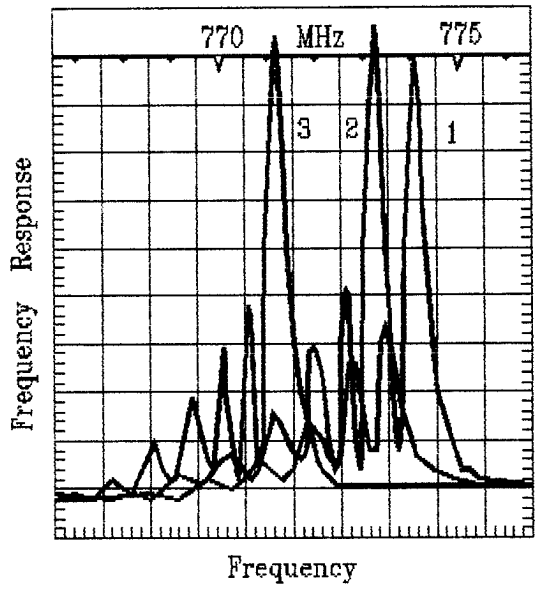


Fig.3.Frequency trimming process for STW resonator. Downshift is (1) 0 MHz, (2) 1MHz, (3) 3 MHz. The ion current density is 0.21 ma/sm^2.

was decreased and the unloaded Q-factor was increased (Fig.4). The parameters of the STW resonator were better than the starting ones for the downshift less 4 MHz. For the downshift more 4 MHz the shape of the frequency response changed and the insert loss increased. At the central frequency peak a poorly observed shoulder appears for the downshift 6.8 MHz were clearly seen for the downshift 11.9MHz (Fig.5). Absolute amplitudes of the satellite frequency response did not change roughly during frequency trimming process.

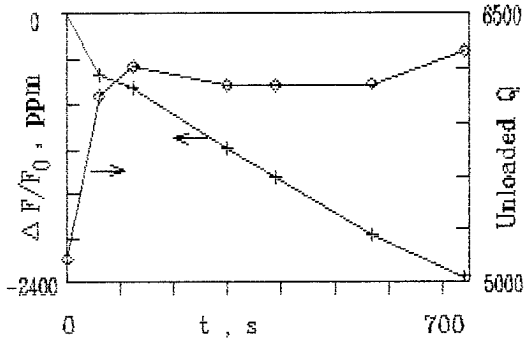


Fig.4.The relative frequency downshift ΔF/F and the unloaded Q-factor for the STW resonator vs time t of frequency trimming.

A behavior of SAW resonators was much the same as that of the STW resonators. Fig.6 shows the frequency response of the unetched SAW resonator. Its starting parameters were:
F0 = 360720 kHz,
ΔF(3dB) = 65 kHz,
SLS = 15 dB,
IL = 6 dB.
After beginning of the ion bombardment the insert loss of the resonator decreased, whereas the unloaded Q-factor increased (Fig.7). For the downshift less 640 kHz all the parameters of the resonator were better than the starting ones. A maximum of the central frequency response was observed for the downshift about 260 kHz. In that case the insert loss was less than 1.5 dB and the unloaded Q-factor was more than 20% in relation to the starting Q-factor. The following downshift caused deterioration of the resonator parameters.

Of a practical significance is a downshift time- dependence for the certain ion current density (Fig.8). Just after the beginning of the ion bombardment a downshift rate was maximum. Then the time-dependence F(t) slowed down and became a linear dependent. Its slope was governed by the ion current density. The conversion from the dependence Δf(Δt)=F(t)-F0 (Fig.9), which was measured with the variety of the ion current density, to a dependence Δf(G), where: G=J*Δt - an exposure of ion bombardment, showed that after the initial region the downshift was linearly proportional to the exposure of the ion bombardment (Fig.10).

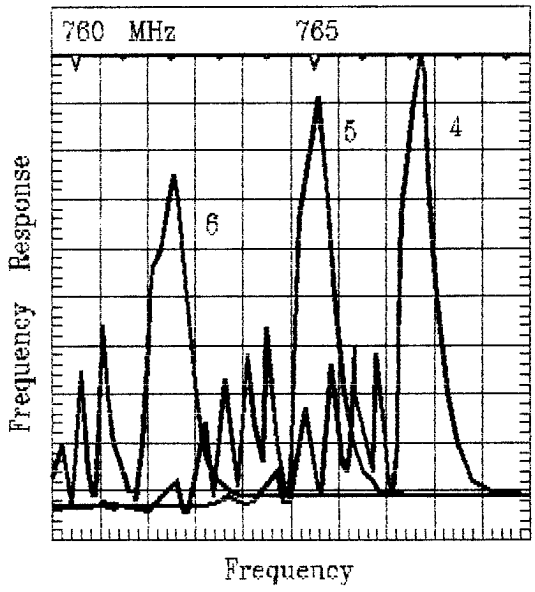


Fig.5.Frequency trimming process for the STW resonator. Downshift is (4) 6.9 MHz, (5) 8.9 MHz, (6) 11.9 MHz. The ion current density is 0.21 ma/sm^2.

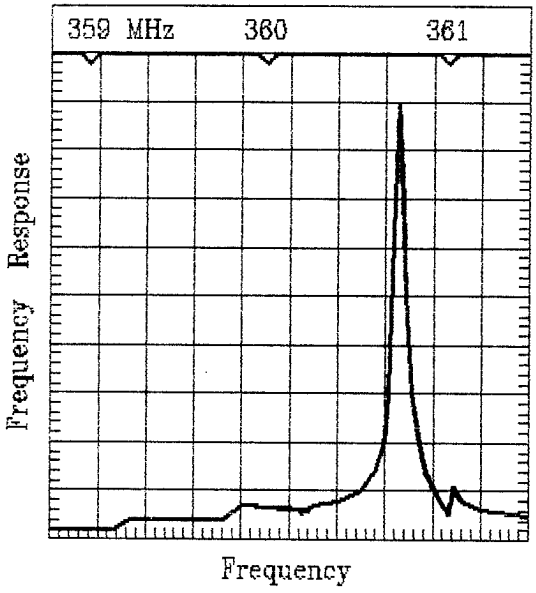


Fig.6.Frequency response of the unetched SAW resonator.

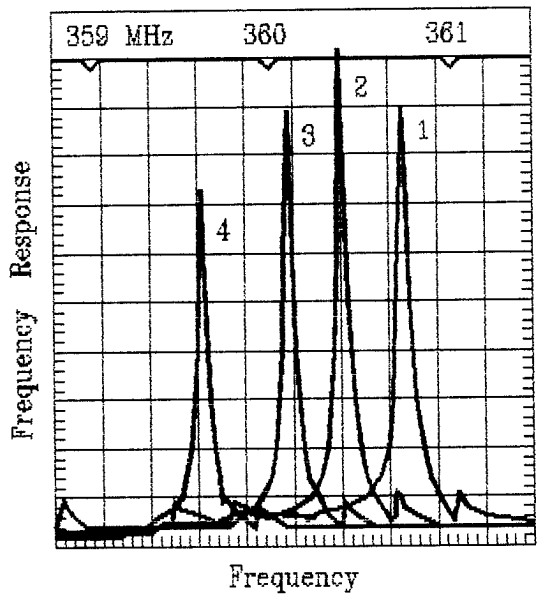


Fig.7.Frequency trimming process for the SAW resonator. Downshift is (1) 0 MHz, (2) 343 kHz, (3) 643 kHz, (4) 1125 kHz. The ion current density is 0.21 ma/sm^2.

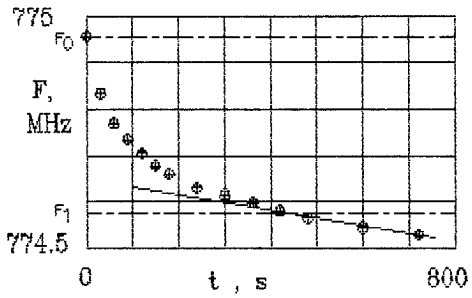


Fig.8.Downshift of central frequency response F vs time. The ion current density is 0.013 ma/sm^2.

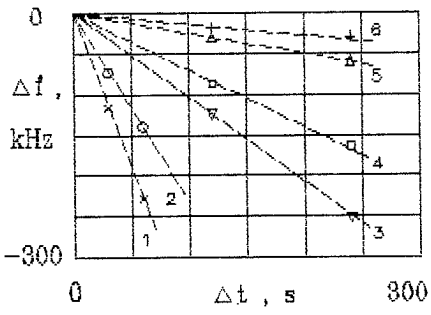


Fig.9.The dependence $\Delta f(\Delta t)=F(t)-F0$ for the variety of the ion current density. The ion density is (1) 0.33, (2) 0.21, (3) 0.106, (4) 0.066, (5) 0.026, (6) 0.013 ma/sm^2.

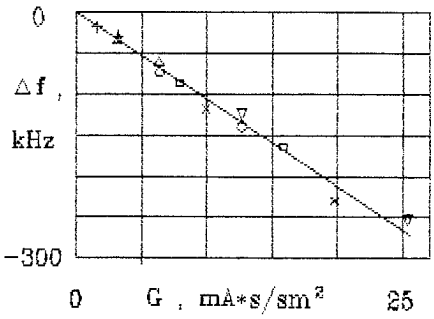


Fig.10.Downshift of frequency for the STW resonator vs exposure G.

The frequency downshift rate with the Xe^+ ion bombardment weakly influenced the O_2 gas partional pressure. The effect of O_2 gas was exhibited in the fact that in the absence or deficient of it the frequency downshift is accompanied by an increase in the insertion loss because of a destruction of Al metallization. A lower limit of the O_2 gas partional pressure, for which a decrease in the insertion loss was not observed, was determined with the ion current density. For the lower density (0.013ma/sm) the value of the lower limit was found to be about $5\cdot10^{-4}$ Pa; for higher density (0.33ma/sm^2) - about $3\ 10^{-3}$ Pa.

The frequency downshift took place with a lighter Ar^+ ion bombardment. However, a higher ion current density was taken in order that the downshift rate achieved a value as large as for the Xe^+ ion bombardment. But, for this Ar^+ ion current density we failed to get the blockage of sputtering of Al metallization by means of O_2 gas supply.

If the surface of the quartz substrate was cleaned from Al metallization of the transducer electrodes by using a liquid etching then a relief was observed. The relief resulted from sputtering of quartz substrate through a mask which was formed from the electrodes on the surface (Fig.11). A measurement, which was made by means an interference microscope, showed that the h depth was linearly proportional to an exposure of the ion bombardment. For the STW resonators and at the (740s*0.21 ma/sm^2) exposure the depth h was found to be 430 Å and the downshift of the frequency - 1.8 MHz.

Discussion

From the fact that after the ion bombardment a hollow appeared on the substrate surface it follows that trimming

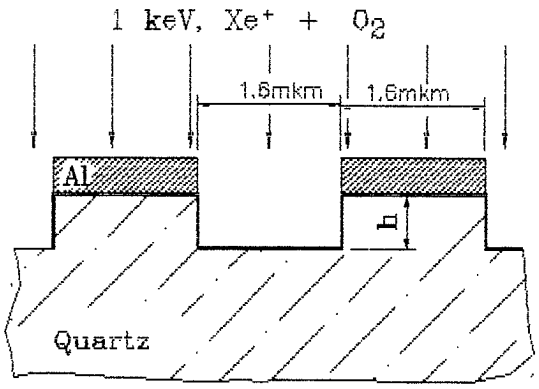


Fig.11.Schematic diagram of the relief formed on the surface.

frequency process is related to sputtering of quartz from interelectrode regions of the transducer. The frequency downshift, the initial region excluded, is linearly proportional of the depth h, which is determined by the total exposure of ion irradiation. A removal of quartz from the interelectrode fields of the transducer displaces the spreading region of the acoustic waves from the surface deep into the substrate. On the one hand,the acoustic waves, driven back from the surface, are dissipated less on the surface decreasing the insertion loss and increasing the unloaded Q-factor. On the other hand, the displacement of the acoustic wave from the surface into substrate decreases an efficiency of the transducer. Therefore, the parameters of the resonators deteriorate for the large downshift of the frequency (>4MHz for STW and >0.6MHz for SAW resonators).

Electrodes sputtering in the frequency trimming process is an unfavorable phenomenon. The use of O_2 gas eliminates this problem. The mechanism of the blockage of sputtering of the Al electrodes was not investigated in this work. It is suggested that adsorbed atoms of oxygen hide atoms of metal from the ions falling on the surface. As the length of an ion race into a solid averages to 1Å for Xe^+ ions with 1 keV energy, only an upper atoms of the surface are sputtered. These atoms are an oxygen atoms. The adsorbed layer of oxygen destructed by ion bombardment is restored by oxygen molecules arriving on the surface from a gas phase. There is a lower threshold of the oxygen pressure for the blockage of sputtering of the metal electrodes. The adsorbed layer of oxygen lacks time for restoration at pressures which less than this threshold. The rate of the restoration (or adsorption rate of oxygen on the Al surface) is a product of two parameters: a pressure and a sticking probability of oxygen. The sticking probability of oxygen on the metallic surface is several orders greater than for quartz or other oxygen containing compounds. Therefore, a masking oxygen layer on the quartz surface is not formed and there is no blockage of sputtering of quartz.

A possibility of this method of precision frequency trimming for great frequency downshifts (1000ppm and more) follow from a possibility of controlling the frequency downshift rate within abroad limit. Its rate is governed by ion current density. As the ion current density is easily operated over a wide range by means of changing the working conditions of the ion source, the rate of the downshift is controlled easily too. For our ion source the ion current density varied in the range from 0.005 to 0.35ma/sm^2 . For the STW resonators this range of the ion current density provided the change of the downshift of frequency from 0.05 to 4 kHz/s. For example, this made possible to downshift the frequency of the STW resonator by 2.5 MHz with an accuracy of 1ppm during 10 min.

Conclusion

The precision frequency trimming of the resonators using Xe^+ ion bombardment is further a development of the plasma chemical method of frequency trimming. This method allows the frequency trimming to be made with a high accuracy (1ppm), in wide range (5000ppm) and in an in-situ monitoring process.

Reference

[1]R.E.Chapman,"Reactive ion etching for fabrication of SAW devices", Vacuum, vol.34, pp.417-424, 1984.

[2]J.Rot,"Chemical sputtering",in book "Sputtering by particle bombardment II" edited by R.Behrisch, Moscow:1986,p.p.134-204.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

APPLICATION OF QUARTZ MICROMACHINING TO THE REALIZATION OF A PRESSURE SENSOR

J.S. Danel, M. Dufour, F. Michel

LETI (CEA-Technologies Avancées)
D.OPT-CEN/G - 85 X - 38041 Grenoble Cédex France.

Abstract

This paper examines the possibility of using quartz for the realization of vibrating beams used as the basic part of a pressure sensor. A comparison with silicon is made, as, from the point of view of their crystalline structure and their high modulus of elasticity, both materials are good candidates for such use. Quartz was prefered because of its simpler technology. We describe the technology used for the micromachining of quartz and based on chemical etching. This knowledge was applied to the realization of a pressure sensor. The complete device, which consists of three sealed quartz wafers, is described. Original techniques for the realization of conductive tracks and for anodic bonding of quartz wafers are also described.

Introduction

The concept of micromachining has emerged from the need to reduce the cost and dimensions of mechanical devices and was made possible by the use of technologies directly derived from integrated-circuit fabrication [1]. Due to this and because of the great amount of knowledge about silicon accumulated to date, most of the work reported on micromechanical devices has been carried out on silicon [2-4]. Nevertheless, the concepts of micromachining and micromechanics are not restricted to this material and other materials can be interesting for the conception and realization of specific microdevices. Another material which could be used for micromechanical devices is monocrystalline quartz, for reliable technologies and a lot of data are available. We compared quartz and silicon for the realization of microstructures, and more especially for use as vibrating beams. Technological problem connected with the design of a resonator were examined. We shall describe here the technologies used for micromachining of quartz. These technologies were used in our laboratory to realize some devices and especially for a study on a miniaturized pressure sensor. For this particular realization, original techniques based on ion implantation were developed in order to overcome packaging problems.

Physical properties of quartz and silicon

Some properties of these materials and of other materials used for their technologies are listed in Table 1. These values are mostly from Landolt-Bornstein and there can be a discrepancy with other sources (for example [5]).

From the point of view of their monocrystalline structure and their high modulus of elasticity, both materials are good candidates for mechanical devices and more specifically for use as vibrating beams. They display different crystalline symmetries, as quartz belongs to a rhombohedral system and silicon to a cubic system, which means that silicon possesses a much higher degree of symmetry. As a consequence some of their physical properties differ completely. Quartz is an insulator and silicon a semiconductor. Thus quartz can be at an advantage over silicon for devices in which no leakage current must exist between the various needed metallizations. The drawback is that the electronic circuit associated with the device is not realized on the same ship. This one-chip conception is clearly an exclusive property of silicon at the moment, but for many applications, it may not be really necessary and gives rise to very tricky compatibility problems.

At a temperature of 573°C quartz undergoes a modification of structure : α-quartz becomes β-quartz with a hexagonal structure, so heating above this point should be avoided.

Quartz was the first piezoelectric crystal to be used widely. This piezoelectric property is mainly used to control the frequency of oscillators and to produce very selective filters [6], but it can very conveniently be used in microstructures for which the basic component is a resonator. For example external data (humidity, strength, acceleration, etc) can be measured through variations in the eigenfrequency of the system. Clearly, as silicon is not piezoelectric, it requires more complex technologies to elaborate resonant devices.

0-7803-0905-7/93 $3.00 © 1993 IEEE

Table 1 : Properties of commonly used materials in silicon and quartz technologies

Material	Density (kg/mm^3)	Young's modulus, or Cij(Pa)	Thermal expansion (K^{-1})	dE/(EdT) or dCij/(CijdT) (K^{-1})	Permitivity
Silicon	2 330	$C11 = 1.66 \times 10^{11}$	2.3×10^{-6}	-85×10^{-6}	11.9
		$C12 = 0.64 \times 10^{11}$		-94×10^{-6}	
		$C44 = 0.79 \times 10^{11}$		-74×10^{-6}	
Quartz	2 650	$C11 = 0.88 \times 10^{11}$	14.3×10^{-6}	-47×10^{-6}	4.5
		$C12 = 0.61 \times 10^{11}$	if $\parallel$ x or y	-3300×10^{-6}	
		$C13 = 0.13 \times 10^{11}$		-700×10^{-6}	
		$C14 = 0.17 \times 10^{11}$	7.8×10^{-6}	-90×10^{-6}	
		$C33 = 1.07 \times 10^{11}$	if $\parallel$ z	-205×10^{-6}	
		$C44 = 0.58 \times 10^{11}$		-166×10^{-6}	
SiO2	2 220	0.7×10^{11}	0.4×10^{-6}		3.9
Si3N4	3 100	1.4×10^{11}	2.8×10^{-6}		7.5
Cr	7 200	1.8×10^{11}	7×10^{-6}		
Au	19 000	0.8×10^{11}	14×10^{-6}		
Al	2 700	0.7×10^{11}	23×10^{-6}		

Excitation and detection of a vibration

The easiest way to set up and detect the vibration of a quartz beam is, of course, to use its piezolectric properties. For micromechanical devices, due to the etching rules, Z-cut wafers are generallly used. With such wafers, only some of the components of the electric field oriented in the plane of the wafer are efficient. Therefore the electrodes of the piezoelectric capacitor must be as perpendicular as possible to the plane of the wafer. This requirement can be met by micromachining a beam with steep vertical sides, and oriented along the so-called mechanical axis. Electrodes evaporated on its sides allow an electric field to be generated along the electrical axis, creating a stress that sets the beam into vibration (figure 1).

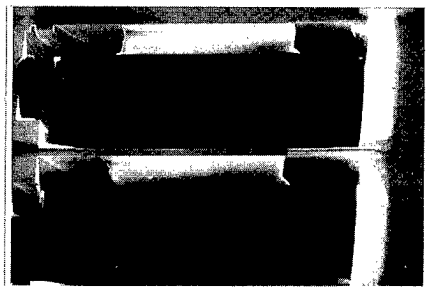


Figure 1 : Excitation of a quartz beam

As silicon is not piezoelectric, a special device is needed to initiate and sustain the vibration of the beam. Several solutions have been tested.

Thermal excitation

Resistors are defined by ion implantation at the end of the beam. At the right frequency, the heating of the resistors due to the Joule effect induces a local stress sufficient to initiate the resonance. Light can also be used to produce a photothermal excitation. Similar results are reported elsewhere [7-10].

Magnetic excitation

A conducting path deposited on the beam is subjected to a permanent magnetic field. If an alternating current flows along this path at the right frequency, the alternating Laplace's force applied to the path, and therefore to the beam, sets up the vibration. This idea can also be used to design vibrating magnetic sensors [11].

Piezoelectric excitation

In this solution, a piezoelectric material ZnO is deposited on adequate points of the beam [12].

Electrostatic excitation

Two configurations were used to perform electrostatic excitation.

In the first one, a metallic electrode is deposited on the beam and the other electrode is in front of it, on another wafer. It is here difficult to reduce and/or control the distance between the electrodes in order to improve the efficiency of the electrostatic force. It is also possible to have both electrodes on the same wafer : one on the beam, the other on the bulk [13].

The second solution involves a capacitor directly realized on the beam. The first electrode is a metallic deposit on the beam, or the bulk itself. The capacitor is achieved by depositing a dielectric capped with a conductive material. The stress produced by the electrostatic force generated between the two electrodes is transmitted to the beam, and is sufficient to initiate the vibration [14].

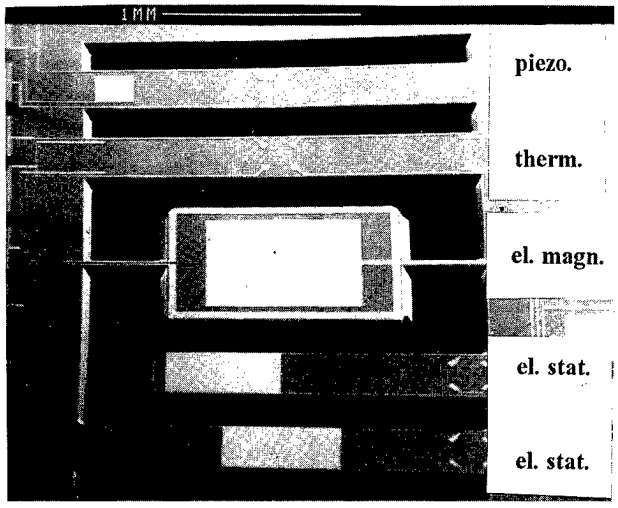


Figure 2 : Excitation of silicon beams

All these solutions have been shown to be efficient, but they are rather difficult to compare. Some conclusions are the following :

- thermal and piezolectric excitations are more efficient than electrostatic and magnetic ones.
- piezoelectric excitation gives rise to very serious technological problems, due to the use of ZnO.
- thermal excitation does not seem very reliable, for there can be uncontrolled frequency shifts due to fluctuations of the power supply.
- when electrostatic excitation is used, the stress in the beam is a quadratic fonction of the applied voltage. So, if the output signal is proportional to this stress (Wheatstone bridge) and is required to be at the same frequency as the excitation, a constant component must be added to the excitation.

Detection can be carried out by the same kind of device as for excitation. For example a piezoelectric configuration can be used. To have a low output impedance, we think that an implanted strain-gauge bridge is interesting for detection : its output is proportional to the voltage applied to the bridge and to the stress initiating the vibration.

Our solution for silicon resonators would be to use an electrostatic excitation and a strain-gauge detection, because of the easiest technological setting, the low-power consumption, the high sensitivity of strain gauges, and the low output impedance. Typical results are given figures 3 and 4.

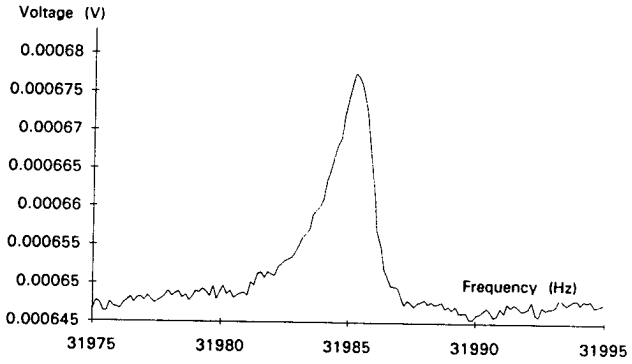


Figure 3 : Typical voltage-frequency curve for an electrostatically excited silicon beam

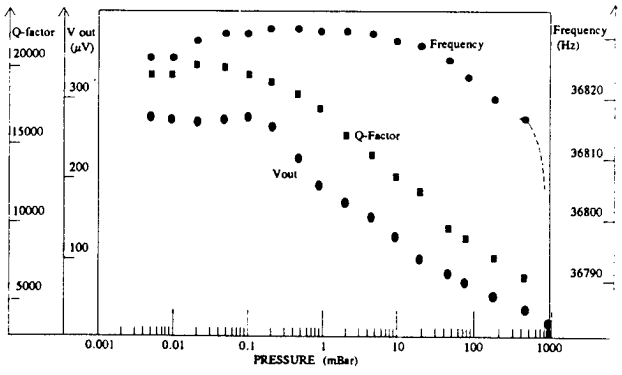


Figure 4 : Influence of pressure on the resonance frequency, Q-factor, and bridge output voltage for a silicon beam.

Q-factors that we measured for quartz and silicon devices were quite similar, around 30×10^3.

For this study, needing the realization of vibrating beams, quartz was finally favoured because of its simplest technology.

Quartz etching technology

Quartz wafers could be machined through dry or chemical etching.

Dry etching (i.e reactive ion etching) is not really satisfactory at the moment to etch deep precise patterns in quartz [15], as the etching depth is still limited to a few tens of microns.

Chemical etching seems actually the best way to machine quartz and this chapter only deals with it.

Procedure

The etchants used are warm mixtures of hydrofluoric acid (HF) and ammonium fluoride (NH_4F). Some additives may be used [16]. The most satisfactory masking material is a chromium-gold thin film obtained by vacuum evaporation or sputtering. Typical thicknesses are : chromium : 100 Å, gold : 1500 Å. The etching patterns are determined in the mask by photolithographic techniques [17].

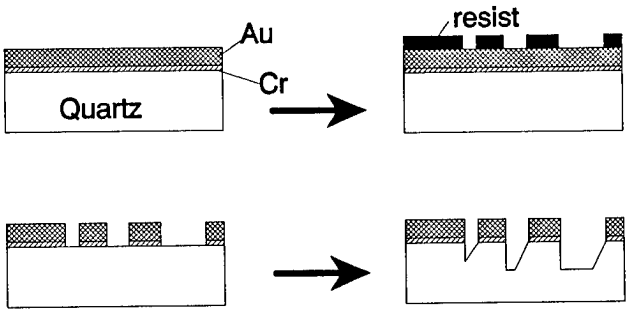


Figure 5 : Quartz technology

As a typical etching temperature is 85°C, protection of the mask must be very good. Lack of protection resulting in defects after etching can be due to several factors :

(i) The quality of the quartz surface itself is not adequate. A mechanically lapped surface can look optically good and can, in fact, be completely amorphous, which will cause the etching to fail. In our experiments, the best result was obtained with double-face polishing (both sides polished at the same time), as we need wafers with both sides polished.

(ii) The process of metal deposition is not adequate. For example, the gold must be applied immediately (< 30 s) after the chromium is deposited , to avoid oxidation of the chromium and lack of adhesion. Best results are obtained when rotating the wafer.

The photoresist used for the photolithography is resistant to the quartz etchant, but it is porous and cannot be used directly to protect quartz or other materials dissolved by this HF/NH_4F etchant. However, several layers of photoresist with different patterns can be deposited and used during the etching process to obtain different etching depths on the same wafer (see principle on figure 6).

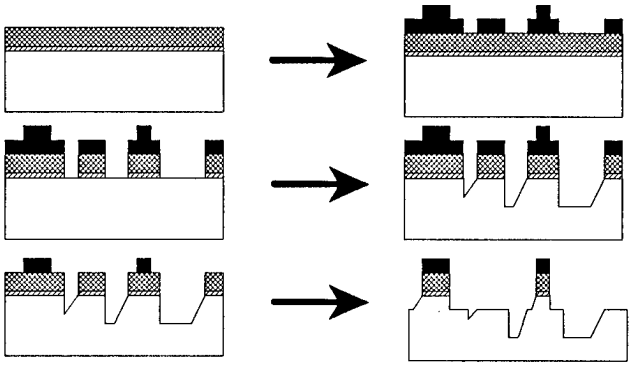


Figure 6 : Principle to obtain several etching depths on a quartz wafer.

Etching rules

The etching speed of a single crystal wafer depends on the crystalline orientation of this wafer [18]. To have a thorough knowledge of the dependance of etching speed on orientation, wafers with different surface orientations must be used. In the case of quartz many planes intervene and it is not at all easy to predict the etching profile obtained with a given mask.

If the etching diagram (i.e the variation of etching rate with direction) is known, a prediction method developed by Jacodine [19] and illustrated in figure 7 can be used.

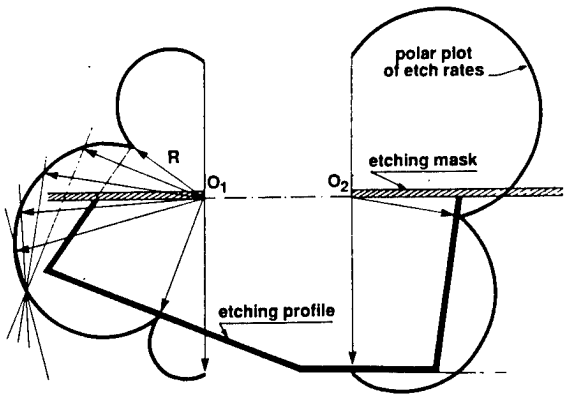


Figure 7 : 2D Wulff-Jacodine prediction method

Points O1 and O2 are the limits of the etching mask. The origin of the etching diagram is placed at O1 and O2, the left side of the diagram being taken for O1 and the right side for O2. The etching profile is the envelope curve of the perpendiculars to the rate vectors R : it is the set of points on these lines that can be reached from O1 and O2 without crossing any other line. Extensions of this method are used for three-dimensional cases, when the etching mask has a complex shape.

An important problem is the obtention of the three-dimensional etching diagram. This can be done from adequate experimental points, by taking into consideration the crystal symmetry [19].

Publications giving complete three-dimensional diagrams are very scarce and, to our knowledge, only Ueda *et al* [20,21] have obtained such diagrams (figure 8).

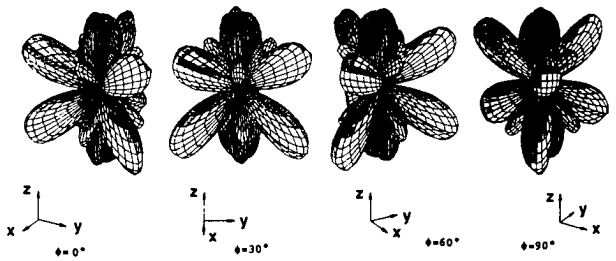


Figure 8 : Example of 3D etching diagrams for quartz.

Researchs are actually being undertaken in several laboratories [22,23,24] to develop simulation methods on computer to allow the prediction of the etching shape resulting from a given mask and even, by inverting the operation, to know the conditions (temperature, composition, cut, mask shape, etc) to obtain a pre-defined structure. Such tools will be very important in the future for the conception of microdevices.

In some cases, a so-called two and a half dimensional method is used. This method deduces empirical etching rules from the observation of profiles on a wafer. We used this method for the design of a quartz accelerometer [15] and the quartz pressure sensor. Experiments were conducted on easily available quartz wafers with Z orientation. The profiles of beams etched by HF/NH4F (proportion 3/2) at a temperature of 82°C were observed. The results shown in figure 9 represent the orientation of the minimum etch-rate facets and their etching rates. Similar results can be found elsewhere [25]. They allow us to predict the etching shape resulting from a mask deposited on a Z-cut wafer. On such a wafer, it is possible to obtain beams with very steep sides oriented along the mechanical or the electrical axes, using double-face etching and appropriate etching times.

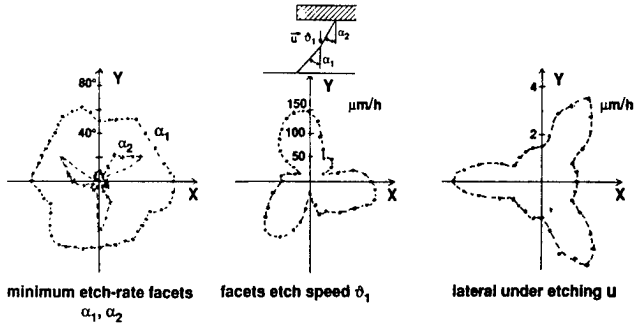


Figure 9 : Results of etching on a Z-cut wafer

A very interesting point about devices obtained by chemical micromachining is that they offer very good surface quality. As a consequence, they display a mechanical strength much higher than that obtained by classical means and almost reaching the theoretical values [15,20,26] ; a microbeam can bear a strength 20 times as large as the one breaking a macroscopic beam (Table 2, figure 10).

Table 2 : Mechanical strength of micromachined beams

Mechanical strength	Tensile	Torsion	Bending
Macroscopic value ($\times 10^7 Nm^{-2}$)	13		9
Micromachined	60	130	170
(Theoretical limit)	(110)	(160)	(230)

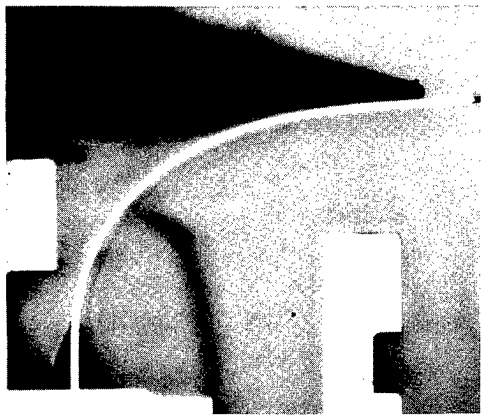


Figure 10 : Bending of a microscopic beam

Quartz metallizations and bonding

Once the quartz wafer etched with the appropriate mask design and the previous-described technology involving several layers of photoresist, adequate conductor tracks and bonding paths must be deposited.

Before doing so, the remains of the masking material are completely stripped off (resist etchant : fuming nitric acid, gold etchant : KI/I_2 or NH_4/I_2, chromium etchant : $Ce(SO_4)_2/((NH_4)_2SO_4)$. The challenge that remains to be solved to achieve a sealed device is to design a package that provides electrically conducting paths from an enclosed structure to contact pads on the outside of the device, across a bond.

Interconnections

The conducting tracks are usually deposited onto the wafer, which causes a problem in that the bonding then has to overcome this step. For devices in which the bonding has not to be hermetic, we realize the conducting paths and metallizations of beams sides by directive evaporation of chromium and gold using a mechanical mask made also in quartz [15].

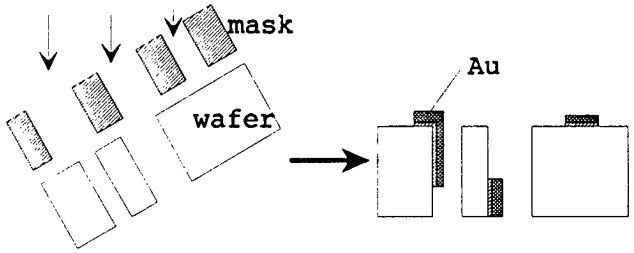


Figure 11 : Remetallization of a quartz wafer.

To solve the problem of hermetic bonding, we decided to bury the conducting tracks in the regions where they must cross the bond. An original technique was developed, based on direct implantation of the conductive layer into the quartz wafer through a mechanical mask. Quartz was implanted with titanium ions with a 2 x $10^{17}cm^{-2}$ dose at 40 keV [27]. The implantation depth is about 800 Å and the implanted layer resistivity is 8 x 10^{-4} Ω.cm, which is sufficient for most applications. The complete remetallization process is shown figure 12.

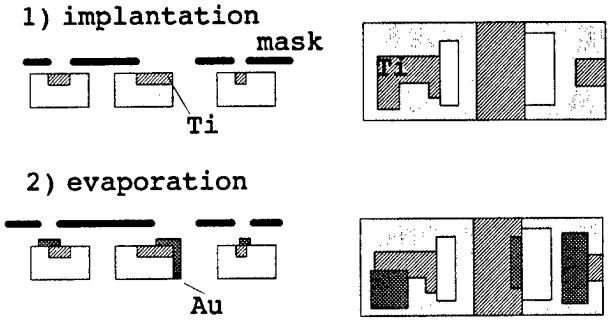


Figure 12 : Remetallization by implantation and evaporation

Resistors are made with this technique without any problem. A very interesting fact is that other metals, such as copper or iron, do not induce enhanced electrical conductivity when implanted [27].

Bonding

Current bonding materials, such as glasses and ceramics, have thermal expansion coefficients that differ from those of quartz. If they are employed to seal two or more quartz wafers, this fact can affect both the sensitivity and the mechanical integrity of the quartz device. Organic adhesives can be used, but long term stability and performance at high temperature are dubious, and their precise deposition onto the wafers is not easy.

Glass fusion bonding was used, but the following points must be taken into account : the glass must have a fusion point below the quartz transition temperature (573°C) and its thermal expansion coefficient should be similar to quartz. With such a technique, we often observed bubble formation in the glass, when sealing under vacuum, and had difficulties in making a correct hermetic sealing, maybe due to a non-adequate apparatus. In this technique, the glass was deposited by serigraphy, which is quite simple. An interesting technique was recently proposed [28] to join quartz to quartz using a modified field - assisted bonding proces. Sufficient reactivity must be present at the quartz/quartz interface well below the transition temperature. To improve the interfacial activity thin films of silicon and Pyrex glass are deposited onto the two adjacent quartz surface (figure 13(a)). Strong, hermetic bonds are obtained for a bonding temperature of 400°C, typical thicknesses of 0.5 µm and an electric field under 10 $keV.cm^{-1}$. Similar configurations can be employed to simultaneously seal three wafers [29]. We developed a variation of this anodic bonding technique

by replacing the silicon layer by a titanium-implanted layer (figure 13(b)).

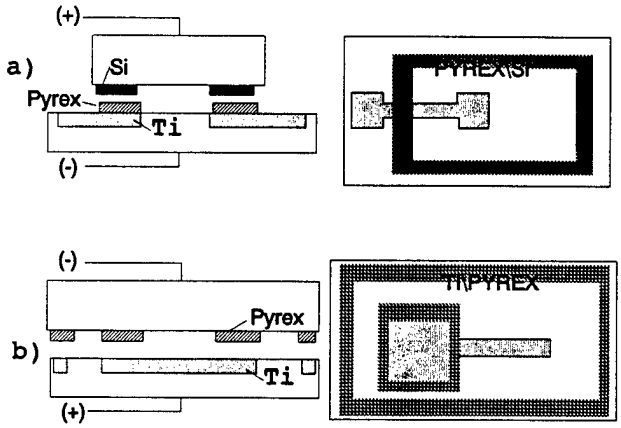


Figure 13 : Anodic bonding configuration for quartz
(a) with Si and glass layers
(b) with Ti implantation and glass layer.

With such a technique, the sealing is hermetic and strong. However some problems still remain. After anodic etching we observe a decrease of conductivity in the implanted layer, but this should not necessarily alter the functionality of the device. More ennoying is the fact that when sealing three etched wafers together the applied voltage had to be quite high (> 50 kV.cm^{-1}) and sometimes produced cracking in the quartz. These problems will have to be solved, but this techique appears quite attractive.

Realization of a pressure sensor

The basic idea is to convert a stress perpendicular to the wafer plane into a parallel stress applied to a resonator lying in this plane. In order to get a high frequency shift, a uniform strain of the resonator beam has to be achieved, as a localized stress would result in reaching the yield strength for a lower frequency shift.

The solution proposed has been studied by finite element simulations and is described below.

Between a pivot and a bending structure where one end of the resonator is clamped, a bending beam is used to pull the bending structure and then the resonator itself. The shape and the size of the whole device is calculated to prevent any of the working mechanical pieces breaking before the stressed resonator has reached the yield strength.

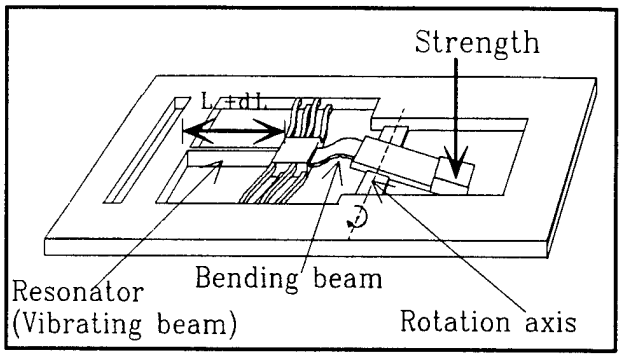


Figure 14 : Principle for stretching a resonant beam

Figure 15 shows such a prototype micromachined by wet chemical etching in a z-cut quartz wafer, using several layers of photoresist to obtain different etching depths on different parts of the device.

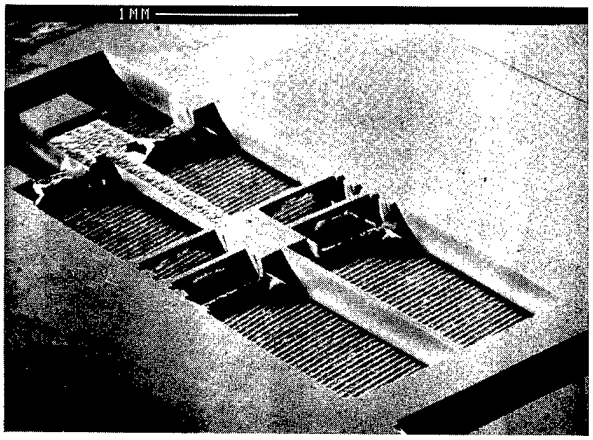


Figure 15 : Prototype device realized on quartz

The metallizations are then realized by implantation of titanium and evaporation of chromium-gold, using adequate mechanical quartz masks.Figure 16 shows the transfer function of such a device.

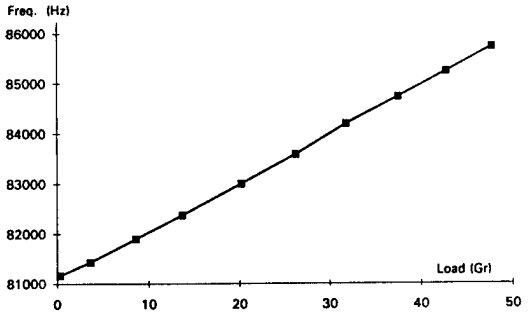


Figure 16 : Transfer function of a quartz resonating sensor.

The frequency span is about 6% of the unloaded resonance frequency, due to breaking of the rotation axis. This miscalculation results from the fact that the etching shapes are more complicated than what was foreseen. Another design would have to be made to avoid this drawback.

The complete device consists of three sealed wafers.

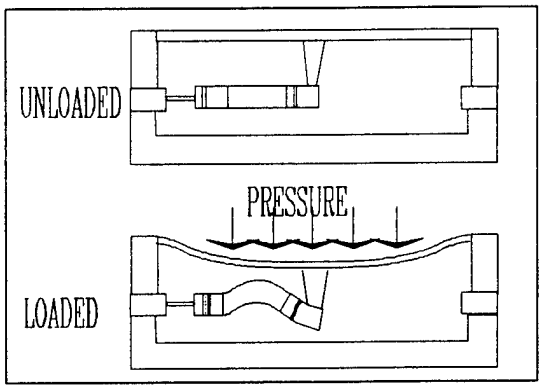


Figure 17 : Diagram of the complete sensor

The middle wafer includes the resonator and the transmission system. The upper wafer is mainly a membrane with a boss that pushes the extremity of the lever of the middle wafer. This membrane is designed to have a lower flexion rigidity than the resonator and the bending system. We therefore avoid the need of precise control of the thickness of the membrane, which can be done only by precise timing. The lower wafer is used for hermetic sealing of the device.

The bonding has been done by two techniques described earlier : either glass-fusion bonding or anodic bonding with Ti-implanted layers. Several problems are still to be solved to obtain a sufficient production yield.

Conclusion and future prospects

This study was undertaken to demonstrate the feasibility of an accurate pressure sensor (accuracy :10^{-3} - 10^{-4}), by using micromachining technology. Quartz was favoured, when compared to silicon, mainly because of a simpler technology. Original etching and packaging technologies were developed for the realization of the device. The use of titanium-implanted layers to solve packaging problems is particularly interesting. Some problems nevertheless remain to be solve with these techniques.

One can find in the litterature many publications on silicon resonators. The use of surface micromachining seems very promising [30,31]. We think, however, that these studies have still a long way to go, and that quartz will remain, for the near future, a very interesting material for the realization of resonant sensors.

References

[1] G. Delapierre, "Micromachining : a survey of the most commonly used processes", Sensors and Actuators, vol.17, n°1, pp.123-138, May 1989.

[2] J.B. Angell, S.C. Terry, and P.W. Barth, "Silicon micromechanical devices", Sci.Am., vol.248, pp. 36-47, April 1983.

[3] K.E. Petersen, "Dynamic micro mechanics on silicon : techniques and devices", IEEE Trans. Electron Devices, vol. ED-25, n°10, pp. 1241-1250, October 1978.

[4] K.E. Petersen, "Silicon as mechanical material", in Proceedings of the IEEE, vol.70, n°5, pp. 420-457, May 1982.

[5] H. Mezger and F.R. Kessler, "Der Deby-Sears-Effekts zur Bestimmung der Elastischen Konstanten von Silicium", Z. Naturforsch., vol. 25A, n°6, pp. 904-908, June 1970.

[6] V. Bottom, in Introduction to Quartz Crystal Unit Design, New York : Van Nostrand Reinhold, 1982, pp. 50-62.

[7] L.M. Zhang, D. Uttamchandani, and B. Culshaw, "Optically powered silicon microresonator pressure sensor", in Proceedings of the 6th Int. Conf. OFS'89, Berlin : Springer, 1989, pp. 470-477.

[8] M.B. Othman and A. Brunnschweiler, "Electrothermally excited silicon beam mechanical resonators", Electron. Lett., vol. 23, n°14, pp. 728-730, July 1987.

[9] W. Riethmuller and W. Benecke, "Thermally excited silicon microactuators", IEEE Trans. Electron Devices, vol. ED-35, n°6, pp. 758-763, June 1988.

[10] D.W. Satchell and J.C. Greenwood, "A thermally excited silicon accelerometer", Sensors and Actuators, vol. 17, n°2, pp. 241-245, May 1989.

[11] R.E. Hetrick, "A vibrating contilever magnetic-field sensor", Sensors and Actuators, vol. 16, n°3, pp.197-207, March 1989.

[12] J.G. Smits, H.A.C. Tilmans, K. Hoen, H. Mulder, J. van Vuuren, and G. Boom, "Resonant diaphragm pressure measurement system with ZnO on Si excitation", Sensors and Actuators, vol. 4, n°4, pp. 565-571, December 1983.

[13] J.C. Greenwood, "Etched silicon vibrating sensor", J.Phys. E : Sci. Instrum., vol. 17, n°8, pp. 650-652, August 1984.

[14] S. Bouwstra, F.R. Blom, T.S.J. Lammerink, H. Yntema, P. Schrap, J.H.J. Fluitman, and M. Elwenspoek, "Excitation and detection of vibrations of micromechanical structures using a dielectric thin film", Sensors and Actuators, vol. 17, n°2, pp. 219-223, May 1989.

[15] J.S. Danel and G. Delapierre, "Quartz : a material for microdevices", J. Micromech. Microeng., vol. 1, n°4, pp. 187-198, December 1991.

[16] J.K. Vondelig, "Fluoride-based etchants for quartz", J. Mater.Sci., vol.18, n°1, pp. 304-314, January 1983.

[17] T. Sonoda and S. Miyazawa, "Production efficiencies enhance tuning-fork-type low-frequency quartz crystal units", J. Electron. Eng., vol. 26, n°271, pp. 44-48, July 1989.

[18] R.B. Heimann, "Principals of chemical etching, the art and science of etching crystals", in Crystals, vol. 8, Berlin : Springer, 1982, pp.214-224.

[19] R.J. Jacodine, "Use of modified free energy theorems to predict equilibrium growing and etching shapes", J. Appl. Phys., vol.33, n°8, pp.2643-2647, August 1962.

[20] T. Ueda, F. Kuhsaka, and P. Yamazaki, "Quartz crystal micromechanical devices", in Proceeding 3rd Int. Conf. Solid-State Sensors and Actuators (Transducer's 85), New York : IEEE, 1985, technical Digest pp.113-116.

[21] T. Ueda, F. Kohsaka, T. Lino, and D. Yamazaki, "Theory to predict etching shapes in quartz and application to design devices", Trans. Soc. Instrum. Contr. Eng., vol. 23, pp. 1-6, 1987.

[22] C.R. Tellier and J.L. Vaterkowski, "A tensorial representation of the dissolution slowness : Application to etched singly rotated quartz plates", J. Mater. Sci., vol. 24, n°3, pp. 1077-1088, March 1989.

[23] C.H. Sequin, "Computer simulation of anisotropic crystal etching", in Proceedings 6th Int. Conf. Sensors and Actuators (Transducers'91), New York : IEEE, San Francisco, 1991, Technical Digest pp. 801-806.

[24] J.S. Danel and G. Delapierre, "Anisotropic crystal etching : a simulation program", Sensors and Actuators, vol. A-31, pp. 267-274, March 1992.

[25] J. Söderkvist, "Two-dimensional etching diagrams for z-cut quartz", J. Micromech. Microeng., vol.2, n°3, pp. 215-217, September 1992.

[26] J.R. Vig, J.W. Lebus, and R.L. Filler, "Chemically polished quartz", in Proceedings 31st. Ann. Symp. Frequency Control, Washington DC : Electronic Industries Association, Fort Monmouth 1977, pp. 131-143.

[27] P. Martin, M. Dufour, A. Ermolieff, S. Marthon, F. Pierre, and M. Dupuy, "Electrical surface conductivity in quartz induced by ion implantation", J. Appl. Phys., vol.72, n°7, pp. 2907-2911, October 1992.

[28] R.D. Watkins, C.D. Tuthill, R. Curlee, D.R. Koehler, and C.F. Joerg, "Field-assisted bonding of single crystal quartz", in Proceedings 11th Quartz Dev. Conf. Exhib.,

Washington DC : Electronic Industries Association, Kansas City 1989, pp. 6-17.

[29] M. Harz, "Anodic bonding for the third dimension", J. Micromech. Microeng., vol. n°3, pp. 161-163, September 1992.

[30] H. Guckel, C. Rystat, M. Nesnidal, J.D. Zook, D.W. Burns, and D.K. Arch, "Polysilicon resonant microbeam technology for high performance sensor applications", in Proceedings IEEE Solid State Sensors and Actuators Workshop, New York : IEEE, June 1992, pp. 153-156.

[31] B. Diem, M.T. Delaye, F. Michel, S. Renard, and G. Delapierre "SOI (SIMOX) as a Substrate for Surface Micromachining of Single Crystalline Silicon Sensors and Actuators", in Proceedings 7th International Conf. on Solid-State Sensors and Actuators, (Transducers'93), June 1993, to be published.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

ENHANCEMENT OF QUARTZ ELECTRICAL CONDUCTIVITY BY ION IMPLANTATION

J.S. Danel, P. Martin, M. Dufour, A. Ermolieff, S. Marthon, F. Pierre, M. Dupuy

LETI (CEA-Technologies Avancées)
D.OPT CEN/G-85X- F 38041 Grenoble Cédex, France

Abstract

There is today an increasing need for solutions to the problem of packaging of microdevices. We have studied a solution based on metallic ion implantation to obtain a buried conductor in order to avoid failures frequently occuring during bonding when the conducting path is a deposited metal layer. Quartz was implanted with titanium and a highly conductive layer was formed near the surface. The implantation-induced changes in quartz were characterized by various means. As a result of implantation quartz was dissociated. Metallic titanium and new chemical compounds including Si, O and Ti were observed. However no conductivity was observed with metallic ions such as Fe and Cu. This suggests a mechanism related to the chemical reactivity of the implanted species with SiO_2. This technique has been developed for the fabrication of miniaturized sensors.

Introduction

The problem of packaging for microdevices has well-known solutions as far as silicon is concerned, for instance anodic or fusion bonding. But this becomes a more serious problem when hermetic sealing must include electrical feedthroughs, and many studies are stumbling on this point. In our laboratory we met this problem when conceiving a pressure resonating sensor, micromachined in quartz, and involving the bonding of three wafers to delimitate a cavity where a resonator stands, the frequency of which shifts according to the variations of the external pressure [1]. Anodic bonding can be used to bond quartz wafers [2], but electrical paths obtained by evaporated metallic thin films are not always compatible with an hermetic sealing, as failures can frequently happen at the joining between the metal and the bonding glass. To solve this problem, we developed a solution based on the modification of the electrical surface conductivity of quartz by direct ion implantation of metallic species. It was expected that a sufficiently conductive layer would be formed at the surface, and that the implantation defects would not affect the properties of the device. Several physical and chemical analyses such as cross-sectional transmission electron microscopy (XTEM), Rutherford backscattering spectrometry (RBS), and X-ray photoelectron spectroscopy (XPS) were conducted in order to get a better understanding of the physical processes involved in the induced electrical conductivity of quartz by ion implantation.

Experiment

The quartz wafers used in these experiments were synthetic Z-cut slabs (i.e. with their surface parallel to the (0001) plane), 300µm thick and with a 1,5 x 1,5 inches square surface. Implantations were performed with relatively high doses of 0.1 - 3.3 x 10^{17} ions.cm^{-2}, at energies of 40 and 150 keV. The metallic implanted ions were Ti^+. The sample surface was tilted 7° to the direction of the beam. The beam current density was approximately 2µA.cm^{-2}. Using an Eaton MC 200 ion implanter, the substrate temperature was kept below 200°C. The samples were not annealed after implantation. Implantations for electrical and physical characterizations were done mainly on full sheet samples.

Some samples however were coated with a 3µm thick resist baked at 200°C for several hours, and shaped with small patterns in order to conduct current-voltage measurements. No heating damage of the resist was observed after ion implantation, and the geometrical aspects of the patterns were well preserved. The resist was then stripped off with fuming nitric acid and gold wires were bonded to the patterned resistors through Cr-Au pads deposited over their ends. Some loss of conductivity was observed, compared with full sheet implantation, perhaps due to a chemical damaging of the near surface during the removal of the resist.

0-7803-0905-7/93 $3.00 © 1993 IEEE

After implantation of 2 x 10^{17} Ti.cm^{-2} at 40 keV, the quartz surface has a metallic luster and has undergone a change in reflectivity. Optical experiment in the visible range showed a variation of reflectivity from 8% for virgin quartz up to 92% for the implanted samples.

Sheet resistance measurements were carried out at room temperature by a four-point probe method. They revealed that the implanted samples were electrically conductive. The ion energy seemed a critical parameter as, for the same implanted dose (2 x 10^{17} Ti cm^{-2}), no sheet conductivity was measurable after implantation at 150 keV in the range of the used apparatus (4,5 x 10^3 Ω/□). At such a high energy, the sample has undergone only a change in color (from colorless to brown).

Results

Figure 1 displays the sheet resistance of quartz implanted with titanium at 40 keV as a function of the dose. The resistivity is measurable only for doses over 1 x 10^{17} cm^{-2} approximately. As the dose increases, the sheet resistance rapidly decreases and then tends towards saturation. The lowest obtained resistance is 70 Ω/□. This is a rather low value, when compared with the one obtained on a 600 Å thick evapored layer of metallic titanium (28 Ω/□).

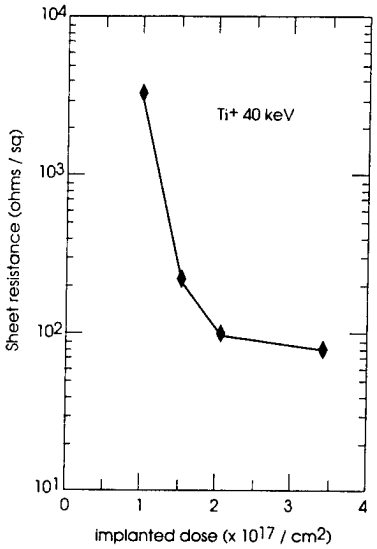


Figure 1: Sheet resistance of quartz implanted with Ti at 40 keV

Furthermore, the electrical conductivity is stable and no significant evolution was observed after storage for some months in air. Current voltage measurements on implanted resistors defined by photolithography follow the Ohm's low, even at high electric fields (up to 16 V, 200 μA).

Grazing-angle (0.1° - 2°) X-ray diffraction (XRD) experiments were carried out, showing that no new crystalline phases due to the reaction between metal and silicon or oxygen (such as silicides or titanium oxides) resulted from implantation. The ion implantation effects were investigated with cross sectional transmission electron microscopy. A micrograph of a sample implanted with a dose 2 x 10^{17} Ti.cm^{-2} at 40 keV is shown in figure 2.

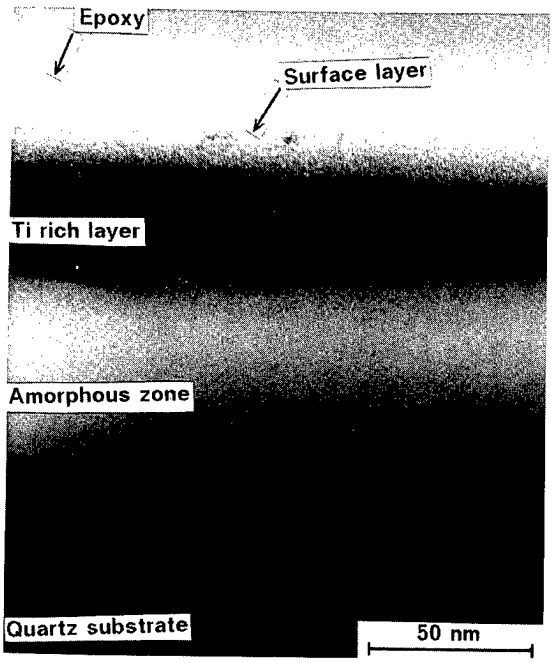


Figure 2 : XTEM micrograph of a sample implanted with 2 x 10^{17} Ti.cm^{-2} at 40 keV

As a consequence of ion implantation, the top surface has been fully amorphized. This result is normal after implantation of such high doses [3]. The measured thickness of this amorphized layer is 1300 Å. A dark region, extending from the surface and 460 Å thick, contains titanium. This region is continuous and no precipitates are observed. The depth at which the concentration of titanium is maximum is estimated at 350 Å.

Simulations using the Monte Carlo TRIM computer code [4] were made, although this code simulates only ballistic mechanisms and does not take

into account high dose implantations effects such as, for example, surface sputtering and materials modifications. The calculated mean projected range of 40 keV T^+ ions is thus 350 Å in quartz and 410 Å in silica. The predicted Ti profiles are in good agreement with the XTEM results. Assuming a gaussian distribution profile, it is expected that the atomic Ti concentration would be as high as a few tenths percent. Adopting the thickness measured by XTEM, the electrical resistivity of the conductive layer is about 5 x 10^{-5} Ω.cm.

RBS experiments were conducted to estimate the implanted structure. A 1 MeV alpha particles beam and a glancing-angle geometry were used to expand the titanium signal but without overlapping the Si signal. Typical spectra are shown in figure 3.

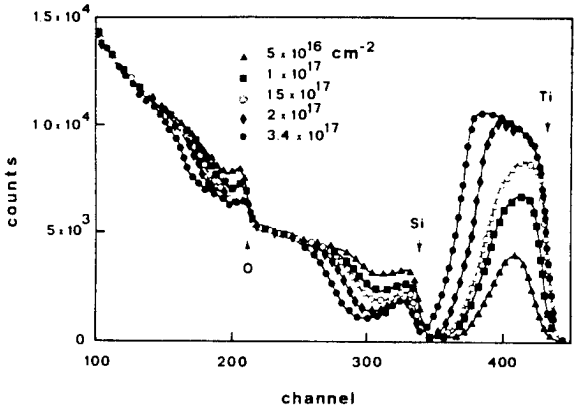


Figure 3 : RBS spectra on samples implanted with 40 keV Ti ions (incident angle : 60°, backscattering angle 165°)

For doses over 1.5 x 10^{17} Ti.cm^{-2}, the Ti-implanted peak is not buried and titanium is detected at the surface, which is a typical feature of surface-sputterring limited implantation profiles. RBS indicates that, in this case, silicon is no longer present at the outset of the surface. Although the elements Si, Ti and O are not uniformly distributed in depth, the resulting structure of the samples implanted at the highest doses was roughly interpreted as being composed of three layers : (1) the top surface is silicon depleted and is composed mainly of titanium oxyde, (2) an intermediate region where the Ti peak has not reached its maximum, (3) the Ti-richest layer which is responsible for the dip in the Si signal. Figure 4 shows the atomic composition of this last layer, deduced from RBS measurements.

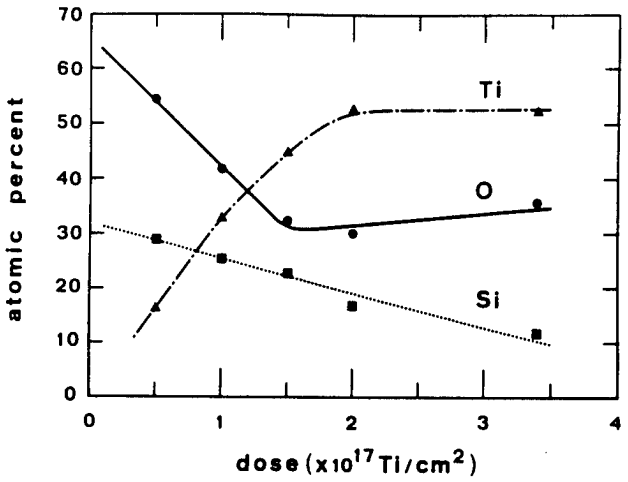


Figure 4 : Plot of the atomic composition in the Ti-richest region vs the implanted dose

The Ti atomic concentration in the compound (TixSiyOz) increases up to 50% with the dose and saturates. This evolution is well correlated with that of the sheet resistance.

Other results were obtained using a surface specific analysis technique such as X-ray photoelectron spectroscopy (XPS). This technique allows an estimation of the bonding states of the elements from the shift of their photoelectron energy peaks. The probing depth estimated from the mean free path of photoelectrons was 20-30 Å. No charge effects were encountered, thanks to the conductive layer. Figure 5(a) shows the in-depth profile of the major elements Ti, Si, O in a sample implanted with 2 x 10^{17} Ti.cm^{-2} at 40 keV.

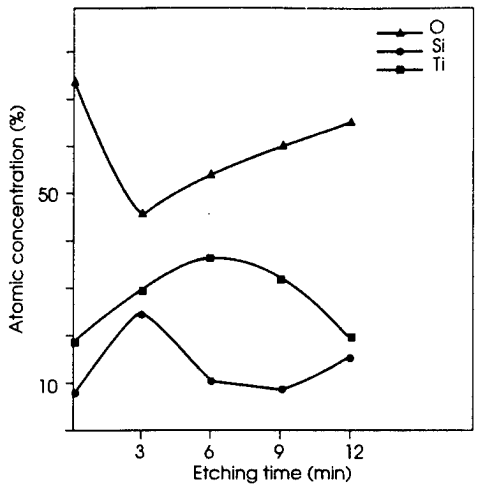


Figure 5(a) : XPS in-depth profile of the elements.

The in-depth profiling of the elements was performed using an argon ion gun for surface sputtering (energy 2 keV, emission current 10 mA, pressure 5 x 10^{-5}Pa, spot size 2 x 5 mm), with an estimated etching rate 90Å/min. A quantification of the atomic composition was made by integration of the photoelectron lines, using the theoretical sensitivity factors.

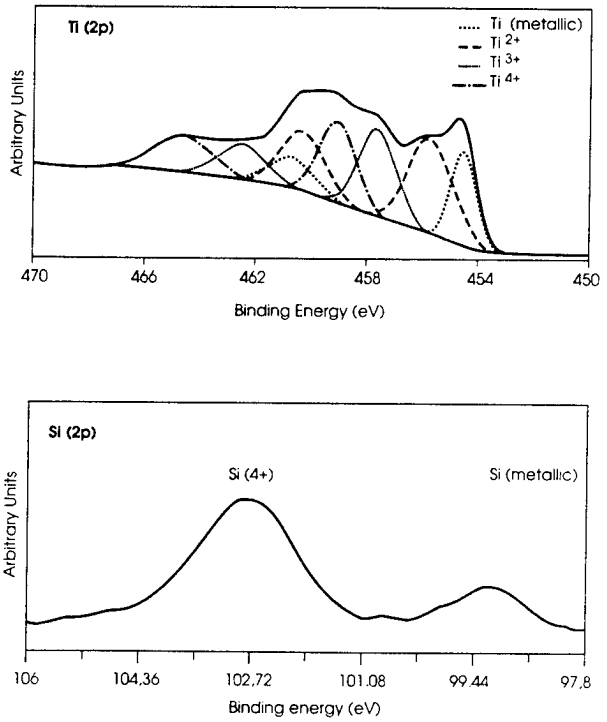


Figure 5(b) : Spectral decomposition of the Ti(2p) and Si(2p) photoelectron lines.

Decomposition of these lines into different components [5] (figure 5(b)) allows us to deduce the percentage of different bonds such as Ti-O(Ti^{2+}), TiO_2(Ti^{4+}), Si-O, measured after different etching times on the same sample (figure 5(c)).

Oxygen presents a peak at the surface and almost all titanium is in the form of TiO_2(Ti^{4+}). The experimental spectrum is quite similar to that of pure TiO_2. On the other hand, silicon is not completely bonded with oxygen as in SiO_2 and 20% of it is now in a metallic form. This region corresponds to the one observed in XTEM and RBS. After 3 min etching, the oxygen concentration decreases and reaches a minimum. TiO_2 is no longer detected. 40% of Ti is under a metallic form with the photoelectrons Ti (2p 3/2) and Ti(2p1/2) at binding energies 455.0 and 460.7 ± 0.2 eV. The residual amount is constituted of understoichiometric titanium oxydes Ti^{3+} and Ti^{2+} (figure 5(c)). After 6 min, the Ti concentration is at a maximum, the Si one at a minimum and the O one is increasing. The Ti/Si ratio is 3.5, rather in good agreement with RBS(3.1). The atomic percentage of oxygen, quite different from RBS, is probably overestimated by XPS. After 9 min, some TiO_2 reappears. The increase of oxygen is due to new Si-O bonds. Finally, after 12 min, near the end of the implanted profile, 90% of Ti is oxidized, 50% of which in the TiO_2 form, Si tends to recover progressively to the SiO_2 state.

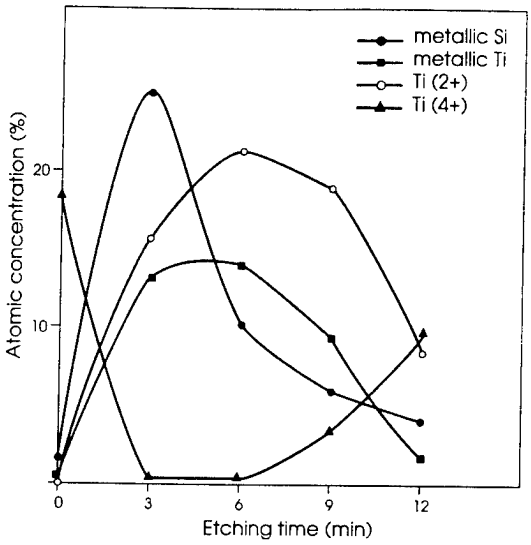


Figure 5(c) : Profile of the chemical bonds : Ti^{4+}, Ti^{2+}, metallic Si and metallic Ti.

A schematic representation of the structure, as revealed by XPS is given in figure 5(d).

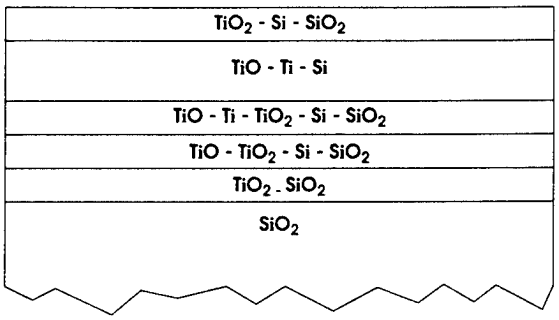


Figure 5(d) : Schematic representation of the structure as deduced from XPS

To summarize, XPS experiments show that quartz is decomposed during ion implantation. They reveal that titanium is in many chemical states, as the metallic forms and various oxides (TiO_2, Ti_2O_3, TiO) were detected along the implanted profile.

Discussion

An induced electrical conductivity is known to appear after high dose (5×10^{16} - 5×10^{17} ions.cm-2) ion implantation in insulators. This has been observed for example in MgO [6], diamond [7], polymers [8], Si_2O_3 [9], etc. In some cases the conductivity was attributed to the presence of metallic precipitates. Another possibility is an interaction between the implanted species and the constituents of the matrix, leading to the synthesis of new compounds. In the case of SiO_2, it is known that it interacts with metals only when chemical reactions leading to formation of silicides and oxides are possible [10] according to :

$$SiO_2 + Mx = MySi + Mx\text{-}yO_2.$$

Titanium is one of the metals that interact readily with SiO_2. Several reactions involving the formation of the silicides Ti_5Si_3, TiSi, $TiSi_2$ and the oxides TiO, Ti_2O_3, Ti_3O_5, TiO_2 exist. Titanium silicides are known for their low thin film resistivities (15 $\mu\Omega$.cm) at room temperature [11].

The induced conductivity of Ti-implanted quartz that we observe experimentally might thus be due not only to the presence of metallic titanium, but also to the existence of compounds having a metallic character, imbricated in an insulating matrix. In order to check if there could be a correlation between the induced conductivity and the chemical reactivity of the implanted metal, implantation with other metallic species, such as ion and copper, were performed. These metallic ions were chosen, as their atomic masses are not too dissimilar from that of titanium, in order to minimize differences in the ballistic processes and in the surface sputtering yield. However, in contrast with Ti, these metals do not react with SiO_2 after thermal treatment [10]. The implantation dose was 2×10^{17} cm^{-2} and the energy 40 keV. No induced conductivity was detected with these ions, the quartz surface showing only a change in colour (brownish with Fe, pink with Cu). These unsucessful results suggest that there is a correlation between the induced high electrical conductivity after ion implantation and the reactivity of the implanted metal, leading to the formation of compounds with a metallic character.

Applications

This technique can be used to join quartz to quartz using a field-assisted bonding process. We have conducted the following experiments : titanium is implanted in a quartz wafer in regions of different shapes defined by photolithography, a thin film of Pyrex glass is deposited on another quartz wafer and the two wafers are bonded at a temperature of 400°C using an electric field under 10 $keV.cm^{-1}$ (figure 6). This technique is similar to the one used by Watkins *et al* [2], but in our case we use implanted titanium instead of deposited silicon.

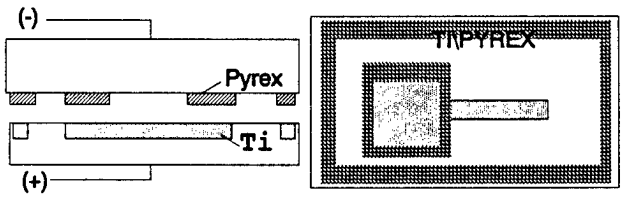


Figure 6 : Anodic bonding of quartz using titanium implantation

Other configurations have been studied to solve packaging problems. For example an adequate Pyrex layer was deposited on conductive tracks made by titanium implantation in a quartz wafer and this wafer was bonded (again by anodic bonding) to another wafer of quartz covered with silicon tracks (figure 7).

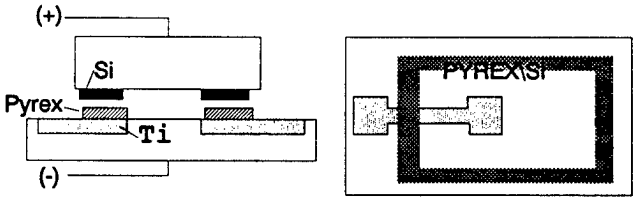


Figure 7 : Other configuration for anodic bonding of quartz

Summary

Quartz was implanted with metallic ions. A conductive layer of relatively low sheet resistance was formed at the surface, provided that the metallic ions, in this case titanium, are implanted at high dose and low energy and react chemically with the substrate. Metallic titanium and various compounds are formed.

They are responsible for this induced conductivity. Ion implantation may be used to form buried ohmic contacts in different devices based on quartz.

Acknowledgments

The authors wish to thank all those who have taken part in these experiments, as well as R. Roux from IBS for suggesting implantation of titanium, R. Eymery and A. Vandrot for performing ion implantation, G. Rolland for XRD analysis and F. Michel for microlithography experiments.

References

[1] M. Dufour, M.T. Delaye, F. Michel, J.S. Danel, B. Diem, and G. Delapierre, "A comparison between micromachined pressure sensors using quartz or silicon vibrating beams", Sensors and Actuators, vol. A-34, n°3, pp. 201-209, September 1992.

[2] R.D. Watkins, C.D. Tuthill, R.M. Curlee, D.R. Koehler, and C.F. Joerg, "Field-Assisted Bonding of Single Crystal Quartz", in Proceedings of the 11th Quartz Dev. Conf. and Exhib, Kansas City, 1989, vol. 1, pp. 6-17.

[3] G. Götz, in Ion Beam Modification of Insulators, edited by P. Mazzoldi and G.W. Arnold, Amsterdam : Elsevier, 1987, p. 412.

[4] J.F. Ziegler, J.P. Biersack, and U. Littmark, in The Stopping and Range of Ions in Solids, New York : Pergamon, 1985.

[5] A. Ermolieff, S. Marthon, P. Martin, F. Pierre, and M. Dufour, "Implantation of quartz with high dose titanium ions as studied by X-ray photoelectron spectroscopy", Solid State Commun., vol. 82, n°7, pp. 517-519, May 1992.

[6] R. Meaudre and A. Perez, "DC current transport in ion-implanted MgO single crystals", Nucl. Instrum. Methods, vol. B32, pp. 75-78, May 1988.

[7] S. Sato, H. Watanabe, K. Takahashi, Y. Abe, and M. Iwaki. "Electrical conductivity and Raman spectra of C^+-ion implanted diamond depending on the target temperature". Nucl. Instrum. Methods, vol. B59/60, pp. 1391-1394, July 1991.

[8] T. Venkatesan, L. Calcagno, B.S. Elman, and G. Eoti, in Ion Beam Modification of Insulators, edited by P. Mazzoldi and G.W. Arnold, Amsterdam : Elservier, 1987, p. 301.

[9] J.G. Perkins, "Conduction properties of microstructure of metals/SiO cermet thin films produced by recoil atom implantation", J. Non-Cryst. Solids, vol. 7, n°4, pp. 349-364, May 1972.

[10] R. Pretorius, J.M. Harvis, and M.A. Nicolet, "Reaction of thin metal films whith SiO_2 substrates", Solid-State Electron., vol.21, n°4, pp. 667-675, April 1978.

[11] R. Kelly, in Ion Bombardment Modification of Surfaces, edited by O. Auriello and R. Kelly, Amsterdam : Elsevier, 1984, p. 79.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MEASURING LIQUID PROPERTIES WITH SMOOTH- AND TEXTURED-SURFACE RESONATORS

S. J. Martin, K. O. Wessendorf, C. T. Gebert, G. C. Frye,
R. W. Cernosek, L. Casaus, and M. A. Mitchell

Sandia National Laboratories, Albuquerque, New Mexico

Abstract

The response of thickness shear mode (TSM) resonators in liquids is examined. Smooth-surface devices, which viscously entrain a layer of contacting liquid, respond to the product of liquid density and viscosity. Textured-surface devices, which also trap liquid in surface features, exhibit an additional response that depends on liquid density alone. Combining smooth and textured resonators in a monolithic sensor allows simultaneous measurement of liquid density and viscosity.

Introduction

The sensitivity of quartz resonators to surface mass accumulation enables their use in a number of sensing applications. The linear change in resonant frequency that occurs with mass accumulation allows the device to function as a general-purpose gravimetric detector or "microbalance." The device is easily instrumented as a sensor by incorporating it as the frequency-control element of an oscillator circuit. Resonators were initially used as thickness monitors in vacuum deposition systems [1]. Later, chemically-sensitive films were added to form gas and vapor detectors [2].

Certain resonator modes permit liquid sensor operation also [3]. Compressional modes couple too strongly to the contacting liquid, generating sound waves that "leak away" acoustic energy and suppress resonance. Shear-modes, however, couple less strongly to the liquid and resonant characteristics are preserved. Thus, TSM resonators can be successfully operated in liquids, provided a specially-designed oscillator circuit is used that overcomes the substantial liquid damping [4]. Device mass sensitivity is preserved in liquids, permitting a number of liquid-phase sensing applications to be addressed [5-7].

TSM resonators can also be used to probe liquid properties. For example, Kanazawa and Gordon have shown that the mechanical interaction between TSM resonators and a contacting liquid results in a change in resonant frequency that depends on liquid density and viscosity [8]. In this paper, we consider how such interactions can be used to probe liquid properties and the role of surface texture in extending this capability.

Smooth-Surface Devices

Fig. 1 shows the cross-sectional displacement profile for the fundamental TSM in an AT-cut quartz resonator. This mode is electrically excited by applying an RF bias to electrodes at the upper and lower crystal surfaces. A thin mass layer that is rigidly bound to the surface moves synchronously with the oscillating surface. When a smooth device is operated in contact with a liquid, the oscillating surface generates plane-parallel laminar flow in the adjacent liquid, as shown in Fig. 1. This "viscously coupled" liquid undergoes a phase lag that increases with distance from the surface.

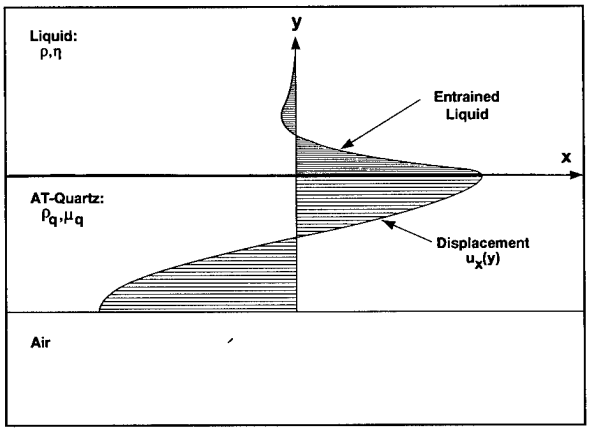


Fig. 1. *Cross-sectional view of a smooth TSM resonator with the upper surface contacted by an ideal mass layer (at the interface) and a Newtonian liquid.*

Fig. 2 shows an equivalent circuit model that describes the near-resonant electrical characteristics of the mass- and/or liquid-loaded TSM resonator [9]. The left branch of the circuit accounts for the "static" capacitance, C_o, that arises between electrodes across the insulating quartz plus parasitic capacitance, C_p, in the test fixture: $C_o^* = C_o + C_p$. The right branch accounts for the "motional impedance" that arises from charges induced on the electrodes by the electrically-excited shear mode resonance. The elements L_1, C_1, and R_1 describe the motional impedance of the unperturbed (without mass or liquid

0-7803-0905-7/93 $3.00 © 1993 IEEE

loading) resonator.

Liquid contact causes an increase in the motional impedance that is accounted for by the motional inductance L_2 and resistance R_2. These can be related to the unperturbed motional impedance L_1 and properties of a contacting Newtonian liquid by [9]:

$$L_2 = \frac{nL_1}{N\pi}\left(\frac{2\omega_s\rho\eta}{\rho_q\mu_q}\right)^{\frac{1}{2}} \quad (1a)$$

$$R_2 = \frac{n\omega_s L_1}{N\pi}\left(\frac{2\omega_s\rho\eta}{\rho_q\mu_q}\right)^{\frac{1}{2}} \quad (1b)$$

where ω_s is the angular series resonant frequency, i.e., where $\omega_s = 1/((L_1+L_2+L_3)C_1)^{1/2}$, n is the number of immersed resonator faces, N is the harmonic number of the TSM, ρ and η are liquid density and viscosity, while ρ_q and μ_q are the quartz density and shear stiffness. (In addition to changing the motional impedance, the liquid dielectric typically affects the parasitic capacitance C_p also.) The motional inductance L_2 represents kinetic energy stored in liquid moving synchronously with the device surface while R_2 represents viscous power dissipation. Energy storage and power dissipation are proportional for Newtonian liquids: $R_2 = \omega_s L_2$.

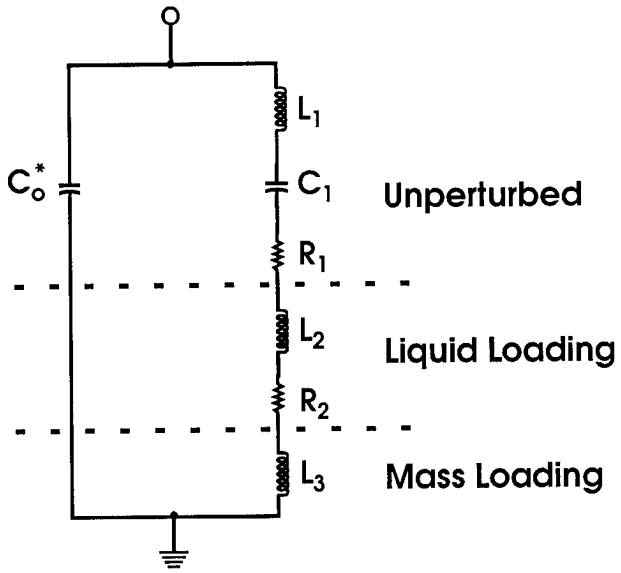


Fig. 2. Equivalent circuit model to describe the electrical characteristics (for f near f_s) of a TSM resonator under mass- and/or liquid-loading conditions.

A surface mass layer that is thin and rigid so that displacement is uniform across the film causes only an increase in motional inductance given by [9]:

$$L_3 = \frac{2n\omega_s L_1 \rho_s}{N\pi\sqrt{\rho_q\mu_q}} \quad (2)$$

where n is the number of coated sides and ρ_s is the areal mass density (mass/area) on each side: $\rho_s = \rho h$, where ρ is the density and h is the thickness of the layer. L_3 represents the kinetic energy of the mass layer moving synchronously with the oscillating crystal surface. Since the film is assumed to be rigid and thus unstrained, no power is dissipated in this process and no motional resistance contribution arises.

The effect of a mass layer or liquid contact on the electrical response of a TSM resonator can be predicted from the equivalent circuit model of Fig. 2. The parameters relating to the unperturbed device (L_1, C_1, R_1, and C_o^*) are determined by fitting the model (with L_2, R_2, and L_3 set to zero) to electrical measurements made on the unperturbed device. Then the response arising from mass- or liquid-loading is determined from the model using element values for L_2, R_2, and L_3 calculated from Eqs. 1 and 2 on the basis of the mass density (ρ_s) or the product of liquid density and viscosity ($\rho\eta$).

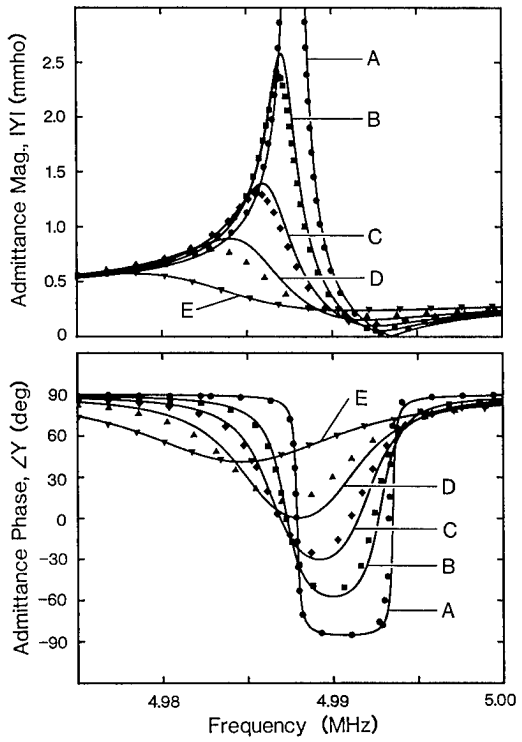


Fig. 3. Resonator electrical admittance measured (points) and calculated (lines) near the fundamental resonance as the density-viscosity product ($\rho\eta$, g^2/cm^4-s) of a contacting fluid increases: (A) air, $\rho\eta = 2 \times 10^{-7}$; (B) water, $\rho\eta = 0.010$; (C) 43% glycerol in H_2O, $\rho\eta = 0.044$; (D) 64% glycerol in H_2O, $\rho\eta = 0.15$; (E) 80% glycerol in H_2O, $\rho\eta = 0.72$.

Fig. 3 shows electrical admittance-*vs.*-frequency measurements (points) made on a liquid-contacted TSM resonator as solution properties were changed. Several glycerol/water mixtures of varying density ρ and viscosity η contacted the device on one side (n=1). As $\rho\eta$ increases, the resonant frequency decreases, while the resonance becomes increasingly damped. The solid lines in Fig. 3 are calculated from the equivalent circuit model with L_2 and R_2 determined from Eqs. 1 using literature values of liquid density and viscosity. The results show that the motional elements described by Eqs. 1 adequately describe the changes in motional impedance caused by liquid contact.

Fig. 4 illustrates the effect of a mass layer on the TSM resonator response. Admittance-vs.-frequency measurements are shown for a resonator before and after addition of a 124 nm gold layer to one side. Responses were measured for the device in air and contacted on the same side by water. The major effect of the gold layer is to *translate* the admittance curves toward a lower resonant frequency without affecting the admittance magnitude or phase angle. The solid lines in Fig. 4 are admittances calculated from the equivalent circuit model using a best-fit inductance L_3=188 μH. From Eq. 2, this L_3 value corresponds to a surface mass density ρ_s=225 μg/cm^2, indicating a gold thickness (ρ=19.3 g/cm^3 for bulk gold) of 117 nm--within 6% of the thickness determined from profilometry measurements (124 nm). Thus, the motional inductance given by Eq. 2 adequately describes the change in motional impedance contributed by a surface mass layer. Moreover, the equivalent circuit model describes the electrical response for *combined* mass and liquid loading (Fig. 4, curve D).

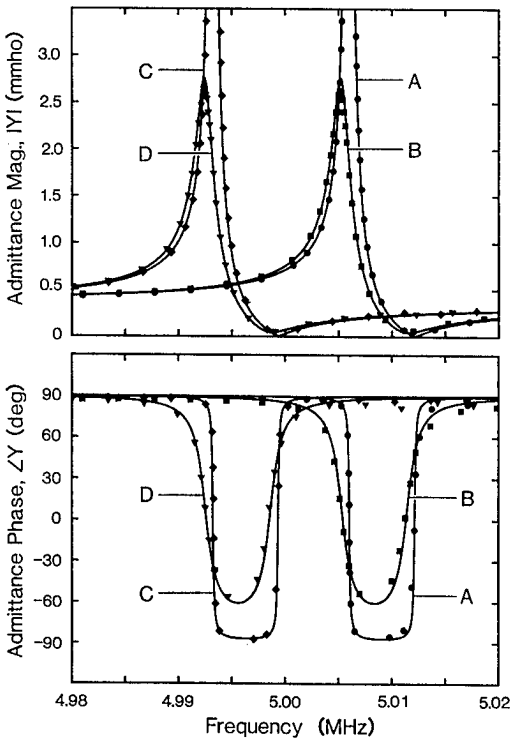


Fig. 4. Resonator electrical admittances measured (points) and calculated (lines) before and after deposition of a 124 nm Au layer. Before Au deposition: (A) in air, (B) in water; after Au deposition: (C) in air, (D) in water.

Changes in the series resonant frequency, Δf_s, and the total motional resistance, ΔR_m, due to mass and/or liquid loading can be determined from the equivalent circuit model of Fig. 2 [9]:

$$\Delta f_s \cong -\frac{(L_2+L_3)f_s}{2L_1} = -\frac{2nf_s^2}{N\sqrt{\rho_q\mu_q}}\left[\rho_s + \left(\frac{\rho\eta}{4\pi f_s}\right)^{\frac{1}{2}}\right] \quad (3a)$$

$$\Delta R_m = R_2 = \frac{n\omega_s L_1}{N\pi}\left(\frac{2\omega_s\rho\eta}{\rho_q\mu_q}\right)^{\frac{1}{2}} . \quad (3b)$$

Eq. 3a indicates that both mass and liquid loading act to decrease the resonant frequency of a TSM resonator. Consequently, measurement of the resonant frequency alone is insufficient to discriminate changes in surface mass from changes in contacting fluid properties. However, if the motional resistance (R_m) is also measured, then $\rho\eta$ is determined from Eq. 3b; this allows the fluid contribution to Δf_s to be determined (second term of Eq. 3a); the remaining frequency shift can be attributed to ρ_s. Thus, mass loading can be quantified simultaneously with fluid properties ($\rho\eta$) if both resonant frequency changes (Δf_s) and crystal damping (R_m) are measured.

With regard to discriminating liquid properties, we note that these enter into the model above only as a product of liquid density and viscosity ($\rho\eta$). This indicates that *a smooth-surfaced TSM resonator is incapable of resolving liquid density and viscosity*. This shortcoming is remedied by including a device with surface texture.

Textured-Surface Devices

Devices with surface texture, either randomly rough or regularly patterned, trap a quantity of liquid in excess of that entrained by a smooth surface [5,10,11]. Vertical features constrain this trapped liquid to move synchronously with the oscillating crystal surface, rather than undergoing a progressive phase lag as occurs with viscously coupled liquid. This trapped liquid thus *behaves as an ideal mass layer* contributing an areal mass density $\rho_s = \rho h$, where ρ is the density and h is now the effective thickness of the trapped liquid layer--dependent upon the vertical relief of the surface texture. If this trapped liquid thickness is small compared to the liquid decay length $\delta=(2\eta/\omega\rho)^{1/2}$ [8], then the surface is considered hydrodynamically smooth and the relative response due to liquid trapping is negligible. If h is comparable or larger than δ, however, then a significant *additional frequency shift arises* from trapping in the textured surface that is dependent only on density and not on viscosity. Thus, a pair of devices, one

smooth and one with a textured surface, allows liquid density and viscosity to be resolved.

From Eq. 3a, the frequency shifts that occur upon immersion of a smooth- (Δf_1) and a textured-surface (Δf_2) device can be written as:

$$\Delta f_1 = -c_1\sqrt{\rho\eta} \tag{4a}$$

$$\Delta f_2 = -c_1'\sqrt{\rho\eta} - c_2 h\rho \tag{4b}$$

where c_1, c_1', and c_2 are constants. If the surface texture is not too large, the contribution to viscous entrainment of liquid is nearly unchanged from the smooth-surface case, i.e., $c_1' \cong c_1$.

From Eqs. 4, the liquid density can be obtained from the difference in responses measured between the smooth and textured devices upon immersion:

$$\rho = \frac{\Delta f_1 - \Delta f_2}{c_2 h} \quad . \tag{5}$$

assuming $c_1'=c_1$.

Having determined liquid density, the response of the smooth device can then be used (Eq. 4a) to determine liquid viscosity:

$$\eta = \frac{(\Delta f_1)^2}{c_1^2\rho} = \frac{c_2 h(\Delta f_1)^2}{c_1^2(\Delta f_1 - \Delta f_2)} \quad . \tag{6}$$

Alternatively, if the motional resistance R_m is measured, liquid viscosity can be determined from this parameter: Eq. 3b indicates that for a smooth device, $\Delta R_m = c_3(\rho\eta)^{1/2}$, from which

$$\eta = \frac{1}{\rho}\left(\frac{\Delta R_m}{c_3}\right)^2 \quad . \tag{7}$$

Dual-Resonator Sensor for Density and Viscosity Measurement

Fig. 5 shows a monolithic quartz sensor that includes smooth and textured TSM resonators to measure liquid density and viscosity. Since these liquid properties (especially viscosity) are temperature dependent, a meander-line resistance temperature device (RTD) is included for measuring liquid temperature. Texture in the form of a surface corrugation is formed on one device by electrodepositing periodic gold ridges on top of the gold electrodes. In order to trap liquid and insure that it moves synchronously with the surface, these ridges are oriented perpendicular to the direction of surface shear displacement--the +X crystalline direction.

In fabricating the dual-resonator device, a Cr/Au (30 nm/200 nm) metallization layer is first deposited on both sides of an optically polished AT-cut quartz wafer. This metallization layer is photolithographically patterned to form the resonator electrodes (both sides) and the meander-line RTD

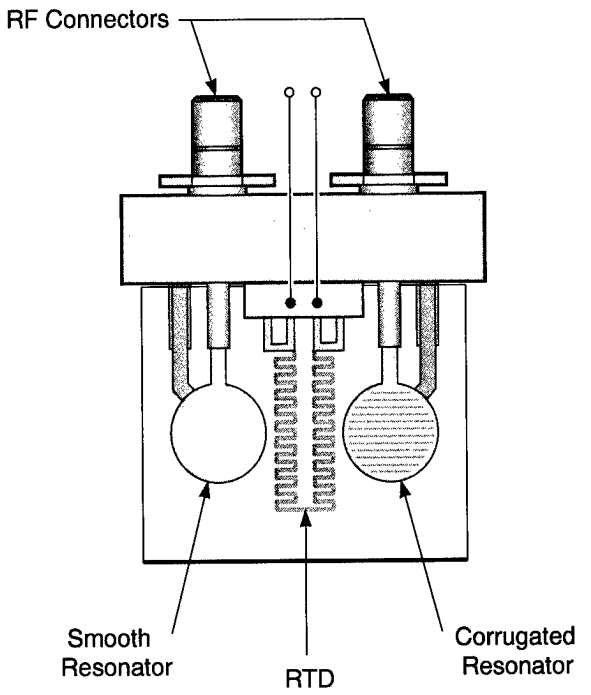


Fig. 5. Monolithic sensor that includes smooth and textured TSM resonators to measure liquid density and viscosity along with an RTD to measure temperature.

(one side). A periodic resist pattern is formed on both electrodes of the resonator whose surface is to be textured. Gold is electrodeposited in the exposed electrode regions to a thickness of 1.5 μm. When the photoresist is removed, a corrugation pattern remains (Fig. 6) with trapezoidal cavities approximately 4.6 μm wide at the base that are well suited for trapping liquid.

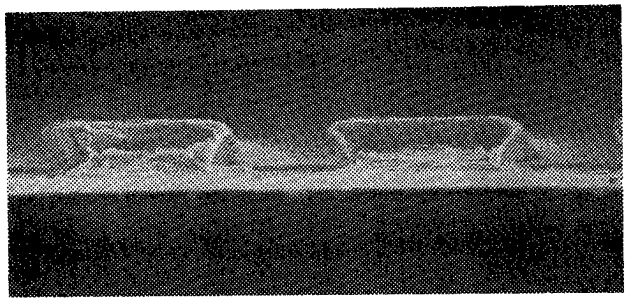


Fig. 6. SEM micrograph of surface corrugation, formed from electrodeposited gold, for trapping liquid at the surface.

The dual-resonator sensor is instrumented for liquid density and viscosity determination as shown in Fig. 7. Each resonator is driven by an independent oscillator circuit that provides two outputs [4]: an RF signal that tracks f_s and a DC voltage proportional to R_m. The RF outputs from the oscillators are read by frequency counters while the DC voltages and RTD resistance are read by multimeters. These signals are

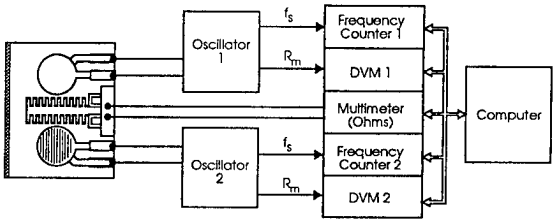


Fig. 7. Schematic of instrumentation for operating dual-resonator sensor shown in Fig. 5 for liquid density and viscosity extraction.

input to a personal computer. The baseline responses are determined by measuring f_s and R_m for each device before immersion. Changes in responses are then measured for the smooth (Δf_1, ΔR_1) and corrugated (Δf_2, ΔR_2) devices after immersion; liquid properties are then determined from Eqs. 5 - 7.

Fig. 8 illustrates the *densitometer function* of the dual-resonator sensor of Fig. 5. The difference in responses (Δf_1 - Δf_2) measured between the smooth and corrugated resonators upon immersion (2 sided liquid contact) is shown vs. liquid density. The response difference is extremely linear with density, following Eq. 5, despite variations in viscosity between the test liquids.

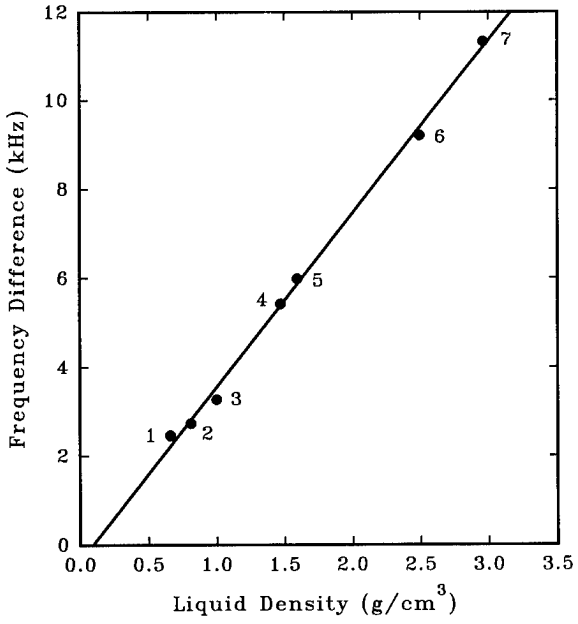


Fig. 8. The difference in responses (Δf_1 - Δf_2) measured between the smooth and corrugated resonators upon immersion vs. liquid density. Liquids tested: (1) hexane, (2) butanol, (3) water, (4) chloroform, (5) carbon tetrachloride, (6) dibromomethane, and (7) tetrabromoethane.

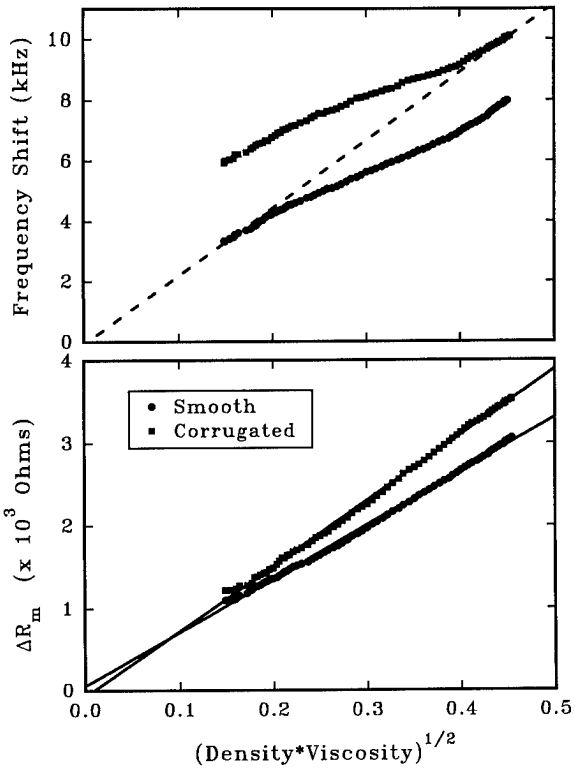


Fig. 9. Changes in oscillation frequencies (upper) and motional resistances (lower) for smooth and corrugated resonators vs. the liquid parameter $(\rho\eta)^{1/2}$. The dashed line indicates network analyzer measurements.

Fig. 9 illustrates the *viscometer function* of the dual-resonator sensor. The variations in oscillation frequency and motional resistance for the smooth and corrugated resonators are shown vs. the liquid parameter $(\rho\eta)^{1/2}$. In the upper plot, the network analyzer data (dashed line) indicate that f_s varies linearly with $(\rho\eta)^{1/2}$, as expected from Eq. 3a. The oscillator follows f_s at low values of damping, but deviates as damping increases, introducing a non-linear response in Δf vs. $(\rho\eta)^{1/2}$. The lower plot indicates that the oscillators are capable of driving the resonators for R_m values up to approximately 3.5 KΩ. R_m increases more linearly with the damping parameter $(\rho\eta)^{1/2}$ and is thus a good alternative to using Δf_s for extracting liquid viscosity.

Fig. 10 shows a "scatter diagram" that compares liquid densities and viscosities extracted from dual-resonator measurements (circles) with literature values (squares). The liquid corresponding to each set of points is indicated in Table I. With 2-sided liquid contact, the crystal becomes damped too severely to sustain oscillation for $\rho\eta > 8$. Since liquid damping increases (Eq. 1b) as $n(\rho\eta)^{1/2}$, limiting liquid contact to a single side extends the $\rho\eta$ range to 32. This data illustrates that liquid density and viscosity can be extracted from measurements made by a pair of quartz resonators--one smooth and one textured. The average differences between the

measured and literature values were 5.3% for density and 19.5% for viscosity.

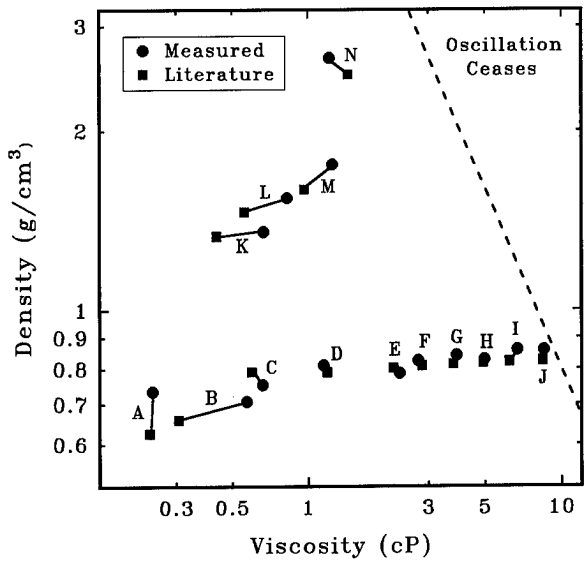


Fig. 10. Scatter diagram comparing the extracted (circle) and literature (square) values of liquid density and viscosity for organic liquids listed in Table I.

Table I. Liquids shown in Fig. 10.

A	n-pentane	H	n-hexanol
B	n-hexane	I	n-heptanol
C	methanol	J	n-octanol
D	ethanol	K	dichloromethane
E	n-propanol	L	trichloroethylene
F	n-butanol	M	carbon tetrachloride
G	n-pentanol	N	dibromomethane

Conclusion

Quartz resonators with smooth surfaces can be operated in liquids to measure surface mass accumulation and the density-viscosity product of a contacting fluid. By measuring both Δf_s and ΔR_m, surface mass can be differentiated from the fluid parameter $\rho\eta$. Surface texture on the resonator traps fluid, which behaves as a mass layer, contributing an additional response proportional to liquid density. Comparing the responses of a pair of resonators--one smooth and one textured--enables liquid density to be extracted. Once density is known, the response of a smooth device yields liquid viscosity.

Acknowledgements

The authors wish to thank Prof. S. D. Senturia of the Massachusetts Institute of Technology for helpful discussions and K. Rice of Sandia National Laboratories for graphical assistance. This work was performed at Sandia National Laboratories, supported by the U.S. Department of Energy under contract No. DE-AC04-76DP00789.

References

[1] Sauerbrey, G. *Z. Phys.* **1959**, *155*, 206-222.

[2] Alder, J. F.; McCallum, J. J. *The Analyst* **1983**, *108*, 1169-1189.

[3] Numura, T.; Minemura, A. *Nippon Kagaku Kaishi* **1980**, 1621.

[4] Wessendorf, K. O. "The Lever Oscillator for Use in High Resistance Resonator Applications," *Proc. of the 1993 Frequency Control Symp.* (IEEE, New York, 1993).

[5] Schumacher, R. *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 329-343.

[6] Ward, M. D.; Buttry, D. A. *Science* **1990**, *249*, 1000-1007.

[7] Hillman, A. R.; Loveday, D. C.; Swann, M. J. *J. Chem. Soc. Faraday Trans.* **1991**, *87*, 2047-2053.

[8] Kanazawa, K. K.; Gordon II, J. G. *Anal. Chem.* **1985**, *57*, 1770-1771.

[9] Martin, S. J.; Granstaff, V. E.; Frye, G. C. *Anal. Chem.* **1991**, *63*, 2272-2281.

[10] Martin, S. J.; Frye, G. C.; Ricco, A. J.; Senturia, S. D. *Anal. Chem.* **1993** (In Press).

[11] Schumacher, R.; Borges, G.; Kanazawa K. K. *Surf. Sci.* **1985**, *L621*, 163.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

TACTICAL BVA QUARTZ RESONATOR PERFORMANCE

Jerry R. Norton

The Johns Hopkins University Applied Physics Laboratory
Johns Hopkins Road
Laurel, Maryland 20723-6099

Raymond J. Besson

École Nationale Supérieure de Mécanique et des Microtechniques
Route de Gray—La Boulie—25030 BESANÇON—France

Abstract

This paper focuses on the performance of tactical BVA quartz resonators, specifically their measured frequency stability. Data are presented demonstrating the resonators' ability to survive harsh environmental conditions. Test results suggest an oscillator using a tactical BVA resonator could be developed that would have very high frequency stability, be relatively small, and have low mass. The tactical resonators evaluated are 10-MHz, 3rd overtone, SC (stress-compensated) cut devices housed in HC-40 enclosures. The Q of the resonators tested was between 1.2 and 1.3 million.

Introduction

Photographs of a conventionally mounted resonator and a BVA resonator shown in Figure 1 illustrate the very different fabrication and assembly techniques used in the two resonators [1, 2]. There are many differences between the two types of resonators, but the BVA has two primary distinguishing features. First, the frequency-determining area of the resonator and its mounting or support system are machined from a single quartz disk. The critical active vibrating area of the resonator is joined to its surrounding support ring by quartz bridges machined during fabrication [3]. Second, metal electrodes used for electrical connection to the resonator are not deposited on the active, vibrating, frequency-determining area of the BVA resonator as they are in the conventionally mounted resonator.

Test Conditions

Resonators were installed in test oscillators and their performance was measured, except where noted. The frequency stability measurements were conducted in a controlled laboratory environment to minimize environmental effects. Some tests were conducted in a vacuum to minimize the effects of barometric pressure and relative humidity changes. Resonator evaluation is an ongoing process that began four years ago with some very early development units. Only eleven resonators were evaluated, but the evaluation period for some was over a year.

Frequency Stability and Phase Noise Results

Allan variance data for the tactical resonator are shown in Figure 2. Four of the eleven resonators evaluated had an Allan variance of $<2 \times 10^{-13}$. The Allan variance data presented here do not assume equal noise sources.

Phase noise ($\pounds_f$) performance for a 10-MHz BVA resonator is shown in Figure 3. Again, equal noise sources were not assumed in the data presented. Phase noise could not be measured on all resonators because many of the developmental resonators were off-frequency.

Twenty-four-hour aging rates for the resonators varied between 1.8×10^{-11} and 3×10^{-10} ; six resonators had an aging rate of $< 1 \times 10^{-10}$. Frequency jumps have not been observed in the 10-MHz tactical BVA resonators tested. Some isolated frequency wander has been observed, but not the very distinct frequency jumps seen in the 5-MHz resonators [4]. The magnitude of the frequency wander of the 10-MHz BVA is much smaller ($<1 \times 10^{-11}$) compared with parts in 10^{-10} for the 5-MHz BVA resonators.

Frequency retrace for resonator SN2303 is shown in Figure 4. Power was removed from the oscillator for a 24-h period, then power was reapplied. Before turning the oscillator power off, a daily aging rate of 5×10^{-11} was

0-7803-0905-7/93 $3.00 © 1993 IEEE

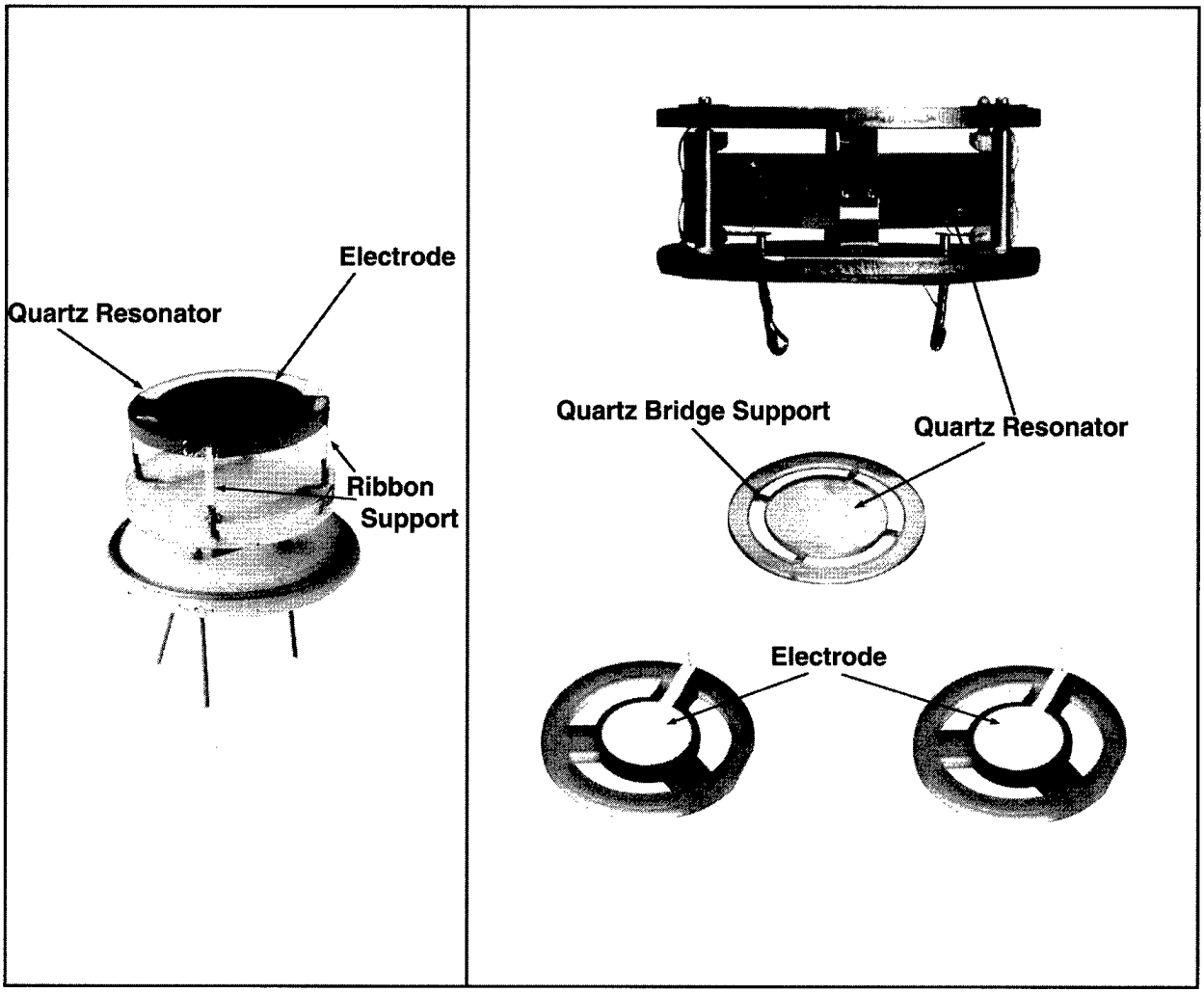


Figure 1. Conventionally mounted quartz resonator (left) and BVA quartz resonator (right).

established. Two days after the oscillator was turned on, there was a residual frequency change or retrace of 4×10^{-10}. A new daily aging rate of 3×10^{-11} was also established two to three days after the oscillator was restarted. These tests were conducted with the resonator and its test oscillator in a vacuum to minimize any effects of pressure and humidity change.

Environmental Effects on Frequency Stability

Vibration

The tactical BVA resonator has many desirable characteristics for spacecraft oscillator applications. One attractive feature is its demonstrated ability to withstand high levels of both random and sinusoidal vibration. Extensive tests have been conducted at 20, 30, 50, and 60 *g* at vibration frequencies up to 2000 Hz. A frequency change of $< \pm 1 \times 10^{-7}$ and a resistance change of $< 2\ \Omega$ resulted from these tests. More recently a resonator was tested at 70 *g* random and sinusoidal without damage. These measurements were conducted on resonators, not in an oscillator. The mounting structure of a conventional 5- or 10-MHz resonator has a mechanical resonance in the frequency range of 300 to 500 Hz. A conventional resonator stimulated at its mechanical resonance, even at low levels, will be severely damaged or destroyed. Therefore, an elaborate and bulky vibration isolation system is required to protect the resonator during spacecraft launch [5]. Vibration test data for the tactical BVA suggest that an oscillator could be designed without using a vibration isolation system, which would greatly reduce the size and mass of the complete oscillator.

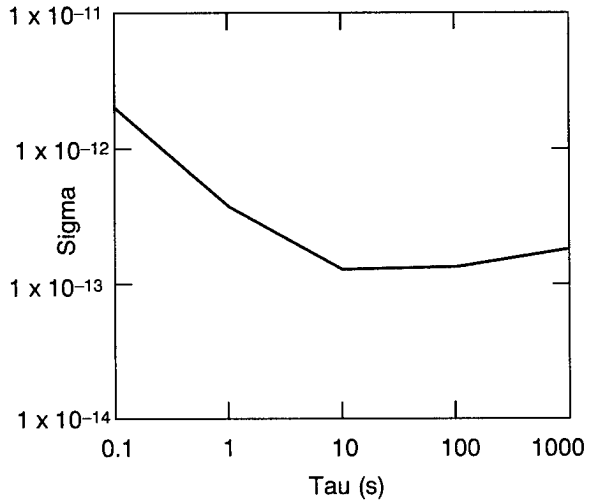


Figure 2. Allan variance of a 10-MHz BVA resonator. Measured data do not assume equal noise sources. Measurement bandwidth = 1000 Hz.

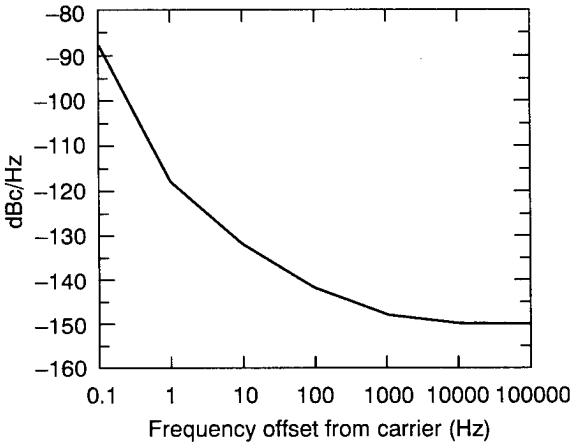


Figure 3. Phase noise of a 10-MHz BVA resonator. Measured data do not assume equal noise sources.

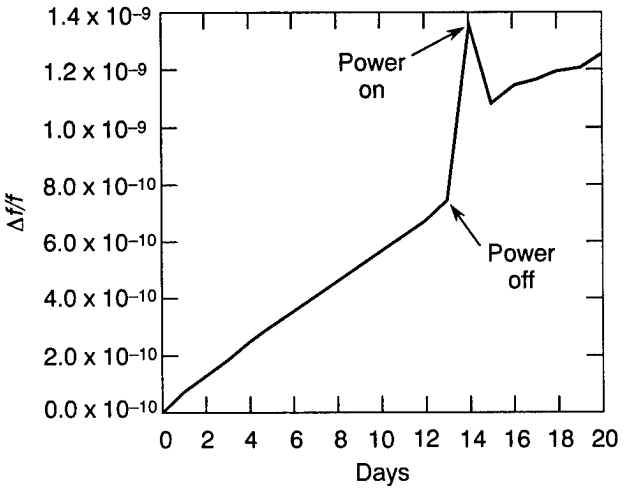


Figure 4. Frequency retrace of tactical BVA resonator SN2303.

Radiation

The operating frequency of a quartz resonator changes when exposed to ionizing radiation. The magnitude and character of the frequency change depend on the radiation dose and dose rate [6]. The radiation-induced frequency change for BVA-type resonators is generally less than conventional electroded resonators [7]. Figures 5 and 6 present the output frequency change caused by radiation for two tactical resonators. Figure 5 presents the radiation-induced frequency change for resonator SN2112 from three 0.6-rad(Si) exposures. Variation in the frequency change was between 2.4×10^{-10}/rad(Si) and 1.3×10^{-10}/rad(Si). Resonator SN2112 was not exposed to a 20-krad(Si) preconditioned radiation dose from a cobalt 60 radiation source. Radiation test data for resonator SN1412 shown in Figure 6 had a frequency change varying between 4.2×10^{-11}/rad(Si) and 1.2×10^{-11}/rad(Si), much smaller than resonator SN2112. Resonator SN1412 was preconditioned with a 20-krad(Si) exposure. A single test of each resonator versus a before-and-after preconditioning test did not give conclusive proof that preconditioning reduced the frequency change for resonator SN1412, but the data are consistent with previous radiation testing of resonators [7].

Radiation shields between the quartz resonator and the radiation source can reduce frequency changes caused by radiation. The tactical resonator provides some shielding; an internal steel cylindrical structure 1 mm thick is used to mount the quartz resonator disk in its enclosure. Figure 7 illustrates the internal mounting structure showing the steel cylinder. From a weight standpoint, the most effective location for a shield is immediately around the quartz resonator, since for a given thickness it adds the least volume of material and, therefore, the least weight.

Acceleration

A quartz resonator's output frequency is also changed by acceleration forces or gravitational forces. As a class, the BVA resonators are less sensitive to acceleration than conventionally mounted resonators, with their long, thin ribbon mounts as shown in Figure 1. The tactical BVA resonators tested had a g sensitivity, using the tip-over test, ranging between 9.2×10^{-11} and 3.4×10^{-10}. This reduced g sensitivity is thought to be partly due to the very rigid symmetrical mounting structure for the frequency-determining quartz resonator.

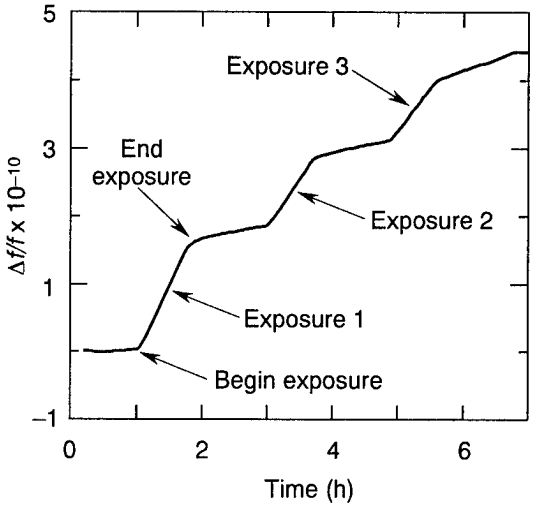


Figure 5. Frequency response of tactical BVA-SC cut quartz resonator SN2112 to radiation. Each exposure was 0.6 rad(Si) for 42 min at a dose rate of 0.014 rad(Si)/min. Resonator not preconditioned.

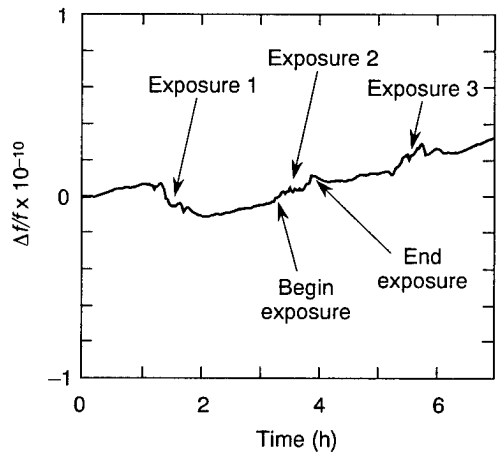


Figure 6. Frequency response of tactical BVA-SC cut quartz resonator SN1412 to radiation. Each exposure was 0.6 rad(Si) for 42 min at a dose rate of 0.014 rad(Si)/min. Resonator preconditioned with a 20-krad(Si) exposure.

Magnetic Sensitivity

The frequency change of resonator SN2807 in a varying magnetic field applied axially to the resonator is shown in Figure 8. The peak frequency change is $\approx 2.7 \times 10^{-12}$/gauss. The radial magnetic sensitivity is $< 9 \times 10^{-13}$/gauss [personal communication, R. Brendel, LPMO-CNRS, 1993]. This series of tests was conducted at the resonator level. The effects of magnetic fields on oscillator circuits can be much greater than the effects on the resonator if care is not exercised in the oscillator design.

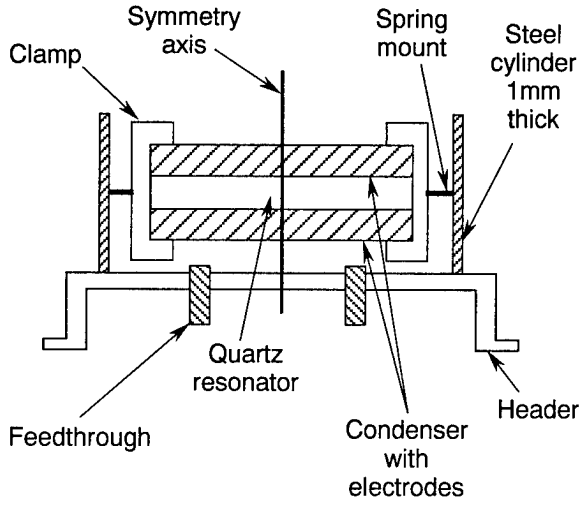


Figure 7. Tactical BVA quartz resonator internal mounting structure.

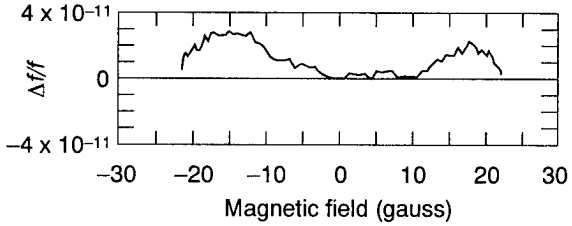


Figure 8. Frequency response of tactical BVA resonator SN2807 to magnetic field. Field applied axially to resonator.

Performance Comparison Between 5- and 10-MHz Resonators

An oscillator designed using the smaller tactical BVA resonator, with its ability to withstand relatively high levels of vibration, is projected to have a size of 353 cm^3 and a mass of 0.32 kg. The estimated input power requirement is 0.8 W.

The data presented in Table 1 clearly show the 5-MHz resonator to have superior performance for Allan variance and phase noise (close to the carrier). Dramatic differences exist, however, in the size and mass of an existing space-qualified 5-MHz oscillator [4] and the projected size and mass of a space-qualified oscillator using the 10-MHz tactical resonator. These differences are due to (1) the difference in physical size of the 5- and 10-MHz resonators as illustrated in Figure 9 and (2) the elimination of the bulky vibration isolation system in an oscillator using a tactical BVA resonator. In applications with critical size, mass, and power constraints, it may be desirable to trade a small degradation in frequency stability for a large reduction in physical parameters, which could be realized by using the 10-MHz BVA resonator in the oscillator design. Also, in applications requiring frequency multiplication to

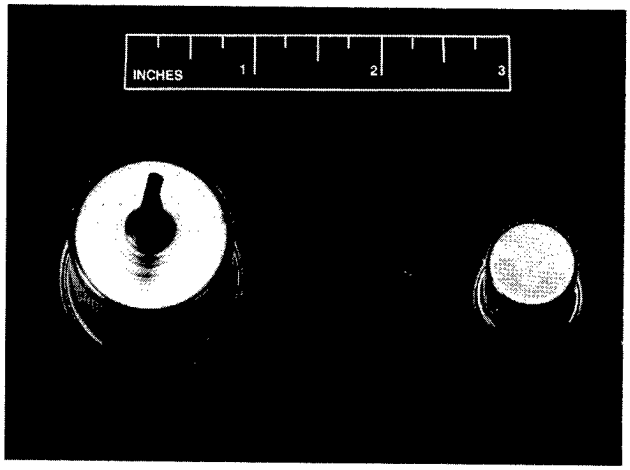


Figure 9. 5-MHz BVA resonator (left) and 10-MHz tactical BVA resonator housed in HC-40 enclosure (right).

microwave frequencies, a 10-MHz oscillator rather than a 5-MHz oscillator may be beneficial.

Conclusion

The tactical 10-MHz BVA resonator is a prime candidate for use in a small, low-mass, high-performance oscillator. A combination of the tactical BVA's demonstrated excellent frequency stability, its superior performance in hostile environmental conditions, and its relatively small size meet the requirements to initiate a new oscillator design.

References

[1] R. J. Besson, "A New Electrodeless Resonator Design," in *Proceedings of the 31st Annual Symposium on Frequency Control*, June 1977, pp. 147-152.

[2] R. J. Besson and M. Mourey, "A BVA Quartz Crystal Oscillator for Severe Environments," in *Proceedings of the 44h Annual Symposium on Frequency Control*, May 1990, pp. 593-596.

[3] R. Delaite, "Special Miniaturised SC Cut Quartz Resonator for Severe Environment Applications," in *Proceedings of the 3rd European Time and Frequency Forum*, March 1989, pp. 296-300.

[4] J. R. Norton, "BVA-Type Quartz Oscillators for Spacecraft," in *Proceedings of the 45th Annual Symposium on Frequency Control*, May 1991, pp. 426-430.

[5] J. R. Norton and J. M. Cloeren, "Precision Quartz Oscillators and Their Use in Small Satellites," in *Proceedings of the 6th Annual AIAA/USU Conference on Small Satellites*, September 1992.

[6] J. R. Norton, J. M. Cloeren, and J. J. Suter, "Results from Gamma Ray and Proton Beam Radiation Testing on Quartz Resonators," in *Proceedings of the 38th Annual Symposium on Frequency Control*, May 1983, pp. 63-72.

[7] J. J. Suter, A. G. Bates, J. M. Cloeren, J. R. Norton, B. Schlueter, and U. Peier, "The Susceptibility of BVA-SC Resonators to Proton Ionization Effects," in *Proceedings of the 3rd European Frequency and Time Forum*, March 1989, pp. 11-21.

Table 1. Performance data for 5- and 10-MHz BVA resonators.

Parameter	Measured 5-MHz BVA	Measured or projected 10-MHz BVA
Aging rate/24 h	2×10^{-12}	1.8×10^{-11}
Allan variance, Tau (s)		
1	9.1×10^{-14}	3.7×10^{-13}
100	6.6×10^{-14}	1.3×10^{-13}
Phase noise		
freq. offset (Hz)		
1	−126 dBc/Hz	−118 dBc/Hz
10	−146 dBc/Hz	−132 dBc/Hz
1000	−155 dBc/Hz	−148 dBc/Hz
Acceleration	$1.5 \times 10^{-9}/g$	$9.2 \times 10^{-11}/g$
Mass	1.76 kg	0.32 kg
Size	1935 cm^3	353 cm^3

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

RESULTS OF THE QUARTZ CRYSTAL MEASUREMENT DATA OBTAINED AT THE LEADLESS RESONATOR MEASUREMENT WORKSHOP HELD IN JAPAN

Yoshimasa Oomura and Yasuaki Watanabe, Tokyo Metropolitan University, Hachioji, Tokyo 192–03, Japan
Mitsuaki Koyama, Nihon Dempa Kogyo Co., Ltd. Kamihirose Sayama-city, Saitama 350–13, Japan

ABSTRACT With the further compactness of electronic devices, the demand for leadless (surface-mounted) crystal units has recently been increasing. A great problem lies in that there is no standard specification for measuring equivalent constants which provide the basis for evaluating the performance of a crystal unit. At present, each crystal unit manufacturer has its own measurement method. To unify the measurement method in Japan and propose the standard measurement method to IEC, the Japan Committee of IEC TC49 / WG6 (sub-committee for crystal unit measurement method) held a workshop in September 1992. This report shows the procedure of the workshop as well as the measurement results, which demonstrate that the proposed standard method to be conducted under the measurement conditions given provides the sufficient measurement accuracy required for practical use.

1. INTRODUCTION

Further compactness and light weight of high technology electronic devices such as portable telephones have carried the trend toward the use of more compact and surface-mounted electronic components. Some leadless (surface-mounted) crystal units are used for generating a reference frequency or used as a clock. With the various requirements of customers such as the design of custom-made crystal units that have become the standard products of each crystal unit maker, the standard specifications such as external dimensions have begun to be discussed here in Japan. As mentioned earlier, there is neither domestic nor international specification for measuring the equivalent electrical constants which determine the specifications of a crystal unit. At present, each crystal unit manufacturer measures them in its own way. Because surface-mounted crystal units are expected to become the mainstream, standardization of the equivalent constant measurement method will have to be accomplished. To meet such requirements, the Japan Committee of IEC TC49/WG6 has been discussing the standard measurement method to unify the domestic measurement method and proposing the international specifications. The main purpose of this report is to show the statistical data which were obtained in the workshop and are to be attached with the proposal. The section 2 shows the specifications of the standard measurement method proposed and the section 3 gives the general description of the workshop. The section 4 shows the statistically processed data based on the measurement data and the evaluation of the results obtained.

2. PROPOSED MEASUR'S SPECIFICATIONS OF SURFACE-MOUNTED CRYSTAL UNITS

2.1 SCOPE

This report specifies the measurement instruments used for measuring series resonance frequency, series resistance and other electrical equivalent parameters accurately.

2.2 SPECIFICATIONS TO BE BASED UPON

The measurement method for measuring the series resonance frequency, series equivalent resistance and other electrical equivalent parameters of a resonator are according to the IEC Pub.444.

2.3 MEASUREMENT METHOD AND THE SPECIFICATIONS OF MEASUREMENT INSTRUMENTS

2.3.1 CONFIGURATION OF MEASUREMENT SYSTEM

The configuration of the electrical circuits of measurement instruments is according to the IEC Pub.444 (PI-circuit method).

2.3.2 CALIBRATION OF MEASUREMENT SYSTEM

The calibration of measurement instruments requires the use of two reliable reference-impedance elements whose impedance imaginary-part (Zi) can be neglected. If the floating capacitance between the terminals is not negligible, the open calibration shall be conducted. This calibration specifies the measurement accuracy (2.3.4).

2.3.3 FREQUENCY RANGE

The frequency range shall be 1MHz - 150MHz.

0-7803-0905-7/93 $3.00 © 1993 IEEE

2.3.4 FREQUENCY ACCURACY

The measurement error shall be within the range specified as shown below in consideration of errors due to calibration, measurement instruments and noises generated at the time of measurement. The following accuracy is required of the measurement instruments.

(1) The measurement accuracy of series resonance frequency(fs) shall be $\pm 1 \times 10^{-6}$.

(2) The measurement accuracy of equivalent series resistance (R1) shall be ±2 ohm or ±10 percent.

(3) The measurement accuracy of parallel capacitance (C0) shall be ±0.2pF.

(4) The measurement accuracy of series inductance and capacitance (L1, C1) shall be ±2 percent.

2.4 MEASUREMENT CONDITIONS

2.4.1 EARTH

The holder of the crystal unit shall be earthed. For a crystal unit which has earth terminals connected to the holder, they can be used for earthing as long as the measurement accuracy specified in the step 2.3.4 is obtained.

2.4.2 LOAD CAPACITANCE

The load capacitance shall be of series.

2.4.3 DRIVE LEVEL

The drive level measured for evaluating the performance shall be 0 dBm in terms of the input level of the measurement instrument.

2.4.4 Measurement environment

The measurement environment for evaluating the performance shall include room temperature, ordinary humidity and ordinary pressure.

2.5 STRUCTURE OF MEASUREMENT INSTRUMENTS

The structure of a measurement instrument used shall not be specified herein except that it should be designed to allow the electrodes, which correspond to the leadwires of the crystal unit to be tested, and the measurement terminals to be securely in contact with each other.

3. EQUIVALENT CIRCUIT CONSTANTS TO BE MEASURED

The following seven items are to be measured.

1. fs: Series resonance frequency
2. R1: Equivalent series resistance
3. L1: Equivalent series inductance
4. C1: Equivalent series capacitance
5. C0: Parallel capacitance
6. r : Capacitance ratio (C0/C1)

4. OVERVIEW OF THE WORKSHOP THE WORKSHOP WAS HELD AT THE TOKYO METROPOLITAN UNIVERSITY (HACHIOJISHI,Tokyo) on the 5th of September 1992.

Actual measurement of the characteristics of crystal units was made by three measurement instrument makers (Japan) and one crystal unit maker (Japan). The four teams brought a network analyzer, measurement fixtures and a process control computer to the workshop before they used their own softwares according to the specified measurement procedure and equivalent circuit constant calculation procedure. The purpose of this workshop is not to specify the standard measurement instrument, but to obtain the data which provide the basis for preparing documents which specify the measurement method, the equivalent constants to be measured and the accuracy required. So the mean of measurement data of each team is regarded as the median and the difference from the median is considered to be the deviation, though the difference from what is called standard value is regarded as the deviation.

4.1 CRYSTAL UNITS TESTED

One of the crystal units shown in Table 1 was used for the test. Their frequencies fall between 12.8MHz and 145MHz. Fig.1 shows the external dimensions. This kind of shape is generally used by most of crystal unit makers as the standard type. Fig.2 shows the resonance frequency versus temperature characteristics around the room temperature, revealing that each crystal unit has the performance within ±2 ppm.

No.	Frequency	Specification
1	12.8MHz	AT-cut Fundamental (NDK)
2	31.3MHz	AT-cut Fundamental (NDK)
3	49.5MHz	AT-cut 3rd overtone (NDK)
4	96.8MHz	AT-cut 5th overtone (NDK)
5	145.0MHz	AT-cut 3rd overtone (Toyocom)

TABLE 1 List of the crystal units used for the test

4.2 MEASUREMENT SYSTEM AND FIXTURES

Fig.3 shows the measurement system used at the workshop. This system is composed of two PI networks allowing transmission. Measurement fixtures made by each maker according to the shapes of crystal units to be tested (See Fig.1) have built-in PI networks. Fig.4(a)-(d) shows the measurement fixtures of each team. For the drawings (a) and (c), the terminals of a crystal unit face upward. For the drawing (b), the contact terminals come in under a crystal unit. For the drawing (d), terminals are to be connected from the side a crystal unit.

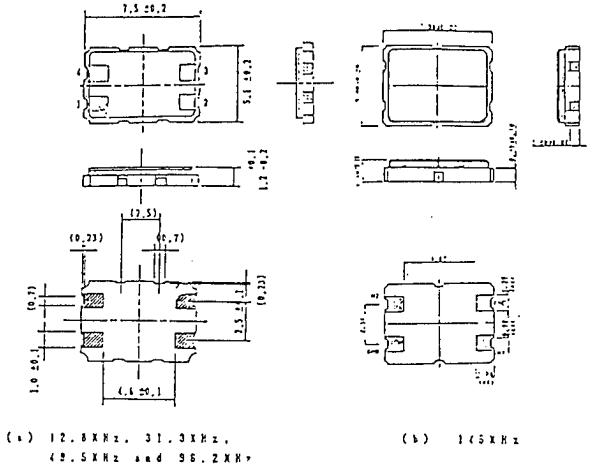


Fig.1 Schemes of test crystals.

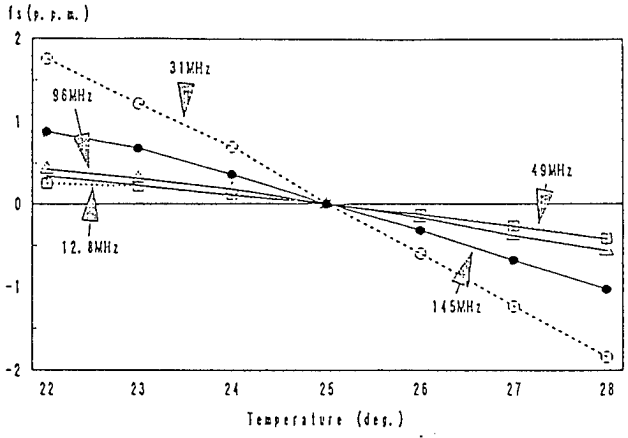


Fig.2 Frequency temperature characteristics of crystal units under tests.

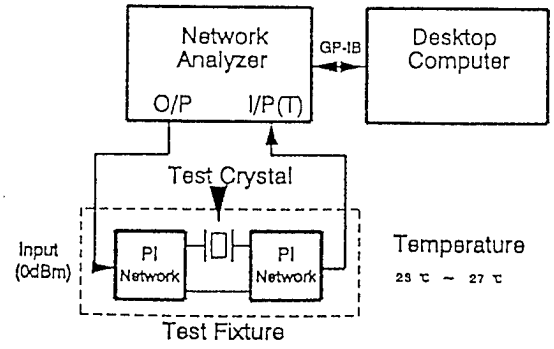


Fig.3 Block diagram of measurement system.

4.3 BASIC ALGORITHM FOR DETERMINING EQUIVALENT PARAMETERS

The electrical equivalent parameters were determined by the admittance circle diagram based on the equivalent circuit consisting of four elements, C1, L1, R1 and C0. C0 measurement was made according to the IEC 49(CO)248 (beingvoted); if the frequency was less than 30MHz, measurement was made on the basis of the impedance corresponding to around 30MHz; if exceeding 30MHz, measurement was made based on the data around the resonance frequency.

4.4 CALIBRATION METHOD

The calibration of the system including the measurement fixtures was conducted through linear interpolation, using two known impedance elements. In other words, a 50 ohm and short bar were prepared for each crystal unit to be tested and calibration was conducted for each measurement.

4.5 MEASUREMENT PROCEDURE

Measurement was made in the following order.

(1) Receiving crystal units to be tested and calibration elements (50/0 ohm)
(2) Calibration of measurement instruments
(3) Measurement is made 15 times.(Each time measurement is over, contact closing/opening is required.)
(4) Re-calibration of measurement instruments
(5) Measurement is made 15 times.

Since improper contact may exert bad influences upon measurement, measured data shall be divided into blocks, each of which consists of three data, and if the resonance frequency or the series resistance of a datum is greatly different from those of the other two data, the former shall be rejected. After the calibration, ten data were statistically processed. The resonance frequency and equivalent parameters are as follows:

1 fs Resonant frequency
2 fr Non-reactive frequency
3 fp Parallel resonant frequency
4 L1 Equivalent Series inductance
5 C1 Equivalent Series capacitance
6 R1 Equivalent Series resistance
7 Rr Non-reactive resistance
8 Co Equivalent Parallel Static capacitance
9 r Co/C1
10 p Power dissipation
11 T Temperature

For the statistical calculation, only the usually used equivalent parameters such as fs, L1,C1,R1, and Co were processed.

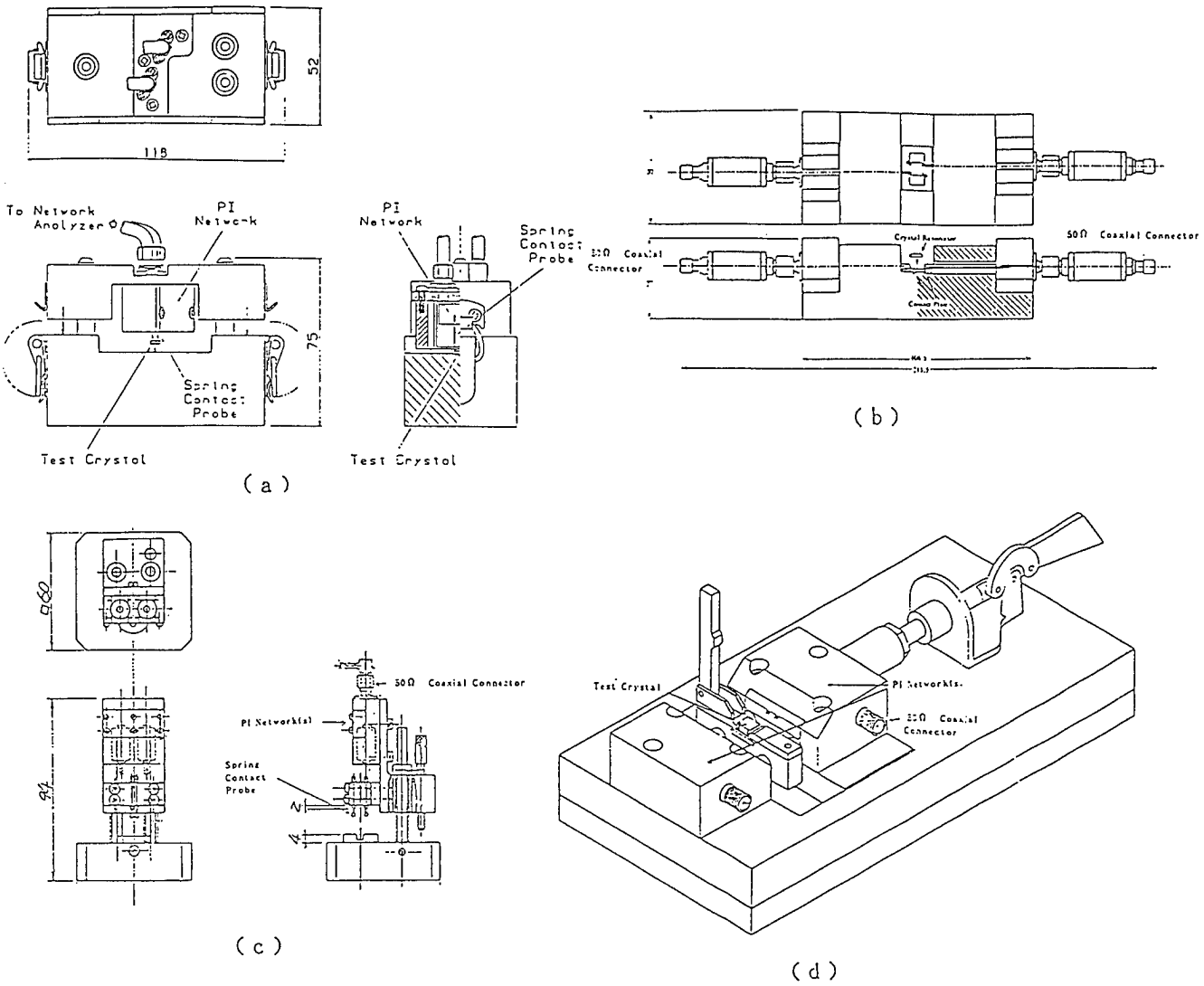


Fig.4 Schemes of test fixtures.

4.6 MEASUREMENT ENVIRONMENT AND ELECTRIC POWER REQUIRED

Measurement was made in a room whose temperature is set at 23 °C to 27 °C . The room temperature was measured by an HP's quartz crystal thermometer: The temperature measured at the time of calibration was regarded as the measurement temperature. To equalize the frequency to be measured by each measurement instrument, a Rb atomic oscillator was used for the calibration each reference frequency source so that the deviation would be within 1E-8 or less. The measurement power was preset so that the PI-circuit input power became 0 dBm, taking into consideration the power dissipation and measurement noises due to the PI networks. The electric power of approximately -30 dBm is applied to crystal units.

5. STATISTICAL CALCULATION METHOD AND RESULTS

The data measured for each constant of a crystal unit were statistically processed as in the following. For the sake of convenience, we will show you the results of only the case (3).

(1) Each team: Calculating the means and standard deviations after each of calibrations to be conducted twice.
(2) Each team: Calculating the means and standard deviations of all the data (first calculation and second calculation included).
(3) All teams: Calculating the means and standard deviations for all the teams for calculating the deviations of the means of each team [case (2)] from the overall means.

The statistical analysis (1) shows the measurement accuracy of each team and the deviation due to calibrations. The analysis (2) reveals the measurement accuracy of each team containing the errors due to calibrations. The analysis (3) shows the deviation of the mean of each team from the overall mean. The data measured by each team were used as they were, without temperature compensation.

5.1 RESULTS OF STATISTICAL CALCULATIONS

Fig. 5 - Fig.8 show the deviations of the means(first calculation and second calculation) of each teamfrom the overall means for the specified parameters.Thesefigures show that the deviation of resonance frequency (fs) from the overall mean is within ±1.32ppm, equivalentseries resistance (R1) within ±1.10 ohm, series reactance(L1) within ±1.73 percent, parallel capacitance (Co) within ±0.37pF According to the above-mentioned analyses, the deviation of the resonance frequency is greater than the value proposed. If it could be able to improve the temperature characteristics of the crystal units themselves, it will allow that the measured value fall within the specified range. A problem lies in a great deviation [eg. case A (145MHz)]. In other cases, the deviation is within ±1 ppm. The deviations of other equivalent constants are small, demonstrating that measurement was made with the accuracy sufficient for practical use.

6. CONCLUSION

We have presented the results of the workshop that was held at the Tokyo Metropolitan University for the purpose of conducting an investigation as to the measurement method and specifications for the surface-mounted leadless crystal units, and of obtaining actual data. Our study shows that unifying the basic specifications (PI circuit method) and the measurement software algorithm (IEC specification proposed) allows high-accuracy measurement to be conducted as shown in the chapter 4, even if different fixtures and measurement instruments are used. We think it necessary to make a study of the influences of the errors due to calibration elements and open calibration, if further measurement accuracy is required in the future.

ACKNOWLEDGMENT

We would like to appreciate all the persons concerned who participated in the workshop from the bottom of our heart.

REFERENCES

[1] IEC Pub.444-1:"Measurement of quartz crystal unit parameters by zero phase technique in a PI-network", 1986

[2] IEC Doc. 49(CO)248:" Methods for measurement of quartz crystal devices for the determination of equivalent electrical paraemeters using automatic network analyser techniques and error correction", 1992

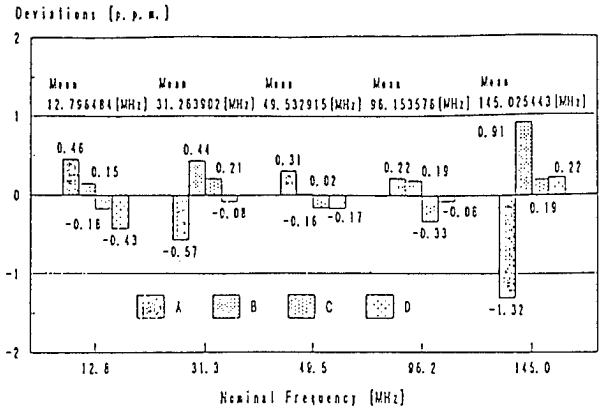


Fig.5 Statistical results of series resonant frequency.

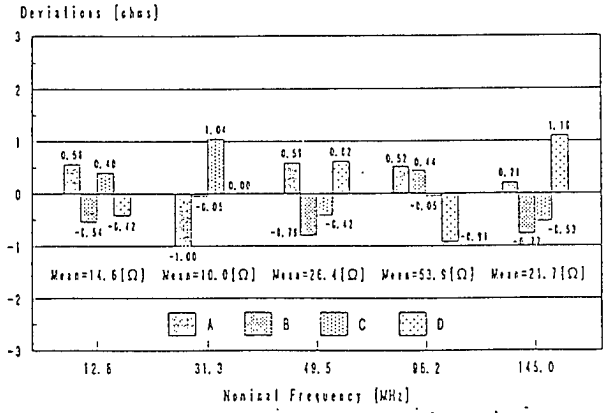


Fig.6 Statistical results of equivalent series resistance.

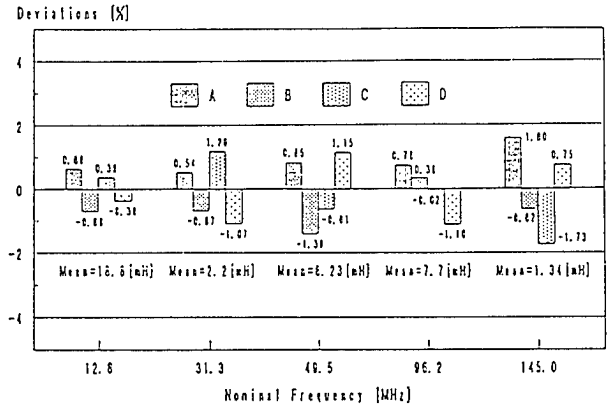


Fig.7 Statistical results of equivalent series inductance.

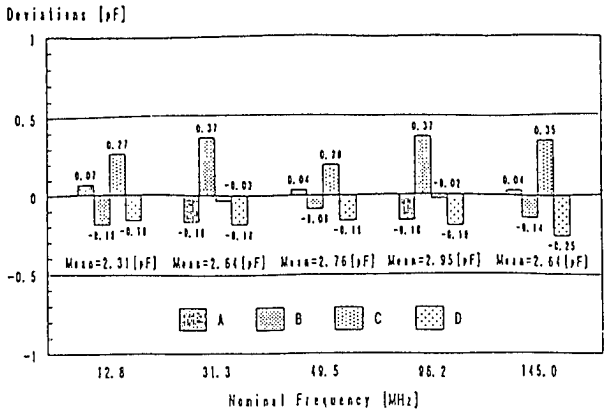


Fig.8 Statistical results of equivalent parallel capacitance.

Appendix

Attendee	
N.Kotani and S.Shiba,	Advantest
T.Uno and M.Uragami	Anritsu
K.Yanagawa and S.Tanaka	Yokogawa Hewlett Packard
M.Koyama	Nihon Dempa Kogyo
M.Kadota and T.Ishiguro	Murata Mfg.
Y.Tanaka	Toyo Communication
M.Hayashi	Meiden
A.Ito	Kinseki
M.Sato	Tokyo Dempa
I.Maruyama and T.Hisanori	Fujitsu
T.Uno	NTT
Y.Oomura and Y.Watanabe	Tokyo Metropolitan Univ.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EIGHT-POLE MONOLITHIC FILTERS USING LITHIUM TANTALATE

Irina Mateescu* and John Kosinski**

*Institute of Physics and Technology of Materials
PO Box MG-7
Bucuresti - Măgurele, România

**US Army Research Laboratory
ATTN: AMSRL-EP-MD
Fort Monmouth, NJ 07703-5601

Abstract: The large electromechanical coupling coefficient of lithium tantalate makes possible the construction of monolithic filters with much wider bandwidths than are achievable using quartz. Fundamental mode, two-pole monolithic filters using lithium tantalate have been reported previously by various researchers. In this paper, we report on the design and fabrication of an eight-pole bandpass filter realized by the capacitive connection of four each third harmonic, two-pole monolithic filters using the 163° rotated Y-cut lithium tantalate substrate. The filter has a nominal center frequency of 32.5 MHz, and attains a bandwidth of approximately 0.5% of center frequency.

Introduction

Lithium tantalate ($LiTaO_3$) is of considerable interest as a candidate material for piezoelectric devices due to its strong piezoelectric coupling. It was first identified as a piezoelectric material in 1949, and afterwards was synthesized in practical form at the Bell Laboratories. Since then, piezoelectric devices using lithium tantalate have found applications in filtering and frequency generation in professional electronics, and, more recently, in consumer electronics [1].

Lithium tantalate possesses an electromechanical coupling coefficient for the fast shear mode which is substantially larger than that of quartz (40% vs. 9%), while at the same time possessing various other desirable features similar to quartz such as high Q, good crystalline perfection, chemical stability, etc. [2,3]. Among the other materials having a strong electromechanical coupling coefficient (piezoelectric ceramics, lithium niobate), lithium tantalate offers an optimum compromise between values of coupling coefficient, Q factor, and temperature performance. This material is thus very interesting for the realization of filters with a large bandwidth.

One approach to the realization of bandpass rf filters is that of the monolithic crystal filter using trapped energy resonators. Working on the fundamental frequency and harmonics in the 10-100 MHz range, monolithic filters on lithium tantalate have realizable bandwidths from 0.4% up to 6% of the center frequency as compared to bandwidths realizable using quartz which range from thousandths of a percent up to 0.3% [2]. For multi-pole responses, there are two possible filter topologies, namely a) deposition of multiple resonator pairs on a single substrate and b) the capacitive connection of a number of separate two-pole monolithic filters. The latter approach is to be preferred since it requires a less exacting processing technology and a simpler design method than the fully integrated approach [3].

In order to realize a wide-band two-pole monolithic filter, the vibrational mode structure of the plate should be dominated by the simple thickness shear modes excited in the constituent resonators, with the individual shear modes coupled at the surface of the plate. The 163° rotated Y-cut is appropriate for monolithic filter structures based on lithium tantalate since the shear mode coupling for thickness-field excitation is large while the other thickness modes are effectively absent. The direction of the particle displacement lies within 3° of the Z' axis so that the mode is close to a pure shear mode [3-5].

The individual resonators of the lithium tantalate monolithic filter may be coupled via either of two modes of vibration, namely thickness twist (TT) or thickness shear (TS). In contrast to quartz (symmetry class 32) where the coupling

0-7803-0905-7/93 $3.00 © 1993 IEEE

direction for TS lies along the X' axis, the coupling direction for TS in lithium tantalate (symmetry class 3m) lies along the Z' axis. For a given electrode spacing, the use of Z' direction TS coupling results in a larger bandwidth than the use of X' direction TT coupling [3].

The equivalent circuit for the lithium tantalate resonator is similar to that for quartz, however owing to the much larger electromechanical coupling the motional inductance is considerably smaller. The resulting lower electrical impedance of such devices (hundreds of ohms) allows a much easier matching of the crystal resonator to the external circuit [6].

Filter Design

The design specifications used in this work are specific to a wide-band bandpass filter required for a modern radio receiver:

Center frequency	f_o	32.5 MHz
3dB bandwidth	df	75 kHz
Insertion loss	A_o	4 dB max.
Passband ripple	A_{max}	0.5 dB
Stop band/attenuation		190 kHz/90 dB
Ultimate attenuation	A_{min}	90dB
Load resistance		500 ohms

The receiver is designed for use with simple frequency modulation and employs a high value of the intermediate frequency in order to facilitate protection at the image frequency. The design of the monolithic filter may be separated into synthesis, analysis, and resonator design steps, each of which is discussed below [7].

Synthesis

The initial design process for the monolithic crystal filter is identical to that used for designing discrete crystal filters. The design specifications, along with the use of third harmonic operation of the lithium tantalate crystal result in the additional design inputs,

Poles	n	8
Capacitance ratio (third overtone)	r	64.004

with the eight-pole response to be realized by the capacitive interconnection of four two-pole monolithic filters. After choosing an appropriate bandpass filter response which achieves the required selectivity, a ladder-type filter design is generated which can absorb the equivalent circuit of the monolithic filter. The conversion from the basic ladder-type filter to the crystal filter is relatively simple because the two-pole devices can be substituted directly into the ladder network [8].

The polynomial design method was chosen for this application, since, for the same degree of filter complexity, the results obtained for monolithic filters designed by the polynomial method are better than those obtained for filters designed using the image parameter method. The polynomial method takes into account the matching resistive load, as opposed to the image impedance in the image parameter method [7]. In this way, the polynomial method allows a rigorous control of the reflection coefficients at the input and output with respect to the terminating impedance. The polynomial method also allows a smaller passband ripple for a given attenuation characteristic. In order to attain the desired shape factor, the Chebyshev approximation was used for the attenuation characteristic.

The structure of the eight-pole wide-band filter composed of capacitively coupled two-pole, symmetrical monolithic crystal filters is shown in Figure 1 [8,9]. The modified eight-pole ladder filter equivalent is shown in Figure 2. The sequence of design steps is as follows [9-11]:

1. Convert the given bandpass filter requirement into a normalized lowpass specification. The network of the lowpass equivalent filter is shown in Figure 3. Given the values of f_o, df, A_{max}, and n, calculate the elements of the lowpass filter using the relations [3,10,11]

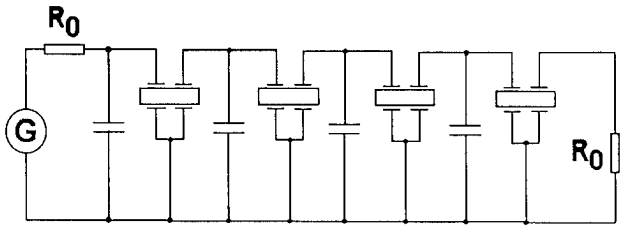


Figure 1. Eight-pole filter structure.

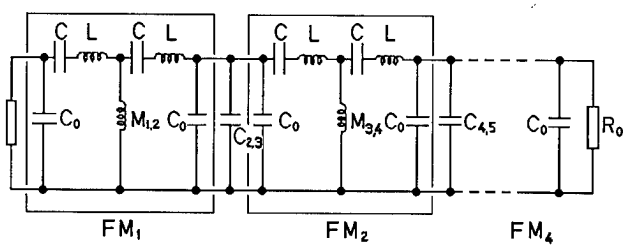


Figure 2. Eight-pole ladder filter.

$$w = \frac{df}{f_o} \quad (1)$$

$$\epsilon = [10^{(A_{max}/10)} - 1]^{(1/2)} \quad (2)$$

$$FC = (1/n)\ \cosh^{-1}\ (1/\epsilon) \quad (3)$$

$$\beta = \ln\ [\coth\ (A_{max}/17.37)] \quad (4)$$

$$Y = \sinh\ (\beta/2n) \quad (5)$$

$$A_s = \sin\ [(2s-1)\pi/2n] \qquad s = 1...n \quad (6)$$

$$B_s = Y^2 + \sin^2(s\pi/n) \qquad s = 1...n \quad (7)$$

$$g_i = \frac{4\ A_{i-1}\ A_i\ \cosh^2(FC)}{B_{i-1}\ g_{i-1}} \quad (8)$$

2. The lowpass filter prototype is transformed into a narrow-band bandpass filter with equal inductances ($l=g_1$) and with reactances independent of frequency in the series arms. Some elements of this network are then computed as

$$\alpha_{2,3} = \frac{g_1}{\sqrt{(g_2\ g_3)}} \qquad \text{for } n>2 \quad (9)$$

$$\alpha_{4,5} = \frac{g_1}{\sqrt{(g_4\ g_5)}} \qquad \text{for } n>4 \quad (10)$$

$$\alpha_{6,7} = \frac{g_1}{\sqrt{(g_6\ g_7)}} \qquad \text{for } n>6 \quad (11)$$

3. The narrow-band bandpass filter is changed into a polynomial filter. The normalized inductance l_q and capacitance c_q of each resonator are computed for $\omega_o=1$ and $R_o=1$ as

$$l_q = q_1/w \quad (12)$$

$$c_q = w/g_1 \quad (13)$$

The coupling coefficient is calculated as

$$k_{i-1,i} = (1/2)\ [B_{i-1}/(A_{i-1}A_i)]^{(1/2)} \quad (14)$$

Due to the symmetry of the filter, $k_{1,2}=k_{7,8}$, $k_{2,3}=k_{6,7}$, and $k_{3,4}=k_{5,6}$.

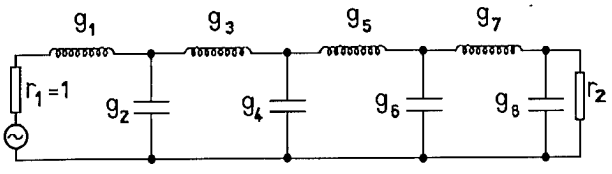


Figure 3. Low-pass equivalent filter.

4. Transform the normalized lowpass parameters into the required bandpass filter using

$$l_o = \frac{R_o}{2\pi f_o} \quad (15)$$

$$c_o = \frac{1}{2\pi f_o R_o} \quad (16)$$

$$L = l_q\ l_o \quad (17)$$

$$C = c_q\ c_o \quad (18)$$

$$C_o = C\ r \quad (19)$$

$$K_{i-1,i} = k_{i-1,1}\ w \quad (20)$$

$$M_{z,\nu} = K_{z,\nu}\ L \quad (21)$$

$$C_{c2,3} = C_{c6,7} = \frac{1}{\alpha_{2,3(6,7)}\ 2\ \pi\ f_o\ R_o} \quad (22)$$

$$C_{c4,5} = \frac{1}{\alpha_{4,5}\ 2\ \pi\ f_o\ R_o} \quad (22)$$

$$C_{2,3} = C_{6,7} = C_{c2,3} - 2C_o \quad (23)$$

$$C_{4,5} = C_{c4,5} - 2C_o \quad (24)$$

The resonance frequencies of the two-pole filters f_{r1} (for filters FM_1,FM_4) and f_{r2} (for filters FM_2,FM_3) are

$$f_{r1} = f_o\ (1 - \frac{1 - [1 - 4\beta_o^2]^{(1/2)}}{4\ \gamma}) \quad (25)$$

and

$$f_{r2} = f_o\ (1 - \frac{K_{4,5}}{2}) \quad (26)$$

with

$$\beta = 2\ \pi\ f_o\ R_o\ . \quad (27)$$

Analysis

The response characteristics of the synthesized bandpass filter were examined by using the values of the circuit elements obtained from the filter synthesis in the psspice circuit analysis program. Figure 4 compares the calculated attenuation characteristic of the filter with the design specification. The results demonstrate that the synthesis procedure is accurate.

Resonator Design

From the relations between the elements of the equivalent electrical

circuit and the physical properties of the monolithic structure, it is possible to calculate the required physical dimensions and geometry of the individual two-pole filters (see Figure 5), namely, the plate thickness t, electrode dimensions l_x and l_z, and electrode spacing d [3,12,13].

The resonance frequency of the plate and the plate thickness are related by

$$f_o = \frac{N_m}{t} \tag{28}$$

where f_o is the resonance frequency of the plate (in MHz), N_m is the frequency constant of the mth harmonic (5.922 MHz-mm for third harmonic 163° rotated Y-cut $LiTaO_3$), and t is the plate thickness (in mm). Using equation (28), we find that the required plate thickness is t=0.1845 mm.

For the special case of 163° rotated Y-cut lithium tantalate, the expression for the resonator motional inductance L is given by [3,6,12,13]

$$L = \frac{0.154 N_m^3}{A\, f_m^3}\left[1 - \frac{0.224 m^2}{N_m^2}\right] \tag{29}$$

where m is the order of the resonance and A is the electrode area (in mm^2). Using the appropriate values for f_m, m, N_m, and L yields the desired electrode area of A=0.927 mm^2. For the square electrodes used here, $l_x=l_z=l_e=0.963$ mm.

Two conditions for separating the unwanted modes have been identified previously, namely [6]

$$A/t^2 \approx 25 \tag{30}$$

and

$$l_e/t \leq 8 \ . \tag{31}$$

These conditions are reasonably satisfied by the indicated electrode dimensions. The diameter of the plates was chosen to be 8.2 mm.

The relation describing the inter-resonator coupling along the Z' direction is [3,12,13]

$$K_{Z'} = A \exp[-B\,(1-2\sqrt{\Delta})\,(d/t)\,\sqrt{\Delta}] \tag{32}$$

where Δ describes the electrode plateback. The constants A and B have been determined experimentally as A=0.008083 and B=-6.417, in good agreement with the results of Hales, et al. [3]. For the third harmonic of lithium tantalate, Δ is given by

$$\Delta = 0.0058 + 1.2R \tag{33}$$

with

$$R = \frac{\rho_e\, t_e}{\rho\, t_1} \ . \tag{34}$$

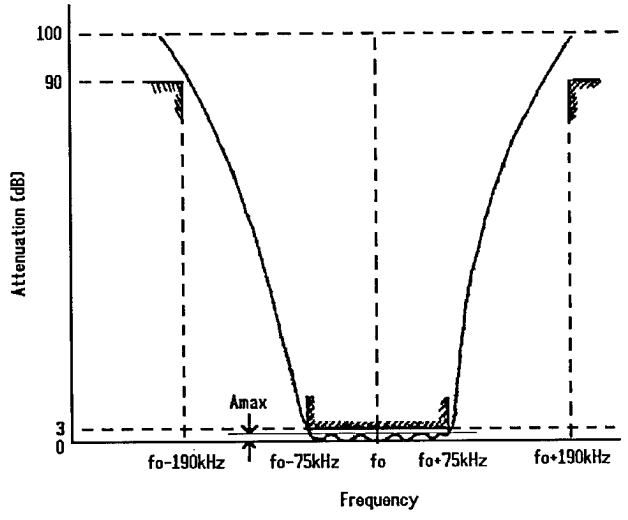


Figure 4. Calculated eight-pole filter attenuation characteristic.

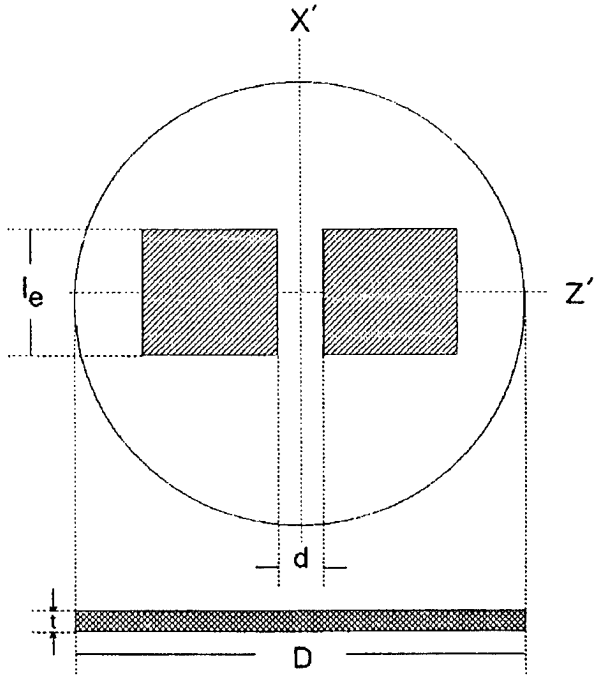


Figure 5. Two-pole monolithic filter layout.

In equation (34), ρ_e is the electrode density, t_e is the thickness of one electrode, ρ is the density of the lithium tantalate plate, and t_1 is one half of the thickness of the lithium tantalate plate. The inter-resonator coupling relations are illustrated in Figures 6 and 7. Figure 6 presents the dependence of the inter-resonator coupling coefficient K_Z, on the electrode spacing d with Δ as a parameter. In Figure 7, the relationship between the electrode spacing and plateback is shown for the two cases of $K_{1,2}=K_{7,8}$ and $K_{3,4}=K_{5,6}$ corresponding to the two cases of FM_1,FM_4 and FM_2,FM_3 respectively.

The electrode thickness is chosen to be 1800 Å in order to obtain a response free of spurious modes. The resulting electrode spacings are then d_1=0.346 mm for the two-pole filters FM_1 and FM_4, and d_2=0.477 mm for FM_2 and FM_3.

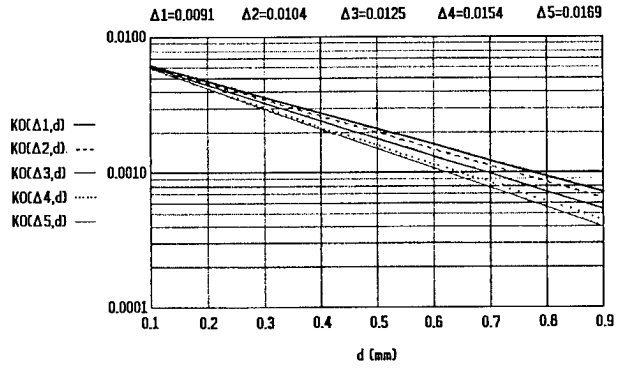


Figure 6. Inter-resonator coupling versus electrode spacing for selected values of plateback.

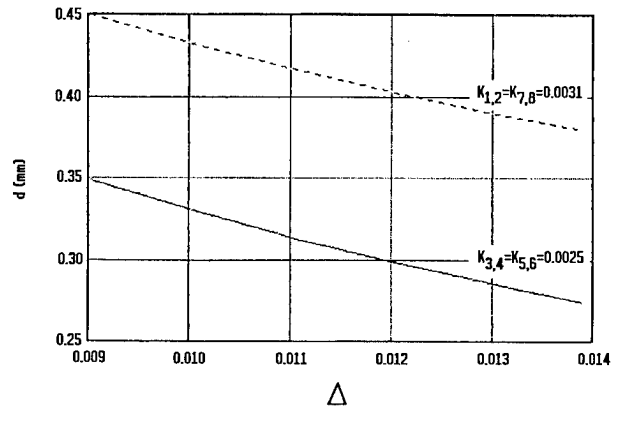


Figure 7. Electrode spacing versus plateback for selected values of inter-resonator coupling.

Experimental Results

The individual two-pole monolithic crystal filters as well as the complete eight-pole filter assembly were fabricated and evaluated at the Institute of Physics and Technology of Materials. Evaporated Cr/Ag electrodes were deposited on the 163° rotated Y-cut plates in an oil-diffusion pumped vacuum system. The electrode thicknesses were established using a MIKI 5 MHz crystal thickness monitor. The electroded plates were then mounted in HC-18 style holders using conductive epoxy, and sealed.

The attenuation characteristic of the eight-pole filter was evaluated using a Wandel-Goltermann test set. Figure 8 compares the measured response to the expected response and the design specification. The critical parameters were measured as

Center frequency	f_o	32.491 MHz
3dB bandwidth	df	72.972 kHz
Insertion loss	A_o	2.8 dB
Passband ripple	A_{max}	2.2 dB
Stop band/attenuation		190 kHz/90 dB
Ultimate attenuation	A_{min}	90dB

The measured response closely approaches the design specifications with the exception of the passband ripple.

Conclusion

The development of a wide-band eight-pole lithium tantalate monolithic crystal filter operating on the third harmonic has been presented in detail. The close agreement between the measured response and

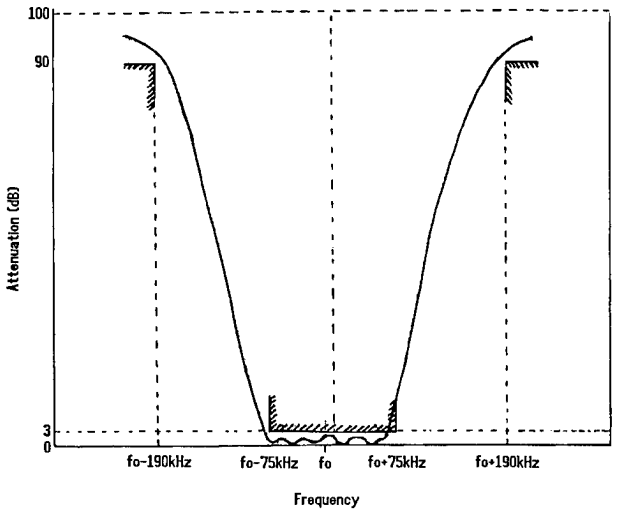


Figure 8. Measured performance of the prototype eight-pole filter.

the design specifications validates the design and fabrication procedures employed here.

Acknowledgments

We are indebted to Dr. Alexandru Şerbănescu for his help in the data analysis and to Constantin Bora and Ioan Mateescu for their assistance throughout this activity.

References

[1] A. Zarka, B. Capelle, Y. Zheng, T. Detaint, J. Schwartzel, C. Joly, D. Cochet-Muchy, "Assessment of Material Perfection and Observation of Vibration Modes in Lithium Tantalate by X-ray Topography," Procedings of the 45th Annual Frequency Control Symposium (AFCS), 1991, p. 58.

[2] S. A. Sakharov, I. M. Larinov, I. V. Issaev, "Monolithic Filters Using Strong Piezoelectrics," Proc. 45th AFCS, p. 181, 1991.

[3] M. C. Hales, J. W. Burgess, "Design and Construction of Monolithic Crystal Filters Using Lithium Tantalate," Proc. IEE, Vol. 123, No. 7, p. 657, July 1976.

[4] J. W. Burgess, M. C. Hales, F. W. Ainger, "Single Mode Resonance in $LiTaO_3$ for Filter Design," Electronics Letters, Vol. 9, No. 11, May 1973.

[5] J. Zelenka, Piezoelectric Resonators and Their Applications. Academia Prague, 1986.

[6] J. W. Burgess, M. C. Hales, R-J Porter, "Equivalent Circuit Parameters for $LiTaO_3$ Plate Resonators," Electronics Letters, Vol. 11, No. 19, p. 449, September 1975.

[7] I. Mateescu, Monolithic Quartz Filters, Ph.D. Dissertation, Polytechnic Institute of Bucharest, 1987.

[8] R. G. Kinsman, Crystal Filters. New York: John Wiley, 1987.

[9] V. M. Kantor, Monolitniie Piezoelectriceskie Filtri. Moskva, 1977.

[10] A. B. Williams, Electronic Filter Design Handbook. New York: McGraw-Hill, 1981.

[11] A. I. Zverev, Handbook of Filter Synthesis. New York: John Wiley, 1967.

[12] W. Soluch, Filtri Piezoelektriczne. Warszawa, 1982.

[13] I. Mateescu, G. Gackowska, G. Korony, and C. Bora, "Monolithic Crystal Filters Using Lithium Tantalate Crystal," Archives of Acoustics, Vol. 16, No.1, 1991.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MODELIZATION OF THICKNESS SHEAR INTEGRATED FILTERS

J.Détaint*, J.Schwartzel*, A.Zarka**, B.Capelle**, C.Joly*.

*FRANCE TELECOM CNET Laboratoire de Bagneux, dept PAB/BAG/MCT/CRL
196, Av.H.Ravera BP 107, 92225 Bagneux France.
**L.M.C.P. Université Paris VI, Paris VII et CNRS 4, place Jussieu
75252 Paris-Cedex France.

Abstract:

Bulk acoustic wave filters are very attractive for filtering in the equipments of radiocommunication systems due to their very good selectivity. Recent technologies allows to obtain plates having fundamental frequencies in the range from 70 to above 300MHz. At these high frequencies, it is possible to obtain filters with extremely reduced dimensions by the integration of several resonators or monolithic cells on the same plate. In order to optimize the design of such filters a model based upon the equations governing the essentially thickness modes was established. We consider structures in which two electroded regions are separated by a "coupling" region. The electroded and the coupling regions are surrounded by an external unelectroded region having an higher cut-off frequency to ensure that the energy can be trapped in the device. Depending on the distance between the electrodes and some other parameters, the resonators can be from strongly (monolithic filters) to negligibly acoustically coupled. The latter case seems to be very interesting at very high frequencies. A semi-algebraical method of solution is employed, making use of series of exact solutions of the equations and of a discretization of the continuity conditions between the different regions. This permits to obtain the eigen frequencies and the eigen-modes. The forced modes of the structures are then computed as linear combinations of the eigen modes. Using the forced modes, the terms of the admittance matrix of the 4 ports structure are calculated so that, the elements of the lattice equivalent circuit are obtained. A computer program was implemented and applied to different types of structures.

1/INTRODUCTION

Up to the recent years, the devices, presenting the most important degree of integration on a single piezoelectric substrate were the acoustically coupled (monolithic) filters. Most of the applications of these filters were situated in the range from 10 to 40 MHz.

One of the most important application of crystal filters is constituted by their use in the radiocommunication equipments. In this field, major evolutions are currently occurring with the progressive replacement of analog systems by digital ones that have much better spectral efficiency and also with the use of higher and higher frequencies. These evolutions induce new requirements for the filters used in such systems which require larger bandwidth, higher centre frequencies and also nearly linear phase responses. Other important trends also related to this application are the research of very small dimensions and of reduced cost.

Most of the systems being currently installed or planed for the next ten years use frequencies in the low gigahertz region. So that, thin film resonators (TFR) filters may be a solution for front end filtering in these systems. Intermediate filters for these systems must have centre frequencies situated in the low hundred megahertz frequency domain. New techniques allow the obtentions of plates vibrating in this frequency range, where the filters have relative bandwidth (0.5% to 5%) fully adapted to this application. Furthermore, due to the reduced dimension of the resonators at these frequencies, it is possible to integrate several resonators and hence complete filters on a plate with very small lateral dimensions. Recently, several very high frequency filters, integrating several acoustically isolated resonators were described and it was shown that such devices have very interesting properties relatively to achievable bandwidth, type of responses, and ease of adjustment [1]. Our purpose in this communication is to elaborate a model of such devices that can permit their optimization either for UHF application (TFR filter) or I.F. filters at VHF frequencies.

2/PRINCIPLE OF THE METHOD:

Several models of monolithic filters were presented sometimes ago [2], [3], [4], [5]. The most recent among these models are based upon finite elements or finite differences solutions of the equations obtained from the theory of the essentially thickness modes established by H.F Tiersten and co-workers [6],[7] or from theories of the same nature [8].

We have shown, for single resonators, that a semi-algebraical method of resolution of the Tiersten equations has the very interesting property to lead to very accurate solutions with a reduced amount of computation [9]. This method uses solutions in the form of series of exact solutions of the equation that verify the continuity and/or boundary conditions at a discrete number of points. The accuracy of the method is the result of a fast alternate numerical convergence of the solutions as the number of terms of the series is increased (Figure 1).

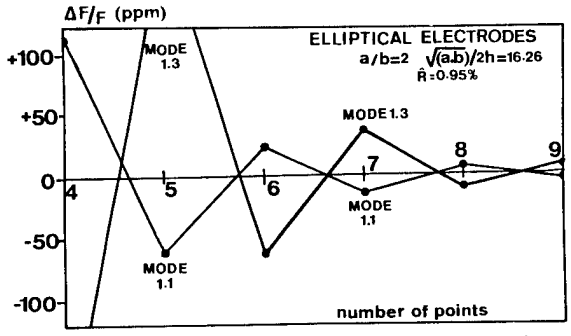


Figure 1: Convergence of the eigen frequencies observed for resonators

0-7803-0905-7/93 $3.00 © 1993 IEEE

Furthermore, this accuracy is obtained for nearly any electrode geometry of practical interest with a very reasonable amount of computation, so it was possible to find optimal geometries for several applications [10]. Last year we have presented the principle of a transposition of this method to consider multi-electrode devices [11], but several features of this first approach were not satisfactory, such as the discretization process used to obtain an equation to the eigen frequencies. Here, we present modified expressions and the first computed results.

In this communication, we consider structures made on a piezoelectric plate of constant thickness having two electroded regions characterised by a mass loading $\hat{R}_2$ and a cut-off frequency f_{ce2} that are separated by a "coupling" region characterised by cut-off frequency f_{ce1} (and a mass loading $\hat{R}_1$ lower than $\hat{R}_2$). The mass loading in the coupling region is supposed to be due either to an insulating deposit or to a metallic one with a short-circuit condition between the two faces. The external region surrounding the two previous ones is unelectroded and has an higher cut-off frequency to ensure that an energy trapping behaviour can occur.

Depending on the distance between the electrodes and several other parameters, the resonators can be from strongly to negligibly coupled. This corresponds respectively to the limiting cases of the two poles monolithic filters and to the case of acoustically isolated resonators integrated on the same plate. A structure which is rather of the first kind is represented in figure 2, whereas the structure of figure 3 is rather of the second kind. In the following, we consider mostly structures that are symmetrical with respect to the two perpendicular axis situated in the plate that lead to the cancellation of the mixed partial derivatives in equation {1} of §3 [7]. In this case only one fourth of the structure and the modes having the required symmetry are to be considered in the analysis. For less symmetrical structures, the principle of the analysis remains very nearly the same, but we need to consider one half or the whole structure and less symmetrical modes.

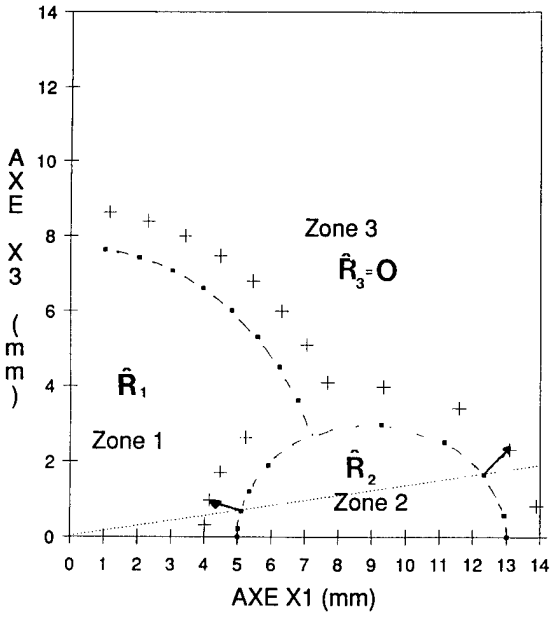


Figure 2: 2 poles structure with a large acoustical coupling.

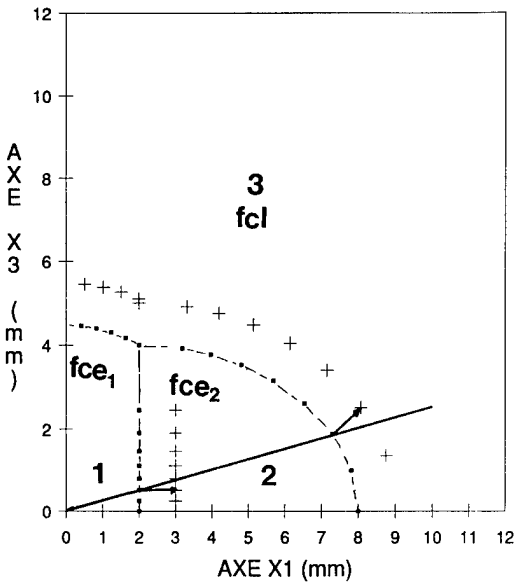


Figure 3: 2 poles structure with a weak acoustical coupling.

3/ANALYSIS OF 2 POLES STRUCTURES

As basis of the model we use the scalar equation governing the lateral dependence of the anharmonics in the vicinity of one overtone (noted n) of the thickness shear. This equation was first established for y rotated plates of class 32 crystals [6] (plates with a 2-monoclinic symmetry), then for plates of arbitrary symmetry [7]. The form given below for this equation is valid for the single resonators, geometrically symmetrical with respect to the plane x_2=0, and having two electrodes between which an excitation voltage equal to $V_0.e^{j\omega t}$ is applied. For the case of structures with more than a pair of electrically active electrodes, a modification of the forcing term is required (see below the paragraph about the forced modes). However since we consider here mostly symmetrical structures with 2 active electrode pairs, and since in this case the bissection theorem (Bartlett) indicates that the consideration of only a half structure is required (one electrically active electrode pair), we will retain this form. The eigen (short-circuit) mode analysis given below is naturally valid independently of such considerations.

$$M_n \frac{\partial^2 \bar{u}_1^n}{\partial x_1^2} + P_n \frac{\partial^2 \bar{u}_1^n}{\partial x_3^2} - \frac{n^2\pi^2 c^*}{4h^2}\bar{u}_1^n + \rho\omega^2 \bar{u}_1^n = \rho\omega^2(-1)^{(n-1)/2}\frac{e_{26}4V_0e^{j\omega t}}{c^{(1)}n^2\pi^2} \qquad \{1\}$$

In this equation, $\bar{u}_1^n(x_1,x_3,t)$ is the lateral dependence of $\hat{u}_1^n(x_1,x_2,x_3,t) = \bar{u}_1^n \sin(n\pi x_2/2h)$. $\hat{u}_1$ arises from a transformation of the usual component of the displacement u_1 and of ϕ the potential , made to replace the inhomogeneous boundary conditions $\phi = \pm V_0 e^{j\omega t}/2$ at $x_2 = \pm h$ in the electroded region, by the homogeneous conditions $\hat{\phi} = 0$. The inhomogeneous term of the equation results from this transformation.

c* is either $\bar{c}^{(1)}$ for the unelectroded regions or $\hat{c}^{(1)}$ for the electroded regions. $\bar{c}^{(1)}$ is the stiffened elastic constant relative to the corresponding one dimensional mode (Eigen-value of the Christoffel Matrix). $\hat{c}^{(1)} = \bar{c}^{(1)}(1 - 8k^{(1)2}/n^2\pi^2 - 2\hat{R})$ is a

constant that includes the electrical and inertial effects of the metallization. $4k^{(1)2}/n^2\,\pi^2$ is approximatively the relative frequency lowering due to the electrical effect of the metallization, $\hat{R}$ is the mass loading. $c^{(1)} = \overline{c}^{(1)}(1-k^{(1)})^2$ is a "pseudo ordinary elastic constant" which is equal to C_{66} for the Y rotated cuts of class 32 crystals. $k^{(1)}$ is the coupling coefficient of the corresponding one dimensional mode. M , P are functions of the material constants and of the plate orientation that can be obtained as described in reference [6] and [7]. In the following, the time dependence in $e^{j\omega t}$ of all quantities , except the excitation voltages, will be ignored.

The important parameters of equation {1} for the fundamental mode of the AT cut of the quartz like materials were computed [11] from the published constants corrected for berlinite and GaPO4 by experimental results as described in [11]. They are given in table I. For the gallium phosphate the coefficients are to be considered as tentative values. Much probably a better approximation for these coefficients can be obtain from the constants recently measured by Krempl et al. [12].

Table I: Energy trapping constants for the AT cuts of the quartz like materials.(M_n, P_n=C_{55} and $\overline{C}$ are in 10^{10} Nm^{-2})

Coefficients of the equation

	M_1	P_1	$\overline{C}$	$\sqrt{M_1/P_1}$
Quartz	11.008	6.922	2.921	1.261
AlPO4	8.718	5.084	2.285	1.309
GaPO4	7.663	3.805	2.310	1.419

EIGEN SHORT CIRCUIT MODES.

The eigen modes (at V=0) of the structure are obtained by the simultaneous solution of the equations for the electroded and unelectroded regions of the structure, taking into account the continuity conditions at the electrode edge.

At V=0 the homogeneous form of the approximate equation for the three regions reduces to anisotropic Helmoltz equations:

$$M_n u^n_{1,11} + P_n u^n_{1,33} + 4\pi^2\rho(f^2 - f^{*2})u^n_1 = 0 \qquad \{2\}$$

with :

$f^* = f_{ce1} = \frac{n}{4h}\sqrt{\frac{c^{(1)}_1}{\rho}}$ for region I. Or:

$f^* = f_{ce2} = \frac{n}{4h}\sqrt{\frac{c^{(1)}_2}{\rho}}$ for region II. Or:

$f^* = f_{cl} = \frac{n}{4h}\sqrt{\frac{\overline{c}^{(1)}}{\rho}}$ for the unelectroded region III.

Using the coordinate transformations :

$$X_1 = \left(\frac{C_{66}}{M_n}\right)^{1/2}.\gamma.x_1; \qquad X_3 = \left(\frac{C_{66}}{P_n}\right)^{1/2}.\gamma.x_3$$

$$X_1 = r\cos t; \qquad X_3 = r\sin t$$

with: $\gamma^4 = \frac{n^2\pi^2}{(h)^4}$

We obtain the isotropic Helmoltz equation:

$$u_{1,rr} + \frac{1}{r}u_{1,r} + \frac{1}{r^2}u_{1,tt} + Au_1 = 0 \qquad \{3\}$$

where:

$$A = \frac{n\pi}{4}\left(\frac{f^2 - f^{*2}}{f^2_{66}}\right) \qquad \{4\}$$

This partial derivative equation is separated into two ordinary differential equations by setting $u_1(r,t) = R(r).T(t)$:

$$r^2R'' + rR' + R(r^2A - \upsilon^2) = 0 \qquad \{5\}$$

$$T'' + \upsilon^2 T = 0 \qquad \{6\}$$

The separation constant υ has to be an integer since the solution of the second equation has to have a periodicity of $2\upsilon\pi$. So that: υ=m.

The (symmetrical) solutions bounded at zero and at infinity are respectively for the different regions:

Region I:

$$u_1 = \sum_{m=0}^{\infty} A_m Be_m(r\sqrt{A})\cos mt \qquad \{7\}$$

Region II:

$$u_1 = \sum_{m=0}^{\infty}\{B'_m J_m(r\sqrt{A}) + B''_m Y_m(r\sqrt{A})\}\cos mt$$

$$= \sum_{i,m} B_i Bs_m(r\sqrt{A})\cos mt \qquad \{8\}$$

Region III:

$$u_1 = \sum_{m=0}^{\infty} C_m K_m(r\sqrt{A'})\cos mt \qquad \{9\}$$

In these relations m is even (for the symmetry). The antisymmetrical solutions with respect to x_1 contains sin(mt) also with m even, the antisymmetrical solutions with respect to x_3 contains cos(mt) with m odd.

J_m, Y_m, I_m, K_m are respectively the Bessel functions of first and second kind and the modified Bessel function of first and second kind, of order m. Be_m is either I_m or J_m depending on the sign of A (relation {4}). Bs_m is successively J_m and Y_m.

Boundary and continuity conditions :

The traction free conditions on the major free surfaces, normal to the thickness are automatically verified as a consequence of the method used to establish equation {1}. On the surfaces limiting the different regions of the structure we have to specify the continuity of u_1 and of its normal derivative.

In this paper we suppose that the amplitude of the vibration mode near the edges of the plate are negligible so that no boundary conditions at the plate edge are taken into account. It was previously shown that, in the case of properly designed devices, their consideration induces a negligible modification of the eigen modes and resonances frequencies. In coherence with this, the solutions bounded at infinity were chosen in region III {9}.

The continuity of u_1 and of its normal derivative on the surface limiting the regions can be approximated [13], by the expression of these conditions at only a discrete number of points (figure 4).

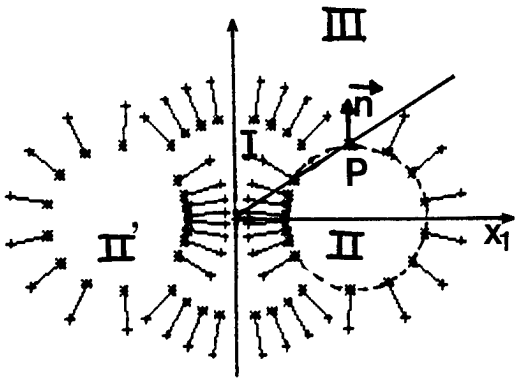


Figure 4: Discretization of the continuity conditions.

For example at a point $P(x_1^P x_2^P x_3^P)$ on the boundary between regions I and II, we have:

$$u_1^I(x_1^P x_2^P x_3^P) = u_1^{II}(x_1^P x_2^P x_3^P) \qquad \{10\}$$

$$n_1 \frac{\partial u_1^I(P)}{\partial x_1} + n_3 \frac{\partial u_1^I(P)}{\partial x_3} = n_1 \frac{\partial u_1^{II}(P)}{\partial x_1} + n_3 \frac{\partial u_1^{II}(P)}{\partial x_3} \qquad \{11\}$$

Many schemes can be used to express these conditions at the different region boundaries and to truncate the series, three were tried numerically, only one will be described here after for the case of a symmetrical structure:

We consider ngap points at the boundary between regions I and III, and nres points between regions I and II and also between regions II and III. This leads to 2ngap+ 4nres conditions; if we chose, as in figure 4, 2ngap=2nres=q, we obtain 3q linear relations between the coefficients A_0, A_2, ..., A_{2q-2},...; B_0, B_2, ..., B_{2q-2}, ...; C_0, C_2,..., C_{2q-2},. ; then we truncate the series solutions for the three regions to q terms.

These 3q relations in 3q unknown constitute an homogeneous linear system in the 3q=6npgap=6npres A B and C coefficients. The determinant of this system has to vanish to have a non trivial (0) solution in these coefficients. This condition constitutes a frequency equation that is solved numerically for the eigen frequencies. For each eigen frequency ($f_{n\mu}$ where n is the rank of the overtone and μ the rank of anharmonic), the A, B and C coefficients of the eigen modes are found as a function of one of them, as the solution of the homogeneous linear system.

On the whole, the eigen (symmetrical) modes for the three region of the structure are given by :

$$u_1(x_1, x_2, x_3) = \hat{u}_1^n = sin(n\pi x_2/2h)\bar{u}_1^n \qquad \{12\}$$

$$\bar{u}_1^{n\mu} = \sum_{i=0}^{i=q-1} D_{2i}^{n\mu} Be'_{2i}\left(r.\sqrt{A_e^{n\mu}}\right) cos(2it) \qquad \{13\}$$

The corresponding potential is:

$$\Phi = \hat{\Phi}^{n\mu} = \left(\frac{e_{26}}{c_{22}}\right)\bar{u}_1^{n\mu}\left\{\sin\left(\frac{n\pi x_2}{2h}\right) - (-1)^{(n-1)/2}\left(\frac{x_2}{h}\right)\right\}$$

Where the D coefficients are either A, B, or C according to the considered region.

The real coefficients of r in the argument of the relevant Bessel or modified Bessel functions (Be') are now given either by the square root of:

$$A^{n\mu} = [n\tfrac{\pi}{4}(f_{n\mu}^2 - f^{*2})/f_{66}^2]$$

or $A'^{n\mu} = -A^{n\mu} = \left[n\tfrac{\pi}{4}(f^{*2} - f_{n\mu}^2)/f_{66}^2\right]$ depending on the sign of A.

The eigen solution must verify the orthogonality relation:

$$\int\int\int_{V_t} \hat{u}_1^{n\mu}.\hat{u}_1^{k\lambda} dV = \delta_{nk}\delta_{\mu\lambda}N^{(n)(\mu)}$$

where V_t is the total volume of the structure and the displacement, is as in the following, defined by piece on the different regions.

COMPUTATION OF THE FORCED MODES:

The modes forced by the voltage $V_{rs}e^{j\omega t}$ applied between the electrodes r and s can be obtained by a linear combination of the short circuit eigen modes. For the "ordinary" potential a supplementary term must be added [14], which results of the definition of the transformed potential.

$$\hat{u}_1 = \sum_n \sum_\mu H_{rs}^{n\mu}\bar{u}_1^{n\mu}\sin(n\pi x_2/2h) \qquad \{14\}$$

The "usual" forced displacement is:

$$u_1 = \hat{u}_1 - \frac{e_{26}x_2V_{rs}e^{j\omega t}}{2hc^{(1)}}$$

While the "usual" potential is:

$$\Phi = \hat{\Phi} + \Phi_0 = \sum_n \sum_\mu H_{rs}^{n\mu}\hat{\phi}^{n\mu} + \frac{x_2V_{rs}e^{j\omega t}}{2h} \qquad \{15\}$$

The term $\frac{x_2V_{rs}e^{j\omega t}}{2h}$ is the "electrostatic" potential (no displacement), between the electrode pair (r,s); by the same convention as for the forcing term of equation {1} it has to be considered equal to 0 outside of an unshorted electrode pair.

Here again we have supposed that only one forcing voltage exists between the electrodes r and s of a (symmetrical) half structure. If several forcing voltages are applied between several electrodes this "static" potential should be replaced by the appropriate "electrostatic" solution which is a function of the different forcing voltages applied on the different electrode pairs.

However, in a first approximation compatible with the present assumptions (essentially thickness modes of thin plates), it is possible to consider a piecewise definition where the "electrostatic" potential is $\frac{x_2V_{rs}e^{j\omega t}}{2h}$ between unshorted electrodes (rs) and zero elsewhere. In this case the H coefficients will contain a summation over the electrode pairs and exciting voltages. But, since only one term of such sums will be retained in the partial derivation made to compute the admittance, we will omit it here after.

Accounting for the fact that the orthogonality of sin $(n\pi x_2/2h)$ has already been expressed in the forms of equation {1} that are relevant for the three regions; we substitute in these equations the lateral dependence of $\hat{u}_1$. For regions I which is either loaded with an insulating film or using a short-circuited conductive films and for region III, $\hat{u}_1$ is not different of u_1. For region (II) which has electrodes on which a potential is imposed, equation {1} contains a forcing term. After this substitution, we multiply these relations by $\bar{u}_1^{k\lambda}$ and integrate respectively on the volume corresponding to the three regions.

$$\sum_n \sum_\mu H_{rs}^{n\mu}(f^2 - f_{n\mu}^2) \iiint_{V_t} \bar{u}_1^{n\mu} \bar{u}_1^{k\lambda} dV$$

$$= \sum_n (-1)^{(n-1)/2} \frac{f^2 e_{26} 4 V_{rs} e^{j\omega t}}{C^{(1)} n^2 \pi^2} \iiint_{V_{e_{rs}}} \bar{u}_1^{k\lambda} dV$$

Using the transformed forms of the orthogonality relation relative to the lateral dependence of the displacement:

$$\iiint_{V_t} \bar{u}_1^{n\mu} \bar{u}_1^{k\lambda} dV = 2h \iint_{S_t} \bar{u}_1^{n\mu} \bar{u}_1^{k\lambda} dS = 2\delta_{nk}\delta_{\mu\lambda} N^{(n)(\mu)}$$

We obtain:

$$H_{rs}^{n\mu} = (-1)^{(n-1)/2} \left\{ \frac{f^2}{(f^2 - f_{n\mu}^2)} \right\} \left\{ \frac{4 V_{rs} e^{j\omega t} e_{26}}{n^2 \pi^2 C_{66}} \right\} \left\{ \frac{\iint_{Se_{rs}} \bar{u}_1^{n\mu} dS}{\iint_{S_t} (\bar{u}_1^{n\mu})^2} \right\} \qquad \{16\}$$

where S_{ers} = electroded surface (rs)

S_t = total surface

Equivalent scheme for 2 resonators integrated structures:

Electrical response :

Using the constitutive equation for D_2 and the definition of the admittance Y_{rs} between electrodes r and s:

$$D_2 = e_{26} u_{1,2} - \varepsilon_{22} \phi_{,2}$$

$$D_2^{rs} = \frac{\varepsilon_{22}}{2h} V_{rs} e^{j\omega t} (k_{26}^2 + 1) + \sum_n \sum_\mu (-1)^{(n-1)/2} H_{rs}^{n\mu} \bar{u}_1^{n\mu} \frac{e_{26}}{h}$$

$$Y_{rs} = \frac{\partial I_{rs}}{\partial \Phi_{rs}};$$

Expressing the current I_{rs} from the time derivative of the charge and hence from the normal component of D (D_2) we obtain:

$$Y_{rs} = \frac{-\partial}{\partial (V_{rs} e^{j\omega t})} \iint_{Se_{rs}} D_2^{rs} dS$$

So that the expression of the admittance Y_{rs} is:

$$Y_{rs} = j\omega \left[\frac{S_{ers} \cdot \varepsilon_{22}}{2h} (k_{26}^2 + 1) - \sum_n \sum_\mu \frac{f^2 4 k_{26}^2 \varepsilon_{22} \left(\iint_{S_{ers}} \bar{u}_1^{n\mu} dS \right)^2}{(f^2 - f_{n\mu}^2) 2 n^2 \pi^2 N^{(n)(\mu)}} \right]$$

or:

$$Y_{rs} = j\omega \left\{ \frac{S_{e_{rs}} \cdot \varepsilon_{22}}{2h} (k_{26}^2 + 1) - \sum_n \sum_\mu \left(\frac{8 k_{26}^2 \varepsilon_{22} \left(\iint_{S_{e_{rs}}} \bar{u}_1^{n\mu} dS \right)^2}{2hn^2\pi^2 \iint_{St} (\bar{u}_1^{n\mu})^2 dS} \right) \right\}$$

$$+ j\omega \left\{ \sum_n \sum_\mu \left(\frac{f_{n\mu}^2 8 k_{26}^2 \varepsilon_{22} \left(\iint_{S_{e_{rs}}} \bar{u}_1^{n\mu} dS \right)^2}{(f_{n\mu}^2 - f^2) 2hn^2\pi^2 \iint_{St} (\bar{u}_1^{n\mu})^2 dS} \right) \right\}$$

Identifying with the admittance of the equivalent scheme including in parallel, one serial arm containing the "static" capacitance and as many LC serial (dynamically) arms as there is modes), we obtain :

The "static" capacitance as:

$$C_{0_{rs}} = C_{Z_{rs}} - \sum_n \sum_\mu C_{rs}^{n\mu} \qquad \{17\}$$

where $C_{Zrs} = \frac{S_{ers} \varepsilon_{22}^S}{2h} (k_{26}^2 + 1)$ is the capacitance at f = 0

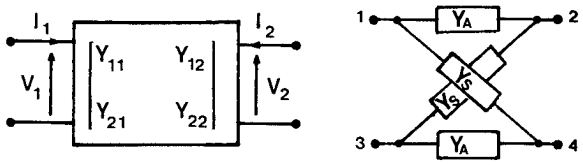
The motional capacitance of the LC serial arm corresponding to the forced mode nμ is:

$$C_{rs}^{n\mu} = \left[\frac{8 k_{26}^2 \varepsilon_{22}}{n^2 \pi^2 2h} \right] \frac{\left(\iint_{Se_{rs}} \bar{u}_1^{n\mu} dS \right)^2}{\iint_{St} (\bar{u}_1^{n\mu})^2 dS} \qquad \{18\}$$

The inductance of the forced mode $n\mu$ is then:

$$L_{rs}^{n\mu} = \frac{1.}{(C_{rs}^{n\mu} 4\pi^2 f_{n\mu}^2)}$$

For symmetrical structures:



$$Y = \frac{1}{2} \begin{vmatrix} Y_A + Y_S & Y_S - Y_A \\ Y_S - Y_A & Y_S + Y_A \end{vmatrix}$$

Figure 5: Symmetrical Equivalent Lattice

Only the symmetrical and the antisymmetrical parts of Y_{13} (Ya and Ys) are required (Bartlett theorem):

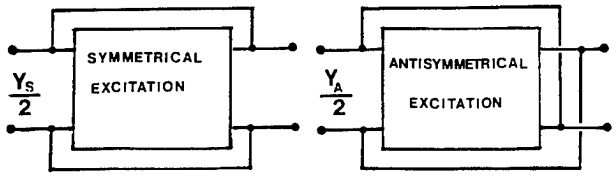


Figure 6: Symmetrical and antisymmetrical excitation.

$$Y_{13} = j\omega \left[\frac{S_e \cdot \varepsilon_{22}}{2h} (k_{26}^2 + 1) - \sum_n \sum_\mu C_{rs}^{n\mu} \right] + j\omega \left[\sum_n \sum_\mu \left\{ \frac{f_{n\mu}}{(f_{n\mu}^2 - f^2)} \right\} C_{rs}^{n\mu} \right]$$

With summation over the symmetrical modes (then Y_{13}=1/2.Y_s) or the antisymmetrical modes (with respect to the axis of the coupling region) (then Y_{13}=1/2.Y_a).

As in the general case for Y_{rs}, Y_{13} and hence Y_s and Y_a can be represented by the scheme of figure 7 where the capacitance and inductance have the same expression as for Y_{rs} but with the consideration of either the symmetrical or the antisymmetrical modes. We obtain thus a total equivalent scheme to the 2 poles structure under the form of a symmetrical lattice already indicated by R.C.Smythe [15].

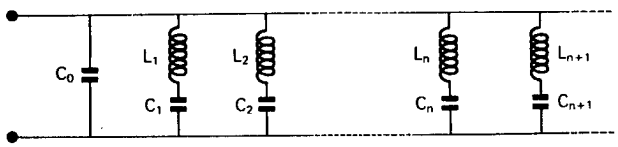


Figure 7: equivalent scheme of one arm of the lattice equivalent to the 2 pôle.

Discussion:

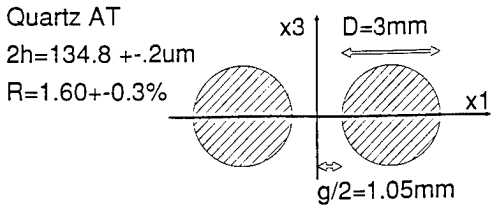
To obtain a good insulation between two resonators put on the same plate:

In the symmetrical case: Ys-Ya has to be very small, so that the symmetrical and antisymmetrical [half] modes have to be as similar as possible.

In the non symmetrical case an advantage can be taken of the possibility to have different resonance frequencies for the 2 resonators.

Comparison of experimental and computed results:

A computed program was implemented. The experimental results were compared to computed ones for different 2 pôles structures having in most cases a weak coupling between resonators. A quite good agreement was observed in nearly all cases. A typical example is given in figure 8 where the parameters employed are displayed. The rate of convergence of the resonance frequency of the fundamental symmetrical mode with the increase of the number of points appears to be quite slower than that was observed for the single resonators. A detailed examination of the numerical coefficients of the truncated series for the different regions has indicated that the rate of convergences of these series is very different between regions, so that we can expect that a different choice of the schemes used for the discretization and truncation of the series, will permit to increase greatly the accuracy, particularly in giving a better representation of region II in which most of the energy of the device is concentrated.



	f.sym.(MHz)
experimental	12.258
computed 4x4pts	12.138
computed 4x6pts	12.277
computed 4x8pts	12.306

Figure 8: Comparison of experimental and computed resonance frequencies in a typical case.

Conclusion

The principle of a method of modelization of the two poles integrated filters was described. Several computer implementations of this model were written and were applied to different two poles filters, the agreement between experimental and computed results is raisonably good. Further work is needed to increase the numerical accuracy and to obtain better convergence properties for the solutions. It is expected that improved schemes of discretization of the continuity conditions will lead to an as good accuracy of the computed resonance frequencies and modes as for the single resonators.

Acknowledgements: The authors acknowledge the assistance of MM. Caumont and LeThuault for the preparation of the many samples used in the experimental part of this study.

References:

[1] J.P.Michel, M.Zafany,T.Soubirous, Y.Billon, D.Mouton, M.Berthe, L.Schwartz, S.Lechopier,J.L.Le Corre Proc. 7th European Frequency and Time Forum (1993)).

[2] H.F.Tiersten J.Acoust.Soc. Amer. vol.62 no 6 p.1424-1430 (1977).

[3] H.Seikimoto Eectron. Communication in Japan vol. 65A,n°12 p10-19 (1982).

[4] L.N.Dworsky Proc. 40th Annual Frequency Control Symposium p.161,167 (1986).

[5] H. Carru, R.Lefèvre, J.P.Aubry, S. LeChopier. Proc.42nd Annual Frequency Control Symp. p.101 (1988).

[6] H.F. Tiersten J.Acoust Soc.Am. vol.59 p.879 (1976)

[7] D.S Stevens, H.F.Tiersten J.Acoust.Soc. Amer.vol.79 n°6 p.1455 (1986)

[8] R.C. Peach IEEE Trans. Ultrason. Ferroel. Freq. Control vol.35 no 5 p.593-611 (1988)

[9] J.Détaint, J.Schwartzel, C.Joly, E.Philippot Proc. 41st Annual Frequency Control Symposium p.314-324 (1987)

[10] J.Detaint, H.Carru, J.Schwartzel, C.Joly, E.Philippot, Proc. 5th.European Frequency and Time forum p.30-40 (1991)

[11] J.Detaint, J.Schwartzel, C.Joly, B.Capelle, A.Zarka, Y.Zheng, Y.Toudic, E.Philippot Proc. 1992 IEEE Frequency Control Symposium p.639-647 (1992)

[12] W.Wallnöffer, J.Stadler, P.Krempl. Proc. 7th European Frequency and Time Forum (1993).

[13] K.Nakamura, M.Shimizu Proc.1976 IEEE Ultrasonics Symposium p.606 (1976)

[14] R. Holland, E.P.Eernisse Design of Resonant Piezoelectric Devices MIT-Press (1969).

[15] R.C.Smythe in Precision Frequency Control E.A Gerber, A.Ballato editors p.202 vol.1 Academic Press (1985).

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

OBSERVATIONS OF THE CRYSTAL PERFECTION AND OF SURFACE ACOUSTIC WAVES IN LITHIUM NIOBATE

A.ZARKA*, B.CAPELLE*, J.DETAINT**, J.SCHWARTZEL** & J.M.HODE***

* L.M.C.P. Université P. et M. CURIE. 4 Place Jussieu. 75252. PARIS CEDEX 05. FRANCE.
** C.N.E.T. Dept. PAB/BAG/MCT/CMM. 196 AV. H. Ravera 92220 BAGNEUX. FRANCE.
***THOMSON- SINTRA. 399 Route des Crêtes.B.P.157.06903. VALBONNE CEDEX. FRANCE.

ABSTRACT

The extended defects in Lithium Niobate ($LiNbO_3$) have been studied by X-ray topography using the white radiation delivered by the D.C.I. synchrotron at L.U.R.E. (Orsay, France). Very few low angle subgrains were observed in slices cut perpendicularly to the X, Y and Y-45° direction.The large intensity of the synchrotron source has permitted an experimental study of the acoustic waves (S.A.W.) in different filters. Furthermore, using the time structure of the radiation of the synchrotron and a synchronous excitation, stroboscopic observations were made.So, it was possible to investigate the interactions of the S.A.W. with the crystal defects, plate edge and an acoustic absorbent. The observation of the displacement fields in high performance filters have revealed interesting properties of apodized transducers and M.S.C..This study shows that the stroboscopic topography is a good technique to observe and analyse the wave fields in surface waves devices.

1. INTRODUCTION

The S.A.W. devices are important components in modern analog signal processing particularly for filtering in television and in radiocommunication equipments. The rapidly growing interest for this latter application is an important factor of the new developments presently made to increase their performances. Among these developments several are relative to the obtention of large wafers of high crystalline perfection.

The first single crystals of significant size were synthetized at the Bell Laboratories [1]. Like many other oxygen-octahedral ferroelectrics, the Lithium Niobate is obtained by the Czochralsky growth method at a very high temperature (1260°C). It can present a quite large non stoechiometry with an usual Li / Li + Nb ratio close to O.47. The crystal structure was determined by several workers [2-3] but very few studies were concerned with the crystal perfection.

Other developments regard the research of S.A.W. device structures which can lead to the better characteristics (reduction of the insertion losses, miniaturization and so on [4]).
Among several piezoelectric materials which can be used for these applications, the Lithium Niobate has the larger coupling coefficients and hence permits to obtain filters with the largest bandwidths and other signal processing devices with a great efficiency.

Due to the extreme complexity of the three dimensional theory which describe enough accuratly the behaviour of the real S.A.W. devices, experiments are needed to understand their characteristics.On the other hand, the influence of the quality of the material on the properties of these devices is not until now very well understood.
The use of synchrotron radiation permits investigations of the quality of this material by X-ray topography. Such research is extremely difficult (or even impossible) with conventional sources, due to the very high X-ray absorption of these crystals.

In this study, we have characterized the crystalline defects of $LiNbO_3$ crystals by X-ray topography using the synchrotron radiation of the D.C.I. storage ring at the L.U.R.E. (Orsay, France). The intensity of the time structure of the X-ray radiation of the synchrotron have permitted precise experimental study of surface acoustic waves.
In this paper, we present different results concerning the perfection of the presently available materials and the results of experiments made to observe the energy distribution of acoustic waves in different devices.

2. QUALITY OF THE CRYSTALS

2.1 Experimental settings

For crystals or filters too thick to be examined in a transmission Laue setting [5], reflection setting [5] with an oblique incidence was used.
In this case the film was adjusted parallel to the sample at a distance (10-20 cm) choosen to obtain in each case a good compromise between the resolution and the separation of the diffraction spots. For the thinner plates the conventional transmission Laue setting was used.

2.2 Experimental results concerning the crystals

In the following figures we represent typical examples of the different observations made.

In figure 1 is presented an X-ray topograph of a Y-cut slice in a crystal grown in the Z direction. This topograph was obtained with the reflection setting and it may be observed that the quality of this crystal is not so good because many subgrains presenting large desorientations are present. Such defects interact strongly with the acoustic waves due to the fact that the acoustic velocity differs from subgrain to subgrain.

0-7803-0905-7/93 $3.00 © 1993 IEEE



FIGURE 1

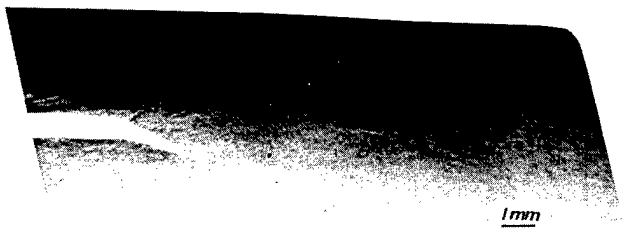


FIGURE 2



FIGURE 3

In Figure 2 is shown another reflection topograph of a Y-45° slice. This topograph as obtained with a diffraction vector having direction close to the normal to the surface. It may be observed a lattice of dislocations and a black and white contrast due to a subgrain boundary. The contrast corresponding to the subgrain boundary is a white line if the relative orientations of the two adjacent subgrains are such that the corresponding diffracted beam are divergent, which gives a white region with no image. In the case of relative orientations giving a convergence of the diffracted beam a darker contrast is obtained owing to the overlapping of the images. This effect can be observed in Figure 2. It may be observed also that the dislocation density is quite large.

Another example is given in Figure 3. The quality of this crystal is enough good and this slice shows two kinds of defects: Straight lines dislocations and surface defects which are due to the imperfect polishing.

Figures 4a and 4b show an example of a slice studied by transmission. It corresponds to an Y+36° cut plate of about 100µm thick. Two types of defects are visible: a subgrain boundary and straight dislocations. In Figure 4b the plate is vibrating in a overtone thickness stretch mode (bulk waves). The central dark part corresponds to the vibrating zone. The subgrain boundary presents another configuration because the corresponding diffraction vector is different than the previous one (Figure 4a).

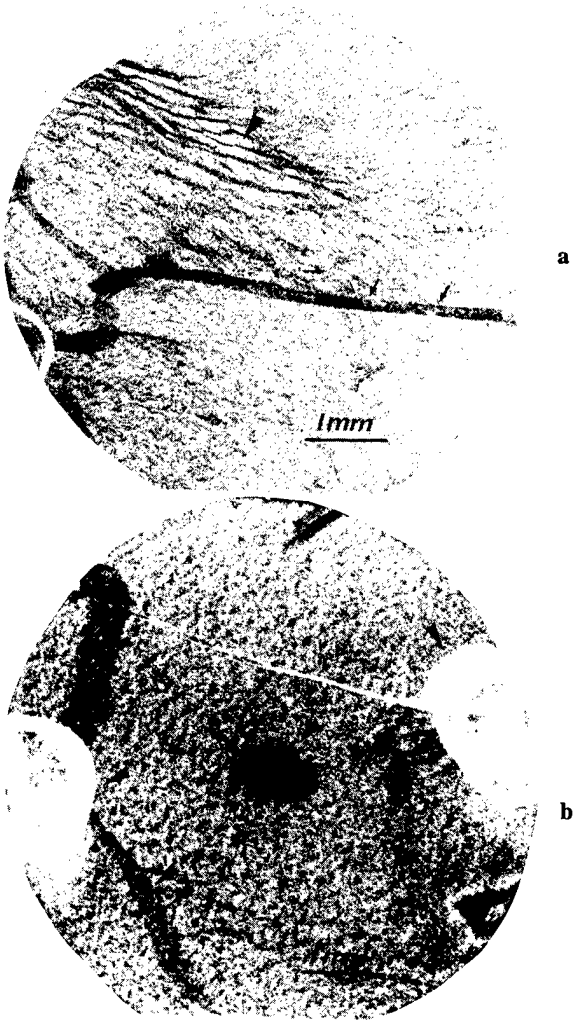


FIGURE 4

In figure 5 is given an X-ray topograph of a X-cut slice. This topograph was obtained with the reflection setting. It may be observed that the quality of this crystal is good and that very few dislocations are present.

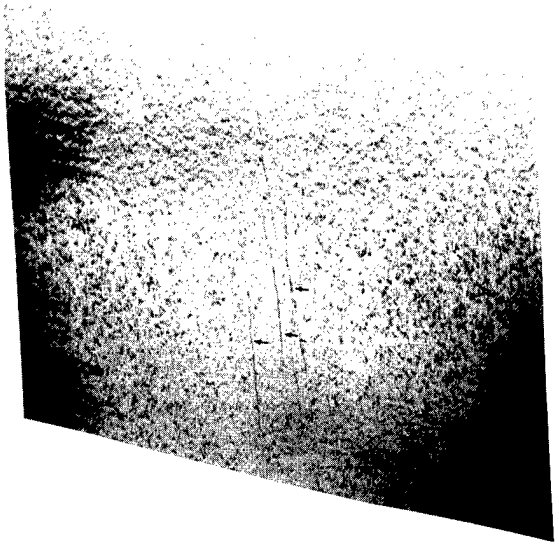


FIGURE 5



FIGURE 6

3. STUDIES OF S.A.W. FILTERS

In Figure 6 is shown an X-ray topograph obtained on a simple transducer. It may be observed in the central part the dark contrast due to the vibrating zone. Due to the setting which correspond to a classical one the obtained image is an integrated one and the S.A.W. cannot be individualized. Nevertheless, the contrast is interesting and it can be observed a slight diffraction of the waves in this zone.

In Figure 7 is shown a stroboscopic topograph of the waves generated by an un-apodized transducer. The wave fronts are straight lines and the waves have constant amplitude in the transversal direction except at the extremities of the aperture. In this plate there is a subgrain boundary which induces a phase shift between the waves which propagate on the two sides of the boundary.

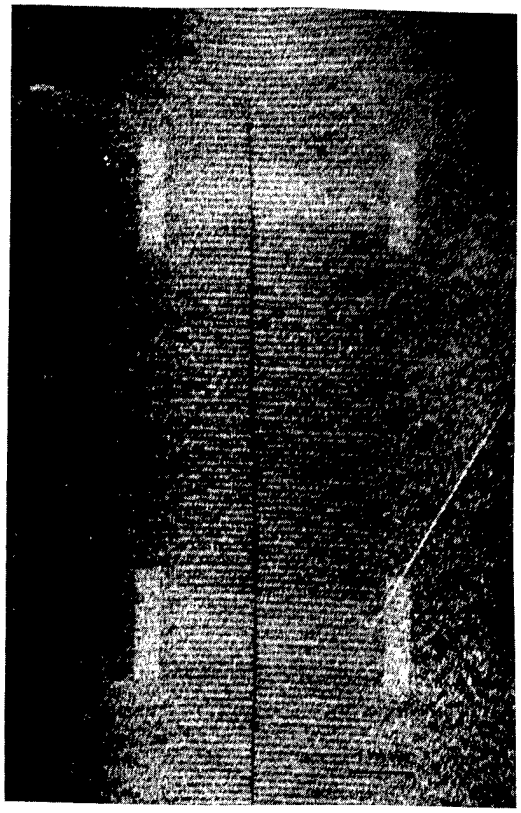


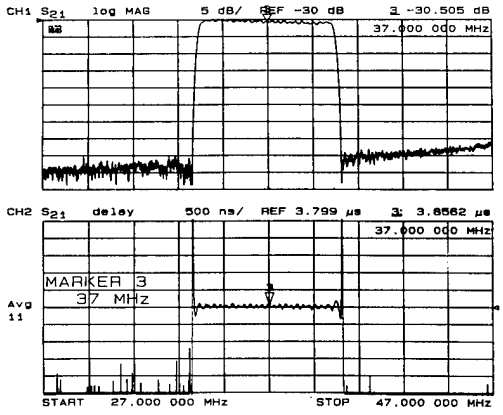
FIGURE 7

In Figure 8a is represented a drawing of a filter with apodized transducers and M.S.C.(Multi-Strip Couplers) [6-7] In Figure 8b it can be observed the response curve of the filters which have been studied by X-ray topography using the synchrotron radiation. In the upper part is shown the amplitude response and in the bottom the group delay.
The representation of the displacement of the Rayleigh waves is shown on the drawing of the Figure 9. The two principal components **Un** and **Up** of the displacement **U** are represented in the saggital plane. Naturally, they are time dependant and the properties of the corresponding Rayleigh waves are given in the following table 1.



(#)From M.Feldman J.Henaff
Masson Paris (1986)

FIGURE 8a



(•) Thomson-Sintra DASM

FIGURE 8 b

Displacement of Rayleigh Waves

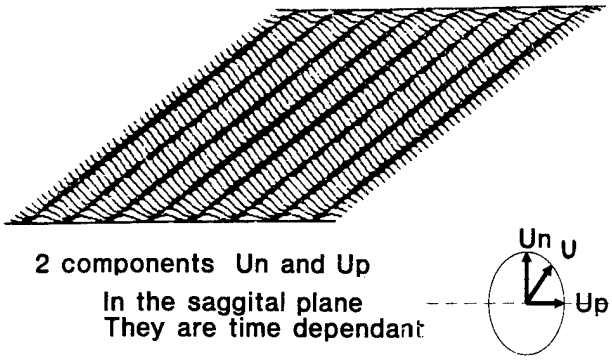


FIGURE 9

VELOCITY (m/s)	3491.1
COUPLING COEFFICIENT	4.86
TEMPERATURE COEFFICIENT	94 ppm
ANGLE OF POWER FLOW	0.0

TABLE 1

This type of filter have been studied by X-ray topography by reflection using the white beam of the D.C.I. synchrotron. This setting is necessary because the whole device constitute a very absorbing mounting (substrate). An example of such topograph taken among the different spots obtained in a Laue pattern is shown in Figure 10. In this figure, it may be observed the homogeneous contrast due to the transducers and the M.S.C.(Figure 10a). In Figure 10b, the filter is in a asynchronous excitation and as in the general case (reflection) only the component perpendicular to the surface is visualised. Due to the fact that the excitation is not synchronous, the contrast of the S.A.W. is integrated and the fringes cannot be resolved individually. Nevertheless, some other fringes due to the displacement of the energy in the transducers are visible. In Figure 10c, the X-ray topograph was taken in the same experimental conditions, but in contrary to the previous case the frequency of the filter was adjusted to be an exact multiple of the frequency of the X-ray beam. This permits to obtain a stroboscopic image [5] (34.862 MHz) where the S.A.W. are individually resolved. Some observations can be done from this topograph: it can be observed that the S.A.W. are not very well confined in the transducer and in other hand that the vibrating energy is also confined in the perpendicular large fringes already observed in the previous topography (Figure 10b). It may be also observed that the contrast of these last fringes is very different for the same level of excitation.

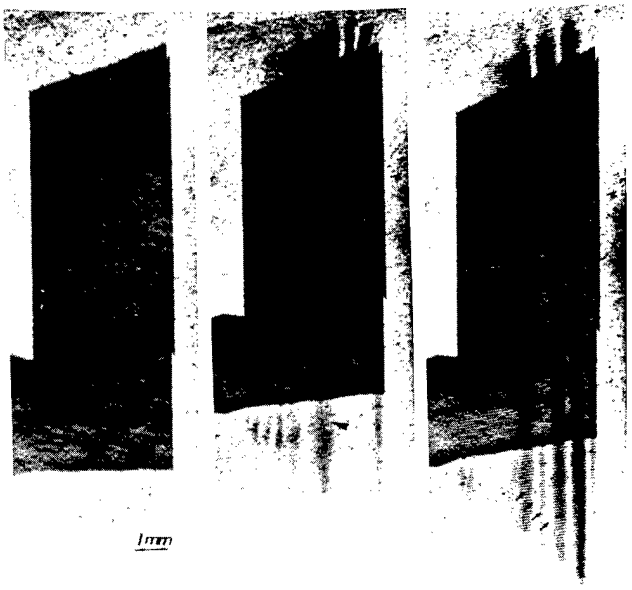


FIGURE 10

Influence of the frequency on the contrast.

The X- ray topographs presented in Figures 11a, 11b and 11c correspond to enlargement of the same zone of the filter. They have been taken with the same experimental conditions (same diffraction vector **g**, same θ_B Bragg angle, same incidence direction) Between these stroboscopic topographs, only the frequency has changed. They were taken respectively at 34.862 MHz (Figure 11a), 38.031 MHz (Figure 11b) and 41.200 MHz (Figure 11c). The principal features which can be observed are the fringes due to the S.A.W. and the very contrasted zones which reveal the flow of the energy in the transducer. So, the radiation diagram of the apodized transducer varies with the frequency.

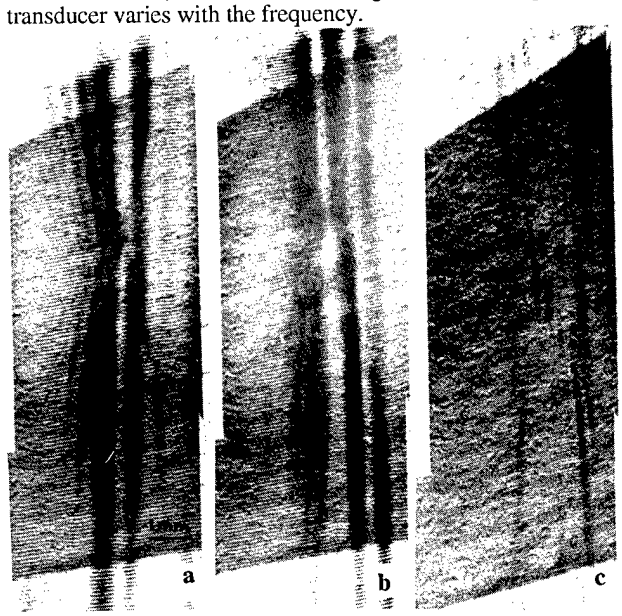


FIGURE 11

Influence of the θ_B Bragg angle.

Variing the θ_B, the contrasts obtained on the topographs change.

In particular, when the θ_B varies, the λ (X-ray wavelenght) changes and the depth of penetration of the X-ray in the filter is more or less important.

An example is given in Figures 12a and 12b. Obviously the contrast depends on the amplitude of the waves inside the filter and is less important when the angle θ is greater (Figure 12a).

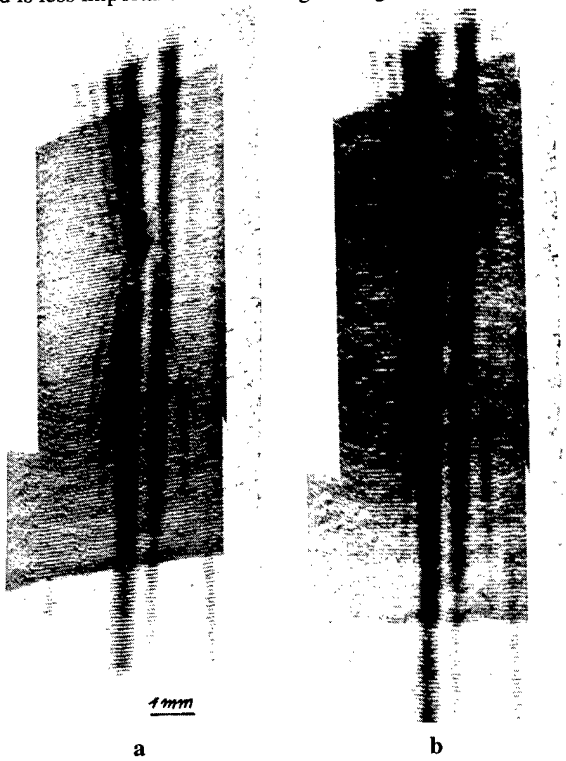


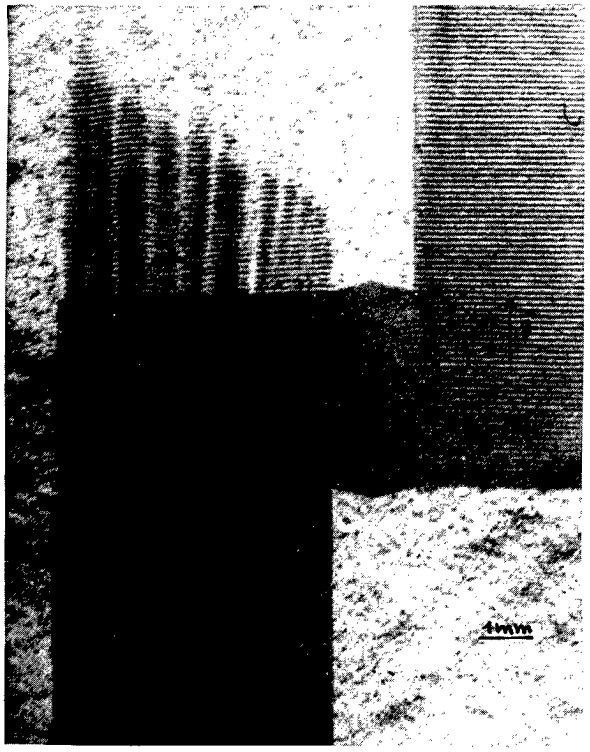
FIGURE 12

Influence of the g diffraction vector.

Depending on the direction of the **g** diffraction vector, different components of the displacement **U** are revealed. In Figure 13a the diffraction vector is parallel to the larger dimension of the transducer. In this case, it can be observed the wave front curvatures. In Figure 13b the visualisation of the modulation is obtained when the **g** diffraction vector is perpendicular. For the intermediary case the different components of the displacement are visible (Figure 13c,where **g** is at 45°).

Influence of the defects.

In Figure 14a is presented a filter where two subgrains exist in the crystal. In the central part, an acoustic absorbent has be deposited. The role of this absorbent is to attenuate the losses by reducing the vibrating amplitude as it can be seen on the topograph. In contrary the two zones with the subgrain boundaries (enlarged in Figures 14b and 14c) show that the S.A.W. present phase shifts due to these defects.



a



b c

FIGURE 13

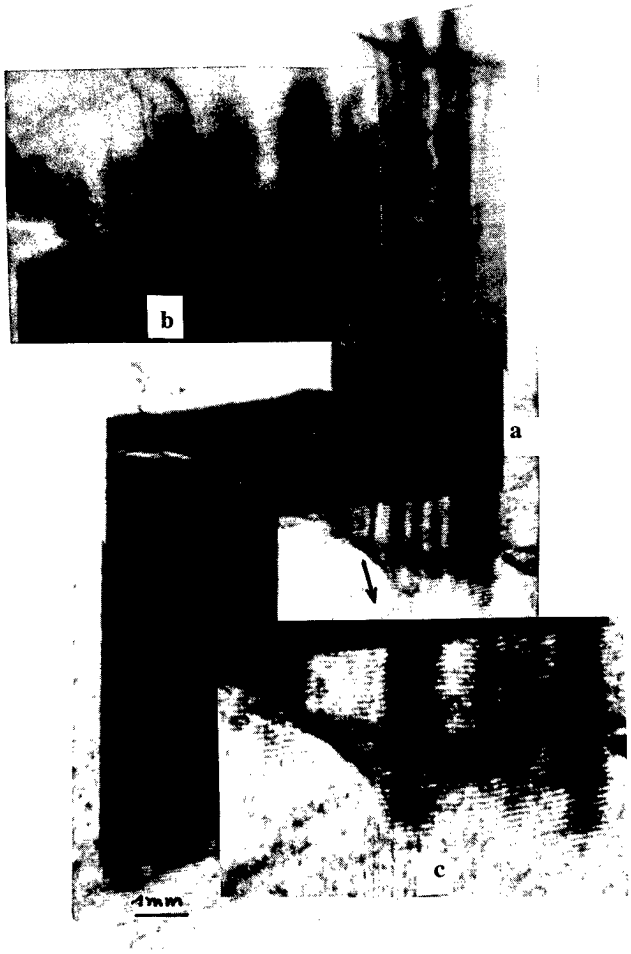


FIGURE 14

4. CONCLUSION

The large intensity of the synchrotron source has permitted an experimental study of Lithium Niobate crystals and of the surface acoustic waves (S.A.W.) in different filters by X-ray topography.
The studies of the apodized transducers by topography present different complicated contrasts which vary with the frequency. Generally these contrasts correspond to a modulation of the wave fronts and the intensity. It was possible to investigate the quality and the interactions of the S.A.W. with the crystal defects, plate edges, and an acoustic absorbent. The amplitude distribution in apodized transducers and in multistrip couplers (M.S.C.) were also observed.

This study shows that the crystal quality of the lithium niobate has been greatly enhanced in the recent years in comparison with those produced some time ago

It was also shown that the stroboscopic topography is a good technique to observe and analyse the wave fields in surface waves devices. It is possible to obtain quantitative informations from densitometry and image processing.

These techniques can be used for devices using others materials.

5. REFERENCES

[1] BALLMAN. A.A. J. Am. Ceram. Soc. (1965) **48**. 112-114

[2] NASSAU. K., LEVINSTEIN. H.J. and LOIACONO. G.M. J. Phys. Chem. Sol. (1966). **27**. 983-996.

[3] ABRAHAMS.S.C., REDDY.J.M. and BERNSTEIN. A. J. Phys. Chem. Sol. (1966). **27**. 997-1012.

[4] HARTEMANN. P. and DIEULESAINT E. Electronic Letters (1969) **5**; 25. 657-658.

[5] CAPELLE. B., DETAINT. J., SCHWARTZEL. J., ZHENG. Y. and ZARKA. A. Proc. IEEE Frequency Control Symposium (1992). 553-560.

[6] FELDMAN. M. and HENAFF. J. Traitement du signal par ondes élastiques de surface. Ed. Masson. (1986).

[7] MAERFELD. C. "Multistrip Couplers". Wave Electronic 2. (1976). 82-110.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DELTA FUNCTION MODEL ANALYSIS OF SSBW SPURIOUS RESPONSE IN SAW DEVICES

Ken-ya Hashimoto and Masatsune Yamaguchi

Department of Electrical and Electronics Engineering, Chiba University
Inage-ku, Chiba-shi 263, Japan

Abstract

This paper describes delta function model analysis of surface skimming bulk wave (SSBW) response by newly defining two physical parameters, K_B^2 and X_c, characterising the SSBW excitation and propagation. Here, K_B^2 corresponds to the electromechanical coupling factor of surface acoustic waves (SAWs) and gives the SSBW excitation strength on the substrate surface. X_c is the critical propagation length representing how SSBW amplitudes decay with propagation. The analysis was applied to the characterisation of an SAW TV-IF filter fabricated on an X-112°Y-$LiTaO_3$ substrate. The result obtained by the analysis was in good agreement with experiments, and showed most clearly how the SSBW response affects the filter performance. The values of K_B^2 and X_c were determined for an SiO_2-film/36°YX-$LiTaO_3$ substrate which may be used for SAW devices, and its SSBW spurious response level was estimated. It was theoretically predicted and experimentally confirmed that the SSBW excitation strength and accordingly the SSBW spurious response is considerably reduced by properly choosing the thickness of the SiO_2-film.

1 Introduction

A large amount of work has been done to develop new substrate materials and/or to find their optimal crystal cuts and propagation directions suitable for surface acoustic wave (SAW) devices, which substantially determine device performance.

Electrical performance and temperature stability of SAW devices are basically characterised by the electromechanical coupling factor and temperature coefficient of delay of substrates. In fact, SAW device substrates have mainly been assessed by these two parameters which can easily be calculated for specified crystal cuts and propagation directions.

Spureious response due to surface skimming bulk waves (SSBWs)[1] should also be minimised in SAW devices. Since SSBWs are not suppressed by conventional saw-cuts fabricated on a bottom surface of the substrate, it is most desirable if one can easily find substrates where SSBWs are not or very weakly launched and detected by interdigital transducers (IDTs). Although several methods have been proposed for the analysis of the excitation and propagation of bulk acoustic waves[2]-[5], it does not seem that they have extensively been applied to search for new substrates with minimised SSBW spurious response.

The authors have previously reported that in piezoelectric materials, acoustic wave excitation including both SAWs and SSBWs is easily estimated by use of the effective permittivity[5][6].

Based upon the effective permittivity, this paper proposes delta function model analysis of SSBW response by newly defining two physical parameters, K_B^2 and X_c, characterising the SSBW excitation and propagation. Here, K_B^2 corresponds to the electromechanical coupling factor of SAWs and gives the SSBW excitation strength on the substrate surface. X_c is the critical propagation length representing how SSBW amplitudes decay with propagation.

First, it was theoretically shown that by determining K_B^2 and X_c for specific crystal cuts and propagation directions, the delta function model analysis[7] is applicable to the estimation of SSBW response. The values of K_B^2 and X_c are also effectively used to make a rough estimate of SSBW spurious response level in SAW devices.

Secondly, the delta function model analysis of SSBW response was applied to the characterisation of an SAW TV-IF filter fabricated on an X-112°Y-$LiTaO_3$ (112-LT) substrate. The result obtained by the analysis was in good agreement with experiments, and showed most clearly how the SSBW response affects the filter performance. The values of K_B^2 and X_c were also determined for an SiO_2-film/36°YX-$LiTaO_3$

0-7803-0905-7/93 $3.00 © 1993 IEEE

(36-LT) substrate which may be used for SAW devices, and its SSBW spurious response level was estimated. It was theoretically predicted and experimentally confirmed that the SSBW excitation strength and accordingly the SSBW spurious response is considerably reduced by properly choosing the thickness of the SiO_2-film.

2 Delta Function Model Analysis of SSBW Response

This section describes the estimation of acoustic wave excitation using the effective permittivity[5] and its application to delta function model analysis of SSBW response. The mathematical treatments are fully described in [6].

2.1 Acoustic Wave Excitation

Figure 1 shows the space coordinate system used in the analysis, where an IDT of infinite length in the Y-direction is placed on the surface of a semi-infinite substrate, and plane acoustic waves propagate in the X-direction. Unless specified, the substrate surface is assumed to be electrically free (unmetallised) in the following analysis.

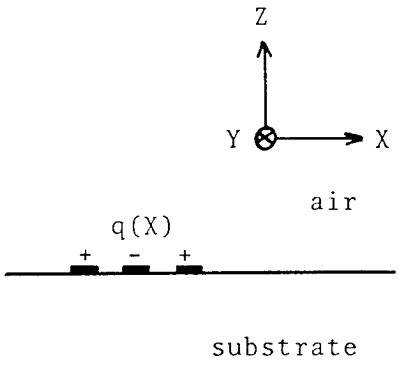


Figure 1: Coordinate system used in the analysis

Provided that the effects of IDT electrode resistance and mass-loading are negligible, electrical surface potential $\phi(X)$ is related with surface charge density $q(X)$ by the Green function $G(X)$,

$$\phi(X) = \int_{-\infty}^{+\infty} G(X-x)q(x)dx, \qquad (1)$$

where,

$$G(X) = \int_{-\infty}^{+\infty} \frac{\exp(-jS\omega X)}{|S|\epsilon(S)} dS. \qquad (2)$$

In eq.(2), S is the inverse velocity in the X-direction and ω is the angular frequency.

The effective permittivity $\epsilon(S)$ in eq.(2) is determined only by elastic, piezoelectric and dielectric constants for specified crystal cuts and propagation directions, and is simply given by[6],

$$\epsilon(S) = -D_f(S)/D_m(S). \qquad (3)$$

Here $D_f(S)$ and $D_m(S)$ are the determinants describing the boundary conditions for free and metallised surfaces, respectively[8]. So $\epsilon(S)$ is readily calculated by slightly modifying SAW velocity calculation programs.

Figure 2 shows $\epsilon(\infty)/\epsilon(S)$ for 112-LT, where $V = S^{-1}$. The pole appearing at $V = S^{-1}$=3,282m/s represents the SAW excitation on a free surface.

Applying the residue theorem to this pole, one can obtain the Green function $G_S(X)$ contributing to SAW radiation as,

$$G_S(X) = \frac{jK_{Sf}^2}{2\epsilon(\infty)} \exp(-jS_f\omega|X|). \qquad (4)$$

In eq.(4), K_{sf}^2 is the electromechanical coupling factor of SAWs on a free surface and given by,

$$K_{Sf}^2 = -2/\{S_f\epsilon(\infty)\frac{\partial\epsilon(S)^{-1}}{\partial S}|_{S=S_f}\}. \qquad (5)$$

The surface potential associated with radiated SAWs is given by substituting $G_S(X)$ into eq.(1).

$G_S(X)$ on a metallised surface is obtained by replacing K_{sf}^2 by the electromechanical coupling factor K_{sm}^2 on a metallised surface, i.e.,

$$K_{Sm}^2 = 2/\{S_f\epsilon(\infty)^{-1}\frac{\partial\epsilon(S)}{\partial S}|_{S=S_m}\}. \qquad (6)$$

where S_m is the inverse SAW velocity on a metallised surface.

Although K_{Sf}^2 and K_{Sm}^2 are almost identical with each other on such typical SAW substrates like 112-LT, 128°YX-$LiNbO_3$, etc., they are much different from each other on SSBW/leaky SAW (LSAW) substrates, for example, on 36-LT.

The Green function $G_B(X)$ contributing to SSBW radiation is also obtained using the similar procedure.

Assume that $\epsilon(S)$ about $S \cong S_B$ (S_B is the inverse SSBW velocity in the X-direction) can approximately be represented in the form of,

$$|S|\epsilon(S) \cong S_B\epsilon_B(\Omega+a)/(\Omega+b), \qquad (7)$$

where $\Omega = \sqrt{\{(S/S_B)^2 - 1\}}$, and a and b are the expansion coefficients. Note here that one does not

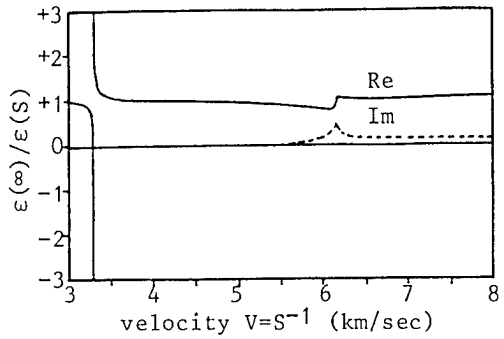


Figure 2: $\epsilon(\infty)/\epsilon(S)$ for 112-LT

have to specify in the analysis whether the relevant SSBWs are either quasi-longitudinal or quasi-shear.

Substituting eq.(7) into eq.(2) and applying the procedure by Yashiro and Goto[4], one may obtain $G_B(X)$ as,

$$G_B(X) \cong -\frac{K_B^2}{2\epsilon(\infty)} H(S_B\omega|X|)U(|X|/X_{cf}), \quad (8)$$

where

$$H(\theta) = \sqrt{(2/\pi\theta)}\exp(-j\theta - j\pi/4), \quad (9)$$

and

$$U(r) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty} \frac{\exp(-t^2)}{1-jrt^{-2}}dt \cong \begin{cases} -j/2r, & |r| \gg 1 \\ 1, & |r| \ll 1 \end{cases} \quad (10)$$

In eq.(8), $K_B^2 = (a-b)\epsilon(\infty)/\epsilon_B$ and $X_{cf} = \lambda_B/\pi a^2$ (λ_B is the SSBW wavelength).

K_B^2 together with the specified propagation length X and surface charge density $q(X)$ gives the radiated SSBW amplitude on the substrate surface, and corresponds to the electromechanical coupling factor of SAWs.

X_{cf} is the critical propagation length representing how the radiated SSBW amplitudes (on the substrate surface) attenuate with propagation. When $X \ll X_{cf}$, the attenuation is simply proportional to $X^{-0.5}$ like deep bulk acoustic waves (DBAWs). For $X \gg X_{cf}$, because of the effect of surface boundary conditions, SSBWs tend to be pushed down into the bulk and the amplitude attenuation on the substrate surface becomes more significant; typically proportional to $X^{-1.5}$. For example, quasi-SH type SSBWs with relatively large X_{cf} are quite unsusceptible to the surface mechanical boundary conditions, and propagate far from their source without much attenuation. On a metallised surface, the critical propagation length is given by $X_{cm} = \lambda_B/\pi b^2$.

2.2 Delta Function Model Analysis

The complex power P of acoustic waves radiated from $q(X)$ on an IDT is given by,

$$P = 2^{-1}\int_{-\infty}^{+\infty} \phi(X)\frac{\partial q(X)^*}{\partial t}dX \quad (11)$$

$$\equiv P_{11} + P_{12} + P_{21} + P_{22}. \quad (12)$$

In eq.(12),

$$P_{ij} = -\frac{j\omega}{2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} G(X-x)q_i(x)q_j(X)^* dx dX, \quad (13)$$

where $q_i(X)$ is the surface charge density on IDT-i. Hence, the transfer admittance Y_{ij} between IDT-i and IDT-j is given by,

$$Y_{ij} = 2P_{ij}/E_iE_j, \quad (14)$$

where E_i is the voltage on IDT-i.

Assume that $q_i(X)$ is approximately given by,

$$q_i(X) = E_i\sum_m^{M_i} A_{im}\delta(X - X_{im}). \quad (15)$$

In eq.(15), $\delta(X)$ is the delta function, M_i is the total number of electrodes of IDT-i, and X_{im} and A_{im} are, respectively, the position and relative excitation amplitude of the m-th electrode. Substituting eq.(15) into eq.(13), one obtains Y_{12} as,

$$Y_{12} = -j\omega\sum_m^{M_1}\sum_n^{M_2} A_{1m}A_{2n}G(X_{1m} - X_{2n}). \quad (16)$$

Replacing $G(X)$ in eq.(16) by $G_S(X)$ in eq.(4), one obtains the well known delta function model which is extensively used for SAW device analysis and design[7].

By the similar way, delta function model for SSBWs is also obtained by replacing $G(X)$ by $G_B(X)$ in eq.(8).

When the insertion loss IL of SAW devices is relatively large, its frequency response is proportional to Y_{12} relative to the internal impedance of electrical source. This suggests that the SSBW spurious response level γ is given by Y_{12} for SSBWs relative to Y_{12} for SAWs, i.e.,

$$\gamma = \frac{f|\sum_m^{M_1}\sum_n^{M_2} A_{1m}A_{2n}G_B(X_{1m} - X_{2n})|}{f_r|\sum_m^{M_1}\sum_n^{M_2} A_{1m}A_{2n}G_S(X_{1m} - X_{2n})|_{f=f_r}|}, \quad (17)$$

In eq.(17), f_r is an arbitrary frequency which specifies the reference level of 0dB; one may conveniently

choose the IL-minimum frequency of the SAW response as f_r.

The delta function model for SSBWs defined by eqs.(8) and (16) is expected to be applicable to SSBW device analysis; similar to the delta function model analysis of SAW devices. Equation (17) also suggests that the SSBW spurious response level in SAW devices could easily be estimated. Note, however, that the delta function model for SSBWs is not suitable for the estimation of Y_{11} and Y_{22}. This is because the approximation employed in eq.(8) is not always valid in the vicinity of an IDT.

3 Comparison with Experiments

3.1 SSBW Spurious Response in an SAW TV-IF Filter

SSBW spurious response in an SAW TV-IF filter fabricated on 112-LT was measured and compared with the theoretical prediction discussed in the previous section.

Table 1 shows K_B^2, X_{cf}, X_{cm} and the phase velocity V_B of slow-shear, fast-shear and longitudinal SSBWs on 112-LT.

Table 1: K_B^2, X_{cf}, X_{cm} and V_B for 112-LT

	slow-shear	fast-shear	longitudinal
K_B^2	2.314×10^{-5}	1.492×10^{3}	7.185×10^{-1}
X_{cf}	1.468×10^{4}	4.561×10^{-7}	2.883×10^{0}
X_{cm}	3.353×10^{5}	4.561×10^{-7}	2.882×10^{0}
V_B(m/s)	3,374	3,691	6,157

Since slow-shear waves on 112-LT are of typical SH-type and less affected by the surface mechanical boundary conditions, both X_{cf} and X_{cm} are very large. However, because of their extremely small K_B^2, they are scarcely launched by an IDT. Although fast-shear waves possess large K_B^2 and might strongly be launched by an IDT, their very small values of X_{cf} and X_{cm} imply that they are unable to propagate very close to the surface. On the other hand, none of K_B^2, X_{cf} and X_{cm} of longitudinal waves is very small. This suggests that longitudinal waves could be observed as SSBWs in practical devices.

Figure 3 compares the measured frequency response of an SAW TV-IF filter with the SAW and SSBW responses calculated by the delta function model. In the calculation, the substrate surface was assumed to be electrically free, and the calculated SSBW response was normalised by the minimum IL of the SAW response (see eq.(17)).

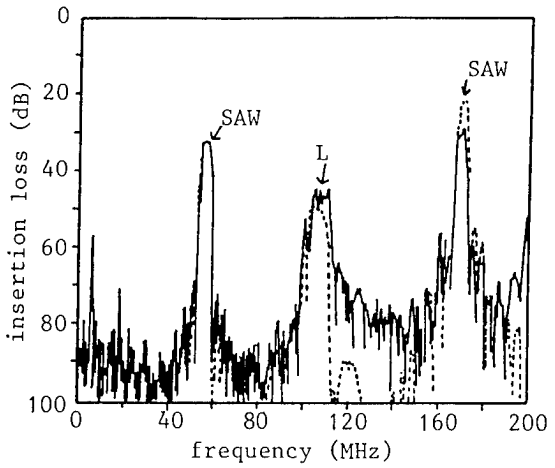


Figure 3: Frequency response of an SAW TV-IF filter fabricated on a 112-LT substrate. Solid line: experiments, and broken line: calculated.

In the figure, L shows the longitudinal SSBW response; it is crystal clear that the experimental response is in good agreement with the theoretical prediction and that the delta function model analysis discussed in the previous section is effective in the estimation of the SSBW response in SAW devices. The result also shows that the effect of both slow- and fast-shear SSBWs upon the SAW response is negligible (less than -110dB). In the present device, the spurious response level due to slow-shear and longitudinal SSBWs is less than -100dB about the trap frequencies; SSBWs scarcely deteriorate the trap performance of the filter.

3.2 SSBW Spurious Response on an SiO_2-Film/36-LT Substrate

When SSBWs and guided modes like SAWs, LSAWs, etc. are both propagation modes, the SSBW excitation strength is simply estimated by their phase velocity difference[9]. If the velocity difference is small, SSBWs are strongly radiated, whereas the SSBW excitation is negligible for a large velocity difference.

Figure 4 shows the phase velocity of LSAWs propagating on a SiO_2-film/36-LT substrate as a function of the thickness h of the SiO_2-film. In the figure, λ_{LSAW} is the wavelength of LSAWs, and V_B is the phase velocity of fast-shear SSBWs which is independent of h.

As can be seen, with an increase in h/λ_{LSAW}, the LSAW phase velocity V_{LSAW} decreases almost monotonically and the velocity difference $V_B - V_{LSAW}$ gets

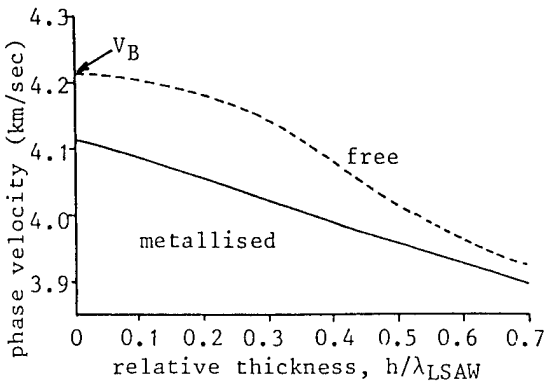


Figure 4: V_{LSAW} as a function of h/λ_{LSAW} for an SiO_2-film/36-LT substrate. Solid line: metallised surface, and broken line: free surface.

large. This suggests that by depositing an SiO_2 film of proper thickness, the excitation of fast-shear SSBWs, i.e., the spurious response level due to fast-shear SSBWs could considerably be reduced[10].

For the qualitative discussion of the SSBW spurious response on SiO_2-film/36-LT substrates, K_B^2, X_{cf} and X_{cm} for fast-shear SSBWs were determined and shown in Fig. 5.

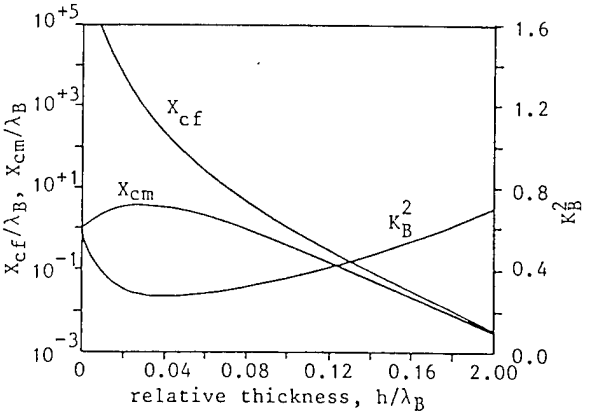


Figure 5: Dependence of K_B^2, X_{cf} and X_{cm} on h/λ_B for fast-shear SSBWs on an SiO_2-film/36-LT substrate.

K_B^2 does not change rapidly with h/λ_B. Although X_{cf} decreases most significantly with an increase in h/λ_B, it is extremely large when h/λ_B is small. This suggests that the SSBW spurious response might not be negligible evenwhen the propagation length is considerably large. On the other hand, X_{cm} is much smaller than X_{cf} for a small value of h/λ_B and decreases quite rapidly with an increase in h/λ_B.

Hence, it is expected that the SSBW spurious response can be suppressed by metallising the propagation surface and depositing the SiO_2 film of the thickness of h/λ_B >0.1-0.2. Although not detailed here, because of higher-order modes of LSAWs appearing for an increased h/λ_B, the SSBW response becomes large when h/λ_B >0.5.

Figure 6(a) shows the frequency response of the delay line employing a 36-LT substrate covered with a 7μm-thick SiO_2 film, whereas Fig. 6(b) for 36-LT substrate without any overlayer. For both launching and receiving unapodised IDTs, the number of total finger pairs is 9.5, the finger periodicity p is 72μm, and the aperture is 25p. The propagation length between the two IDTs is 150p and the propagation surface is metallised by an Al film.

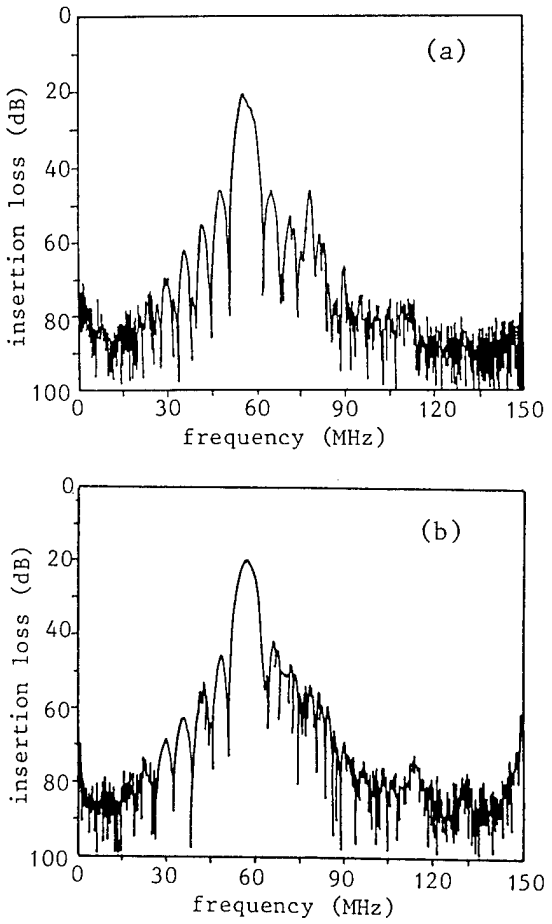


Figure 6: Frequency response of delay lines on 36-LT. (a): 7μm-SiO_2 film overlay, (b): without SiO_2 film overlay.

In Fig. 6(a), the main response at about 56MHz is attributed to LSAWs and takes the form of sinc function. Compared with Fig. 6(b), it can be seen that the zero points above the main response are clearly observed because of the suppressed fast-shear SSBWs.

The main response in Fig. 6(a) is a little distorted. This is caused by the interference between

LSAWs propagated along the surface and DBAWs which are radiated from LSAWs with propagation, reflected at the bottom surface and detected by the receiving IDT. The DBAW response of this kind could completely be suppressed by saw-cuts fabricated on a bottom surface. The response appearing at about 40MHz and 80MHz are attributed to SAWs and longitudinal SSBWs, respectively.

The experiments were also performed for the devices covered with SiO_2 films of the thickness of 2, 3, 4 and 5μm. When h is very small, as is theoretically predicted, the spurious response due to fast-shear SSBWs increases with h and the zero points above the main LSAW response are not clearly observed.

The fast-shear SSBW response was significantly suppressed for $h/\lambda_B(=0.1)$ smaller than theoretically expected. This may be because the elastic constants of the SiO_2 film are somewhat different from those of fused quartz used in the analysis. In fact, the shift of the main LSAW response towards lower frequencies was also more significant.

Note that the use of thin film layers is not always effective in suppressing SSBW spurious response. The analysis suggests that an SiO_2 film, for example, deposited on a 128°YX-$LiNbO_3$ substrate increases slow-shear SSBWs.

4 Conclusions

Using the effective permittivity, we have proposed delta function model analysis of SSBW response by defining two physical parameters characterising the SSBW excitation and propagation: one corresponds to the electromechanical coupling factor of SAWs giving SSBW excitation strength on the substrate surface, and the other is the critical propagation length representing how SSBW amplitudes decay with propagation.

The proposed delta function model analysis of SSBW response was applied to the characterisation of an SAW TV-IF filter. The result obtained by the analysis was in good agreement with experiments, and showed that the proposed analysis is most effective in characterising filter performance accompanied by SSBW spurious response. The values of the two parameters were also determined for an SiO_2-film/36-LT substrate to estimate its SSBW spurious response level. The theoretical prediction was in good agreement with experiments, and it was shown that the SSBW excitation strength and accordingly the SSBW spurious response is considerably reduced by properly choosing the thickness of the SiO_2-film.

The proposed delta function model analysis of SSBW spurious response could most effectively be applied to search for substrates suitable for SAW devices.

Acknowledgements The authors are greatly indebted to Mr.Y.Yasuhara of Toshiba Corp. for supplying SAW devices. It is also a pleasure to thank Mr.Y.Aono for carrying out the experiments.

References

[1] M.Lewis, 'Surface skimming bulk waves, SSBW', Proc. IEEE Ultrasonics Symp., 1977, pp.744-752.

[2] R.F.Mitchell and R.E.Read, 'Suppression of bulk wave radiation from surface acoustic wave devices', IEEE Trans. on Sonics and Ultrasonics, vol.SU-22, 4, pp.264-270, July 1975.

[3] R.S.Wagers, 'Plate mode coupling in acoustic surface wave devices', ibid., vol.SU-23, 2, pp.113-127, March 1976.

[4] K.Yashiro and N.Goto, 'Analysis of generation of acoustic waves on the surface of a semi-infinite piezoelectric solids', ibid., vol.SU-25, 3, pp.146-153, May 1978.

[5] R.F.Milsom, N.H.C.Reilly and M.Redwood, 'Analysis of generation and detection of surface and bulk acoustic waves by interdigital transducers', ibid., vol.SU-24, 3, pp.147-166, May 1977.

[6] K.Hashimoto and M.Yamaguchi, 'Analysis of acoustic waves launched from an interdigital transducer by means of effective permittivity', J. Faculty of Engrg., Chiba Univ., vol.41, 2, pp.9-17, 1990.

[7] R.H.Tancrell and M.G.Holland, 'Acoustic surface wave filters', Proc. IEEE, vol.59, 3, pp.393-409, March 1971

[8] J.J.Campbell and W.R.Jones, 'A method for estimating optimal crystal cuts and propagation directions for excitation of piezoelectric surface waves', IEEE Trans. on Sonics and Ultrasonics, vol.SU-15, 4, pp.209-217, July 1968

[9] M.Yamaguchi and K.Hashimoto, 'Simple estimation for SSBW excitation strength', J. Acoust. Soc. Jpn. (E), vol.6, 1, pp.51-54, January 1985

[10] K.Sato, Y.Fujiwara and K.Hashimoto, 'SiO_2/36°Y-X $LiTaO_3$ surface acoustic wave resonator', Jpn. J. Appl. Phys., vol.29, Suppl.29-1, pp.157-159, 1990

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DESIGN, FABRICATION AND PERFORMANCE OF 2 GHz SURFACE TRANSVERSE WAVE RESONATORS

SERGEY DENISENKO[*] *and IVAN D. AVRAMOV*[#]

[*] *Institute of Semiconductor Physics*
13 Lavrentyev Avenue, 630090 Novossibirsk, Russia

[#] *Institute of Solid State Physics*
72 Tzarigradsko Chaussee Boulevard, 1784 Sofia, Bulgaria

Abstract

This paper presents results from 2 GHz surface transverse wave (STW) resonant devices fabricated with a line width of 0.6 μm which is considered to be the upper limit at which reproducible line widths are obtained with conventional photolithography. Extended cavity single mode STW resonators with an insertion loss of 10 dB and a record value of 3800 for the unloaded Q, low-loss, low-Q resonators and two-pole resonator filters for voltage controlled oscillators (VCO) with a tuning range of up to 0.1% of the resonant frequency were designed, fabricated and tested. The device realization was performed on the highly productive electron beam lithography (EBL) system ZBA-21S of Carl Zeiss, Germany using a direct-write process. The device characteristics show that STW are ideally suited for low-cost fabrication of high performance resonant devices with insertion losses well below 10 dB at 2 GHz on temperature stable quartz substrates.

Introduction

Since the first successful high-Q resonator, using STW was realized in 1987 [1], the interest in these shear horizontal (SH) surface acoustic waves has been permanently increasing. Recent investigations on SH-wave resonant devices have led to the important conclusion that in simple two-port resonator configurations STW allow a control over the resonant Q in wide limits without or with insignificant sacrifice in insertion loss [2], [3]. This unique feature comes from the fact that metalization strongly influences the localization of the wave energy with respect to the substrate surface and the overall behavior of the SH-wave, propagating under a metal strip grating [4], allowing an additional degree of freedom in the design of resonant devices. This phenomenon was successfully applied to the design of single and multimode high-Q resonators with an insertion loss of 7 dB and unloaded Q values as high as 70% of the material limit for surface acoustic waves (SAW) [2] as well as combined mode resonator filters (CMRF) with an insertion loss of 3-5 dB and a 1 dB bandwidth of 0.2% of the resonant frequency at 1 GHz [5], [6].

Here we have decided to check by how far STW demonstrate their real advantages in devices, fabricated with a line width of 0.6 μm which nowadays can reproducibly and costeffectively be fabricated with most of the available conventional photolithographic techniques. The frequency limit for SAW devices, fabricated with this line width on quartz is 1300 MHz, while devices of the same line width, but using the SH-wave mode, exceed the 2 GHz limit. Since no adequate photolithographic equipment was available at our laboratories we used direct-write (EBL) for the device fabrication. Three different types of SH-wave resonant devices were optimized, fabricated and tested. These were extended cavity single mode high-Q resonators, low-loss, low-Q resonators and 2-pole resonator filters. The design and fabrication procedure as well as results will be described in the next sections.

Design of 2 GHz STW Resonant Devices

The design of the resonant devices, described here, is very similar to the designs investigated in [2], [3], [5] and [6]. The layout schematics of the extended cavity single mode resonator and the low-Q resonator and the two-pole resonator filter are shown in Fig. 1 a) and b) respectively. The

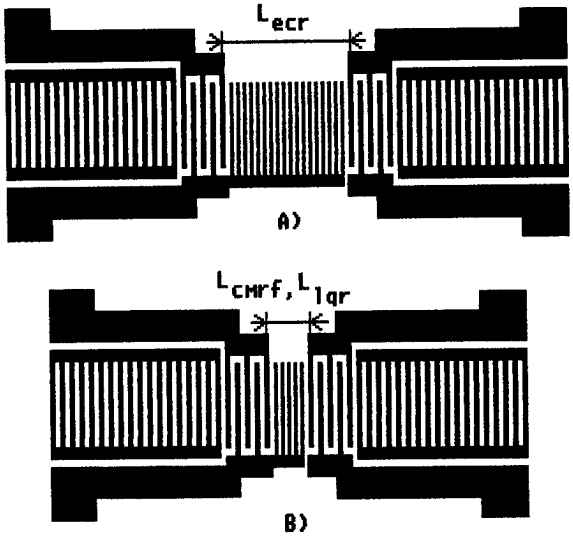


Fig. 1. Layout schematics of the:
a) extended cavity single mode resonator,
b) low-Q resonator, and the two-pole resonator filter.

design parameters, common for all devices, were optimized by means of a modified COM algorithm [2], [6] and are as follows:

0-7803-0905-7/93 $3.00 © 1993 IEEE

- acoustic wavelength λ: 2.6 μm
- fingerwidth: 0.6 μm
- gapwidth: 0.7 μm
- number of finger pairs per interdigital transducer (IDT): 70
- number of strips per grating: 800
- acoustic aperture: 40 λ

The only difference between the designs in Fig. 1 a) and b) is in the metal thickness and the length of the center waveguide grating which was designed with a period slightly lower than the transducer/reflector one to be transparent for the travelling wave while providing trapping of the wave energy to the surface [4]. In the low-Q resonator and two-pole resonator filter the number of strips in this grating was 12 and in the extended cavity resonator it was about 130 strips long. The layout configurations were optimized by the COM algorithm used in [2], [5] and [6]. This algorithm was initially developed for Rayleigh wave resonators [7] and does not model STW device characteristics ideally. After proper modification, though, the model was found to predict the frequency responses of STW devices below the center frequency of the stopband with sufficiently high accuracy. The COM optimization of the device layouts allowed us to choose proper design parameters and spacers and predict the device insertion loss, loaded Q, sidelobe suppression, the positions of the two cavity resonant peaks in the two-pole resonator design and finally the overall device frequency responses. The design procedure was performed in the same way as described in [6] and [8].

EBL System ZBA-21S (Carl Zeiss, Germany)

This direct-write EBL system has a shaped electron beam with a controlled format size from 0.2 to 6.3 μm with a step of 0.1 μm. The accelerating energy for the electrons is 20 KeV and the maximum beam current density is 3 A/cm^2. The pattern data preparation is performed on a "Kulon-1" computer station which is used to prepare the codes for direct controlling the computer of the EBL system. This computer is of a ES-100/25 type and has its own software for fully automatic operation of the EBL system and sequential exposure of up to 13 different 5-inch wafers or 7-inch photomasks.

The entire pattern field is divided in elements consisting of working fields, subfields and rows of single stamps of the shaped electron beam exposure. The pattern division of the system is shown in Fig. 2.

In our case the devices were smaller than the working field of 3.2×3.2 mm^2 and therefore it was not necessary to move the specimen holder during the exposure of a single device. The large working field dimension, necessary for that, could conveniently be handled by using the double E-beam deflection system. A magnetic deflector is used to perform the large beam deflections and an electrostatic system performs the fine beam deflection in the subfield. Thus the double deflection system strongly reduces the exposure time. With a current density of 0.7 A/cm^2 the exposure time for a single STW resonator did not exceed 1.5 min. We used coupling of fields and

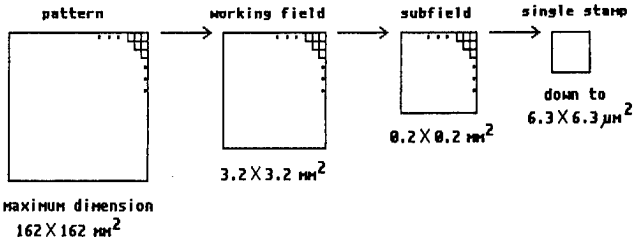


Fig. 2. Pattern field division of ZBA-21S.

subfields only when exposing the bus bars and when multiplying the devices on the substrate. The coupling accuracy for subfields and working fields was 0.1 and 0.15 μm respectively. The multiplication scheme of the STW devices on a 22×9 mm^2 AT-cut quartz substrate is shown in Fig. 3.

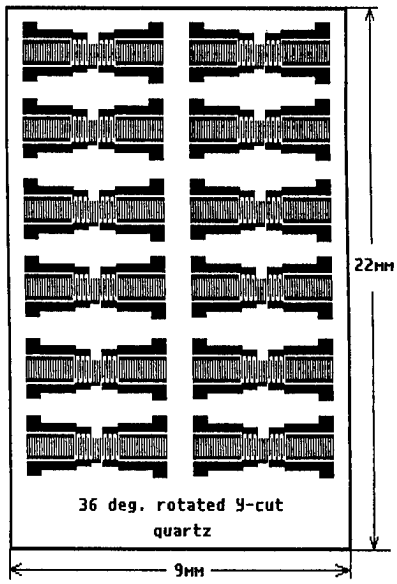


Fig. 3. STW device multiplication on the substrate.

Fabrication Process

The quartz substrates were processed according to the following steps:

1. Cleaning
2. Al-Cu metalization of the polished surface
3. Spinner coating of the metalized surface with a positive PMMA-type electron resist of about 0.3 mm thickness
4. Resist baking at T=170 deg. C for 40 min.

6. Resist developing in a (1:3) mixture of metiletilceton and isopropyl alcohol for 30 sec.

7. Thermal resist treatment at T=100 deg. C for 15 min.

8. Wet etching of the blank metal surfaces with an acid mixture of H_2PO_4:H_2NO_3:CH_3COOH:H_2O (15:7:3:1) for 45 sec.

9. Removing the resist rests in heated metil-etilceton.

For the EBL exposure the quartz substrates were placed in a specially designed universal specimen holder which was capable of holding any kind of substrates with a 10 to 40 mm length. Since we used a negative lithography the gaps of the resonators had to be exposed. The exposure line width , we used, was in the range of 0.3 - 0.6 μm with a pitch of 1.0 - 1.2 μm. The exposure dose was about 65 $\mu C/cm^2$. Fig. 4 shows a scanning electron microscope (SEM) picture of the IDT of one of the devices and Fig. 5 is a SEM picture of the

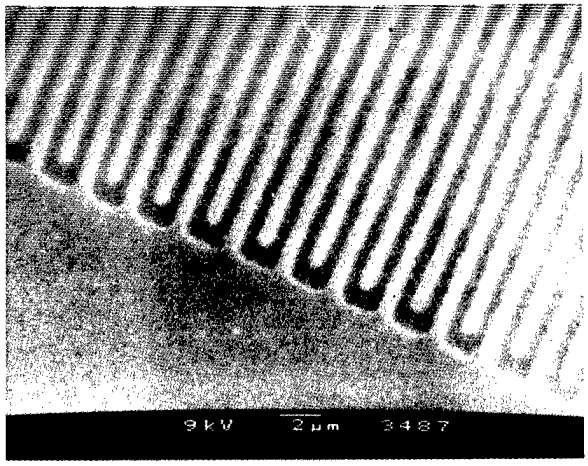


Fig. 4. SEM picture of the IDT of one of the devices.

grating. From these two pictures it is clearly evident that the electrode edges are not very uniform. The reason for this is the fact that we used wet chemical etching which is not the best etching method for line widths below 1 μm. In some of the devices we observed a substantial electrode underetching, especially when we increased the etching time. We believe that the finger edges could be significantly improved using a lift-off or a dry etching process.

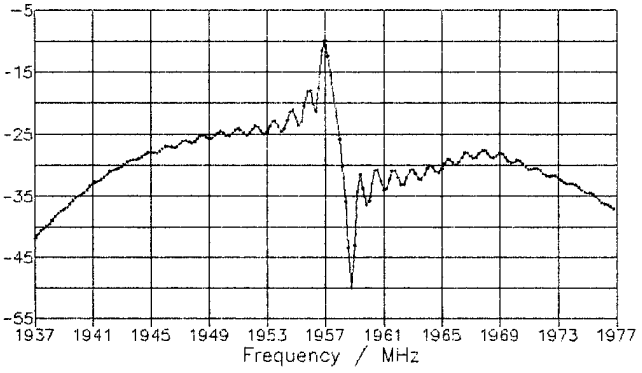
Results

Extended Cavity, Single-Mode, High-Q STW Resonator

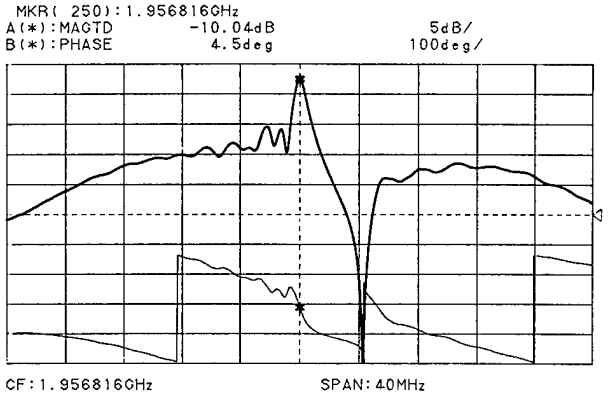
First we realized the device from Fig. 1a). Here we intended to achieve the highest possible device unloaded Q with a relatively simple layout configuration. The COM predicted frequency response of this device is shown in Fig. 6a) and Fig. 6b) is the experimental device frequency and



Fig. 5. SEM picture of the STW reflector grating.



A)



B)

Fig. 6. Characteristics of the high-Q device:
a) predicted COM response
b) measured frequency and phase response

phase response, measured in a 50 Ohm system without matching. The device insertion loss is 10 dB and the loaded Q is 2600. This values correspond to an unloaded Q of 3800 which is as high as 71% of the material limit for SAW and, to the best of our knowledge, is the highest Q achieved with a 2 GHz STW resonator so far. The comparison with the data obtained from the best 2 GHz SAW resonator, reported to date, which achieved an unloaded Q of 3300 at 2052 MHz with a line width of 0.45 μm and a gap width of 0.314 μm [9], clearly shows the real advantage of STW at GHz frequencies. In our opinion the unloaded Q, we achieved, could readily reach 90% of the material limit for SAW by using a better etching process. This assumption is consistent with the results of Bagwell and Bray, described in [1]. A further improvement will hardly be possible since the film resistance will become a serious limiting factor.

Low-Q Resonator

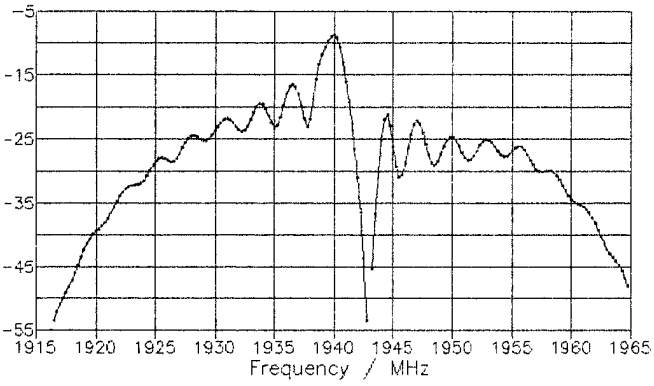
The key to achieving sufficient tuning range and high data rates in a STW resonator stabilized VCO with direct frequency modulation (FM) is the bandwidth of the acoustic device and its group delay value at the carrier frequency [6]. There are two mechanisms responsible for low device bandwidth and low group delay values:

1. As short as possible resonant cavity and
2. Increased metal thickness.

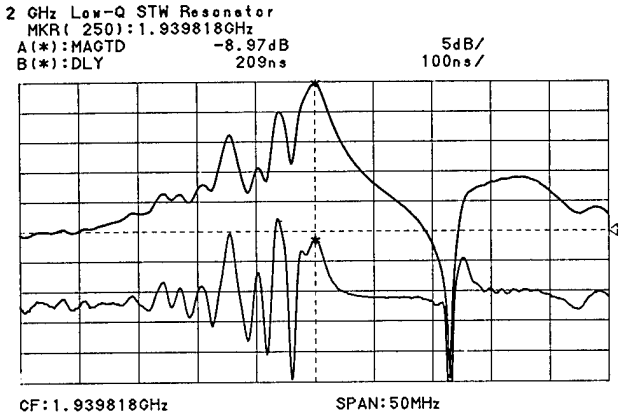
Here we applied both mechanisms to the short cavity length device shown in Fig. 1 b). Unfortunately, we did not have adequate equipment for measuring the metal thickness precisely but in this design the metal thickness used was approximately twice as high as that of the high-Q resonator discussed in the previous section.

The predicted COM response of the low-Q SH-wave resonator is shown in Fig 7a). Its measured frequency and group delay responses are shown in Fig. 7b). The comparison of the theoretical and experimental responses indicates a good agreement for the frequency range around and below resonance. The COM algorithm for Rayleigh wave device modeling does not give satisfactory results in the frequency range above resonance, especially in the vicinity of the upper stopband edge (see the position of the notch) and above the stopband. Still one can rely on this algorithm for proper adjustment of the spacers, the cavity length and the other topological design parameters.

The device, characterized in Fig. 7b), has an insertion loss of 9 dB and a loaded Q of 1070 which would provide a tuning range of at least 1.5 MHz in a VCO, stabilized with this resonator. Actually the best devices demonstrated an insertion loss of 8 dB, which is by 2 dB less than the loss of the high-Q device. This result is consistent with the data presented in [3] and can be explained with the fact that the low-Q device has lower resistive losses due to the higher metal thickness. Unfortunately, the devices with the 8 dB loss had a slightly lower bandwidth than desired. Also here the group delay is very low, only 209 ns. According to previously published data [6] this resonator is a very elegant solution for achieving FM data transmission rates of hundreds of Kbits/s.



A)



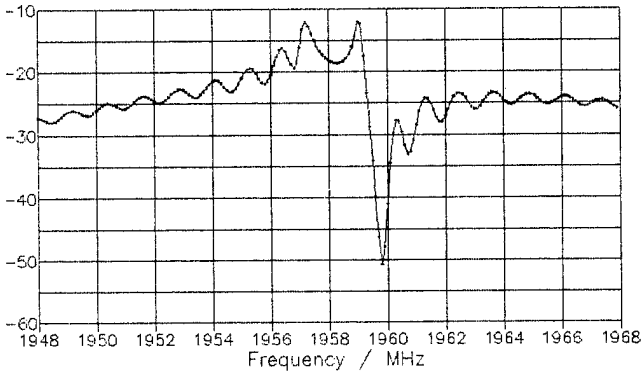
B)

Fig. 7. Characteristics of the low-Q resonator:
a) COM predicted frequency response
b) measured frequency and group delay responses.

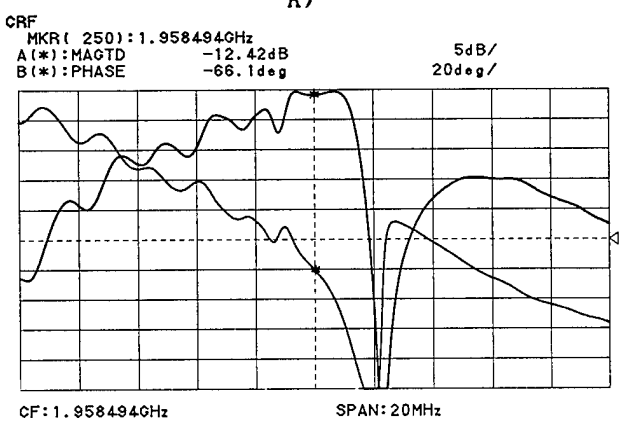
Two-Pole Resonator Filter

In some cases, e. g. for low-noise synthesizer applications, even higher VCO tuning ranges, and therefore larger device bandwidths may be necessary. As shown in [6] increased bandwidths can be obtained with two-pole resonator filters. The design procedure of such devices is well documented in [6] and [8]. Here we used again the configuration in Fig. 1b). The difference between the low-Q resonator, described in the previous section, and the two-pole filter is only in the spacers and the metalization. The spacers were optimized again with the COM algorithm and the metal thickness was again relatively thin to achieve better control over both longitudinal modes (poles), giving the shape of the frequency response. The theoretical and experimental data of this device is shown in Fig. 8 a) and b) respectively. The 3 dB device bandwidth is 2.7 MHz, its 1 dB bandwidth is 2.2 MHz and the insertion loss is 12.4 dB. The reason for this

higher loss value is again the increased metal film resistance and the increased device bandwidth due to both poles with equal magnitude (see Fig. 8). Also with this device we would expect an improvement of at least 2 dB if a better etch process is used.



A)



B)

Fig. 8. Characteristics of the two-pole resonator filter:
a) predicted COM response
b) measured frequency and phase responses.

Summary and Conclusions

We have discussed the design, EBL-fabrication and performance of 2 GHz low-loss SH-wave resonant devices, realized on temperature stable AT-cut quartz with a line width of 0.6 μm which can be fabricated at a low cost using conventional photolithographic techniques. After careful evaluation of the results and comparison with previously published data we came to the conclusion that at 2 GHz STW perform better in resonant devices with respect to device Q, insertion loss and freedom in the control over the resonant bandwidth with the metalization parameters. In our opinion 2 GHz STW devices can achieve unloaded Q-factors as high as 90% of the material limit for SAW and an unmatched insertion loss of 7-8 dB in a 50 Ohm system if Cu-doped Al metalization is used. The main limiting factor to achieve even better values is the metal film resistance. STW resonant devices are generally more sensitive to tolerances in the metalization parameters than their SAW counterparts. That is why careful control over the metal thickness, the finger-to-gap ratio and the finger edges will be necessary, especially if the device specifications are tight. Never the less we feel that GHz range low-cost STW resonant devices will find a variety of applications in the near future.

Acknowledgements

We would like to thank Drs. A. Aseev and I. B. Yakovkin from the Institute of Semiconductor Physics in Novossibirsk, Russia for the support of this work. We gratefully acknowledge the invaluable help of Dr. A. S. Kozlov, M. Dolgov, M. Kachanova, G. Ostanina and A. Mayer from the same institute who contributed to the device fabrication and T. Gavrilova and L. Litvin for the SEM photographs. Finally we direct our special thanks to the Organizing Committee of the IEEE 1993 Frequency Control Symposium and SAWTEK Inc. in Orlando, Florida who made the presentation of this paper possible by supporting and organizing the authors' trip to the United States.

References

[1]. *T. L. Lagwell and R. C. Bray, IEEE 1987 Proc. Ultrasonics Symposium, pp. 319-324.*

[2]. *I. D. Avramov, Proc. 1992 IEEE Freq. Control Symp., pp. 391-408.*

[3]. *L. Almar, B. Horine, J. Andersen, Proc. 1992 IEEE Ultrasonics Symposium, pp. 51-56.*

[4]. *B. A. Auld and D. F. Thompson, Proc. 1987 IEEE Ultrasonics Symposium, pp. 305-312.*

[5]. *I. D. Avramov, Electronics Letters, February 1991, Vol. 27, No. 5, pp. 414-415.*

[6]. *I. D. Avramov, Proc. 1991 45th Frequency Control Symposium, pp. 230-238.*

[7]. *C. K. Campbell, C. B. SAW, IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control, Vol. UFFC (34), No. 3, May 1987, pp. 357-367.*

[8]. *I. D. Avramov, A. S. Kozlov, Proc. 10th European Conf. on Circuit Theory and Design, ECCTD-91, 2-6 Sept, 1991, Lyngby, Denmark, pp. 729-738.*

[9]. *L. L. Pendergras, L. G. Studebaker, IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control, Vol. UFFC (35), No. 3, May 1988, pp. 372-379.*

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN ANALYSIS OF SELF-SUPPORTED CIRCULAR STRUCTURES FOR REDUCED STRESS SENSITIVITY IN SAW DEVICES

S. Ballandras, E. Bigler

Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon - 32,avenue de l'Observatoire - 25044 Besançon Cedex - France

Abstract

This paper is devoted to the study of the mechanical stress sensitivity of Surface Acoustic Wave (SAW) devices built on self supported circular quartz structure (BVA type). The purpose of this approach is to reduce this sensitivity by using the symmetry properties of circular structures commonly used in the design of Bulk Acoustic Wave (BAW) devices. These structures are coupled with planar stress compensated SAW quartz cuts that are the equivalent for SAW of the well known SC-cut of bulk waves. Since the BVA-like structure chosen for the present study does not permit analytical resolutions of quasi-static mechanical stress problems, a finite element modelisation has been used and a number of computing tools have been implemented in order to couple the finite element results with a perturbation method based on variational equations. The theoretical prediction of the sensitivity of these SAW-BVA devices to symmetrical radial compression and to normal bending effects presented in this paper could yield new kinds of optimized resonator structures insensitive to both acceleration effects and thermoelastic effects.

1 - Introduction

Reducing G-sensitivity and stress sensitivity of Surface Acoustic Wave (SAW) devices is a critical topic in the design of high stability surface wave resonators. Nevertheless, one can take advantage of a lot of work devoted to the mechanical stress sensitivity of Bulk Acoustic Wave (BAW) devices. These studies have been performed following two different approaches. One of these approaches consists in using anisotropic properties of the propagation substrate (mainly quartz plates) coupled with particular geometries of this substrate. It has been implemented by E. P. Eernisse to demonstrate the existence of BAW quartz cuts with very low sensitivities to mechanical perturbations [1]. One of these cuts (the SC-cut) being also temperature-compensated has been extensively used in the design of very high stability BAW resonators and has become an industry standard.

This work is supported by DGA/DRET under contract # 90/074

The second approach is based on the use of mechanical structures designed to minimize the influence of mechanical perturbations on the high frequency wave propagation. This can be achieved by using particular fixation conditions of the propagation substrate that permit to separate it from the area where the stresses induced by the mechanical perturbation are the most important. Such a structure has been implemented by R. J. BESSON [2] in the case of BAW resonators. This structure called BVA (Boîtier à Vieillissement Amélioré) is composed of a self sustained circular quartz plate connected to an outer ring by quartz bridges. The outer ring itself may be supported in 2 or 4 points. The position and the number of the quartz bridges depend on the mechanical properties of the cut used. An optimized design allows to reduce the effects of the mechanical perturbation at the center of the plate where elastic waves propagate and to take advantage of the SC-cut low sensitivity to symmetrical stress effects [3].

The methods of investigation used by E. P. Eernisse for BAW devices can be transposed to the case of SAW quartz cuts. Two crystalline orientations have been found exhibiting low sensitivities to both stress and temperature effects : B. K. Sinha's STC cut [4] defined by the cut angles Φ, θ and the propagation direction ψ respectively equal to 0°, 41.8° and 48.2° and the TG cut [5] defined by the angles (Φ = 0°, θ = -10°, ψ = 30°) studied by G. Théobald *et al.*

Self supported circular quartz resonators (BVA - type) have been proposed by Besson *et al.* [French patents n° 7601035, 7616289, 7717309, 7802261, 7828728, 7835631, 7918553 and corresponding patent or patents pending in other countries]. Until now, they received little attention in practical SAW device applications.

The purpose of the present study is to use such a self sustained BVA-like structure coupled with the TG or the STC cut in order to reduce the effects of mechanical perturbation on temperature-compensated SAW devices. A theoretical analysis of the mechanical properties of such devices has been implemented. Since BVA-like structures are too complicated for establishing analytical models, a finite element package has been used to evaluate stresses, strains and displacement gradients induced by quasi-static mechanical perturbations. These data have been used to calculate a perturbation tensor that can be injected

0-7803-0905-7/93 $3.00 © 1993 IEEE

in a perturbation method developed by H. F. Tiersten [6] based on variational equations describing small elastic fields superposed on a bias [7]. This perturbation method is a very efficient tool to compute the frequency shift of a SAW device due to a mechanical bias. The first part of the paper is devoted to the description of the theoretical background used for the numerical computation.

Then, different results of calculation performed on 2D quartz disks are presented which have permit to valid the approach. Finally, computing results of BVA-type SAW delay lines submitted to symmetrical radial compression and normal bending effects are presented. The calculations have been performed for (ST,X) cut, TG cut and STC cut and a qualitative analysis of the results is proposed in order to highlight the possibilities to design BVA-type SAW resonators with a low sensitivity to stress and temperature perturbations.

2 - Perturbation method

The perturbation method implemented for the present study consists in calculating the frequency shift of an elastic wave device due to an external perturbation (force, temperature, etc.). In the most general case, the perturbation equation will be written as follows

$$\frac{\Delta F}{F_o} = \frac{\int\int\int_v \overline{\overline{H_{ikjm}}}\, \frac{\partial u_i^{o*}}{\partial a_k}\frac{\partial u_j^o}{\partial a_m}\, dv}{2\rho_o\,\omega_o^2 \int\int\int_v u_\ell^o\, u_\ell^o\, dv} \qquad (1)$$

where $\Delta F/F_o$ is the relative frequency shift, ρ_o the mass per unit volume in the natural state [10], ω_o the angular frequency, u^o_i the dynamic mechanical displacements in the natural state, a_k the material coordinates, v the volume where the wave propagates and $\overline{\overline{H_{ikjm}}}$ the perturabation tensor. This tensor relates the statical stresses $\overline{T_{km}}$, strains $\overline{S_{uv}}$, and displacement gradients $\partial u_i/\partial a_r$ due to the perturbation and the elastic properties of the substrate. When limited to first order effects, $\overline{\overline{H_{ikjm}}}$ will be written

$$\overline{\overline{H_{ikjm}}} = \delta_{ij}\,\overline{T}_{km} + C_{ikjmuv}\,\overline{S}_{uv} + C_{pkjm}\frac{\partial \overline{u}_i}{\partial a_p} + C_{iqkm}\frac{\partial \overline{u}_j}{\partial a_q} \qquad (2)$$

where C_{ikjm} represents the linear elastic coefficients, C_{ikjmuv} the non-linear elastic coefficients and δ_{ij} is the Kronecker delta.

In the case of simple stress problems, analytical models can be established and the value of $\overline{\overline{H_{ikjm}}}$ will be known in each point (a_1, a_2, a_3) of the propagating domain [11, 12]. But when the geometry of the substrate is too complicated to resolve the mechanical perturbation problems by analytical means, the stresses, strains and displacement gradients will be evaluated by using finite element methods. A finite element package can be used to know the values of stresses, strains and displacement gradients in a certain number of points which are the center of mass of the elements in the case of the MODULEF package used for the present study. Then, a mean value of $\overline{\overline{H_{ikjm}}}$ can be calculated for each element (e) and will be expressed as follows

$$\begin{aligned} \overline{\overline{<H_{ikjm}(e)>}} &= \delta_{ij}\,\overline{<T_{km}(e)>} + C_{ikjmuv}\,\overline{\overline{<S_{uv}(e)>}} \\ &+ C_{pkjm}\,\overline{\overline{<\frac{\partial u_i}{\partial a_p}(e)>}} + C_{ikqm}\,\overline{\overline{<\frac{\partial u_i}{\partial a_q}(e)>}} \end{aligned} \qquad (3)$$

Now, in order to inject eq. (3) into the perturbation equation, the integrals of eq. (1) must be discretized. But in the case of Rayleigh surface waves, an expression of eq. (1) can be obtained mixing discretization and analytical resolution of these integrals. Assuming that all the elements of the SAW device mesh are cylindrical and that the volume of these elements is equal to S(e)*(h_2 - h_1) where (h_2 - h_1) is the thickness of the element e, the following form of equation (1) will be obtained

$$\frac{\Delta F}{F_o} = \frac{\sum_{r,s=1}^{4}\left\{A_r^* A_s\; {}^o u_i^{o(r)*}\; {}^o u_j^{o(s)}\, n_k^{(r)*}\, n_m^{(s)} \sum_{e=1}^{N}\left[\overline{\overline{<H_{ikjm}>}}\, A\right]\right\}}{2\rho_o V_o^2 \sum_{r,s=1}^{4}\left\{A_r^* A_s\; {}^o u_\ell^{o(r)*}\; {}^o u_\ell^{o(s)} \sum_{e=1}^{N}[A]\right\}} \qquad (4)$$

$$\text{with } A = \frac{S(e)\left(\frac{\omega_o}{j V_o}(n_2^{(r)*} - n_2^{(s)})\, a_2\right)_{h_1}^{h_2}}{j\left(n_2^{(r)*} - n_2^{(s)}\right)}$$

where A_r are the relative amplitudes of the Rayleigh wave, ${}^o u_i^{o(r)}$ the dynamic mechanical displacement amplitude and $n_k^{(r)}$ the propagation coefficients in accordance to the usual harmonic analysis of the Rayleigh wave propagation in natural state [13]. V_o is the velocity of the SAW and N the number of elements taken into account for the calculation of $\Delta F/F_o$.

Eq. (4) has been implemented in a FORTRAN program and used for both 2D and 3D finite element models. In the case of 2D models, the value of h_2 has been fixed to 0 and the value of h_1 to the lower real number accepted by the computer in order to simulate a calculation performed on a semi-infinite substrate occupying the space defined by $a_2 < 0$ (Fig. 1).

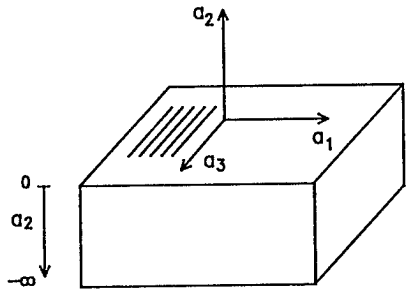


Fig. 1 : Definition of the axes for the semi-infinite substrate where the SAW propagates.

3 - Finite element analysis

3.1 - 2D Models

To test the coupling between finite element analysis results and the perturbation method described previously (Sec. 2), two 2D models of a circular plate submitted to symmetrical radial compression have been implemented. The choice of this plate geometry can be justified by the fact that the existence of theoretical and experimental data obtained during previous works [9] concerning the SAW sensitivity to radial compression permits to check easily the validity of the present results.

Another reason for this choice is that 2D static linear elastic calculations are very easy to perform with a finite element package, particularly when the degree of the interpolation polynomials does not exceed 2. Moreover, in the case of a thin disk submitted to in-plane stresses, the effects of the perturbation will be assumed homogeneous over the whole thickness of the disk. This hypothesis also justifies the implementation of 2D models, at least for the first tests.

Following this approach, a first 2D mesh of the quartz disk has been realized using triangular elements together with Lagrangian interpolation polynomials of degree 1. The finite element package MODULEF used for this study permits to take into account the crystal anisotropy. Its modular design is also suitable to access easily to subsidiary results necessary for the calculation of displacement gradients $\partial \bar{u}_i / \partial a_p$. A viewgraph of the mesh is reported in Fig. 2.

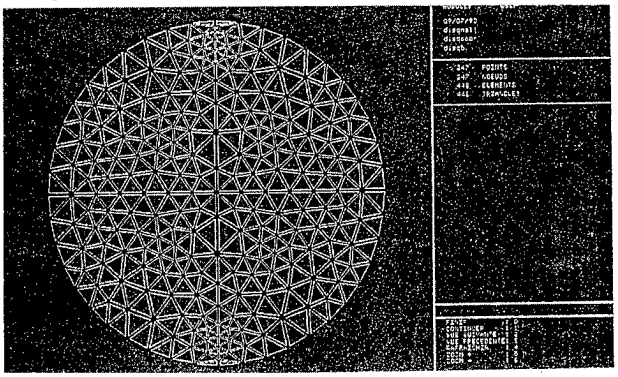


Fig. 2 : Mesh of a 2D quartz disk coupled with interpolation polynomials of degree 1

To simulate correctly the fixation conditions of the previous experiments [9], the statical displacements $\bar{u}_i$ have been set to 0 at the bottom of the disk and a vertical compressive force has been applied at its top. In this point, only the vertical displacements could take place freely (that is also in accordance with the experimental conditions). The stresses, strains and displacement gradients are known at the center of mass of each element of the mesh. Since the mechanical perturbation is assumed to act principaly on the elastic wave at the center of the quartz disk, only the 8 triangles at its center, forming a regular octagon, have been taken into account for the perturbation calculation. The second model implemented also uses triangular elements but with Lagrangian interpolation polynomials of degree 2. More elements have been taken into account for the calculation in order to increase the precision of the model (Fig. 3).

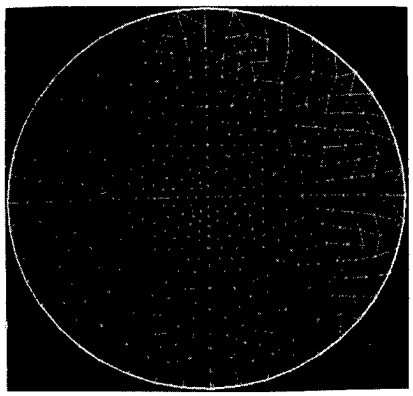
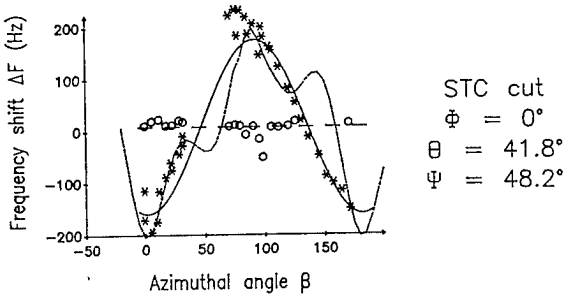
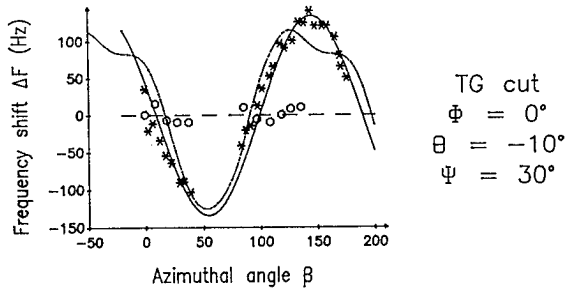
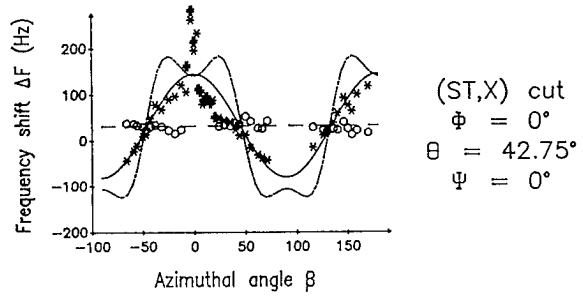


Fig. 3 : Mesh a 2D quartz disk coupled with interpolation polynomials of degree 2

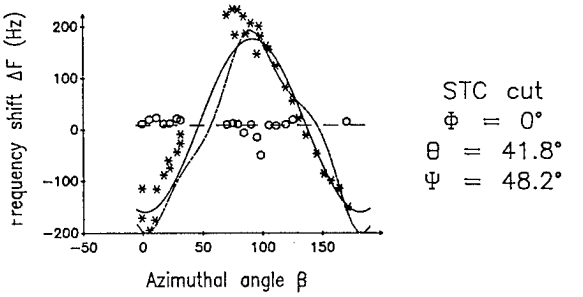
For both cases, the 6 planar elastic coefficients C_{11}, C_{13}, C_{15}, C_{33}, C_{35} and C_{55} have been rotated for each azimuthal angle β which is the angle between the radial force direction and the propagation direction Ψ. After calculations. the stresses, strains and displacement gradients are rotated with an angle $-\beta$ to bring them back in the SAW axes. The results of these models are presented in Fig. 4 for the (ST, X) cut, STC cut and TG cut. They are super-imposed on classical isotropic model results [9] and on experimental data.

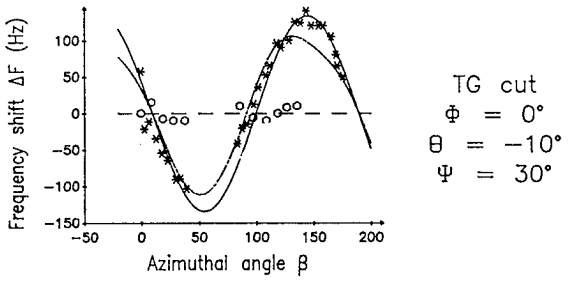


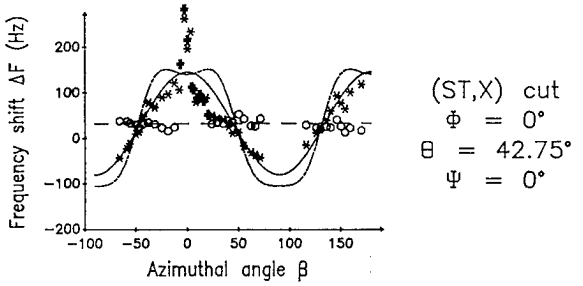




a) Interpolation polynomials of degree 1







b) Interpolation polynomials of degree 2

Fig. 4 : 2D models of radial in-plane compression comparison between classical theoretical model (solid lines), experimental measurements (o and ∗) and finite element analysis predictions (dashed lines).

In the case of models using Lagrangian interpolation polynomials of degree 1, strong oscillations appear for STC cut and (ST, X) cut sensitivities to radial compression versus the azimuthal angle β. The magnitude of these oscillations is less important in the case of the TG cut.

Nevertheless, a relative coherence between theoretical and experimental data can be underlined, particularly the phase of the frequency variation vs. β is rather well described by the finite element models.

The precision of the theoretical predictions is increased by using interpolation polynomials of degree 2. This can be explained by the fact that the continuity of the results is one degree higher that in the previous case. This implies a better description of the elastic properties of the crystal. The number of elements taken into account was also higher than in previous calculations, but the present model is not very sensitive to the variation of this number (Fig. 5).

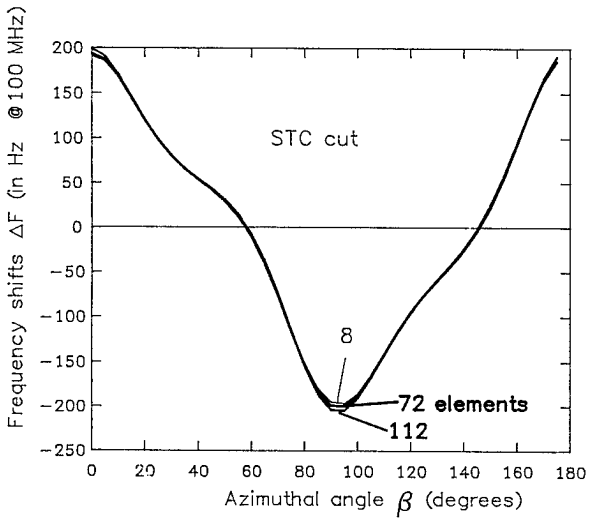


Fig. 5 : Sensitivity of the perturbation calculation to the number of element taken into account (case of the STC cut)

Finally, the validity of the approach is demonstrated, particularly with the interpolation polynomials of degree 2. This perturbation method will be now applied to the case of 3D structure modelisations.

3. 2 - 3D models - BVA-type structures

A structure based on a self-sustained circular quartz plate (Fig. 6) has been studied. The quartz plate is connected to an outer ring by 4 quartz bridges. This ring itself is supported in 4 mounting points. It is expected that a proper choice of the cut and the position of the quartz bridges could yield interesting compensation effects. This choice will be conditionned by the 2D theoretical results and by experimental data concerning diametrical in-plane compression. Assuming that the main effects induced by normal bending perturbation will be traction or compression stresses, strains and displacement gradients, to the

design of SAW devices with a low sensitivity to both acceleration and thermoelastic effects will be now investigated by coupling a BVA-type structure (Fig. 6) with the TG cut or the STC cut.

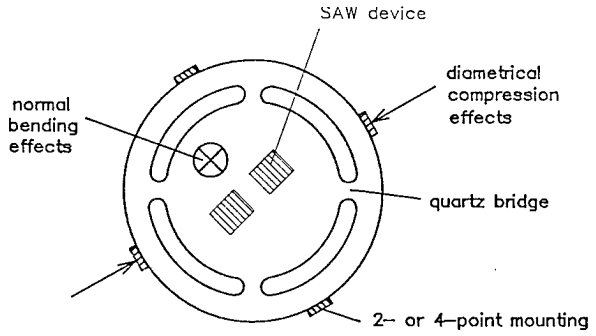


Fig. 6 : Self-sustained circular quartz plate for SAW application

Following this approach, the quartz bridges will be positionned each 90° around the plate. One of the bridge will be placed at an angle corresponding to the azimuth β for which the cut presents the lowest sensitivity to compressive stresses (Fig. 4). In the case of the TG cut, the value of this angle is close to 20°. For the STC cut and the (ST,X) cut this angle can be set in the region of 45°



Fig. 7 : Mesh of the 3D self-sustained circular quartz plate

Moreover, as shown in Fig. 8, the TG cut exhibits an interesting compensation phenomenon for normal bending effects within an extended range of angular position β. This phenomenon offers a certain degree of freedom to satisfy another criteria. For example, since both STC and TG cuts are low sensitive to dynamical thermal effects [9], it is expected that an optimum position of the 4-point mounting of the outer ring will be found allowing to reduce simultaneously the G-sensitivity and the thermal sensitivity of the SAW device.

3D models of this structure have been simulated (Fig. 7) using the position of the four quartz bridges previously defined. The 3D elements used to compute the stress distribution are pentahedric (triangular based prisms). The SAW propagating domain has also been meshed (Fig. 7) and 64 elements have been taken into account for the perturbation calculation. The density of elements has been increased near the region of the bridges in order to ensure a better continuity of the solution. For diametrical compression effects, the force uniformly distributed along the thickness of the plate (1 mm) is applied along a pair of opposite faces. The width of these faces corresponds to a 6° arc. In the case of normal pressure effects, 4 faces on the outer ring are fixed and the pressure is applied uniformly on the whole upper surface of the device. The azimuth β is defined in that case as the angle between the propagation direction Ψ and the fixed mounting points. In both cases, β will be used as a parameter for optimization. The results are presented in Fig. 8. It is shown in this figure that a minimum of sensitivity to both criteria can be found for the (ST,X) cut, as well as for planar-stress compensated STC cut and TG cut. The advantage of these new cuts over the (ST,X) cut being that two positions will satisfy the minimum of sensitivity to diametrical compression instead of one.

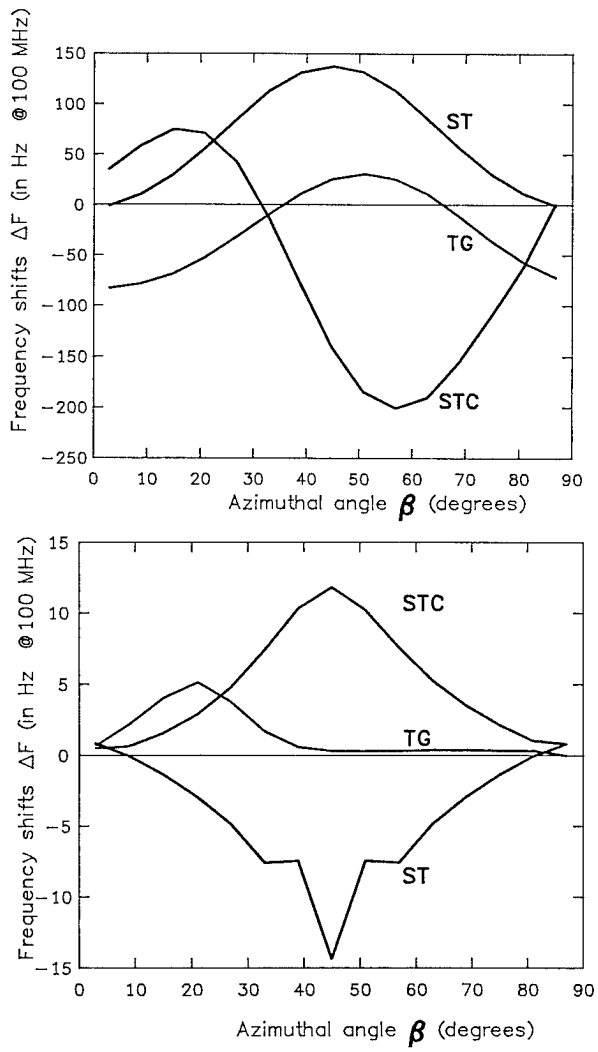


Fig. 8 : Sensitivity of BVA-type SAW devices to mechanical perturbations
(a) diametrical in-plane compression
(b) normal hydrostatic pressure

4 - Conclusion

An analytical calculation has been implemented to evaluate theoretically the sensitivity of SAW devices to mechanical stress perturbation. This model is based on the coupling between variational equations and a finite element analysis, tested both with 2D and 3D models. In the case of 2D models, the efficiency of this perturbation method has been demonstrated with Lagrangian interpolation polynomials of degree 2. The results obtained with 3D models use interpolation polynomials of degree 1 only to reduce computing time. An analytical integration of the well-known Rayleigh wave parameters is performed along the thickness of the plate to simplify computations.

A precise simulation of more complex kinds of waves (resonant Rayleigh waves, or other kinds of waves) could be performed also by a full numerical integration with interpolation polynomials of a higher degree, at the cost of an increased computing time which is not really a problem with the dramatic increase of CPU speed in modern workstations.

Nevertheless, it is possible to investigate the sensitivity of BVA-type SAW devices to radial compression and normal bending effects. First results show that the TG cut seems to be well-suited for this application. Experimental tests will be necessary to confirm these theoretical results.

Acknowledgements

The authors would like to thank Pr J.-M. Crolet, director of the Laboratoire de Calcul Scientifique (LCS Besançon, France) where all finite elements calculations have been performed.

References

[1] E.P. Eernisse, "Calculations on the stress-compensated (SC-cut) quartz resonator", Proc. 30th Ann. Freq. Cont. Symp., 1976, pp. 8-11.

[2] R.J. Besson, "A new electrodeless resonator design", Proc. 31st Ann. Freq. Cont. Symp. (1977), pp. 147-152.

[3] R.J. Besson, U.R. Peier, "Further advances on B.V.A. quartz resonators", Proc. 34th Ann. Freq. Cont. Symp., Ft Monmouth, may 1980.

[4] B.K. Sinha, "A stress and temperature compensated orientation and propagation direction for surface acoustic wave devices", IEEE Trans. Ultrason. Ferro. Freq. Cont., vol. UFFC-34,n° 1, pp. 64-74 (jan. 1987).

[5] H.F. Tiersten, J.T. Song and D.V. Shick, "A continuous representation of the acoustic surface wave mode shape in arrays of reflecting grooves", J. Appl. Phys., vol. 52, n° 4, pp. 1154-1161 (Aug. 1987).

[6] H.F. Tiersten and B.K. Sinha, "A perturbation analysis of the attenuation and dispersion of surface waves", J. Appl. Phys., vol. 49, n° 1, pp. 87-95 (jan. 1978).

[7] J.C. Baumhauer and H.F. Tiersten , "Nonlinear electroelastic equations for small fields superposed on a bias", J. Acoust. Soc. Am., vol. 54, n° 4, pp. 1017-1074 (oct. 1973).

[8] E. Bigler, G. Théobald, D. Hauden, "Stress-sensitivity mapping for surface acoustic waves on quartz", IEEE Trans. on Ultrasonics, Ferro-electrics and Frequency Control, vol. 36, n° 2, pp. 57-62 (jan. 1989).

[9] S. Ballandras, E. Bigler, "Surface acoustic wave devices with low sensitivity to mechanical and thermoelastic stresses", J. Appl. Phys., vol. 72, n° 8, pp. 3272-3281 (october 1992).

[10] Physical Acoustics, Principle and Methods, Part A : Wave Propagation in Fluids and Solids", R.N. Thurston, vol. I, Warren P. Mason Ed., Academic Press, 1965.

[11] B. K. Sinha, H.F. Tiersten, "On the influence of a flexural biasing state on the velocity of the propagation of surface waves", Wave Motion, vol. 1, pp. 37-51 (1979).

[12] D. Janiaud, "Modélisation de l'influence d'une accélération sur la fréquence des résonateurs à quartz", Thèse, Université de Besançon, 1978, pp. 63-67.

[13] K.A. Ingebrigtsen, "Surface waves in piezoelectrics", J. Appl. Phys., vol. 40, n° 7, pp. 2681-2687 (1969).

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

WITHDRAWAL WEIGHTED FAN-SHAPED SAW TRANSDUCERS

Bausk E.V., Yakovkin I.B.

Institute of Semiconductor Physics,
Siberian Branch of the Russian Academy of Sciences
Novosibirsk 630090, Russia

Abstract

Withdrawal weighted (WW) fan-shaped interdigital transducers (IDT) are used here to design surface acoustic wave (SAW) filters with a good shape factor, flat passband frequency response (FR), and out-of-band rejection of 50 dB and more. Utilization of "stepped" functions as a weighting functions for fan-shaped IDTs ensures a simple filter structure synthesis procedure that does not take a long time on a personal computer. Three modifications of SAW filters with 2% bandwidth, in-band ripple of 0.1 dB, and out-of-band rejection of 50-55 dB have been fabricated on ST-quartz.

To realize surface acoustic wave filters with a good shape factor and a flat passband of frequency response, fan-shaped interdigital transducers are regarded as the most suitable ones. An uniform electrode length affords a low level of secondary effects such as diffraction and phase front distortion. Unfortunately, SAW filters including two unweighted fan-shaped transducers can not have out-of-band rejection values better than 40 dB. To improve the rejection and to keep the advantages mentioned above, we use withdrawal weighted fan-shaped IDTs in SAW filters. In the present paper a simple synthesis procedure of withdrawal structure for fan-shaped IDTs is proposed.

For analysis of filters including fan-shaped IDTs, a division of transducers into strips parallel to the direction of SAW propagation is used [1]. A particular strip is supposed to have a constant structure period λ_o and a time response $A(t,\lambda_o)$. The filter frequency response can be written as:

$$H(f)=\int_{\lambda_o\,min}^{\lambda_o\,max} FA_{in}(f,\lambda_o)\, FA_{out}(f,\lambda_o)\, d\lambda_o. \qquad (1)$$

Here $FA(f,\lambda_o)$ the Fourier transform of $A(t,\lambda_o)$ is the frequency response of a strip. The δ-function model [2] gives for WW transducers the following expression

$$A(t,\lambda_o) = \sum_{n=1}^{N} A_n\, \delta(t-n\lambda_o/2v) \qquad (2)$$

where A_n is an amplitude of SAW source located on the n-th electrode, N - the number of electrodes, v - SAW velocity. The frequency response H(f) calculation of the filter with specified transducers structures can be easily performed by means of numerical-type integration. A straight synthesis procedure, that optimizes the function H(f) for parameters $A_1 \ldots A_n$ and determines the electrodes to be removed, is, however, much complicated because of the great number of variables in Eq.(2) and the complexity of the expression (1) after substitution of $A(t,\lambda_o)$.

The conventional way for synthesis of WW transducers with parallel electrodes [3] includes a choice of some weighting function $A'(t,\lambda_o)$ followed by its approximation by the expression (2). The Fourier transform of $A'(t,\lambda_o)$ shall satisfy requirements to transducer FR. In case of fan-shaped WW transducers the above method does not guarantee a good result when $A'(t,\lambda_o)$ is a smooth function such as the Hamming or Kaiser one. The substitution of the Hamming function into the expression (1) gives the rejection value of 80 dB and more (dotted curve in Fig.1) but after replacing of $A'(t,\lambda_o)$ by $A(t,\lambda_o)$ the rejection becomes much worse. The solid curve in Fig.1 is the calculated frequency response of the filter having two identical Hamming-weighted fan-shaped transducers with the "three fingers per period" configuration (in each group consisting of three electrodes the middle one has the positive polarity and the two outer ones have the negative polarity) [4]. Each group is considered as a single SAW source with the amplitude equal 1 (if the group is present in IDT) or 0 (if the group is

0-7803-0905-7/93 $3.00 © 1993 IEEE

absent). We use such IDT configuration because of the advantage of a greater electrode width then in transducers with split electrodes. Fig.1 illustrates the significant disparity between the two curves.

We suggest to use "stepped" functions $A_s(t,\lambda_o)$ as weighting functions for fan-shaped transducers:

$$A_s(t,\lambda_o)=\begin{cases} S_4, & \text{for} \quad |t|\leqslant t_4, \\ S_3, & \text{for} \quad t_4<|t|\leqslant t_3, \\ S_2, & \text{for} \quad t_3<|t|\leqslant t_2, \\ S_1, & \text{for} \quad t_2<|t|\leqslant t_1, \\ 0, & \text{for} \quad t_1<|t|. \end{cases}$$

Here $t_m=l_m\lambda_o/v$, $2t_1v=2l_1\lambda_o=N\lambda_o/2$ is the transducer length. The Fourier transform of $A_s(t,\lambda_o)$ has the simple form:

$$FA_s(f,\lambda_o)=\sum_{m=1}^{4} 2(S_m-S_{m-1})t_m * \\ * \operatorname{sinc}[2\pi(f-v/\lambda_o)t_m] \qquad (3)$$

where $S_o=0$. At specified values of $S_1 \ldots S_4$ and t_1 the weighting function $A_s(t,\lambda_o)$ has three variables only: t_2, t_3 and t_4. Because of the simplicity of the expression (3), H(f) can be readily calculated if $FA_s(t,\lambda_o)$ is substituted into Eg.(1). Due to the small number of variables a "stepped" function optimization procedure does not take a long time on a personal computer to give the best coincidence between H(f) and a desired filter frequency response. In case of a "stepped" weighting function a complicated withdrawal procedure, such as the one shown in [3], is unnecessary. The "stepped" function can be most simply applied to weighting fan-shaped transducers with a "three finger per period" configuration. For example, when $S_4=1$, $S_3=0.75$, $S_2=0.5$, $S_1=0.25$, the weighting procedure is as follows: no electrode is removed from the central region of the transducer (here $|x|\leqslant t_4v$ and x is the electrode coordinate in the direction of SAW propagation), each fourth group, consisting of three electrodes, is taken out of the neighboring areas on the left and on the right ($t_4v<|x|\leqslant t_3v$), from the next region ($t_3v<|x|\leqslant t_2v$) - two of each four groups are removed, and finally, at the edges of the transducer only one of each four groups is retained. In Fig.2 two functions H(f) are presented. The dotted curve has been calculated after substitution a "stepped" function $A_s(t,\lambda_o)$ (the same for the input and output IDTs) in Eq.1. The approximation of $A_s(t,\lambda_o)$ by the expression (2) has given the solid curve. The agreement between the curves in a passband proximity is here much better then in Fig.1.

Fig.3 shows the measured FR of a filter fabricated on ST-quartz. The filter includes two identical "stepped"-weighted fan-shaped transducers, both $200\lambda_o$ long with the "three fingers per period" configuration. The measured response is in a good agreement with the solid curve in Fig.2.

Fig.4 presents the calculated (1) and measured (2) frequency responses of a filter with two transducers each of that has the same "three finger per period" configuration and a $200\lambda_o$ length but different "stepped" weighting functions: values of t_2, t_3 and t_4 for the input and output IDTs are different. In this case the synthesis procedure is more complicated then for two identical IDTs because of the greater number of variables but the out-of-band rejection is better.

A disadvantage of the mentioned filters is undesirable overshoots at the frequencies $f_o \pm f_o n/4$ (where f_o - center frequency of passband, and n is an integer). The overshoots are caused by the withdrawal periodicity of four electrode groups. To reduce the overshoots we have used in a next filter modification the output IDT with the "two fingers per period" configuration [4]. In transducers with the "three fingers per period" configuration removing a single group does not disturb a SAW source distribution in the neighboring electrode regions because the outer electrodes in each group have the same polarity. Due to this feature, coefficients A_n in Eq.(2) are equal to 1 or 0 only. In IDTs with the "two fingers per period" configuration removing any electrode or electrode group changes source amplitudes on neighbors substantially. Coefficients A_n may take a number of values [3]. This gives another withdrawal procedure then in the previous case but it is simple too if the "stepped" function is used. As a result, the output transducer withdrawal periodicity is different from the input one, and undesirable overshoots are loca-

ted at other frequencies. Fig.5 shows the measured response of a filter that has the same input IDT as in the previous filter and the output IDT with a "two fingers per period" configuration. A significant decrease of the undesirable overshoots is observed.

All measured filter frequency responses have an in-band ripple of about 0.1 dB.

References

[1] M.R. Daniel and J. de Klerk, "Acoustic radiation Measurements and Calculations for Three Surface Wave Filter Designs," in Proceedings of the IEEE Ultrasonics Symposium, 1973, p.449-455.

[2] R.H. Tancrell and M.G. Holland, "Acoustic Surface Wawe Filters," in Proceedings of the IEEE Ultrasonics Symposium, 1971, p.48-64.

[3] K.R. Laker, E. Cohen, T.L. Szabo and J.A. Pustaver, "Computer-Aided Design of Withdrawal-Weighed SAW Bandpass Filters," IEEE Trans., vol. CAS-25, N.5, p.242-251, 1978.

[4] S.J.Kerbel, "Design of Harmonic Surface Acoustic (SAW) Oscillator Without External Filtering and New Data on the Temperature Coefficient of Quartz," in Proceedings of the IEEE Ultrasonics Symposium, 1974, p.276-281.

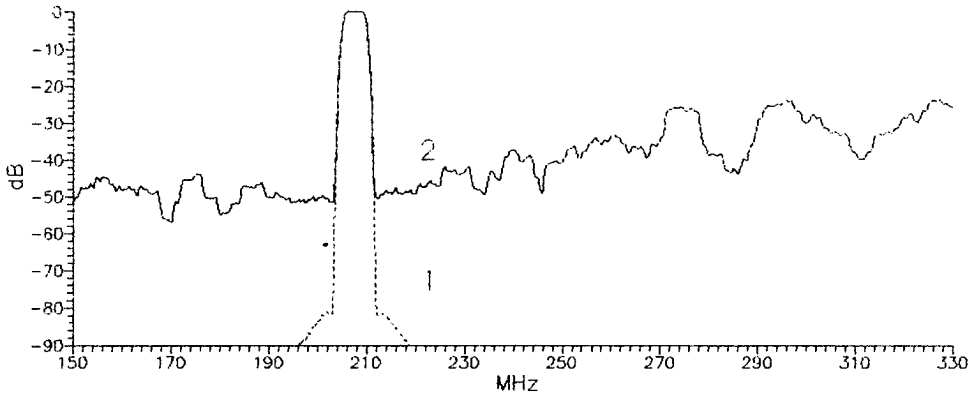


Fig.1. Filter frequency responses calculated after substitution into Eq.(1) the following functions:

1 - $A'(t,\lambda_o)$ the Hamming function;

2 - $A(t,\lambda_o)$ the withdrawal approximation of $A'(t,\lambda_o)$.

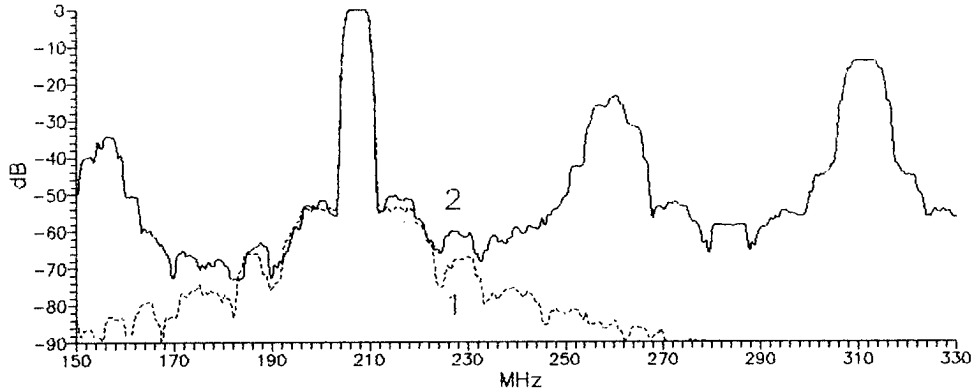


Fig.2. Filter frequency responses calculated after substitution into Eq.(1) the following functions:

1 - $A_s(t,\lambda_o)$ the "stepped" function;

2 - $A(t,\lambda_o)$ the withdrawal approximation of $A_s(t,\lambda_o)$.

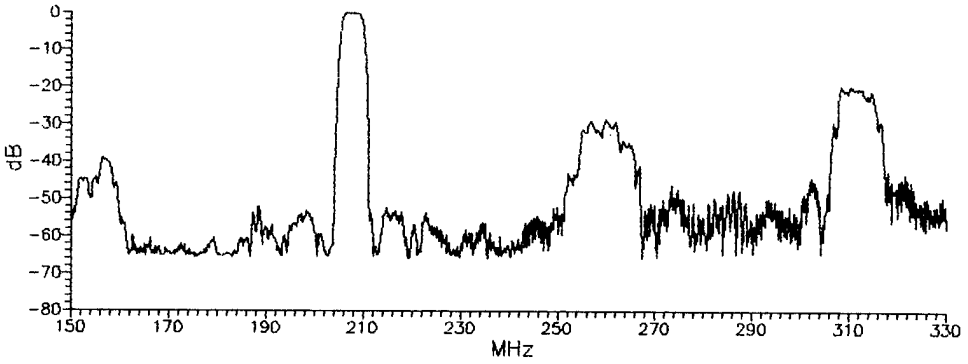


Fig.3. Measured FR of the filter having two identical IDTs with the "three fingers per period" configuration.

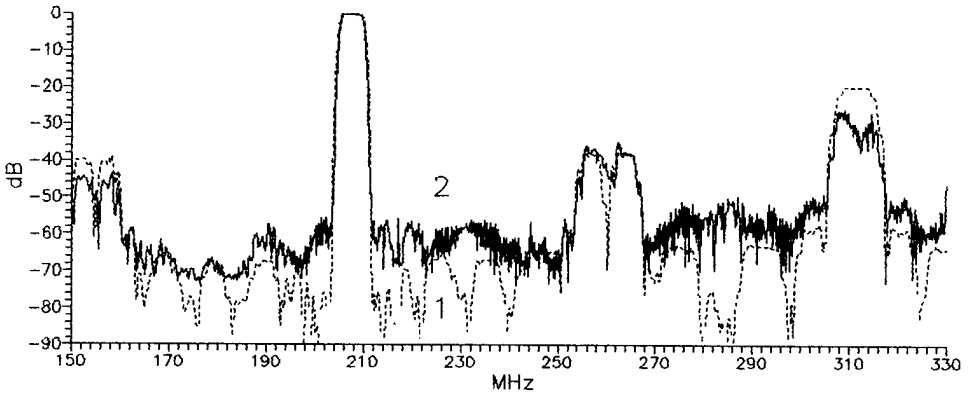


Fig.4. Calculated (1) and measured (2) FRs of the filter having the "tree fingers per period" configuration for both IDTs but different weighting functions.

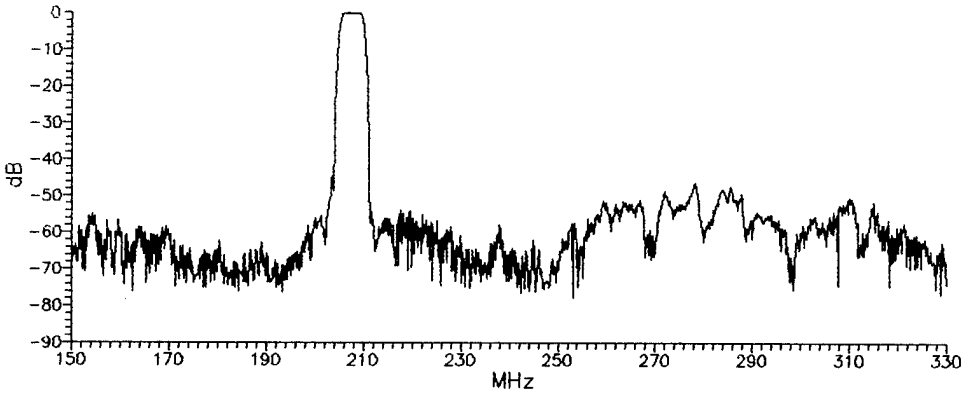


Fig.5. Measured FR of the filter input and output IDTs of that have different weighting fuctions and electrode configurations.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DEFORMATION-SENSITIVE CUTS FOR SURFACE ACOUSTIC WAVES IN α-QUARTZ

R.M. TAZIEV, E.A.KOLOSOVSKY, and A.S.KOZLOV

The Institute of Semiconductor Physics
Russian Academy of Sciences, Siberian Branch
13,Lavrentyeva,Novosibirsk,630090,Russia

Abstract

The influence on SAW velocity of external static forces is computed in a case of circular plates (membranes) Pressure sensitivity of the SAW velocity as a function of α-quartz crystal anisotropy are presented by sensitivity contour mapping from two independent variables: direction of SAW propagation and cut angles of α-quartz. Theoretical sensitivity values are compared with experimental ones in the case of circular (membrane) plates. Both experimental and theoretical results are in good agreement.

Introduction

Some authors have investigated static and dynamic temperature and force effects to avoid its influence on SAW device stability [1-3]. Several cuts for α-quartz which both temperature and different kinds of stresses compensated have been proposed. One of them so-called "STC-cut" ($\Phi=0^{\circ}$; $\theta=41,8^{\circ}$;$\Psi=46,89^{\circ}$) can be obtained by B.K.Sinha [4] and another cut so-called "TG-cut" ($\Phi=0^{\circ}$; $\theta=-10^{\circ}$; $\Psi=30^{\circ}$) introduced by G.Theobald et al [5]. The problem of propagation of SAW in pre-strained medium is transformed into separated problems: calculation of the bias and calculation of the wave propagation by using linearized equations. When the solid is subject to a homogeneous bias, the equations of motion and boundary conditions in terms of effective elastic constants can be solved in a straightforward manner. However, when the biasing state is inhomogeneous the effective elastic constants are position dependent and a direct solution of the problem is not possible. In the later case, a perturbation procedure more suitable for computation of small changes in the elastic wave velocity. This method allows one to investigate the influence of inhomogeneous stress gradients on the surface wave propagation when the active length of SAW device expands over region of significant stress gradients.

Moreover, a major problem in such studies is to determine a biasing deformation state of the plate, which is strongly influenced by the plate geometry and mounting supports of the device. Last time for explaining some discrepancies between theoretical and experimental sensitivities of SAW devices to acceleration and dynamic temperature effects have been employed the finite-element method (FEM) to obtain the real stress distribution in the port of plate where SAW device fabricated [6-7].

In this paper we obtain a general expression relating the fractional velocity shifts of SAW to a pressure loading of a surface in the case of thin circular arbitrary anisotropic plates. The velocity shifts are calculated by the perturbation method [8]. All necessary independent displacement gradients induced by external forces are derived analytically in any point of the membrane surface. So, we can predict the frequency shift for a SAW resonator at any point of the surface of thin circular plates. The problem of pressure sensitivity will be considered from a general point of view as a location of SAW resonators on the surface of plate which orientation and direction of SAW propagation are choosing to achieve an extremal sensitivity. For circular quartz plate results are given by contour mapping of the pressure sensitivity of a SAW resonator occurring in the the center of plate as function of the crystal anisotropy. There are some cuts with high pressure sensitivity in quartz. One of them more suitable for sensor applications is the Z-cut with a SAW propagation direction along the X axis. Further we optimize the locations of two SAW resonators on the surface of a membrane to achieve an extremal pressure sensitivity for dual channel differential pressure sensor.

SAW propagation in a pre-strained medium

The propagation of surface waves in presence of quasi-statically deforming states is governed by the equations of motion for small dynamic fields superposed on a bias. When reffered to the reference (or undeformed) coordinate axes

0-7803-0905-7/93 $3.00 © 1993 IEEE

fixed to the solid, the equations of motion in presence of a static bias may be written in the form [9]:

$$\rho\ddot{u}_j = P_{ij,i}\,, \qquad D_{i,i}=0, \qquad (1)$$

with boundary conditions on the surface of crystal ($a_3=0$)

$$P_{3j}=0, \qquad D_3+i\varepsilon_0|k|\phi=0, \qquad (2)$$

Here $P_{ij}=(C_{ijkl}+\hat{C}_{ijkl})u_{k,l}+e_{kij}\phi_{,k}$,

$D_i=e_{ikl}u_{k,l}-\varepsilon_{ik}\phi_{,k}$,

$\hat{C}_{ijkl}=T_{ik}\delta_{jl}+C_{ijklmn}E_{mn}+C_{ijnl}W_{l,n}+$
$+C_{inkl}W_{j,n}$

where $\hat{C}_{ijkl}$ are effective constants that depend on the biasing state, C_{ijkl}, e_{ikl}, ε_{ik} are second-order elastic, piezoelectric, and dielectric constants, respectively, ρ- denotes the mass density, T_{ik}, E_{mn}, $W_{j,n}$ are the biasing stress, strain, and displacement gradient component, respectively, ϕ, D_i, P_{ij} are electric potential, electric displacement vector, and Piola-Kirchhoff stress tensor, respectively, C_{ijklmn} is the third-order nonlinear constants. Other nonlinear constants are ignored. It should be noted that $\hat{C}_{ijkl}$ has lower symmetry than C_{ijkl}.

The solution for surface waves propagating in arbitrarily anisotropic, piezoelectric substrates satisfying the wave equations and boundary conditions may be written

$$u_j=\sum_{m=1}^{4} C^m A_j^m \exp(ip_m k a_3)\exp[ik(a_1-Vt)],$$

$$\phi=\sum_{m=1}^{4} C^m A_4^m \exp(ip_m k a_3)\exp[ik(a_1-Vt)],$$

where u_j, ϕ are the complex representation of the real mechanical displacement and electric potential for propagation in the natural direction a_1 with a_3 normal to the surface.

A perturbation method [8] is used to solve (1)-(2) and leads to the first-order relative velocity change

$$\frac{\Delta V}{V_o} = \frac{\int_V u_{i,j}^{*}\,\hat{C}_{ijkl}\,u_{k,l}\,dV}{2\rho\omega_o^2\int_V u_i^{*}\,u_i\,dV}\,, \qquad (3)$$

where u_i is the solution of the unperturbed surface wave with velocity V_o and the angular frequency ω_o.

Pressure loading of membrane

There are different desings of pressure sensors. Here we consider a circular piezoelectric plate of thickness h and in radius R with rigidly fixed boundary , which is subjected to uniform pressure on one its surfaces. Because the plate is thin, a plane-stress distribution assumption is made. The static strains on the surface of membrane due to application of a uniform pressure P are [10]:

$$E_{11}=\frac{3}{8}\left(\frac{R}{h}\right)^2\frac{P}{C}\left\{1-3\left(\frac{x}{R}\right)^2-\left(\frac{y}{R}\right)^2\right\},$$

$$E_{22}=\frac{3}{8}\left(\frac{R}{h}\right)^2\frac{P}{C}\left\{1-3\left(\frac{y}{R}\right)^2-\left(\frac{x}{R}\right)^2\right\},$$

$$E_{12}=-\frac{3}{8}\left(\frac{R}{h}\right)^2\frac{P}{C}\left(\frac{x}{R}\,\frac{y}{R}\right),$$

where C is the following combination of elastic constants:

$$A_{ij}=C_{ij}-C_{i3}C_{j3}/C_{33},\quad i,j=1,2,6,$$

$$C=(3(A_{11}+A_{22})+2(A_{12}+2A_{66}))/8.$$

Other strain components are obtained from surface free condition $P_{3j}=0$:

$$\begin{pmatrix} C_{33} & C_{34} & C_{35}\\ C_{43} & C_{44} & C_{45}\\ C_{53} & C_{54} & C_{55}\end{pmatrix}\begin{pmatrix} E_{33}\\ 2E_{23}\\ 2E_{13}\end{pmatrix}=-\begin{pmatrix} C_{31} & C_{32} & C_{36}\\ C_{41} & C_{42} & C_{46}\\ C_{51} & C_{52} & C_{56}\end{pmatrix}\begin{pmatrix} E_{11}\\ 2E_{22}\\ 2E_{12}\end{pmatrix}$$

The expression (3) is usually a function of nine displacement gradients. But the plane-stress assumption and no rigid rotation condition in the center of plate, where the SAW devices is placed ,may be reduce expression (3) to a function only six strain or stress coefficients.

For numerical calculations we used the material constants for α-quartz have been obtained in [11]. For symplicity we assume a quasi-ponctual SAW propagation in small volume of plate port submitted to an homogeneous mechanical bias. Then using (3) and (4) we can obtain the following relation for a local change of SAW velocity

$$\frac{1}{A}\frac{\Delta V}{V_o}=a_o+a_1\left(\frac{x}{R}\right)^2+a_2\left(\frac{y}{R}\right)^2+a_3\left(\frac{x}{R}\,\frac{y}{R}\right), \qquad (5)$$

from which we can predict the frequency

shift of SAW resonators in any point of surface of plate. Here the cofficients a_i depend on the material constants both explicitly and implicitly through the SAW properties and may be computed for each crystal cut and propagation direction of the SAW, A is the such normalization coefficient that (5) is valid for any membrane dimensions with the same constants a_i, and $A=P*(R/h)^2$.

To obtain the extremal values of sensitivity of SAW resonator we must find a maxima and a minima of the function (5). It may be easy shown that they exist in the center of membrane (x=0,y=0) and in the contour line of membrane which extrema coordinates are determined from relation which gives us two roots

$$tg^2\theta + 2(a_1-a_2)/a_3 tg\theta - 1 = 0,$$

here $x=R\cos\theta$, and $y=R\sin\theta$ describes a contour line of membrane. So we must choose only two points among three ones to obtain a maxima and a minima magnitudes of function (5) on the surface of membrane. If we consider a case of one SAW resonator sensor, we must choose only the maxima value of (5) to obtain high sensitivity.

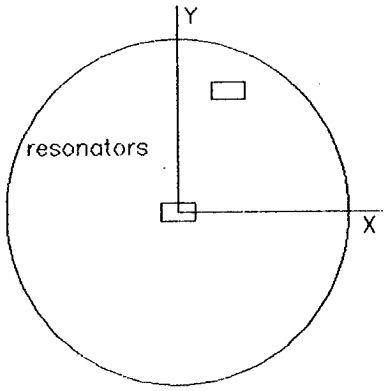


Fig.1. A schematic diagram of quartz membrane with dual channel differential SAW resonator structure with the same direction of SAW propagation.

But if we consider a case of dual channel differential sensor, we must choose both maxima and minima values which difference allows us to reach a high sensitivity. The dual channel differential scheme of a sensor is more preferable than one channel scheme because the first allows one to avoid or sufficiently to decrease a temperature influence on a SAW sensor (see Fig.1).

Fig.2 gives the contour mapping of pressure sensitivity (5) for SAW resonator occuring in the center of quartz membrane as a function of singly rotated Z-cut angle of quartz plate θ and direction of SAW propagation angle Ψ. Here for numerical calculation the following parameters have been used: h=0,25mm R= 20mm ,P=1 kPa, and to obtain the real value of sensitivity DV/V we must multiply it to 1.E-05. As one can see there are exist cuts of quartz which have higher sensitivity than well-known ST,X α-quartz with Euler's angles (Φ=0; θ=132,75; Ψ=0). One of them is the Z-cut with the direction of SAW propagation along the X axis of quartz.

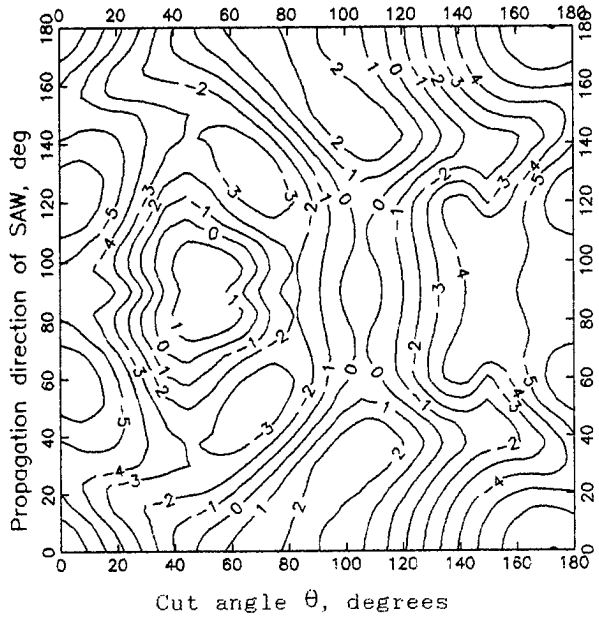
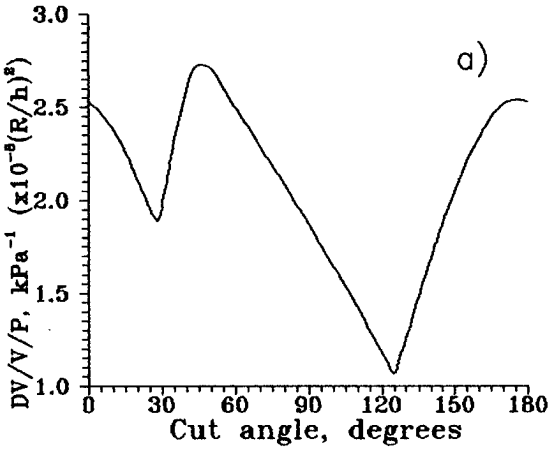


Fig.2. A contour-mapping of pressure sensitivity of a SAW resonator which are placed in the center of α-quartz circular plate.

Fig.3a gives the pressure sensitivity for differential scheme of two SAW resonators as a function of cut angle of quartz plate θ and direction of SAW propagation along the X-axis of quartz. There are some cuts with higher pressure sensitivity then ST,X quartz. When we change a cut angle of plate the positions of two resonators on the surface of plate move to achieve an extremal pressure sensitivity. The breaks on the sensitivity curve are explained by choosing of alone maxima differential value of (5) amoung three other values.For Z-cut the positions of two SAW resonators are following: one of them has coordinates (R,0), and other takes place in the center of membrane with coordinates (0,0).

Fig.3b,c,d give the pressure sensitivity for Z, ST, and Y- cuts as a function of direction of SAW propagation on the surface of membrane, respectively. As one can see there are some directions of SAW propagation with high sensitivity to pressure loading of circular plates. For ST-cut of quartz the extremal sensitivity exists at the angle Ψ=50 degrees with locations of two SAW resonators



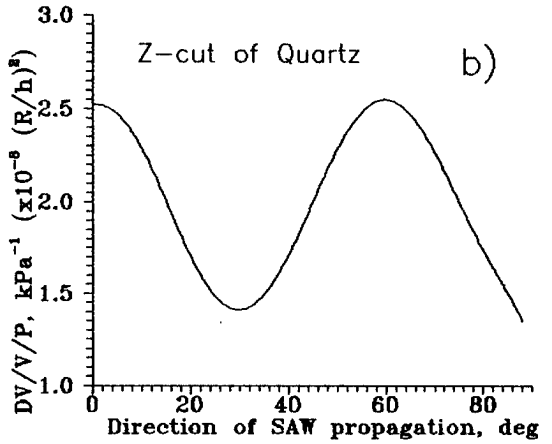
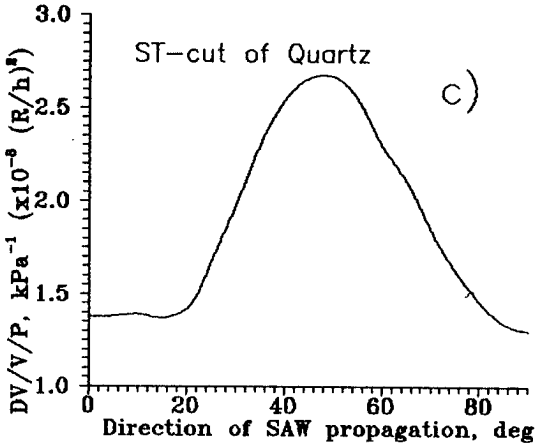


Fig.3a,b. a) Numerical sensitivity to pressure loading as a function of cut angle of α- quartz circular plates. b) Numerical sensitivity to pressure loading for Z-cut of quartz as a function of direction of SAW propagation.



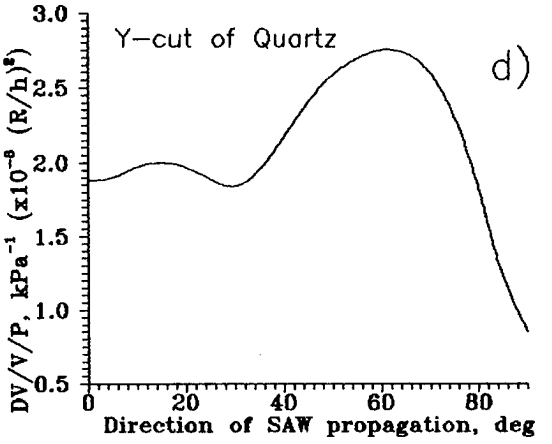


Fig.3c,d. Numerical sensitivity to pressure loading for ST and Y-cuts of quartz as a function of direction of SAW propagation, respectively.

at the coordinates: one of them has coordinates (0,R), and another- (R,0). It should be mentioned that at this direction the extremal sensitivity value is insensitive to small errors of location of the centers of SAW resonators on the surface of plate. This cut is more suitable for sensor application because the direction of Ψ= 47 degrees is the SAW temperature compensated direction. The maxima value of sensitivity in the Y-cut is realized in the direction of SAW propagation of Ψ=60 degrees where the electromechanical coefficient of SAW is negligible.

It should be noted that the mean (true) sensitivity of a sensor are usually obtained by averaging over the path length and beam width of real structures of two SAW devices. Nevertheless to find the extremally high sensor sensitivity cuts in quartz we assume that a quasiponctual SAW propagation is sufficient suggestion for our purpose. But in the case of a SAW resonator structures this problem may be more complicated because necessary to take into account a spatially dependence of SAW reflection coefficient from groove or strip periodic structures on the surface of plate subjected to inhomogeneous bias. We believe that influencing of inhomogeneousity of biasing field distribution on a change of SAW resonator frequency shift is small in our experimental devices.

For verification of our calculation results we compared them with experimental characteristics of different pressure sensors with rotated Y-cut quartz membranes. Each membrane contains two SAW resonators (see Fig. 1). One of them is placed in the center of the membrane. The SAW propagates in the X axis direction. The other resonator is located at some distance from the former one with the same SAW propagation direction. Their relative frequency shifts depend on external pressure. Calculated and experimental difference of these relative frequency shifts and some other membranes parameters is represented in Table 1.

Dimensions of resonator electrode structure which frequency is equal to 360 MHz are 0.5x4 mm. The same parameters of

Table I.

Comparison between experimental and culculated pressure sensitivity of qwartz membranes

Membrane, N	1	2	3	4	5
Thickness, mm	2	2	1.1	0.38	0.25
Diameter, mm	9	9	12	12.7	10
Distance between resonator centres, mm	4	4	4.8		3.8
Resonator central frequencies, MHz	310	310	360	130	105
Substrate , Quartz	Y-cut	ST,X - cut	ST,X - cut	Y-cut	Y-cut
Experimental relative sensitivity, (1/Pa)*10 $^{-12}$	60	20.2	90	2700	1400
Theoretical relative sensitivity, (1/Pa)*10 $^{-12}$	75.4	24	100	2770	1300

resonator with frequency 310 MHz are - 0.5x5 mm. Membranes of number 1, 2, 3 are fabricated in our laboratory. For membranes of number 4, 5 we used experimental data from review [12]. Because of absence of information about the distance between resonator centers of membrane № 4 we compare here only predicted and measured relative frequency shifts of central resonator. Table 1 shows a good agreement between numerically calculated and experimental data.

Conclusion

A simple model have been presented to predict a pressure sensitivity of SAW resonator located on any point of surface of quartz circular plates. This allows us to obtain an extremal pressure sensitivity and location coordinates of two SAW resonators for dual channel differential sensor. Theoretical results are verified by experimental data obtained for different cuts of α-quartz. Experimental tests are going on to confirm these theoretical results. It has been found that ST-cut membrane with the direction of SAW propagation of Ψ=50 degrees have to be high pressure sensitivity and temperature stability simultaneously.

References

[1] B.K.Sinha, W.J.Tanski, T.Lukaszek, and A.Ballato," Influence of biasing stresses on the propagation of surface waves,"J.Appl.Phys.,vol. 57, pp. 767-776, Feb.1985.

[2] G.Theobald, D.Hauden," Theoretical analysis of dynamic thermal effects in SAW devices,"Proc.38th Ann.Freq. Contr. Symp. IEEE, cat.84CH2068-8, 1984, pp. 294-299.

[3] S.Ballandras, E.Bigler," Experimental study of SAW quartz cuts with low sensitivity to dynamical thermal effects," Electron.lett.,vol.27, pp. 1343-1345, July 1991.

[4] B.K.Sinha," A stress and temperature compensated orientation and propagation direction for surface acoustic wave devices," IEEE Trans.Ultrason. Ferroelect. Freq.Contr. vol. UFFC-34, pp. 64-74, Jan.1987.

[5] E.Bigler,G.Theobald, and D.Hauden," SAW quartz cuts with low stress and temperature sensitivity," Elect.lett. ,vol.23, pp.514-516,May 1987.

[6] B.K.Sinha and S.Locke," Acceleration and vibration sensitivity of SAW devices,"IEEE Trans.Ultrason. Ferroelect. Freq.Contr.,vol. UFFC-34,pp. 29-38, Jan.1987.

[7] S.Ballandras and E.Bigler," New results for the characterization of SAW sensitivity to mechanical stresses using Finite element analysis," IEEE Proc. 1991 Ultrason. Symp., vol. 1, December 1991.

[8] H.F.Tiersten and B.K.Sinha," A perturbation analysis of the attenuation and dispertion of surface waves," J.Appl.Phys., vol. 49, pp. 87-95, Jan.1978.

[9] H.F.Tiersten and J.C.Baumhauer," Nonlinear electroelastic equations for small fields superposed on a bias," J. Acoust. Soc. Amer., vol. 54, pp. 1017-1034, Oct.1973.

[10]S.G.Lehnitsky," The solution of a plane- stress problem for anisotropic elastic ellipsoidal plate," Reports of Academy of Sciences of the USSR, vol. 15, pp. 527-530, 1937.

[11]R.Bechmann,A.D.Ballato,and T.J.Lukaszek," Higher order temperature coefficients of the elastic stiffness and compliances of alpha-quartz," Proc. IRE, vol.50, pp.1812-1822, 1962.

[12]W. W. Malov, Piezoresonance transdusers. Moskow, Atomenergoizdat, 1989, pp. 245-249.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

PRECISION SAW PRESSURE SENSORS

YU.N.VLASSOV, A.S.KOZLOV, N.S.PASHCHIN, I.B.YAKOVKIN

Institute of Semiconductor Physics, Siberian Branch of the Russian Academy of Sciences, 13 Lavrentyev Avenue, 630090 Novosibirsk, Russia.

ABSTRACT

The design and performance of two types of SAW pressure sensors have been considered. SAW rezonator oscillators with metal strip gratings designed for operation at 200 and 360 MHz have been used in both type devices. The sensitivity and the nonlinearity of both sensor with a sensitive element shaped as an integral crystal diaphragm and sensor with mechanically loaded quartz plates are measured and discussed. Recomendations are made for optimal configuration of SAW pressure sensor according to pressure and dynamic range demands.

INTRODUCTION

Due to their high sensitivity to the infinitesimal deviation of the propagation conditions surface acoustic waves (SAW) are very suitable for the design of the whole class of solid state miniature sensor that main features: sensitivity, dynamic range, precision, selectivity, reaction time and so on are correspondent to the modern standards. The fabrication based on microelectronic technology and the easy compatibility with digital information systems make such sensors a very interesting objects for further investigations.

To estimate quantitively the sensitivity of SAW-sensor or its transfer function one must solve a differential equations system, that describes a mechanical displacement of the material particle near solid state surface, shifted from the equlibrum state by a measurand. It is desirable to have opportunity for the influence estimation of nonuniform biasing field both in a space and in a time.

The deviations of the propagation conditions resulting, for example, in the distance or phase velocity disturbances, when external forces are applied to the SAW device substrate, change significantly such output parameters as phase response or delay time. Recently, some reports have appeared that describe high perfomance pressure sensors based on SAW resonator oscillators with a one or two acoustic channels [1-4]. Such sensors are very promising because of their reported sensitivity and their adaptability to a broad spectrum of applications and pressure ranges.

In this paper we present the design and realization of SAW osciilator system for pressure sensor applications. The SAW resonator devices have been fabricated in quartz substrates and the electronic circuitry including two channels amplifier and mixer has been designed as a special plate.

THE SENSITIVE DEVICE

THEORY

If we determine the SAW pressure sensor sensitivity as a relative phase change caused by the single pressure, we can express the sensitivity by two terms:

$$S=(d\phi/\phi)/dp = (dl/l - dv/v)/dp$$

Here dl/l is the substrate deformation along SAW propagation direction and dv/v is the phase velocity deviation. Unfortunately, the exact estimation of the SAW pressure sensor sensitivity or its transfer function determination may be accomplished only as a result of the simultaneous decision of a some differential nonlinear equation complicated systems. The general problem of the determination of the SAW propagation characteristics demands the solution of the equations for nonlinear solid state mechanics, i.e. introduction of material (Lagrange) coordinate system with the subsequent transformation from the material to space frame description. An initial nonlinear system was linearized [5,6] and divided into two parts. The first of them is a statical part, that describes the static mechanical field distributions on the substrate. The second is responsible for dynamic (SAW) deviations and the values of material constants introduced into dynamic equation coefficients became the functions of the static biasing field.

As the first step we determined the static strain distribution in the SAW pressure sensor sensitive element consisted of the thin circular quartz diaphragm, having radius (R) and thickness (H), whose edges are claimed. We used the isotropic approximation and suggestion that strains are not so strong variated on the SAW device length a = l/R,

0-7803-0905-7/93 $3.00 © 1993 IEEE

where l is the resonator cavity dimension. Fig. 1. illustrates the calculated static average lengthly strain distribution as the function of the SAW device displacement (y = x/R) from the diaphragm center. Calculations were made for three different SAW device length (a = 0.2, 0.5, 0.8). All nondiagonal strains are zero if the SAW device position is symmetrical to the plane passing through the diaphragm center. All diagonal strains have their maximum magnitude in the center of diaphragm. One of them (11) decreases monotonically to zero, another two change their sign in the point y =0.5 -0.7 when SAW device displacement increases to y = 1.0. The opposite sign maximum values of these strain occur at y = 0.85 - 0.99 for different a. It is clear from Fig. 1 that maximum sensitivity can be achieved if SAW device is situated in the diaphragm center or near its edge and it has the minimal length. But we must take into account that resonator length give us the frequency stability.

The phase velocity disturbations can be determined only numerically. It is impossible now to get a simple interpretation for the calculations that we have accomplished.

DESIGN AND FABRICATION

The first type SAW pressure sensor has a sensitive element shaped as an integral quartz diaphragm with two SAW resonator's structures on its surface. The diameter and thickness of the diaphragm depend on the demand of dynamic and pressure range. One resonator is in the diaphragm center. The second device is placed in a parallel direction by a some distance from the first one. SAW propagation direction is coinside with X-axis direction. Both resonators are determining the frequencies of the SAW oscillators in the each channel. When the pressure is applied to the diaphragm its surface get the concave form. Such propagation condition deviations result in a central frequency shift. Each SAW resonator is placed on the diaphragm surface in such manner that frequency shift due to applied pressure has an opposite sign. Both SAW oscillator's output signals are mixed and difference frequency, that usualy lies in the range from 100 to 1000 KHz, is proportional to applied pressure and slightly deviated by the temterature. In this two resonator configuration the temperature deviation of the output frequency is more smaller than similar deviation within each resonator channel. We used a thin film aluminium thermoresistor deposited on the diaphragm surface for the further compensation of the output signal temperature deviation. Both resonators and thermoresistor, all etched by the same photolithographic process in a 0.15mkm aluminium film. A special processor to calibrate output frequency - pressure response accounting the temperature deviation was used.

The construction of the second type SAW pressure sensor we studied is shown schematically in Fig. 2. In this sensor the pressure to be measured is applied to the intermediate membrane (1) to bend it. The displacement of the membrane center through the steel rod (2) transmites to the acoustoelectronic sensitive element (3). It is ST,X-cut quartz substrate having its dimensions 15 x 8 x 0.3 mm^3 , that is leaned along its short edges on the rectangular hole. The hole dimensions are 12 x10 mm^2 .The SAW resonator and thin film thermoresistor are fabricated on the plate surface. Under the influence of pressure the plate is deformed and the resonator frequency changes. Its frequency shift and the thermoresistor value are processed in the same manner as in the previous sensor. The second oscillator uses a mechanically unloaded SAW resonator as a frequency control element. The material choosing is due to its small temperature coefficient and high sensitivity of SAW velocity to the static strains.

The short description of the fabrication sequence of SAW pressure sensor sensitive element is the next:

-the basic material is Y-cut (or SΓ-cut), X direction quartz, optically polished

-10 nm thick vanadium layer, and

-300 nm thick aluminum layer, deposited by vacuum evaporation

- etching of the aluminum pattern

DEVICE PERFOMANCE

Attenuation and phase versus frequency scans of the each sensitive element were obtained using a vector voltmeter equipped with a frequency synthezator directed by the personal computer. The results of such a measurement are given in Fig.3 (a - d). Each resonator has three resonant peaks on its frequency response, and the phase in each feedback loop is adjusted in such manner, that the resonant frequencies of the medium and edge oscillator, respectively, differ by the frequency distance between the peaks from each other. The central frequencies of the resonators are about 197 MHz.

The main features of three different samples concerning to the first type SAW pressure sensor are given in Table 1. These parameters are measured at room temperature. The correspondent responses are shown in Fig. 4 and Fig. 5. Fig.

Table 1

Parameters	Sensor 5	Sensor 7	Sensor 9
Diaphragm diameter, mm	30	30	30
Diaphragm thickness, mkm	200	200	200
Pressure range, KPa	0 - 16	0 -16	0 - 16
Sensitivity, Hz/Pa	4.00	3.91	3.42
Nonlinearity, %	0.56	0.89	0.92
Mean error, %	0.28	0.22	0.25
Threshold pressure (at 10 Hz accuracy), Pa	2.5	2.6	2.9
Central frequency, MHz	197.1	197.4	197.2

4 presents the difference frequency(kHz) - pressure(kPa) response for three experimental sensors with a Y-cut quartz diaphragm. Their measurement accuracy at constant temperature is equal to 0.4%. The sensitivity varies from 3.4 to 4.0 kHz/kPa and the nonlinearity of the response doesn't exceed 0.5%, as it can be seen from Fig. 5, where nonlinearities of the same three sensors are shown.

The main feature of second type SAW pressure sensor is a strain transformation due to the fact that dimensions of the membrane and quartz element are strongly different while their center displacement is the same. Thus we have an opportunity to increase sensor sensitivity by using of the large diameter intermediate membrane. We have studied the next second type sensor:

Sensor No 2, it has the vacuum steel membrane;
Sensor No 3, has the melted quartz membrane. Its diameter is 40 mm and its thickness is 300 mkm.

Fig. 6 and Fig. 7 show the relationship between the frequency shift and the magnitude of the measured pressure at constant temperature. The main discrepancies of these sensors are the higher sensitivity and higher hysteresis as compared with the first type sensor. The reason of the such hysteresis value is the material property, that is not good as monocrystal quartz.The application of the membrane made from special steel with low hysteresis allows, to our oppinion, to decrease this magnitude to 0.1%, the lowest boundary of the hysteresis in this type sensor. The main properties of two second type sensor are given in Table 2.

Table 2

Parameters	Sensor No 2	Sensor No 3
Sensitivity, Hz/Pa	9.7	27
Pressure range, KPa	100 - 116	0 - 20
Hysteresis. %	0.15	0.15
Pressure threshold (at 10 Hz frequency accuracy), Pa	1	0.4

In the all cases the measured pressure was controlled by the water manometer. The error of pressure determination does not exceed 0.15%. The difference frequency sensitivity to the oscillator supply voltage vas estimated as 4 kHz/V. The voltage stability as large as 20 - 50 mV was suffucient for all our measurements.

It is well known that bulk wave oscillator frequency can be measured up to 0.1 -0.01 Hz. We can measure the difference frequency of our sensor up to 10 - 100 Hz because of different reason (electromagnetic influence, acoustic noise, etc.). The minimal threshold is achieved in the sensor No 3. It is equal 0.4 Pa.(The correspondent frequency accuracy is 10 Hz). This magnitude is higher than another published results for the same or another type pressure sensor.

CONCLUSIONS

SAW pressure sensors we studied have such attractive properties as high sensitivity, low threshold, design and technology simplicity, noncomplicated electronic equipment without any thermocompensation, that doesn't influence on sensor stability, small dimensions and frequency output.

They have the higher temperature and long time zero drift as compared with bulk wave resonator pressure sensor. The strain transformatoin sensors have the additional disadvantage. It is a high level hysteresis.

The experimental data allow us to conclude that the sensitivities of both types of SAW pressure sensors is sufficiently high and they could be used in pressure measurement equipment.

REFERENCES

1. P.Das, C.Lanzl, and H.F.Tiersten. A pressure sensing acoustic surface wave resonator.-1976 IEEE Ultrasonics Symp. Proc. NY, 1976, p.306 - 308.
2. D.E.Cullen, G.K.Montress. Progress in the development of SAW resonator pressure transducers.-1980 IEEE Ultrasonics Symp. Proc. NY, 1981, p. 696 - 701.
3. D.Hauden, M.Planat, J.J.Gagnepain. Nonlinear properties of surface acoustic waves. Application to oscillators and sensors.- IEEE Trans on Sonics and Ultrason., 1981, v. SU-28, No 5, p.342 - 348.
4. N.Pashchin, I.Yakovkin. A design principles for SAW pressure sensors. - Vibrotechnique (selection from higher school transaction), Vilnus, 1986, No 2(53), p.85 - 91, (in Russian).
5. J.C.Baumhauer, H.F.Tiersten. Nonlinear electroelastic equations for small fields superimposed on a bias.-Journ. Acoust. Soc. Amer., 1973, v.54, No 3, p.1017 -1034.
6. R.Brendel. Material nonlinear piezoelectric coefficient for quartz.- Journ. Appl. Phys., 1983, v.54, No 9, p.5339 - 5346.

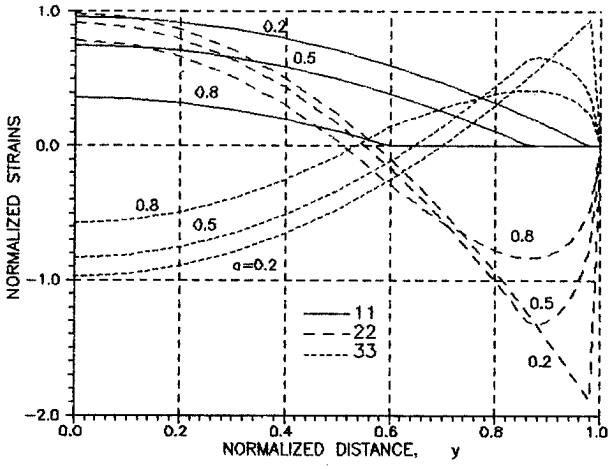


Fig. 1. Average lengtly normalized strains versus SAW resonator normalized displacement plot for the circular diaphragm. Parameter a is the normalized resonator cavity length.

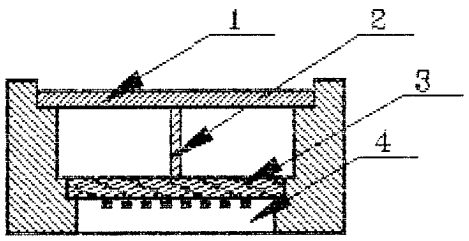
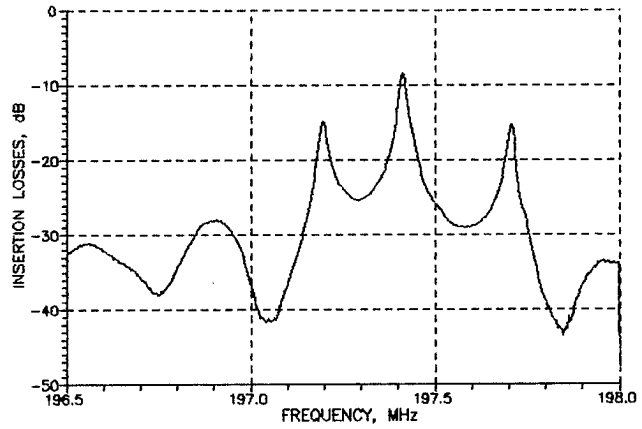
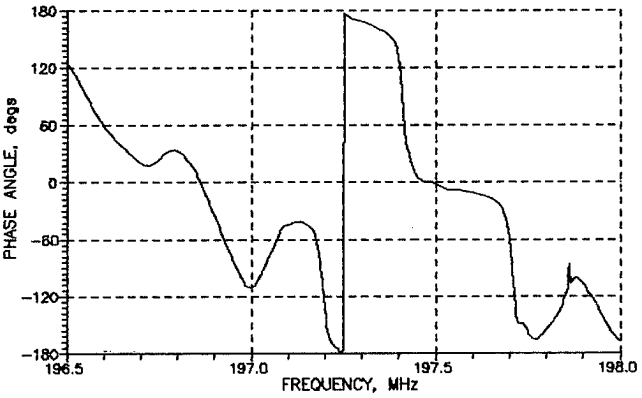


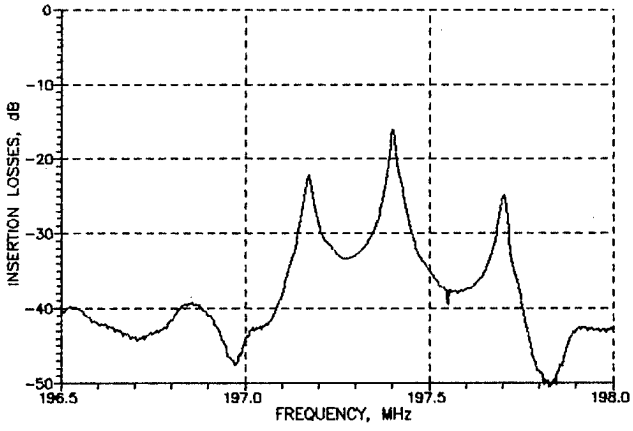
Fig. 2. Second type SAW pressure sensor schematic wiev.



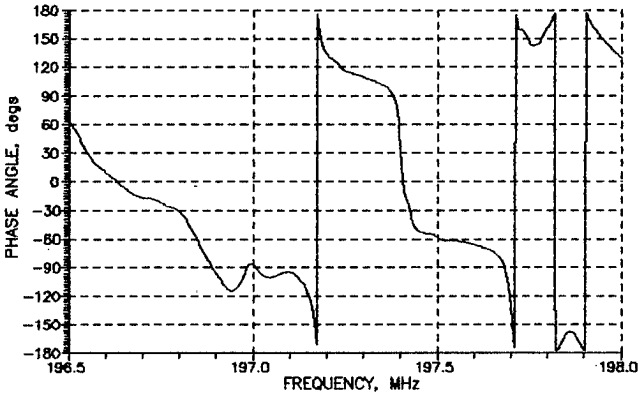
a)



b)



c)



d)

Fig. 3. Amplitude (a,c) and phase (b,d) versus frequency scans for central (a,b) and edge (c,d) resonators.

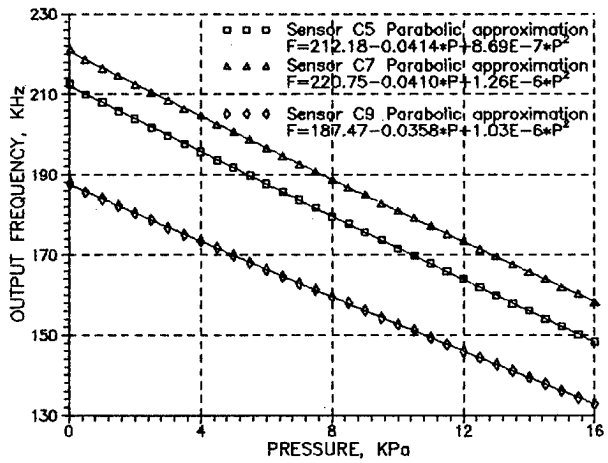


Fig. 4. Frequency shift dependence on pressure for the sensor with a monocrystal quartz diaphragm.

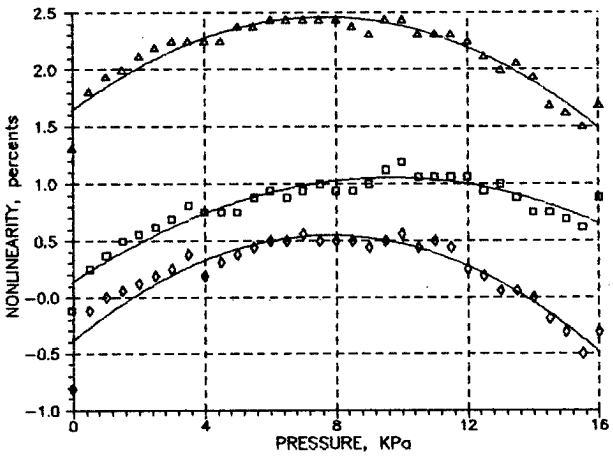


Fig. 5. Nonlinearities of the sensors characteristics, shown on fig. 4.

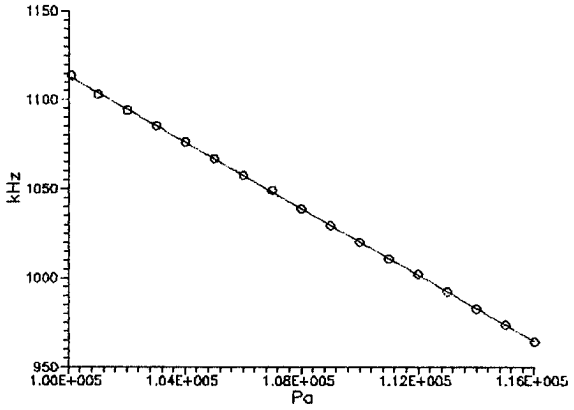


Fig. 6. Frequency shift dependence on pressure for the sensor with a vacuum steel membrane.

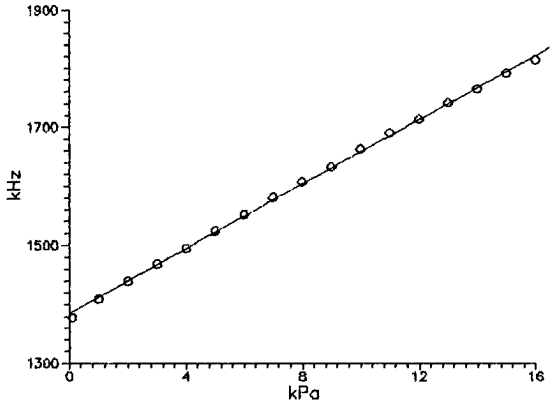


Fig. 7. Frequency shift dependence on pressure for the sensor with a melted quartz membrane.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Voltage Controlled S-TCXO's Employing NS-GT Cut Quartz Crystal Resonators

Hirofumi KAWASHIMA and Kenji SUNAGA

Seiko Electronic Components Co., Ltd.
Tochigi-shi Japan 328

ABSTRACT

This paper describes the frequency temperature behavior of S-TCXO's(Self-Temperature Compensated Crystal Oscillators) employing a new shape GT cut quartz crystal resonator(NS-GT cut resonator) controlled by voltage. An object of this paper is to clarify the frequency temperature behavior of voltage controlled S-TCXO's employing a NS-GT cut resonator. Another object of this paper is to show aging characteristics of further long term for about two years and three months of S-TCXO's.

First, the circuit construction of a voltage controlled S-TCXO(VC-S-TCXO) is represented, which consists of a CMOS inverter, feedback resistance Rf, gate and drain capacitances Cg and Cd consisting of Cd_1 and Cd_2 of the VC circuit, drain output resistance Rd, capacitance Cv of variable capacitor diode, resistance Rc and a NS-GT cut resonator. In addition, S-TCXO's are so designed that the frequency deviation in oscillation is within +/-2.5ppm in a wide temperature range of -30 to +85 °C.

Next, relationships of frequency deviation in oscillation and frequency temperature behavior versus voltage change ΔVc are represented, so that the VC-S-TCXO's employing a NS-GT cut resonator are found to be successfully obtained with frequency deviation within $+/-4\times10^{-9}$ directly after the voltage control in a wide temperature change of -30 to +85°C.

Finally, aging characteristics for about two years and three months are described. Frequency shift of eight samples is from 0.061ppm to 0.146 ppm(mean:0.098ppm) for about one year and three months after the pre-aging for the about one year.

§1. Introduction

For use in communication equipment and consumer products as a source of frequency standard, a quartz crystal oscillator excellent in frequency temperature behavior and simultaneously, its miniaturization and low consumption power have been also required. As an oscillator to satisfy these requisites, it was reported[1] at the last symposium by the authors that a S-TCXO using a NS-GT cut resonator[2],[3] is optimum because it shows very good frequency temperature behavior without external temperature compensation by thermister and so forth, and also because it has a little current consumption of approximately 500 μA. Specifically, a S-TCXO with frequency deviation of +30/-150ppb over a wide temperature range of -30 to +85°C was successfully obtained at best without any external temperature compensation. In addition, thermal hysteresis and aging for about one year of the S-TCXO's were also introduced.

In this paper, the frequency temperature behavior of a voltage controlled S-TCXO (hereafter we call it VC-S-TCXO) is described. The VC-S-TCXO where comprises a NS-GT cut resonator is performed utilizing a S-TCXO with frequency deviation being within +/-2.5ppm over a wide temperature range of -30 to +85°C. The objectives of this paper are to clarify the frequency temperature behavior of VC-S-TCXO's employing a NS-GT cut resonator and also to show aging characteristics of further long term for about two years and three months of S-TCXO's

A circuit construction is, first, represented. In particular, various relationships of capacitance Cv of a variable capacitor diode versus control voltage Vc in a circuit that is called the "VC circuit", is examined in detail. As a result, the capacitance-temperature behavior in the VC circuit is found to be hardly affected by a change of control voltage Vc of 1 to 3V. Furthermore, a VC-S-TCXO is fabricated under the above condition, a relationship of frequency deviation $\Delta f_+ / f_+$ in oscillation versus control voltage Vc is measured accurately and the relationship is also approximated theoretically by an equation of the high degree. An error of the predicted and measured values is shown, that is to say, the error determins probably the precise of the VC-S-TCXO.

0-7803-0905-7/93 $3.00 © 1993 IEEE

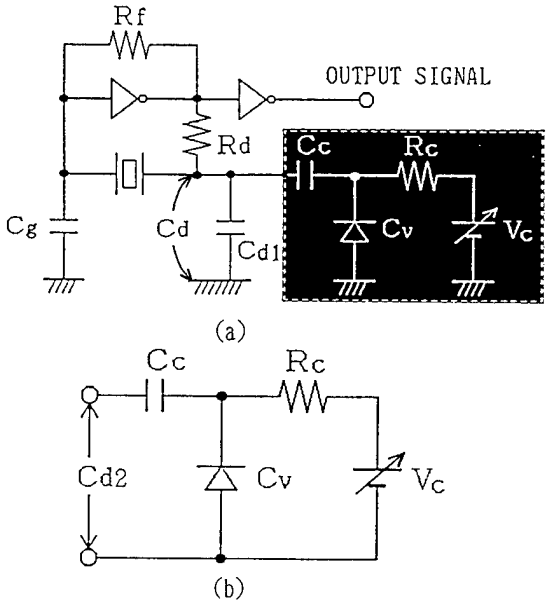


Fig. 1 Circuit construction(a) of VC-S-TCXO (voltage controlled S-TCXO) and the VC circuit(b).

Finally, some examples of frequency temperature behavior for VC-S-TCXO's are shown, so that the VC-S-TCXO's are found to have excellent frequency stability more than 4×10^{-9} directly after the voltage control in an absolute value over a wide temperature range of -30 to $+85$℃.

§2. Circuit Construction

Figure 1 represents a circuit construction(a) of a voltage controlled S-TCXO(VC-S-TCXO) and a VC circuit(b). The VC-S-TCXO consists of a CMOS inverter, feedback resistance Rf, gate and drain load capacitances Cg and Cd comprising Cd_1 and Cd_2 of the VC circuit, drain output resistance Rd and a NS-GT cut resonator. Also, the VC circuit consists of capacitance Cv of a variable capacitor diode, capacitance Cc, resistance Rc and control voltage Vc, as is shown in Fig. 1(b). Accordingly, capacitance Cd_2 is given as

$$Cd_2=\frac{CvCc}{Cv+Cc} \tag{1}$$

From eq. (1) Cd_2 and its temperature coefficient are derived theoretically. Total drain load capacitance Cd is, therefore, given as

$$Cd=Cd_1+Cd_2 \tag{2}$$

An oscillation frequency in the VC-S-TCXO can be easily calculated from the electrical equivalent circuit as well as the way in the references[4], [5]. The derivative procedure of the oscillation frequency is simply described here.

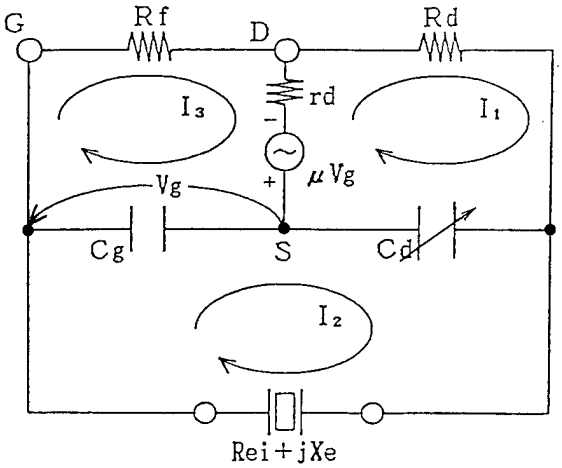


Fig. 2 Electrical equivalent circuit of VC-S-TCXO.

Figure 2 represents an electrical equivalent circuit of the VC-S-TCXO shown in Fig. 1(a). A CMOS inverter can be replaced by drain resistance rd and voltage source μVg (μ: amplification ratio, Vg:gate voltage). Electric currents I_1, I_2, I_3 flow in each loop. From the circuit an oscillation frequency f_+ for principal vibration is given as

$$f_+=\left[-\frac{1}{2}-\left\{(f_1'^2+f_2'^2)+\left((f_1'^2-f_2'^2)^2+4Ks^2f_1^2f_2^2\right)^{1/2}\right\}\right]^{1/2}$$

where, (3)

$$f_i'^2=\left(1+\frac{1}{r_i}-\frac{a}{1+a-h}\right)f_i^2$$

($\because i=1$ for principal vibration and 2 for subvibration)

$a=C_0/C_L$, $b=C_m/C_L$

$r_i=C_0/C_i$ $(i=1,2)$

$$Ks^2=Km^2\left(1+\frac{b}{1+a-h}\right)^2$$

$Km^2=C_1C_2/C_m^2$

$h=(1+\mu)/[4Rf(rd+Rd)\pi^2f_+^2CgCd]$

f_1, f_2: Uncoupled frequencies of principal- and sub-vibrations

C_0: Shunt capacitance

C_m: Coupling capacitance

C_1, C_2: Motional capacitances of principal- and sub-vibrations

and then load capacitance C_L is

$$C_L=\frac{1-\frac{C_0}{Cd}\frac{2R_i}{rd+Rd}}{\frac{1}{Cg}+\frac{1}{Cd}\left(1+\frac{R_i}{rd+Rd}\right)}$$

R_1, R_2: Series resistances of principal- and sub-vibrations

As one can understand from eq. (3), an oscillation

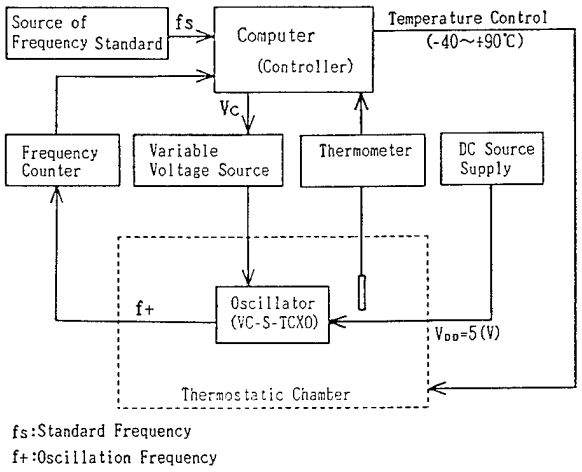


Fig. 3 Block diagram of a method for controlling voltage Vc

frequency f_+ is given as a function of load capacitance C_L. An oscillation frequency for the VC-S-TCXO can be adjusted by controlling drain capacitance Cd.

§3. Control method of voltage

Figure 3 represents a block diagram of a method for controlling voltage Vc. An output frequency of a VC-S-TCXO is compared with a standartd frequency of base station. When there takes place the difference between both frequencies, Vc is so controlled that the difference reaches zero approximately.

Table 1 represents electrical equivalent circuit parameters of two NS-GT cut resonators that are used in the VC-S-TCXO's. A series resistance R_1 and a quality factor has a value of 54.8 and 55.0Ω, and 263,000 and 264,000, respectively. These values are sufficiently satisfactory as a resonator. In addition, capacitance ratio r_1 which plays an important role in this paper, is 493 and 496. Frequency deviation to load capacitance C_L get enough large because capacitance ratio r_1 is sufficiently small.

§4. Results and Discussion

The results for VC-S-TCXO's employing NS-GT cut resonators are discussed according to the following procedure. First, a relationship of capacitance Cv of a variable capacitor diode and its first order temperature coefficient versus control voltage Vc is shown. Second, an examination on a VC circuit is performed, in particular, the VC circuit are so designed that it is mostly independent on control voltage Vc.

Table 1 Values of electrical equivalent circuit parameters used in the VC-S-TCXO's.

No.		f (MHz)	$R_{1,2}$ (Ω)	$L_{1,2}$ (H)	$C_{1,2}$ (fF)	C_0 (pF)	$r_{1,2}$	$Q_{1,2}$ ($\times 10^3$)
1	Prin.	2.0997	54.8	1.09	5.26	2.59	493	263
	Sub.	1.8941	353	2.76	2.56		1012	93
2	Prin.	2.0997	55.0	1.10	5.23	2.59	496	264
	Sub.	1.8942	258	2.72	2.60		996	125

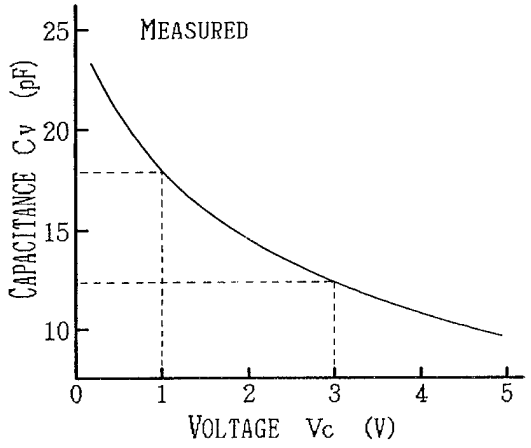


Fig. 4 Capacitance Cv of variable capacitor diode versus voltage Vc.

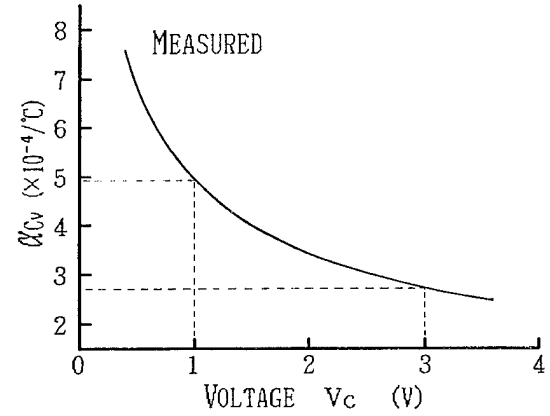


Fig. 5 The first order temperature coefficient $\alpha_{cv}(\dot{C}v/Cv)$ of variable capacitor diode versus voltage Vc.

Third, an experimental and theoretical relationship of an oscillation frequency f_+ versus control voltage Vc is examined and the error of both relationships is found to influence markedly the accuracy of the VC-S-TCXO's.
Fourth, VC-S-TCXO's measured actually are shown when temperature is varied from −30 to +85 ℃. Finally, aging characteristics of long term over two years and three months are also shown.

4.1 Variable capacitor diode

Figure 4 represents a relationship between capacitance Cv of a variable capacitor diode and

control voltage Vc. As Vc is varied from 1 to 3 V, Cv decreases. The variation of Cv becomes large when Vc is small.

Figure 5 represents a relationship of the first order temperature coefficient α_{cv} ($=\dot{C}v/Cv$, $\cdot\equiv\partial/\partial T$ T:temperature) of Cv versus Vc. As is apparent from Fig. 5, it is easily understood that α_{cv} is very dependent on Vc. Therefore, such design of the VC circuit is naturally required as capacitance Cd_2 is independent upon temperature when Vc is varied. The variation of Vc=1 to 3 V causes that of $\alpha_{cv}=4.95$ to 2.70×10^{-4}/℃. It is needless to say that the variation of α_{cv} has slight influence upon the precision of VC-S-TCXO's.

4.2 VC circuit

As described in the previous Article 4.1, it is, therefore, of great importance to design the VC circuit where capacitance Cd_2 is independent upon temperature. The design of an independent circuit upon temperature can be performed by selection of capacitance Cc connected in series to Cv. The condition of a circuit independent upon temperature is calculated from eq. (1).

Figure 6 represents a relationship between the first order temperature coefficient α_{cd2} of Cd_2 and Cc whose first order temperature coefficient α_{cc} is from -8.43×10^{-4}/℃ to -8.54×10^{-4}/℃ slightly dependent on the value of capacitance Cc, when Vc is varied from 1 to 3 V. The increase of Cc causes a change for α_{cd2} of a negative value to a positive value. When Vc is 2 V, α_{cd2} reaches zero at Cc=36.04pF. As one can see from Fig. 6, α_{cd2} varies from 0.308×10^{-4}/℃ to -0.107×10^{-4}/℃, when Vc is changed from 1 to 3 V at Cc=36.04pF. These values are negligibly small.

Figure 7 represents capacitance-temperature behavior of Cd_2, when Vc is varied from 1 to 3 V. As is obvious from Fig. 7, capacitance deviation of Cd_2 to voltage change of 1 to 3 V is less than 0.25 percent in a temperature range of −30 to +85℃. Thus, the VC circuit is successfully obtained, which is mostly independent upon the temperature.

In addition, as was reported in the references [4], [5], the first order temperature coefficient α_+ for a NS-GT cut resonator is very dependent on the largeness of a value of load capacitance C_L. Such a value of C_L must be, therefore selected as α_+ is independent upon a change of C_L so as to raise the precision by control voltage Vc.

Figure 8 represents a relationship between variation $\Delta\alpha_+$ of α_+ and load capacitance C_L for a NS-GT cut resonator. It is easily understood that α_+ is heavily dependent on C_L when C_L has a very small value such as 10 to 4pF, but when C_L is

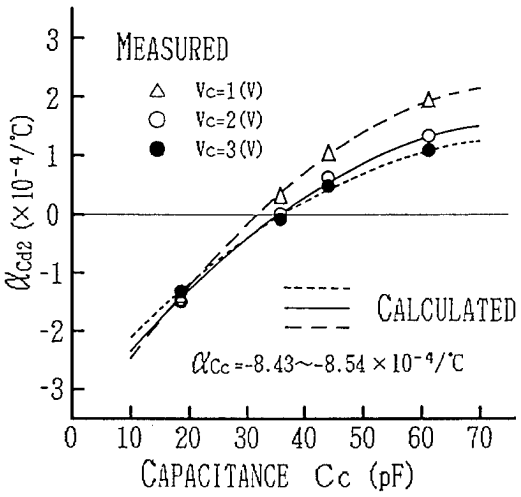


Fig. 6 Relationship between $\alpha_{cd2}(\dot{C}d_2/Cd_2)$ and Cc when Vc is taken as a parameter.

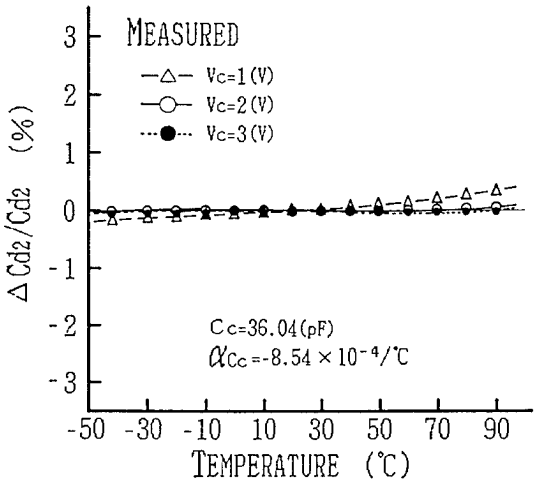


Fig. 7 Capacitance-temperature behavior when Vc is taken as a parameter.

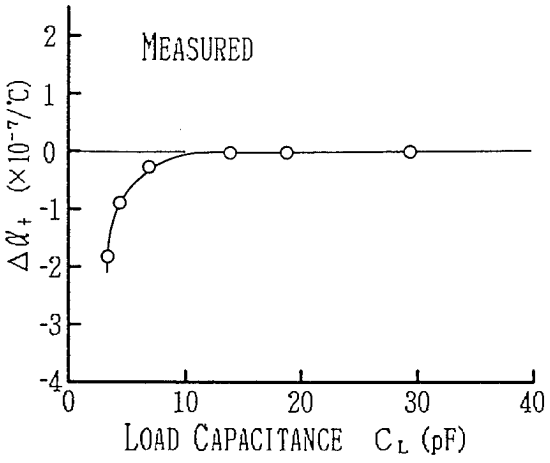


Fig. 8 Relationship between $\Delta\alpha_+$ of α_+ and C_L.

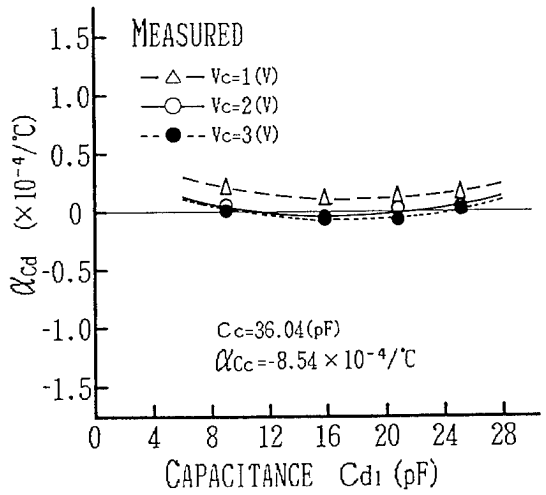


Fig. 9 Relationship between the first order temperature coefficient α_{cd}($\dot{C}d/Cd$) of Cd and Cd_1 when Vc is taken as a parameter.

larger than 15pF, α_+ is mostly independent upon C_L. From this result, C_L is preferably selected with a value larger than 15pF.

Figure 9 represents a relationship between the first order temperature coefficient α_{Cd} of Cd and Cd_1 whose first order temperature coefficient α_{Cd1} is 9.6×10^{-6}/℃ at maximum for Cc=36.04pF and $\alpha_{Cc}=-8.54\times10^{-4}$/℃, when Vc is varied from 1 to 3 V. As was shown in Fig. 1, Cd_1 is connected in parallel to Cd_2 of the VC circuit.
A variation of α_{Cd} is less than 1.2×10^{-5}/℃ to a change of Cd_1=9.00pF to 25.15pF. The variation is found to be very small.

4.3 The error by control voltage Vc

Figure 10 represents an experimental relationship between frequency deviation of oscillation and control voltage Vc for a VC-S-TCXO with Cg=29.57pF, Cd_1=19.81pF and Cc=36.04pF so designed that the frequency deviation to the nominal frequency is zero at Vc=2 V(Cd=30.36pF at Vc=2 V). This relation is shown by the circles and when Vc is varied from 3 to 1 V, frequency deviation is within +3/−4ppm.
Therefore, in this case, Vc makes it possible to control the frequency deviation $\Delta f_+/f_+$ within +/−2.5ppm that is principally caused by temperature variation. In order to get the frequency deviation larger than that for the above condition, a small value of Cd is needed, and the small value is performed by Cd_1. The black circles represent the relationship for Cd_1=17.97 pF, Cg=31.55pF and the same VC circuit as the previous one (Cd=28.52pF, C_L has a constant value of 15pF that is the same value as the previous C_L). In this case, when Vc is varied from 3 to 1 V, the frequency deviation is within +4/−5ppm, the deviation is found to be large.

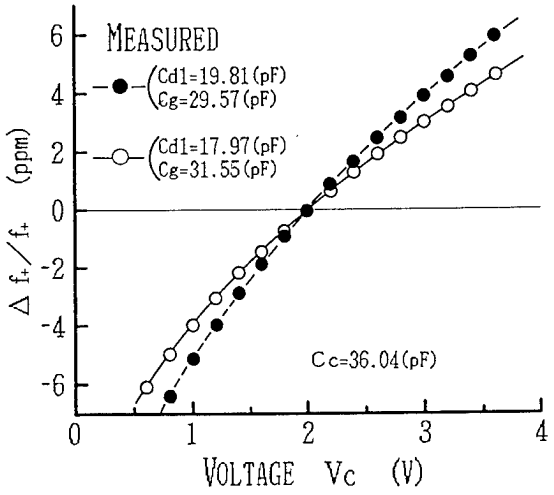


Fig. 10 Frequency deviation $\Delta f_+/f_+$ versus control voltage Vc.

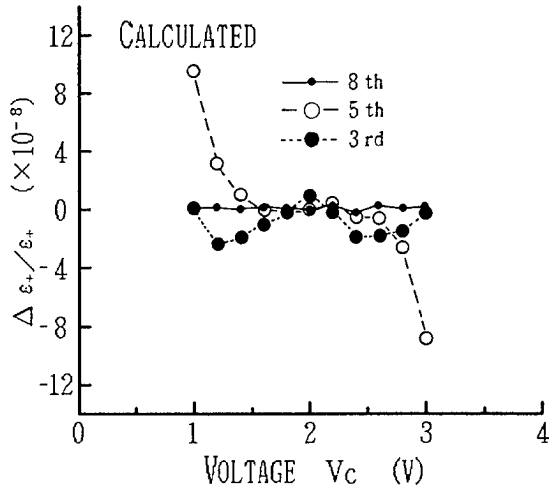


Fig. 11 Relationship of the error between the measured and calculated values versus voltage Vc.

Figure 11 represents an error between the measured values of oscillation frequency and the values calculated from the measured values, when Vc is varied from 1 to 3 V. As is shown in Fig. 11, the error $\Delta\varepsilon_+/\varepsilon_+$ differs remarkably by the degree of the equation used. As one can see from Fig. 11, the equation of the eighth degree gives a very small error of the measured and calculated values. The error is actually less than 4×10^{-9}. As a consequence, a VC-S-TCXO can be expected with excellent frequency stability to temperature by controlling Vc.

Figure 12 represents frequency temperature behavior of a VC-S-TCXO employing a NS-GT cut resonator that has a comparatively large value for the first order temperature coefficient α_+ at +25℃. Since the VC circuit is designed mostly independent upon temperature, even when Vc is

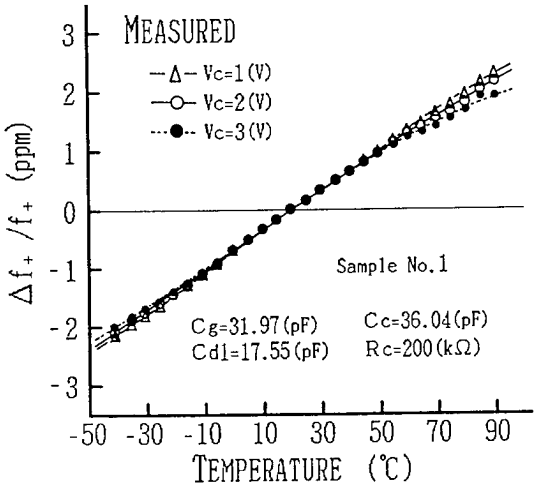


Fig. 12 Frequency temperature behavior.

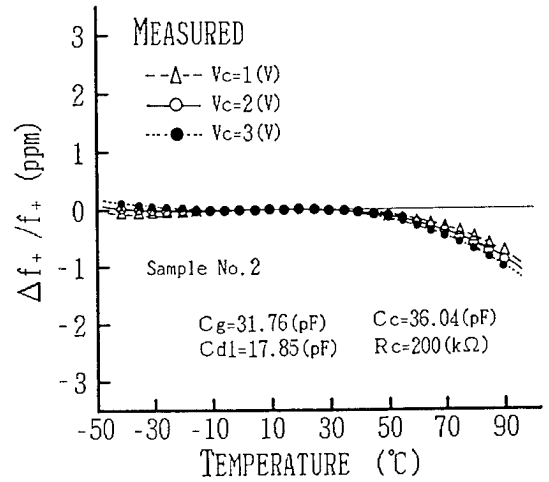


Fig. 13 Another example of frequency temperature behavior.

varied from 1 to 3 V, the frequency temperature behavior is also found to be mostly independent on the control voltage Vc.

Figure 13 represents another example of frequency temperature behavior of a VC-S-TCXO, when Vc is varied from 1 to 3 V. α_+ at +25℃ is zero approximately. Similar to the frequency temperature behavior in Fig. 12, it is mostly independent on Vc.

4.4 Frequency temperature behavior of VC-S-TCXO's

Figures 14, 15, 16 and 17 represent frequency temperature behavior of VC-S-TCXO's controlled by Vc. Control voltage Vc and the speed of temperature excursion were performed under one of three conditions that

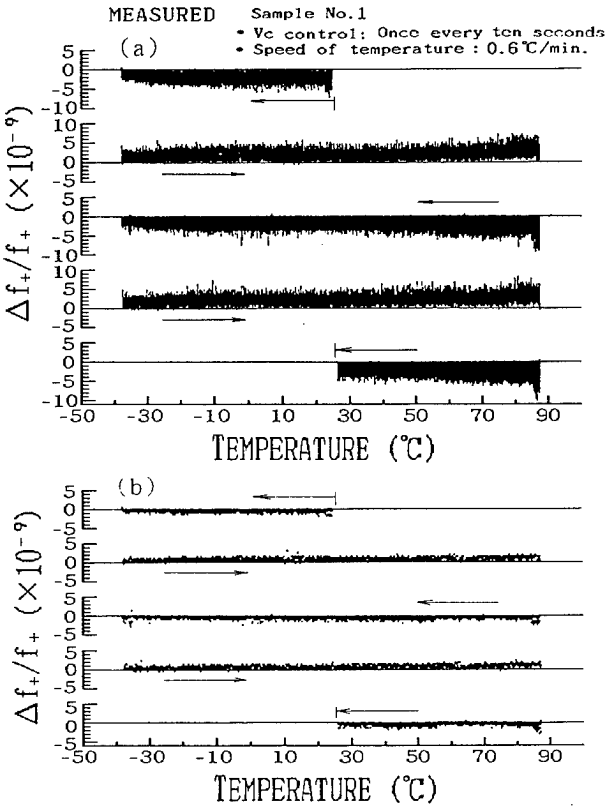


Fig. 14 Frequency temperature behavior of a VC-S-TCXO with a NS-GT cut resonator of sample No. 1.

(1) Vc is controlled once every ten seconds and the speed of temperature excursion is 0.6℃/min. (3℃/5mins.).

(2) Vc is controlled once every thirty seconds and the speed of temperature excursion is 0.6℃/min. (3℃/5mins.).

and

(3) Vc is controlled once every ten seconds and the speed of temperature excursion is 1.2℃/min. (6℃/5mins.).

In addition, temperature excursion is performed in a temperature range of −40 to +90℃. First, starting temperature is +25℃, then to low temperature of −40℃, and next, back to high temperature of +90℃. Temperature excursion of two cycles is performed under three conditions. A VC-S-TCXO with one of two samples shown in Table 1 is measured under each condition.

Figure 14 represents the frequency temperature behavior of a VC-S-TCXO with a NS-GT cut resonator of sample No. 1, which, as shown in Fig. 12, has a comparatively large value for α_+. Voltage Vc is controlled once every ten seconds and the speed of temperature excursion is 0.6℃/min. (3℃/5mins.). The upper figure(a) represents the temperature characteristics directly before and after the control of Vc. Temperature

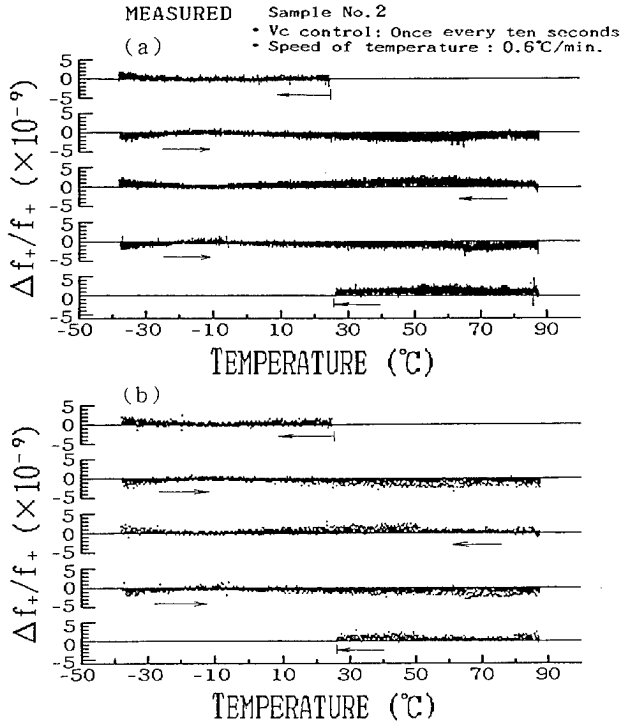


Fig. 15 Another example of frequency temperature behavior of a VC-S-TCXO with a NS-GT cut resonator of sample No. 2.

excursion of two cycles was performed. The frequency deviation is within $+/-8.1\times10^{-9}$ in a temperature range of -30 to $+85$℃. While the lower figure(b) represents the temperature characteristics directly after Vc was controlled. The frequency deviation is within $+2.9/-2.4\times10^{-9}$ in a temperature range of -30 to $+85$℃ for temperature excursion of two cycles. By controlling Vc, the frequency stability is found to be improved markedly.

Figure 15 represents the frequency temperature behavior of a VC-S-TCXO with a NS-GT cut resonator of sample No. 2 whose resonator has excellent frequency temperature behavior more than that of sample No. 1. Measurement condition is the same as that in Fig. 14. As is shown in the upper figure(a), the frequency deviation is within $+3.3/-3.8\times10^{-9}$ directly before and after the control of Vc, while, as is shown in the lower figrure(b), the frequency deviation is within $+2.9/-3.8\times10^{-9}$ directly after Vc was controlled. Thus, the frequency stability after voltage Vc was controlled, is found to be independent upon the frequency temperature behavior of resonators such as samples No. 1 and No. 2.

Figure 16 represents the frequency temperature behavior of the VC-S-TCXO for Vc control of once every thirty seconds and 0.6℃/min., which is the same as that in Fig. 14. In this case, a time interval of Vc control is different from that

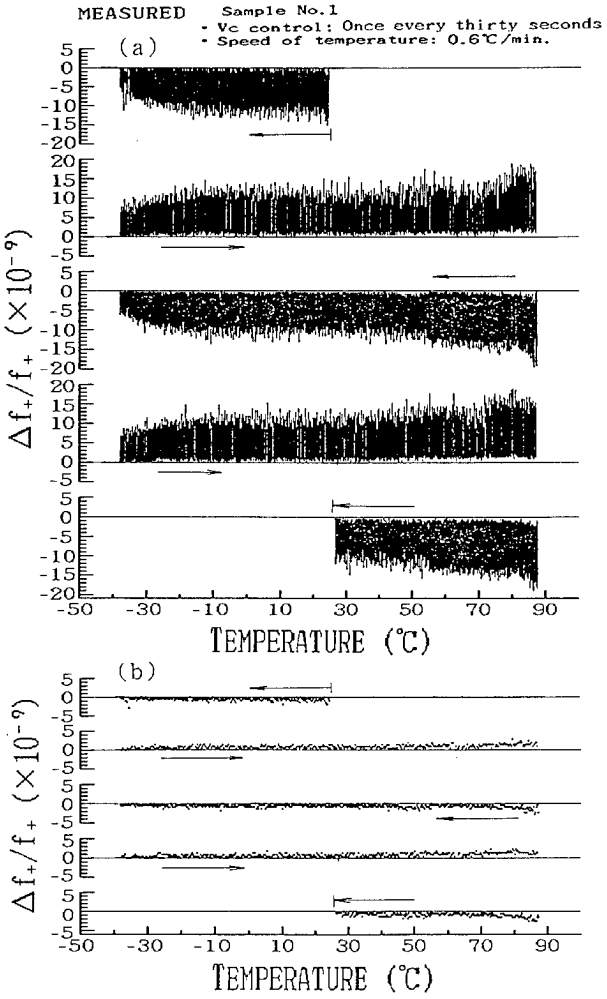


Fig. 16 Frequency temperature behavior of the same VC-S-TCXO as that in Fig. 13 for Vc control of once every thirty seconds.

of once every ten seconds in Fig. 14. As is shown in the upper and lower figures(a), (b), directly before Vc is controlled, the frequency deviation in the figure(a) is within $+20.0/-16.2\times10^{-9}$, while directly after Vc was controlled, it is within $+/-2.9\times10^{-9}$ in the figure(b). Frequency control for the control interval of Vc of once every thirty seconds is found to be performed at the same precision as that for once every ten seconds.

Figure 17 represents the frequency temperature behavior of the VC-S-TCXO for Vc control of once every ten seconds and the speed of temperature excursion 1.2℃/min., which is the same as that in Fig. 14. The speed of temperature excursion is different from that of 0.6℃/min. in Fig. 14. Directly after Vc was controlled, the frequency deviation in the figure(b) is within $+/-3.8\times10^{-9}$.

As described in Figs. 14-17, it is conceivable that these values obtained directly after voltage

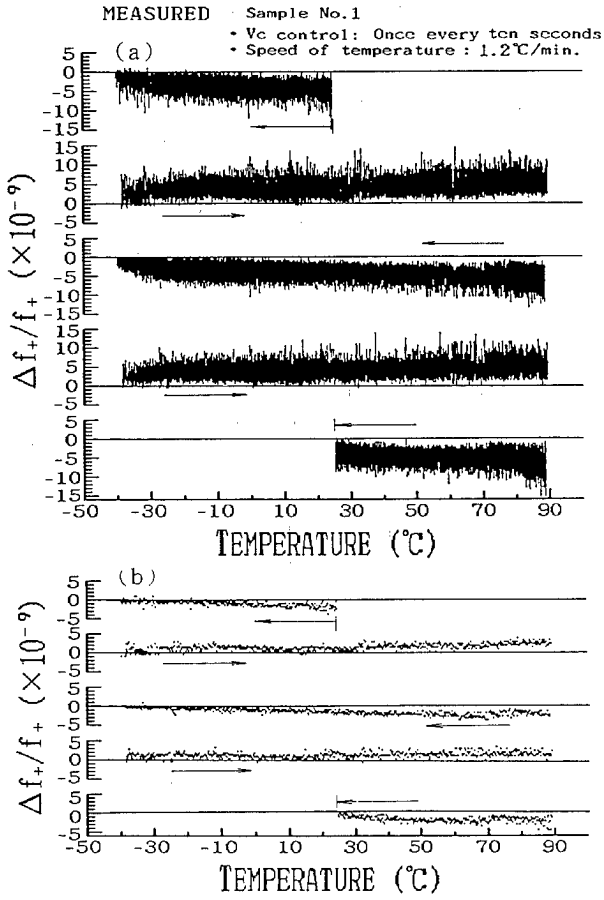


Fig. 17 Frequency temperature behavior of the same VC-S-TCXO as that in Fig. 13 for the temperature speed of 1.2℃/min.

control under each condition may be within a measurement error, because the error of the measured data and the values calculated from the measured data of oscillation frequency versus Vc is, as represented in Fig 11, about 4×10^{-9} at maximum. For the present conditions, it is easily understood that measurement accuracy is independent upon an adjustment interval by Vc and the speed of temperature excursion. It is naturally predicted that there is a relationship between the speed of temperature excursion and an adjustment interval by Vc, because it decides the bandwidth of frequency deviation directly before and after the control of Vc.

4.5 Aging characteristics

Figure 18 represents aging characteristics for S-TCXO's of eight samples whose electrical characteristics were shown in the reference [1] for about two years and three months. As is apparent from Fig. 18, the oscillation frequency is saturated gradually with time. Frequency shift of eight samples is from 0.309 to 0.549ppm (mean:

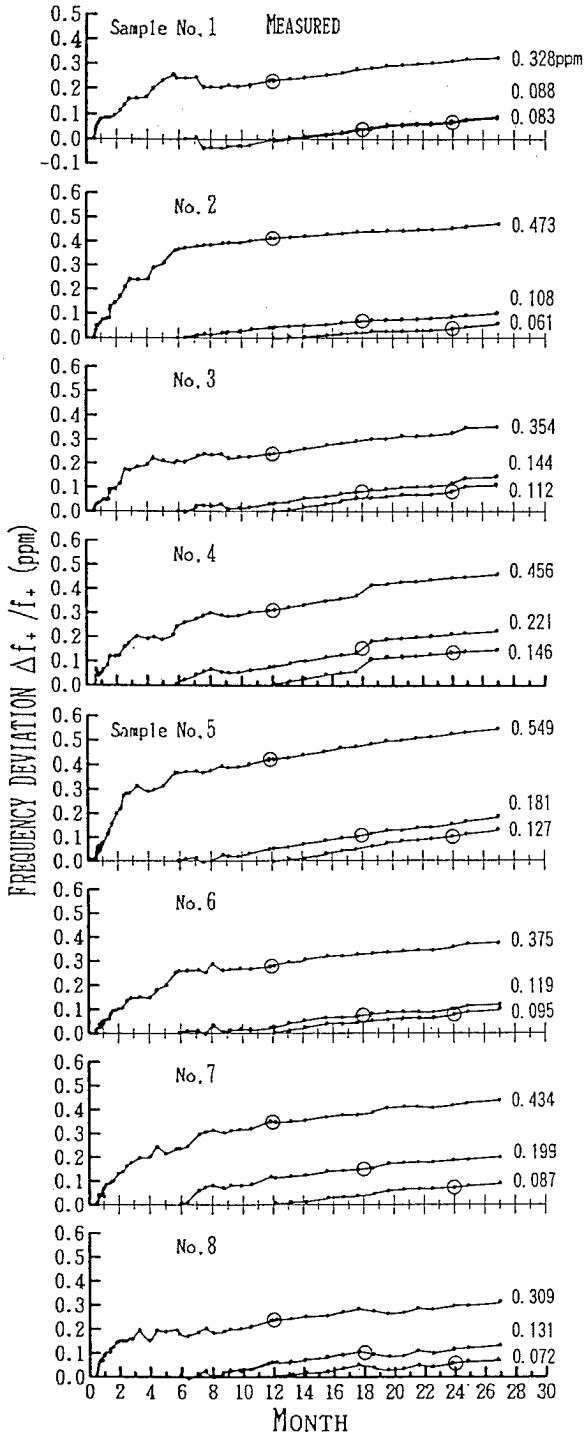


Fig. 18 Aging characteristics.

0.410ppm) for about two years and three months after the pre-aging for two weeks, from 0.088 to 0.221ppm (mean:0.149ppm) for about one year and nine months after the pre-aging for about six months and from 0.061 to 0.146ppm (mean:0.098ppm) for about one year and three months of the pre-

aging for about one year. Also, the first, second and third circles indicate frequency shift for one year after the pre-aging for two weeks, six months and one year, respectively. Thus, frequency shift of the S-TCXO's decreases gradually with time, it is essentially understood that, as shown in Fig. 18, the S-TCXO's are also very excellent in frequency aging.

§5. Conclusions

For the purpose of clarifying the frequency temperature behavior of voltage controlled S-TCXO's (VC-S-TCXO's) employing a NS-GT cut quartz crystal resonator and showing aging characteristics of further long term for about two years and three months of S-TCXO's, a study and an examination were performed.

First, the circuit construction of the VC-S-TCXO was represented. In particular, a detailed examination was performed with respect to a Vc circuit which plays a very important role to determine basically the accuracy of the VC-S-TCXO's. That is to say, the Vc circuit is so designed that its capacitance is mostly independent upon temperature when voltage Vc is varied. Therefore, load capacitance C_L that determines the oscillation frequency is mostly independent on temperature.

Second, a relationship between Vc and oscillation frequency was examined in detail. This relationship is of great importance because it decides the frequency stability of the VC-S-TCXO's by means of control voltage Vc, and the measured data is approximated by the equation of the eighth degree using it, so that the measured data was found to agree well with the values calculated from the equation of the eighth degree, namely, an error between them is less than 4×10^{-9}.

Third, VC-S-TCXO's of two samples were actually examined under all or one three conditions of a time interval of control voltage Vc and the speed of temperature excursion. Temperature excursion of two cycles was performed here. As a consequence, it was shown that the VC-S-TCXO's were successfully obtained with frequency deviation less than 4×10^{-9} in an absolute value after voltage Vc was controlled.

Finally, aging characteristics were examined for about two years and three months, so that frequency shift was found to be from 0.061 to 0.146ppm (mean:0.098ppm) for about one year and three months after the pre-aging for about one year. Thus, it was shown that the present VC-S-TCXO's were very excellent in frequency temperature behavior and also aging characteristics.

A coming subject will be to clarify if a relationship of oscillation frequency to voltage Vc is dependent upon time or not.

References

[1] H. Kawashima and K. Sunaga: "Frequency Temperature Behavior and Its Hysteresis of S-TCXOs Using New Shape GT Cut Quartz Crystal Resonators", Proc. the 1992 IEEE Frequency Control Symp., pp. 434-441(1992).

[2] H. Kawashima: "Variational Analysis of GT Cut Quartz Crystal Resonators with the Supporting Portions at the Ends", Proc. the 40th Annual Symp. on Frequency Control, pp. 193-200(1986).

[3] H. Kawashima: "An Analysis of Frequency Temperature Characteristics and Electrical Equivalent Circuit Parameters of a New shape GT Cut Quartz Crystal Resonator", IEEE Ultrasonics Symposium Proc., 1, pp. 465-470(1989).

[4] H. Kawashima: "An Analysis of Oscillation Frequency Characteristics in a CMOS Oscillating Circuit Using a Coupling Quartz Crystal Resonator", Proc. the 44th Annual Symp. on Frequency Control, pp. 585-592(1990).

[5] H. Kawashima and K. Sunaga: "Temperature Compensated Crystal Oscillator Employing New Shape GT Cut Quartz Crystal Resonator", Proc. the 45th Annual Symp. on Frequency Control, pp. 410-417 (1991).

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MANUFACTURING HYBRID TCXOS

Bob Achille, Tim Collins and Dennis Marvin

Motorola
1301 E. Algonquin
Schaumburg IL 60196
708-576-1000

The Saber oscillator is a thick film, temperature compensated crystal oscillator, TCXO, that was designed in the mid-eighties to provide Motorola with a small, inexpensive, highly reliable TCXO to operate as a reference oscillator for the synthesizers in its growing land mobile and cellular telephone markets. Earlier papers have described the crystal and compensation design, but their basics will be reviewed here to provide the background for a review of the manufacturing issues encountered. [1,2,3]

It is the intent of this paper to describe several of the issues encountered in the manufacturing process and to relate how these were addressed. Over its life, millions of Saber Oscillators have been manufactured and it is still in volume production.

INTRODUCTION

Motorola is the world's largest manufacturer of land mobile and cellular telephones. As such, it saw the need to develop an internal source of state of the art, inexpensive TCXOs to serve as reference oscillators for these products. Historically, Motorola's TCXOs had been constructed from numerous discrete parts inserted into printed circuit boards. Compensation was achieved by removing parts and replacing them with new parts selected by an algorithm.

The changing culture at Motorola dictated that the product be designed with manufacturability in mind. This means:

1. Minimizing the number of any redo loops.
2. Minimize the chance of human error in product assembly.
3. Eliminate any waste in the manufacturing process.
4. Simplify the manufacturing process.

With the goals of achieving higher quality, improved manufacturability, smaller size and lower costs, virtually the entire process had to be replaced.

In order to meet the needs of this market, the following oscillator requirements were established.

1. Only fundamental crystal frequencies would be used.
2. Frequency stabilities of 1 to 5 PPM would be required.
3. Performance would be maintained over the standard EIA temperature range.
4. The supply voltage would be 5 volts.
5. Current drain would be minimized.
6. A sinewave output of approximately 1 volt.
7. A frequency adjust range would be provided to allow for ten years of aging.
8. Phase noise necessary to meet radio performance through the 900 MHz bands would be achieved.
9. The lowest cost, small package consistent with radio performance would be used.

The oscillator is shown in Figure 1. The manufacturing practice is based on surface mount technology assemblies and the design incorporated high density without sacrificing cost.

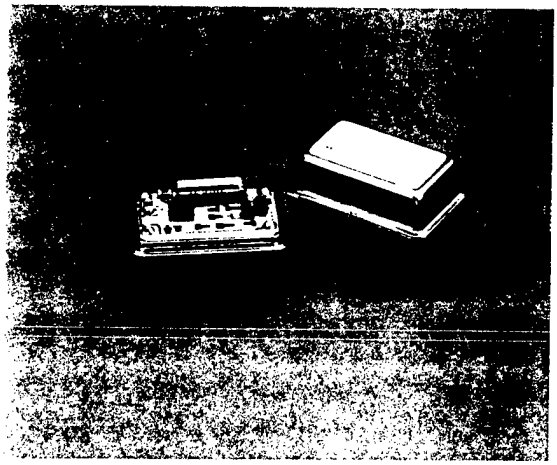


Figure 1

0-7803-0905-7/93 $3.00 © 1993 IEEE

ELECTRICAL DESIGN

OVERVIEW

Figure 2 is a block diagram of the Saber TCXO. The design is based on a compensation and regulator IC developed a few years earlier. It was also necessary to integrate the oscillator and buffer amplifier circuits in order to reduce the size of the circuit to fit in the small package.

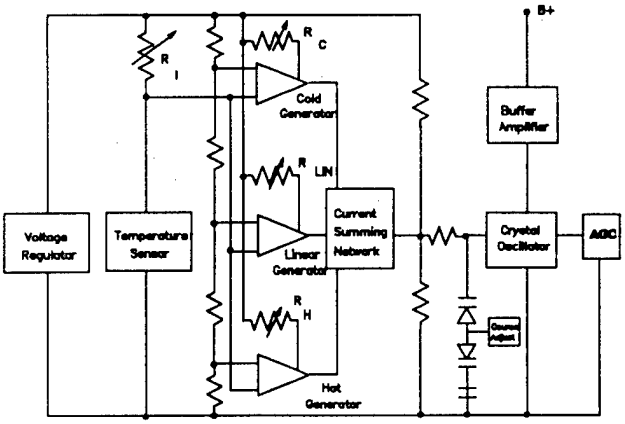


Figure 2

Feedback capacitors of moderate size capable of operation over a frequency range of 10 to 20 MHz were employed. This range was adequate to cover the frequency range required by the synthesizers and the limited frequency range available with the small crystal package. The capacitors also proved to be small enough to be integrated easily.

The output buffer amplifier is connected as a common base with a single current flowing through the oscillator and buffer. This is done in order to minimize the total current drain. Its collector circuit is arranged in such a manner that output levels can be obtained to match the levels of the various synthesizer chips by a laser trim option.

It was decided to integrate an AGC system for more consistent output vs temperature, improve starting time as well as mitigate any DLD issues with the crystals. Done with discrete parts, this would require too many components for a small oscillator.

A small, relatively high performance, low cost crystal was also needed. An existing design utilizing a strip crystal, cantilever mounted in a 3X8 mm tubular package was adapted. The design was being used for looser tolerance products such as clock oscillators and second IF reference oscillators. The design process was enhanced to better design out coupled modes and tighter process controls were added to the manufacturing process.

Three frequency adjustments would be necessary to manufacture the product. In addition to the frequency - temperature compensation, a correction was necessary for all of the initial reactance tolerances for the parts and the make tolerance of the crystal, and a modulation / AFC function / aging adjustment were needed. Reliability concerns, manufacturing complexity and labor costs dictated that their would be no removal and replacement of parts. A two-varactor scheme was used, with the tolerances corrected at the center tap and the f - T compensation and AFC performed at the ends of the varactor network. Suitable hyperabrupt varactors were used.

OSCILLATOR LOOP

A simplified schematic of the oscillator loop is given in Figure 3. To achieve a wide frequency range and a good waveform output, a Colpitts configuration was used. In this configuration, the bias of the oscillator transistor is controlled by an AGC network.

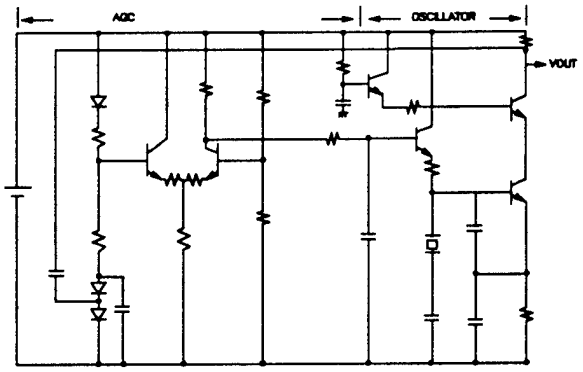


Figure 3

As is well known, crystals have a starting resistance phenomenon to different degrees. One of the design goals of the program was to design an oscillator with low output power to match the input characteristics of a particular synthesizer chip. At a low output level the oscillator transistor current and circuit gain will be quite low, which will result in the potential for a "starting resistance" problem. To make this product manufacturable required that the oscillator be capable of starting all crystals manufactured for this product. This does not mean just in the factory but also in the field. A study of crystal starting resistance characteristics helped to determine the required drive level to overcome the crystal's starting resistance characteristics. The use of an AGC circuit in the design of the oscillator allowed overcoming the starting resistance phenomenon. This network also had an added benefit in that by making the bias to the AGC circuit laser trimmable the gain of the oscillator could be adjusted to obtain the desired RF level for a particular customer.

TEMPERATURE COMPENSATION

Temperature compensation is achieved by taking advantage of the I-V characteristics of the three differential amplifiers, each independently adjusted to operate over a specific temperature range, approximately -30°C to +10°C, 10°C to 60°C, and 60° C to 85°C. The use of the differential amplifier creates an adjustable collector current versus the differential base voltage slope with a curvature which approximates the inverse of a crystal slope. One of the differential amplifier's inputs is tied to a fixed voltage reference which sets the nominal bias temperature point for the differential amplifier. The other input is connected to a temperature sensing network, the slope of which changes -6 mV/°C. Both the voltage reference and the slope generator combine to form the differential input for the differential amplifier. As can be seen in Figure 4 through Figure 6, the circuit can be adjusted to vary the slope of the compensation voltage versus differential input voltage. By varying the emitter current of the differential amplifier the output current slope is varied. Figure 7 shows the combination of the desired compensation slopes for each differential amplifier. The various currents are summed and applied to a load resistor to create a voltage. The output of this network is applied across a varactor which varies the reactance presented to the crystal.

So, in effect, the IC approximates a third order voltage versus temperature function that compensates for the crystal's f-T characteristic, Figure 8. Hyperabrupt varactors for both the frequency trim as well as the compensation bias voltage were chosen due to their linear capacitive reactance versus voltage characteristics as well as their high voltage sensitivity.

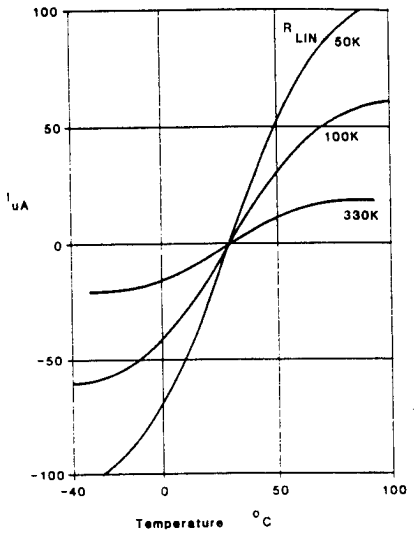


Figure 4

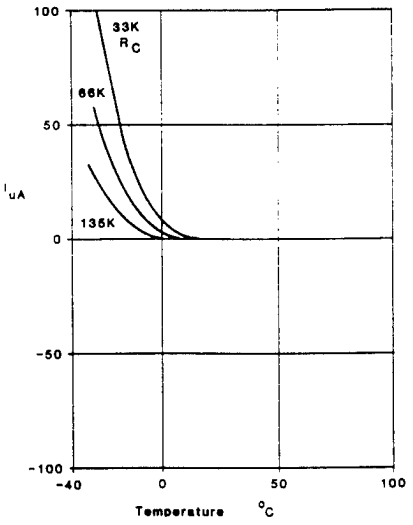


Figure 5

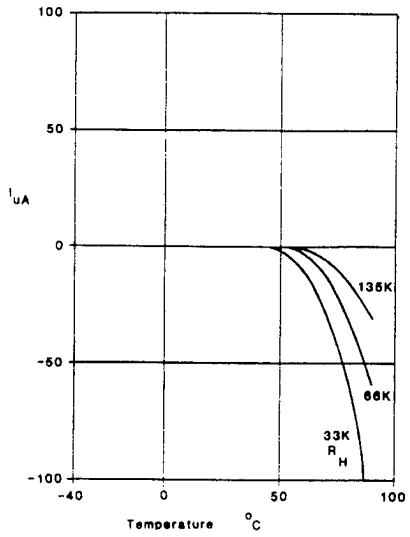


Figure 6

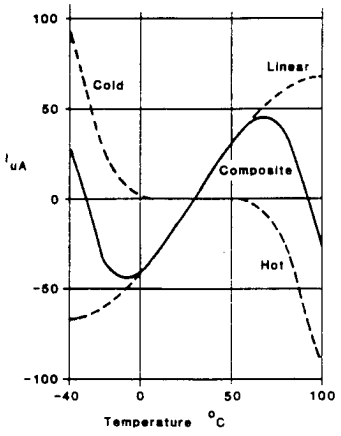


Figure 7

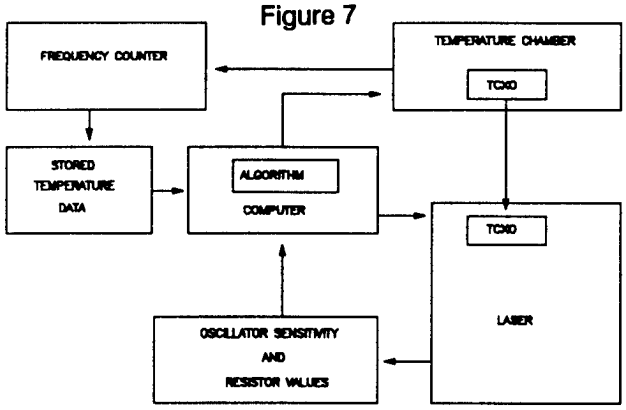


Figure 8

Two patents have been granted for the frequency - temperature design, 3,970,966 and 4,254,382.

COMPENSATION PROCEDURE

The IC allows compensation of a wide range of crystal angles as well as frequencies with a low parts count. The temperature compensation network requires a thorough characterization of the variations in the crystal's performance due to manufacturing variations. These variations not only include crystal make tolerance, lot to lot variations of Cm, but also the compensation components tolerances and temperature characteristics.

At the beginning of the program a decision had to be made as to how temperature compensation (TC) would be performed, i.e. at room temperature or over temperature (real time) . Some of the factors which had to be considered were the cost of performing temperature compensation. The TC system needed to be a low cost system and still not jeopardize the performance of the compensation. A tradeoff could be made between the size of the oscillator and the complexity of the compensation system. If the oscillator had to bring out the bias voltage for each of the differential amplifiers, the size of the package would be quite large. One of the original design goals was to make the package as small as possible. At the time of the design, the package standard in the "clock" oscillator industry had become the four pin DIP "data clock" package. If all the features of the oscillator could be packaged into that size then the cost of the packaging could be kept low since the cost of the data clock package was less than $0.10. As can be expected, real time compensation would require more pinouts then the typical 4 pins being purchased by manufactures of clock oscillators, and each pin added to the header of the oscillator not only would increase the cost of the header but also increase the complexity of the soldering operation of the header to the substrate. "Real time" temperature compensation is much more costly for this configuration, compared to room temperature compensation since the temperature system would have to be capable of forcing the correct bias voltage for all three differential amplifier on each oscillator in the temperature chamber at the same time. Based on all these variations the best choice was to perform temperature compensation at room temperature. A block diagram of the system is shown in Figure 9.

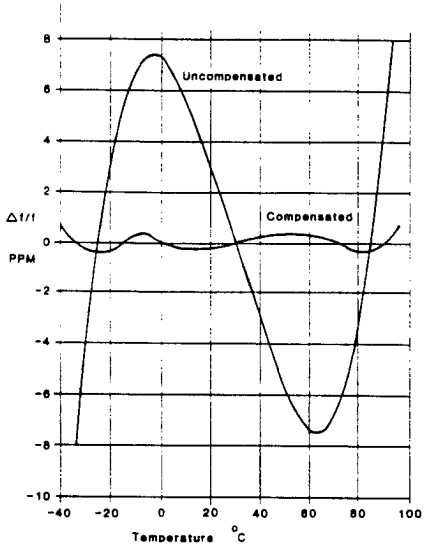


Figure 9

To determine the correct emitter current for each of the differential amplifiers the temperature compensation system measures the temperature, frequency and the differential amplifiers current required to put the oscillator onto frequency at the desired temperature. To reduce the complexity of the compensation system, the actual currents for each differential amplifier are not determined in the compensation system. As can be expected, any error created while determining the desired current for the differential amplifiers will directly effect the frequency versus temperature performance of the oscillator. These errors include the voltage versus temperature characteristics of the on-chip regulator. Since the differential amplifiers current source is fed by the regulator, the temperature characteristic of the regulator can also have a detrimental effect on the differential amplifier's transfer characteristic.

To improve the temperature compensation yields it was necessary to look much more closely at the performance of the room temperature compensation algorithm. Originally the algorithm was based on a relatively small sample size. To reduce the errors in the algorithm, a large number of oscillators were characterized over a long period of time. Using this technique, the variations in IC process as well as the variations in the manufacture of the thick film resistors were included in a new improved temperature compensation algorithm.

A tradeoff between assembly cycle time and scrap was also made. Originally, the algorithm had been quite conservative, as an overtrimmed compensation resistor could no longer be corrected and the unit had to be scrapped. By making the compensation algorithm much more aggressive the number of oscillators which were capable of 2.0 ppm temperature performance had increased to a point where the elimination of a verification before weld could be made. By doing this, the oscillator could go directly from temperature compensation to weld then to final verification without having to perform an additional TC run to determine how well the compensation algorithm performed the temperature compensation. It was determined that it cost less to throw away the small number of oscillators which did not meet the 2.0 ppm specification after weld then to verify all the oscillator prior to weld.

COMPENSATION RESISTORS

A custom top hat resistor was required to match the wide range of resistances required for compensation, while maintaining a minimum size. This resistor used high resistor inks for the upper part of the resistor together with inks of a lower resistance for the brim to get a laser trimmable resistor which could cover a wide range of resistances. Since no similar resistors of this type had been used, a process capability study had to be performed to determine what were the characteristics of such a resistor. A substrate was designed which contained all the different resistor geometries which were being considered. The test vehicle was designed to expose the several geometries to the same environments during testing. Some of the variables in the test included screen mesh sizing which allowed determining what size mesh would give the best results definition. To determine if any diffusion occurs on a two ink top hat resistors, the process steps which would be performed on it were performed but only one resistance ink was used. This was compared to resistors which were of the same geometries but did not use the same process steps. The typical characteristics of a laser trimmable thick film resistor, i.e. temperature cycle, humidity, temperature shock as well as aging were evaluated. The results of the study determined that the two ink top hat could be used to obtain a wide range of trimmed resistances without negatively affecting the performance of the oscillator. It also showed that it was possible to manufacture two ink top hat resistors reliably.

Patent 4,626,822 was granted to the two ink top hat resistor.

MECHANICAL CONFIGURATION

The form factor of the oscillator is that of the dual-in-line package, DIP. The package consists of a metal cover, a hybrid, and a metal header or base. The hybrid comprised of a ceramic, alumina, substrate with thick film technology. The three components are assembled and hermetically sealed. A suitable hermeticity level was required to prevent silver migration from occurring in the circuitry. The four leads from the header are the pinouts for the customer radio board.

All devices that are shipped to the customer require an identification by means of a label. The application of the label for the standard factory used a marking process on the lid which was then laser marked with the proper ID. The marking process had a curing schedule of 170 ^{O}C for 10 minutes which caused the crystal component to shift in frequency by as much as 2 ppm, thus creating catastrophic failures. The root cause for the failure was attributed to the plating of the crystal's shield or lid: Cu undercoat plus 20/80/ Sn/Pb alloy with solidus temperature of 183 ^{O}C.

The decision was made to eliminate the standard process and to use a semi automated label system with equivalent overall cost for the process. The label was qualified using a B vs. C test, label vs. laser mark, using an accelerated life testing (ALT) and a 95 ^{O}C/ 95% relative humidity (RH) for 240 hours.

Another key decision that was made in the development of the oscillator was containment of assembly until statistical data can be gathered on the components to assure proper clearance between the crystal package and the metal cover. In worst case analysis, it was analyzed that interference will occur thus a height gage was added to the process which is used as an integral feature on an existing transfer tool. The calculated specification for the height was based on a 3-sigma value of the nominal dimensions of the components thus assuring clearance in the assembly.

The decision risk was lessened due to the mounting of the components on the board. Typically, "rattling" space is necessary in an assembly during vibration and mechanical shock. The rigid mounts employed for all components on the ceramic substrate resulted in micro inches of deflection during the required specifications for shock and vibration.

The crystal required a lead forming process, with the height being a critical parameter. A team was formed to assure process capability from the design with zero defects. A six-up tool was finalized which yielded crystals that can be flush mounted onto the substrate 100 % of the time.

Statistically and based on the level of total defects per units (tdu) gathered over time, the height gage is no longer a requirement.

THICK FILM HYBRID PROCESS

The thick film process that was used for the product represent those of our current factory and our sister division AIEG, see Figure 10. This design did add some new materials in the process, namely, dielectric material for crossovers from the circuit design, a low temperature solder for the crystal assembly, and a second conductor pass solely for the low temperature solder. It was shown under temperature cycling that failures were occurring in the oscillator at the low temp solder/conductor 1 joints. The failure mode was that of the conductor pads lifting from the substrate. A B vs C qualification run was done on the oscillator to test out materials compatibility on the hybrid. With the guidance of the ink supplier, conductor 2 was selected in order to reduce the leaching effect causing the adhesion failure. Leaching tests and data from the ALT proved that the materials were incompatible thus with proper material selection provided a reliable product.

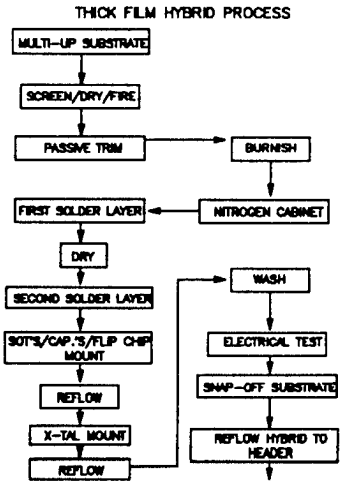


Figure 10

The hybrid was also designed to be used in a multi-up assembly, Figure 11, rather than a 1-up due from the cost analysis. Due to the geometry of the resistors and the fine lines and spaces required for the metalization runners, the screen was evaluated and specified for good printing resolution.

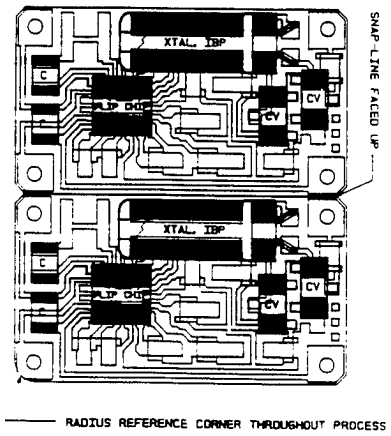


Figure 11

From the first prototype builds of this new hybrid, it was apparent that tight process control must be maintained. Therefore, all of the thick film inks had the following variables for control:

1. Viscosity
2. Delta of wet thickness
3. Dried thickness
4. Printing alignment

Also, all resistor passes underwent a sample test run thru the firing furnace for resistor measurement.

The goal for this new oscillator was to be first in class in design and manufacturability. Due to the fine line/pitch design required for the Flip Chip, the hybrid was built in a clean room environment.

The emphasis of the product is based on the Flip Chip. The technology was not new and is shown in Figure 12. The wafer processing is the same or identical to any other IC die except solder bumps are created for connection to the ceramic instead of the typical wire bonding. The metallurgy entails sputtering of Titanium/Tungsten (Ti/W) bimetal, the wafer is then electroplated with Copper/Lead/Tin (Cu/Pb/Sn) forming the bumps. Cleaning of the photoresist, etching of the bimetal are done followed by a reflow in a reducing atmosphere causing the solder to form spherical bumps about 0.004" high and 0.008" diameter.

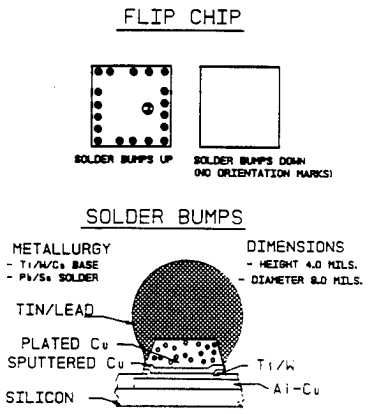


Figure 12

The flip chips are shipped with solder bumps up, the chips are prepared using an inverter with orientation of the waffle pack chamfer. The orientation of the IC's were maintained by the supplier per our specifications in order to eliminate human error which will cause defects.

Two reflow processes were required for the oscillator, one for the flip chip assembly which also included the capacitors and the varactors. The second reflow is for a low temperature solder which is required due to the possible damage of the crystal package at temperatures of greater than 300 ^{o}C from the first reflow.

In order to provide proper wetting of the solder to the metalization, oxide buildups on the conductor for the flip chip were removed by means of a mechanical burnishing process. There was a limit setup for maximum time between the burnish process and the first solder printing process. But it was later determined that once burnished and stored in a reducing environment, the storage time is indefinite.

The soldering process was done in two steps on a screen printer, with a fine-mesh screen used for solder. The fine-mesh is required to assure high printing resolution for the flip chip's fingers. Solder two was applied on the other components except the flip chip's pads using a standard mesh screen. Flux was used to hold the flip chip onto the substrate during the reflow process. The other components

were mounted onto the thicker solder paste to assure proper joint strength. To maintain product consistency, a solder weight control chart was used as a variable.

The units were then reflowed in a Nitrogen atmosphere furnace to eliminate oxidation with a profile as shown in Figure 13. The profile was critical in the process especially during the reflow and cool down zones. The time above the 310 ^{o}C was designed to be 40 seconds maximum to avoid leaching of the IC pads. The cooling rate from 310 to 280 ^{o}C was designed to be under 10 seconds to assure proper formation of the bumps, thus assuring a 2-3 mils gap between the IC and substrate for flux removal during the cleaning process.

TEMPERATURE PROFILE

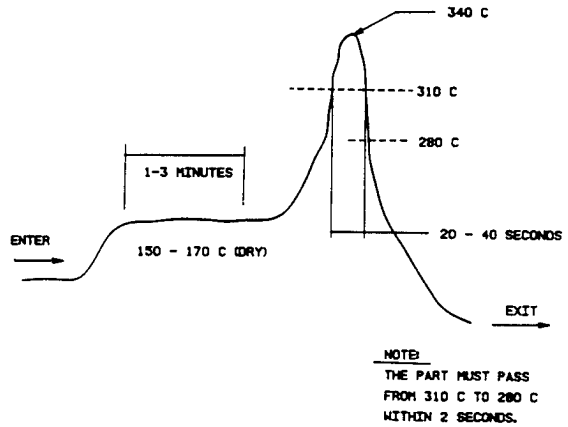


Figure 13

The control parameters of a good profile are shearing of the solder bumps at the die surface and no conductor peeling caused by leaching. The shear test was done only as an engineering tool in order to set the profile controls for the production line. The joint strength was a result of a torque calculation per [4]

Ti = (Y/sqrt 3) *a*b*Ri (1)

and was used as the limit in the experimental studies for the profile settings.

The low temperature solder process was done using a dispensing application and a conduction reflow system. The conduction reflow was necessary to minimize the temperature being experienced by the crystal thus eliminating frequency shifts to the oscillator resulting in rejects.

A cleaning operation was required for all assemblies. The solvent that was used was Methylene Chloride in a vapor degreaser and was used mainly to remove the flux underneath the IC. The current process is using a non-ozone depleting material to clean the flux. In the beginning stages of the program, cleaning was done using the ultrasonic energy to assure adequate scrubbing for flux removal. One of the findings was that ultrasonic cleaning caused structural damage of some components due to energy level. This was eliminated by controlling the energy level in the solvent.

The multi-up assemblies are separated using a "snap" fixture, followed by hand soldering onto the bases. For ease of manufacturability, special fixtures were designed to assure repeatable electrical parameters verification at the laser trim process.

The units are sealed by means of resistance welding. This welding method has been proven over the years to be very reliable and repeatable.

CONCLUSION

This oscillator has been in production for the past eight years and has required limited engineering support upon its release to production. With the introduction of the Motorola "six-sigma" goal in new product development, this "Hybrid TCXO" have reached the six-sigma goal in its maturity.

REFERENCES

[1] T. Keller, D. Marvin and R. Steele, "Integrated Circuit Compensation of A.T. Cut Crystal Oscillators," Proceedings of the 34th Annual Symposium of Frequency Control, 1980, pp. 498-503

[2] J. P. Tomase and L. N. Dworsky, "Analysis and Design of Coupled Mode AT Rectangular Resonators," Proceedings of the 34th Annual Symposium of Frequency Control, 1984, pp. 141-149

[3] L.N. Dworsky, "AT Quartz Strip Resonators," Proceedings of the 34th Annual Symposium of Frequency Control, 1987, pp. 295-302

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE NEW METHOD OF STATISTIC PIESEWISE-LINEAR INTERPOLATION AND ITS APPLICATION TO DTCXO CREATION

A. KOSYKH, V. BAGAEV, B. IONOV, A. LEPETAEV, S. ZAVIALOV
Omsk politechnic institute, 644050, Mira Pr. 11, Omsk, Russia
A. VASILIEV, *Microsystem SSI, Moskow, Russia*

SUMMARY

Thermocompensated crystal oscillators (TCXO) are widely used as sources of high stable oscillations in miniature radio equipment. Their stability and tuning properties can be considerably improved by using of digital methods of compensating action forming. Different methods are compared. Simple and effective method of piecewise linear interpolation allowing to reduce markedly memory volume is offered. While analyzing the method particular attention was paid to spectral characteristics of output signal.

The suggested method, called statistical, can be used in different types of TXO. One-chip LCI miniature DTXO is described. The second oscillator using this method is precision one with output frequency 100 MHz. Its features: employment of double mode oscillator and frequency synthesizer. Detailed technical data of the oscillators and test results are given.

INTRODUCTION

Thermocompensated crystal oscillator (TCXO) are widely used as sources of high-stable oscillations in small-sized, self-contained radiotechnics devices. They are more efficient and have less readiness time than ovenized models, but they are less stable than the latters. This parameter can be essentially improved by applying digital methods of compensation influence.

Some firms produce digitally thermocompensated crystal oscillators (DTCXO) providing temperature stability 0.3...1 ppm in the temperature range $(-60...-30 - +60...+85)^{o}C$. Further improving frequency stability is prevented by several factors such as thermotransducer and reference resonator ageing; dynamic temperature effects [1]; static errors of compensation function (CF) reproducing. This article is devoted to the last factor but simultaneously other problems are considered and, in particular, decrease of ageing effect. It is show that a complex approach can essentially help to improve DTCXO parameters.

1. METHODS OF COMPENSATION FUNCTION FORMING

Accuracy of compensation function forming strongly determines thermocompensation static error that is why this problem is actual now. When the compensation function synthesizer (CFS) is not restricted by complexity and consumed power, the synthesis problem is solved rather simply. But in fact DTCXO investigator is strictly limited and is compelled to use the most simple solutions giving in the end limited stability at no-so-high level. Let us consider in brief the most frequently used technical solutions.

1.1. Stepped approximation of CF

CF stepped approximation is the most frequently used in DTCXO when operating temperatures range is divided into zones where compensation function is constant and its value changes in steps when it passes from one zone to another. Most easily is to split operating temperatures range into zones of equal length. Temperature quantization is carried out this time by means of analog-digital converter (ADC) or by pulse counter (if thermotransducer signal is frequency) with word length n ,and CF values in direct form are stored in memory (ROM) with output word length m. It is shown in [2] that such approximation method gives thermocompensation error having two independent components: at the expense of argument quantisation and at the expense of output signal quantization. If thermotransducer curve and compensated oscillator frequency control curve are close to linear then the thermocompensation error is determined as follows:

$$\delta f = \alpha f(T) \frac{\Delta T}{2^{n+1}} + \frac{\delta f(\Delta T)}{2^{m+1}} \qquad (1)$$

where $\alpha f(T)$ - temperature dependence of reference resonator 1-st order frequency temperature coefficient; ΔT - operating temperature range; $\delta f(\Delta T)$ - relative oscillator retuning in the temperature range. If the curves of thermotransducer and the oscillator are non-linear then the compensation error increases (or device expenses increase at the same stability) because of non-optimum quantization in zones with small steepness. One and the same stability can be reached at different m and n combinations entering (1). By minimization ROM capacity $k=2^{n} V$ optimum parametres can be found from system solution:

$$\begin{cases} 2^{m} = \delta f \cdot (1 + m \cdot \ln 2)/2 \\ n = \log_2 \left(\dfrac{\alpha f(\Delta T)}{2\delta f(\Delta T)} \dfrac{2^{m}}{m \ln 2} \right) \end{cases} \qquad (2)$$

0-7803-0905-7/93 $3.00 © 1993 IEEE

Values n_{opr}, m_{opr} and k_{opr}, calculated according to expression (2) when AT-cut resonators with optimum orientation at given operating temperatures range are used, are shown on Fig.1, 2 and 3. Analysis of Fig 3 shows that to obtain instability $1* 10^{-7}$ or better in a wide temperature range a great ROM capacity is needed ($>10^4$ bit). This factor restricts DTCXO frequency stability with such CF approximation method on level of (3 5)10 that is confirmed by parameters of manufactured oscillators.

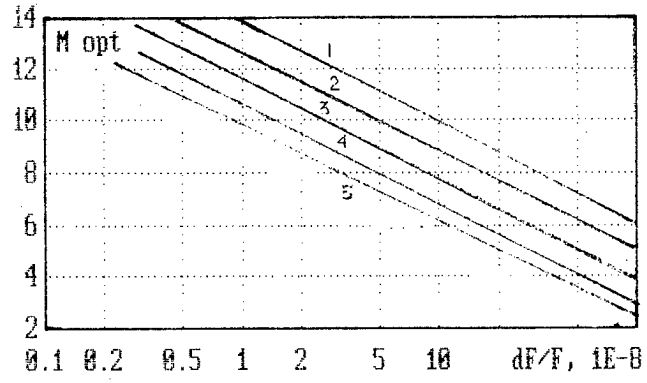


Fig.1. Optimal ADC word length dependence vs.stability and operating temperature range

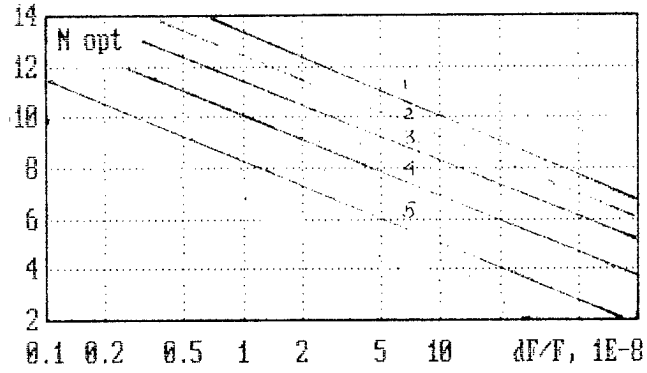


Fig.2.Optimal DAC word length dependence vs stability and operating temperature range

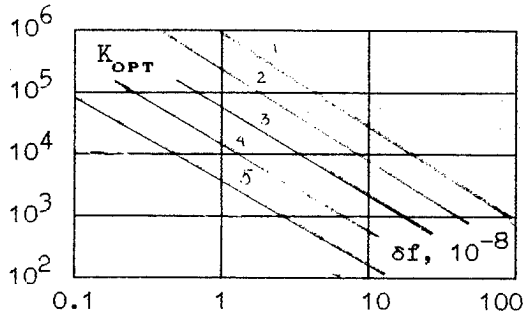


Fig.3. ROM capacity dependence vs. stability and operating temperature range

Table to Fig.1 - Fig.3

operating temperature range,°C				
1	2	3	4	5
-60 +90	-40 +80	-30 +70	-20 +60	-10 +50

1.2. Polynomial approximation of CF

Some high stable DTCXO [3,4] use a microprocessor to form CF and here polynomial approximation can be applied:

$$U_c = U_o + U_1T + U_2T^2 + U_3T^3 + \ldots + U_nT^n \quad (3)$$

where U_I - coefficients, T - temperature code. This method has a fault that polynom order increases when requirements to compensation accuracy grow. It is especially difficult to describe oscillators CF by polynoms when their resonators have local F-T curve irregularities. In particular such irregularities take place in two- mode resonators due to spurious coupling between oscillation of different modes. Besides, mere algorithmic difficulties appear due to truncation errors when computing series (3) with low processor. Fig. 4 shows dependence of real CF approximation error versus polynom order. Moreover, computation of series (3) is possible only by microprocessor whose application is not expedient in all types of DTCXO.

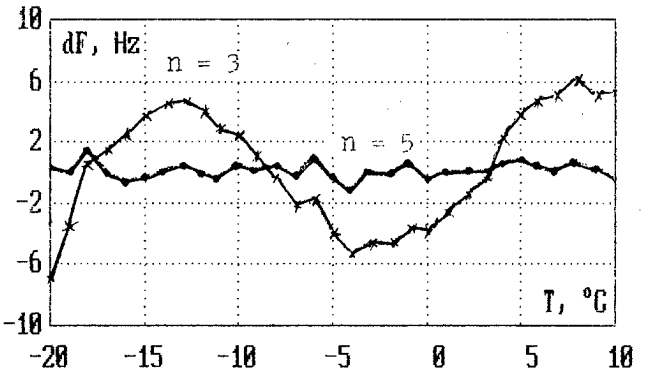


Fig.4.Decompensation curve vs.polynom order

1.3 Piecewise - linear CF approximation

This method of CF interpolation is used in DTCXO not so often although it is well investigated theoretically [2]. Ifinterpolation straight lines are drawn through nodes T_j, T_{j+1} and so on, standing on equal distances by the value of interpolation interval $\Delta\theta$, then the thermocompensation error on the j-interpolation portion (in the vicinity of T_j) is described by the expression

$$\delta f(\xi)=3a_3\Delta\theta(\xi^2-\xi)[-\frac{a_2}{3a_3}+T_j+\frac{\Delta\theta}{3}(\xi+1)] \quad (4)$$

where a_2 and a_3 - CF expansion coefficients; $\xi=(T-T_j)/\Delta\theta$ - relative position inside the interpolation interval.

View of compensated oscillator F-T curve with piecewise-linear CF approximation is presented on Fig.5. If to consider that $a_2/3a_3$ is a

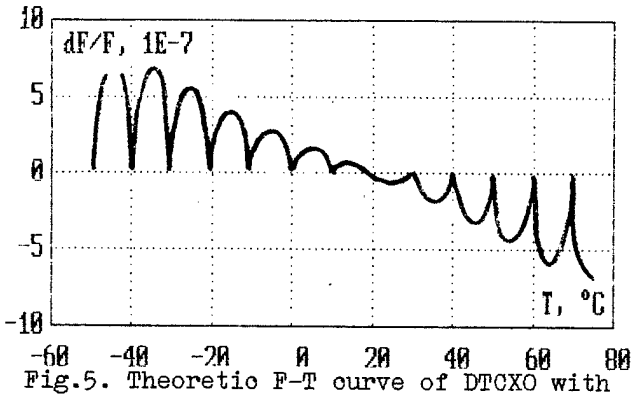


Fig.5. Theoretic F-T curve of DTCXO with piecewise-linear interpolation

temperature of CF bending (the point where its second derivative is equal to zero) and if to be interested only in error maximums, then the maxima envelop equation will take this form:

$$\delta f = \pm \frac{3}{8} a_3 \Delta\theta^2 T^* \qquad (5)$$

where $T^* = T_i - a_2/3a_3$

The error is maximum on the borders of operating temperatures range. The dependence of thermocompensation error from interpolation interval value Q and from width of operating temperatures range (indirectly it is expressed by T -T) is shown on Fig.6. Consideration shows that instability 1 10 (and even 1 10) in wide temperature range can be reached with a small volume of the stored data. This method looks to be the most flexible.

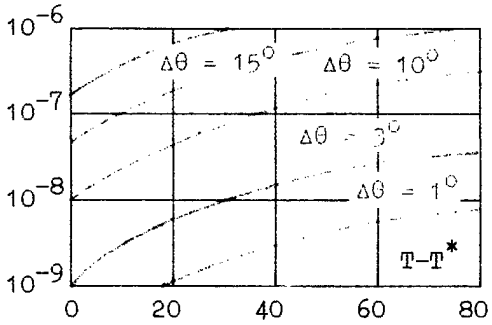


Fig. 6. Decompensation curves vs. operating and interpolation interval length

2. METHOD OF STATISTIC PIECEWISE-LINEAR COMPENSATION FUNCTION INTERPOLATION

Rave application of CF piecewise linear interpolation is explained, first of all, by high complexity of compensation device [4 ,e.t.c]. Growth of device expenses on interpolator realization cannot be compensated by ROM capacity reduction. We developed and successfully tested on several DTCXO a new method of piecewise linear interpolation distinguished by exceptional simplicity and effectiveness. We named this method statistical because exact value of CF is obtained by means of time averaging of readings train made with low accuracy.

2.1. The method groundation

Suppose that the output signal of the thermotransducer is voltage, and for its conversion into code an ADC is used, which has constant quantization increment U.In this case we can know only the quantization interval U where the signal came, but its position inside this interval is unknown. If to modulate the measured signal by periodic function having uniform probability density of distribution values in the limits of the period (for example by a triangle one) and having an amplitude equal to the width of interval [5], then the situation will change (See Fig. 7). Passing this signal through non-linear threshold ADC-ROM-DAC, and, after filtering resulting pulse sequence by of low pass filter (LPF), on its output CF is realized in the form

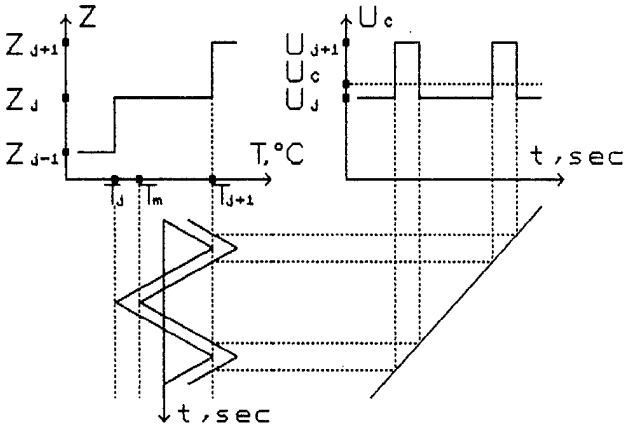


Fig.7. Statistic piecewise-linear interpolation method eluciditation

$$U_c = U_j\left(\frac{T_{j+1} - T}{T_{j+1} - T_j}\right) + U_{j+1}\left(\frac{T - T_j}{T_{j+1} - T_j}\right) \qquad (6)$$

where T- temperature of medium; U_{j+1} and U_j compensating voltage at temperature T_{j+1} and T_j respectively. Expression (6) corresponds to La-Grange interpolation for a case of two nodes.

The great interest presents a case when thermotransductor coming signal is frequency [6]. In this case DTCXO block diagram looks as on Fig 8. The counts number of pulses on the input during interval that came from the thermosensivite oscillator

$$N' = \tau_c f(T)$$

In general case N'is not equal to a whole number and due to non-correlation of frequency $f_T = 1/\tau_T$ and of gate oscillator frequency $f_c = 1/\tau_c$, the registration of pulses N or N+1 is made alternative mode. Here N equals Int (N'). The condition of counter N+1 realization probability

$$p_{N+1} = (\tau_c \bmod \tau_T)/\tau_T$$

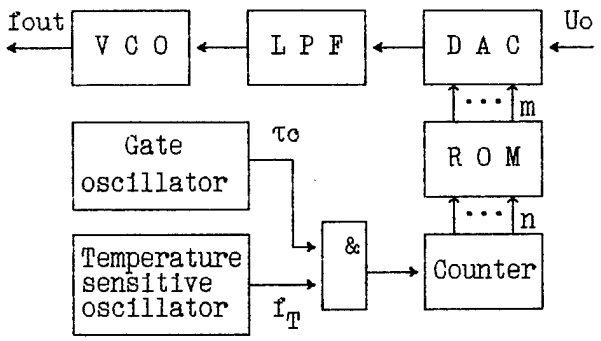


Fig.8 DTCXO with statistic picewise - linear interpolation block diagram

and condition of counter N probability

$$P_N = 1 - P_{N+1}$$

where $\tau_T = 1/f_T$

Periodic change of counter condition from address N to address N+1 causes alternating change of compensating voltage of LPF output. Constant component of the pulse sequence on the output will be equal

$$U_C = U_N \cdot P_N + U_{N+1} \cdot P_{N+1}$$

wich is complete analog of expression (6).

2.2. Parameter calculation of statistical thermocompensating devices

Initial values for parameter calculation are required frequency stability, operating temperatures range T and F-T curve of the used resonator. It needs to determine the rate n of a counter address, the rate m of the DAC and the count time t of a address counter. If to specify thermocompensation errors components at the expense of piecewise-linear interpolation df and at the expense of output signal quantization df , then minimum admissible values m and n can be determined as

$$\begin{cases} n > \frac{1}{2} \log_2 \left(\frac{3}{8} \frac{a_3 \Delta T^2 T^*}{\delta f_2} \right) \\ m > \log_2 (\Delta f / \delta f_1) - 1 \end{cases} \qquad (7)$$

Here to determine n equation (5) was used, and Δf is oscillator retuning range Time of count τ_c can be found by solving the system:

$$\begin{cases} M_1 = \tau_c f(T_1) \\ M_2 = \tau_c f(T_2) \\ M_1 - M_2 < 2^n \end{cases}$$

where $f(T_1)$ and $f(T_2)$ - values of thermosensitive oscillator output frequency at low T_1 and high T_2 temperatures; M_1 and M_2 - quantity of pulses coming to counter for the time τ_c . The system solution is:

$$\tau_c \leqslant 2^n / [f(T_1) - f(T_2)]$$

Usually M >> 2 and in this case the counter overflows K = Int[t f(T)/2] times. When using thermosensitive oscillator with quasy-linear F-T curve with steepness S, calculation time is determined as

$$\tau_c \leqslant 2^n / (\Delta T \cdot S)$$

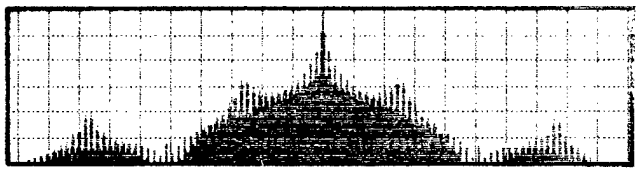
2.3. Spectral characteristics of DTCXO

Often spectral characteristics of oscillator output signal are determinative parameters. It is particulary interesting to investigate output spectrum of digitally thermocompensated oscillators because it contains specific components and is poorly studied.

DTCXO output signal spectrum with statistical formation of control operation is rather complicated. Its theoretical analysis can be fulfilled by two ways.

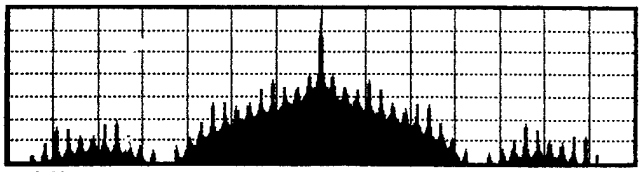
The first way. Main time characteristics of pulse sequence on DAC output are determined at given position within interpolation interval T..T, and then this signal spectrum is defined. Knowing filter characteristics, spectrum of control operation can be computed, and then, using apparatus of Bessel's functions, it is possible to pass to spectrum of frequency-modulated signal. This way, however, gives satisfactory results only for having simple spectrum and small modulation indexes control signals. In real case the calculations become to be complicated due to appearance of infinite number spectrum components.

The second way of spectrum computation consists in determination of signal form on the input of a controlled oscillator in time domain, and then output signal spectrum is determined by means of algorithm of Fast Fourier Transfer (FFT). In this case it is convenient to use computer for numerical process simulation. Fig.9 - Fig.10 show computation results, and Fig.11 - Fig.12 show results of experiment investigation of output signal spectrum at different position within interpolation interval (different values of probability



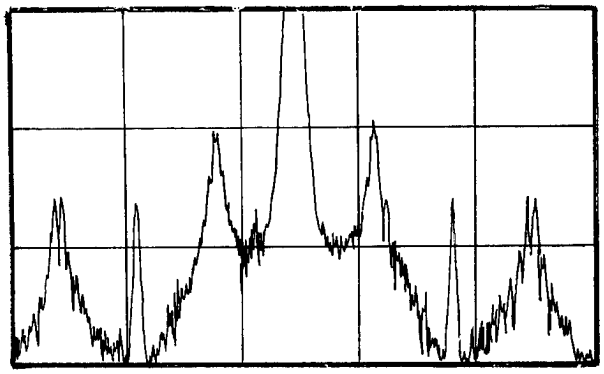
-1kHz +1kHz

Fig.9. Theoretical DTCXO output signal spectrum for impulse τ_c=1.8 mS, P=0.48.



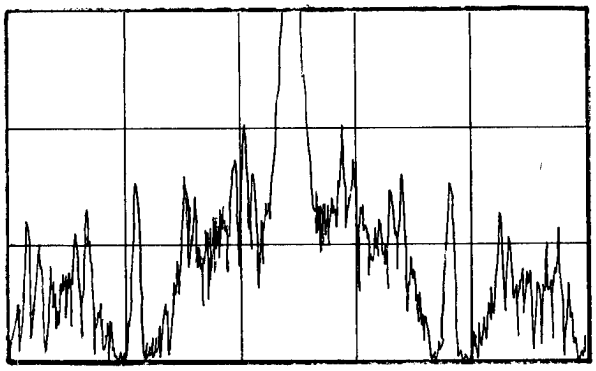
-1kHz +1kHz

Fig.10 . Theoretical DTCXO output signal spectrum for impulse τ_c=1.8 mS, p=0.31.



-1kHz +1kHz

Fig.11. Experimental DTCX) output signal spectrum for impulse τ_c=1.8 mS, p=0.48.



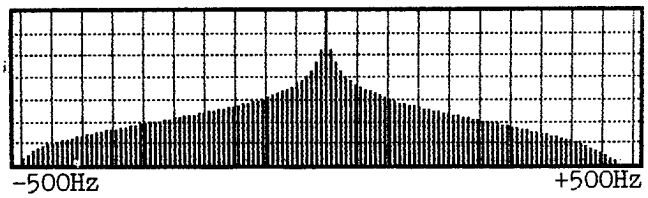
-1kHz +1kHz

Fig.12. Experimental DTCXO output signal spectrum for impulse τ_c=1.8 mS, p=0.31.

P_N).

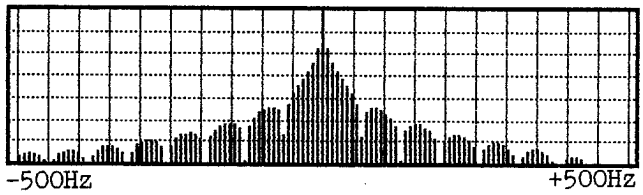
The results shows that at change of P_N (at temperature changing) the spectrum will alter strongly. The closer P_N to full or to a fraction aliquotness (0; 1; 1/2; 1/4; etc), the purer spectrum is. At P = 0 the spectrum does not contain additional components. At P = 1/2 the spectrum contains components with frequency f = 1/2 etc. To improve the distinguishing of spectrum components under statistical interpolation simulation we increase the pulsation level comparatively with real condition (Fig.11-Fig.12).

It is interesting to compare signal spectrums of DTCXO with statistical and with stepped CF forming at one and the same temperature and at one and the same temperature frequency stability. At stepped interpolation the width of interpolation zone mast be many times smaller, and the count time according to (10) must be the same times longer. In this case the amplitude of pulsation on filter output decrease, and the pulsation period increases. Fig.13 - Fig.14 show typical spectral characteristics of such DTCXO. It is easy to see that when t increases, the spectral components level in remote zone decreases. In the vicinity of the carrier



-500Hz +500Hz

Fig.13. Typical output signal spectrum: count time τ_c = 2 mS, relative pulsation amplitude δU = 1, pulsation period τ_c/P = 0.1 s.



-500Hz +500Hz

Fig.14. Typical output signal spectrum: count time τ_c = 16 mS, relative pulsation amplitude δU = 0.125, pulsation period τ_c/P = 0.1 s.

the spectrums have resembling form. Similar results can be achieved with optimum choice of filter parameter. Thus, improvement of temperature stability of frequency without growth of ROM volume can be obtained either at the expense of some aggravation of spectral characteristics, or at the expense of increase of filter time constant. Technically this solution is not very complex.

3. MINIATURE DTCXO WITH PIECEWISE - LINEAR STATISTICAL-TYPE INTERPOLATOR

3.1. Circuit and construction.

Since statistical piecewise-linear interpolator permits to reduce markedly equipment expenses it deserved to be used in extraminiature DTCXO. With this aim we developed special LCI containing practically all DTCXO units (Fig.15). The microcircuit

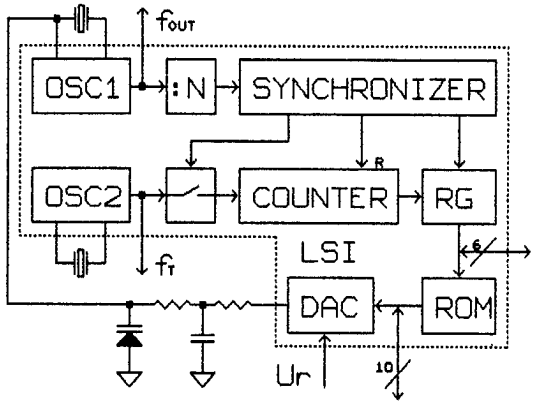
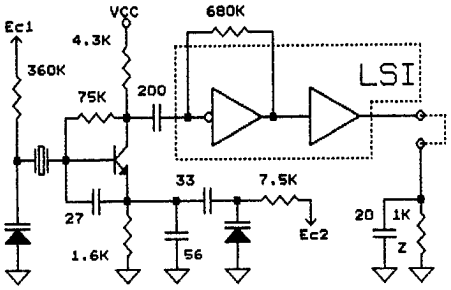
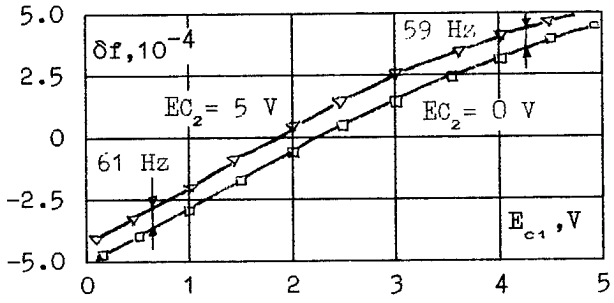


Fig.15. Miniature DTCXO block-diagram

is manufactured by CMOS - technology and has a 15x15 mm 42-pin surface mount package. Yet it has one-time programmed CMOS memory with 64x10 configuration. Built-in 10-digit DAC has a voltage output. Estimation by formulas (7) show that here

stability $1x10^{-7}$ can be obtained over a wide temperature range. Using Y-cut resonator with 400 Hz/ $^{\circ}$ C steepness as thermotransduser gives, according to (8), measuring time cycle: $t < 5.3x10^{-4}$ sec., which corresponds to 1.8 kHz pulsation frequency. The microcircuit is capable of discrete switching count time t_c , allowing to employ transducers with different steepness and to vary operating temperature range. Reference and thermosensitive resonators were placed in separate miniature enclosures. Outer buffer or outer controlled oscillator can be added to the circuit to suit output oscillation spectrum, load-carrying capability and overall dimensions. Inverter, being basis of OSC2, is employed here as a buffer.

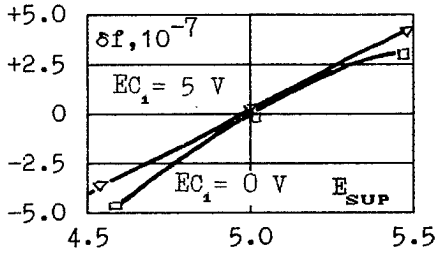




frequency vs. control voltage

Z	0	1k+20pF
δf	0	$<1\cdot10^{-7}$
I_C,mA	2.6	4.5
U_{OUT},V	5.0	3.5

loading characteristics



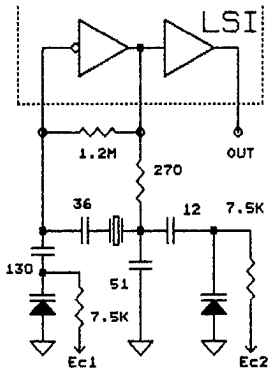
frequency vs. supply voltage

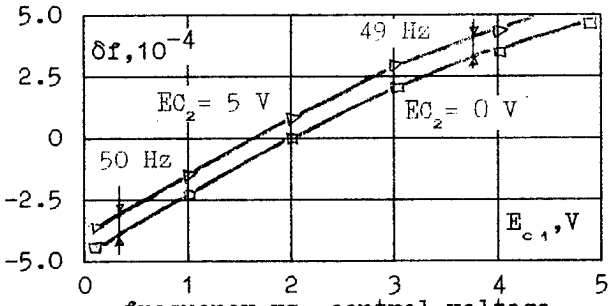
Fig.16. Miniature DTCXO oscillators investigation results

3.2. Test results

Two types of oscillators were tested:
- inner inverter-based controlled oscillator;
- outer transistor-based ontrolled oscillator.

Fig.16 and Fig.17 shows scillator circuits and their parameters. Fig.18 shows typical F-T curve of one of the oscillators. Great importance was attached to investigation of output signal noise parameters. Fig.19 shows Allan variances -averaging

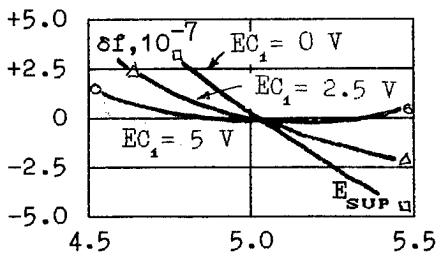




frequency vs. control voltage

Z	0	10 K	1 K
δf	0	$1.5\cdot10^{-7}$	$4\cdot10^{-7}$
I_C,mA	1.5	1.8	3.0
U_{OUT},V	5.0	4.8	3.5

loading characteristics



frequency vs. supply voltage

Fig.17. Miniature DTCXO oscillators investigation results

time relationship. Fig.20 shows typical output signal spectrum. The obtained results show that

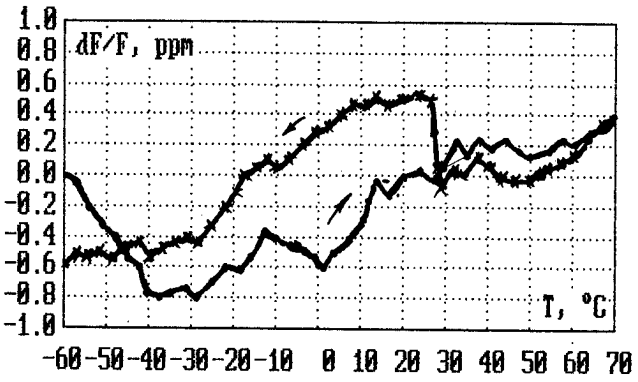


Fig.18. Miniature DTCXO F-T curve

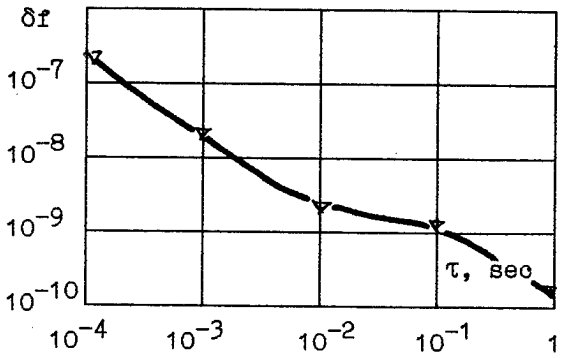


Fig.19. Miniature DTCXO output signal Allan variances vs. averaging time

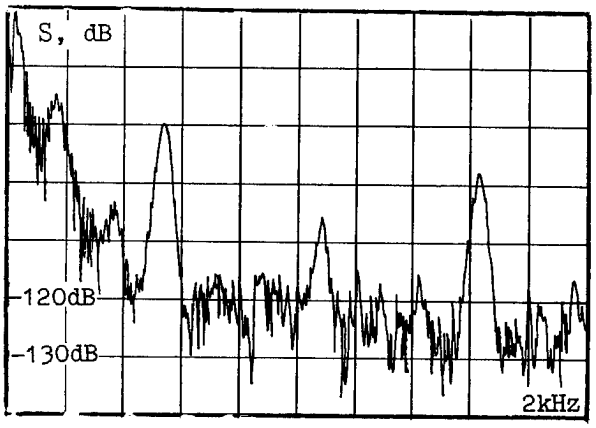


Fig.20. Miniature DTCXO output signal spectrogram

application of statistic piecewise-linear interpolation does not deteriorate output signal parameters and considerably simplifies the circuit. Oscillator parameters have been tabulated:

Output frequency	10 MHz
Operating temperatures range	-40...+85 C
Frequency stability over the temperature range	3 10
Supply voltage	5.3...15 V
Consumption current	5 mA

4. PRECISION 100-MHz DTCXO.

The proposed method of forming CF can be successfully used to design a DTCXO of higher class. In some devices it is desirable to have an oscillator featuring 50...100 MHz output frequency, high temperature and long - term stability (better than 1 10), as well as small consumed power. Traditional DTCXO version built according to the circuit similar Fig.10 cannot provide such parameters by some reasons:

- use of DTCXO resonator excited on the 1-st harmonic to 100 MHz will not provide high long-term stability;
- use of high stable resonator excited on higher harmonics will provide oscillator electronic trimming insufficient for compensation of temperature drifts;
- ageing of reference voltage source badly affects frequency stability;
- temperature - dynamic distortions due to temperature gradient between thermotransduser and reference resonator will be visible on the general background.

DTCXO with quite new level of parameters can be realized according to the circuit on Fig.21. The circuit is featured with using double - mode crystal oscillator with double - rotated cut resonator and frequency synthesizer.

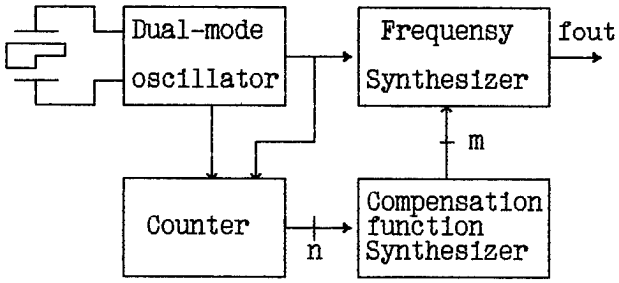


Fig. 21. Precision DTCXO symplified block diagram

4.1. Double-mode oscillator.

It is known that high temperature - dynamic stability of DTCXO can be obtained by application of resonators having double - mode excitation. It is result of ideal space matching of a resonator and a thermotransducer. Moreover, suitable for double - mode excitation resonators usually have smaller temperature-dynamic sensitivity than AT-cut resonators. At first, great hopes of this sort were associated with SC-cut resonators excited on modes B and C. It is attractive that mode C has low temperature - dynamic coefficient and mode B has high F-T curve steepness. Attempts of practical employing such resonators, but, revealed their essential defects: change of equivalent resistance over a temperature range: activity falls and abrupt frequency hoppings at certain temperatures (sometimes over very narrow interval). Cause: interaction and energy pumping

out of one oscillation into another under definite frequency relationship. These defects proved to be so serious that some investigators began to believe this method to be non - perspective and commenced to investigate resonators excited on the 1-st and the 3-rd mechanical harmonics of mode C[8] By the way, much earlier, Soviet scientist L. Marianovsky offered this method, but with employing AT - cut resonators [9]. Not going into detailed comparison of the methods, which can be subject of a separate article, we consider the resonators with modes B and C to be more perspective. This conclusion was supported to great extent by investigations of I. Abramzon and A. Dikidji who developed resonators without falls of activity [10], satisfying our requirements. A few versions of two-frequency oscillators are known. The most simple of them are circuits with separated non-linearities and mode filtration. Usually such circuits contain many selective loops [4] making them non-technological. In our opinion, the most perspective is the method of automatic amplitude balance of excited modes, developed in our laboratory by V. Samoilenko. The method is based on the mechanism of asynchronous stimulation of oscillations due to presence of zones having positive second derivative in the oscillatory curve of the oscillator. Having auto-bias circuit and cut-off angle less than 90, oscillator can generate mild two-frequency oscillations. These oscillations are stable, but, only in narrow zone of regimes. Since oscillator transistor operates in non-linear regime, the output oscillation spectrum contains except excited modes their combinative components. At high level of both generated frequencies, output spectrum will be "dirty". Controlling transistor operation by direct current, the level of less active oscillation in output signal spectrum can be changed to a great extent. Yet, operation regime at fixed bias is not stable. At supply changing and at other distabilizing actions the level of combinative frequencies can sharply, grow or, on the contrary, the oscillator can pass to one -frequency regime. To stabilize the regime, V. Samoilenko offered to introduce into the circuit a channel of automatic balancing of the level of excited modes [11] (Fig.22).

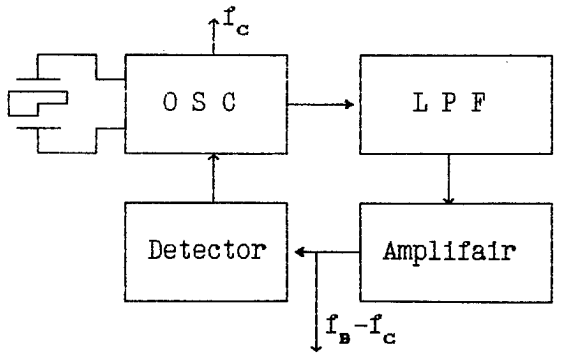


Fig. 22 . Dual-mode oscillator block diagram

Multifrequency signal is derived of one of oscillator points having less loading influence to the f stability. High frequencies are suppressed by the filter, and the most low - frequency component with frequency $f_B - f_C$ is amplified on 40...60 dB and is detected. The detected signal controls oscillator operating regime by direct current. Such circuit can provide stable two - frequency oscillation with suppression of secondary components till -60 dB. Amplified and filtered signal with frequency $f_B - f_C$ can be used as a signal of thermotransduser. V. Samoilenko oscillators assembled according to a modified Colpitts circuit are very critical to a relationship of equivalent resistance of excited modes. We have offered two-transistor circuit [12] which has greater advantages and is capable to excite resonators with larger interval of equivalent resistances R_B/R_C ratio.

4.2. Frequency synthesizer

Not every circuit of frequency synthesizer will be optimal for use in DTCXO due to a number of specific requirements. The most essential among them are:

- output frequency mast be not less than 100 MHz;
- frequency retuning range: $(0.5 \div 2) \cdot 10^{-4}$;
- grid spacing: $(1 \div 2) \cdot 10$
- output signal spectrum must be maximum pure;
- high quick-action of switching is not needed;
- small non-linearity of control curve is admissible;
- consumed power and overall dimension must be minimum;
- synthesizer must operate with statistic CF synthesizer non-violating interpolation law.

Frequently such requirements are met by combination of two-level and PLL synthesizers. PLL synthesizer used in DTCXO [13] does not provide interaction with statistical CF synthesizer since the process of division coefficients setting is synchronized with output frequency but not with temperature count interval. Besides, the law of synthesizer retuning is very nonlinear causing non rational use of divider rateness. We have proposed synthesizer circuit [14] markedly getting over these chortcomings. Fig.23 shows block diagram containing such synthesizer. This DTCXO consists mainly of PLL-circuit firming stable (in ideal) frequency f_{OUT} from reference frequency f_R, unstable in temperature range. Mixer of PLL system operates on harmonics of reference oscillation using multiplication factor 10. In static mode the input frequencis of a phase detector (PHD) must be equal

$$(f_{OUT} - 10f_R)/N_3 = f_{SYNT}$$

Trimming output frequency f_o must be fulfilled by changing value of synthesized frequency f_S so that at any values $f_R(T)$ the output frequency f_o to be constant. Synthesizer frequency f_S is formed of reference by removal of definite number of pulses.

To decrease pulse modulation the obtained sequence in divided (by frequency) into N times. A number of removed pulses depends on code coming from ROM output.

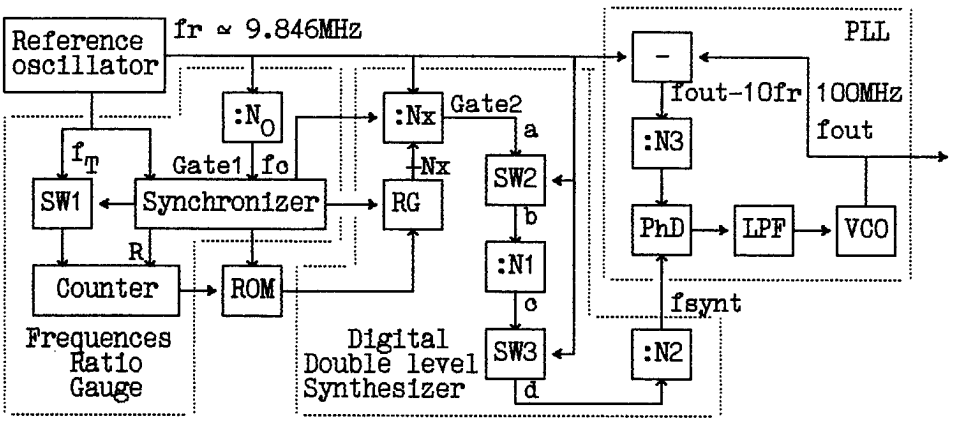


Fig. 23. Precision DTCXO block diagram.

4.2.1. Design relationships.

Division coefficients of dividers (Fig.23) are strigtly associated with DTCXO output parameters. Rate value of synthesized frequency $\Delta f_{SYNT} \approx f_R/N_2$. Range of this change is $f_{OUT} = f_R/(N_1 \cdot N_2)$. Frequency trimming increment $H_{SYNT} = f_R/(N_1 \cdot N_3 \cdot N_4)$, where N_4 is the most possible value of N_X. Division coefficient N defines time of temperature frequency count (its selection procedure was given earlier); $N_o = Mf_R/S\Delta T$, where M is a number of interpolation zones. The table below gives needed number of zones at operating temperatures -40 ÷ +80 °C.

M	16	32	64	128	256
δf	$4 \cdot 10^{-7}$	$8 \cdot 10^{-8}$	$2 \cdot 10^{-8}$	$0.5 \cdot 10^{-8}$	$1.2 \cdot 10^{-9}$

Possibility of discrete changing N_o must be provided to work in various temperature ranges using different thermotransducers.

To improve filtration conditions, phase detector operating frequency f must be maximum high. But it cannot be extremely high, because here its forming accuracy decreases and phase modulation increase. Optimum values f_S are 150 ÷ 600 Hz at f_R= 9984 kHz, f_{OUT}= 100 MHz, S= 400 Hz/°C, M= 256. Division coefficients will be N_o = 49152, N_2 = 65536, N_4 = 65536; N_1 = 16, N_3 = 1024. Phase detector operating frequency f_S=148 Hz, count time is 5 ms.

4.4.2. Making compromise between PLL dynamic and filtering parameters

Presence of piecewise-linear interpolation system originates specific components in the spectrum of control operation (See part 3). Besides, so-called "fractionness interference" is typical to such synthesizers. It causes necessity of careful study of the control signal filtration channel whose parameters will affect, among others process of frequency and temperature-dynamic distortion formation. To compute dynamic and filtering parameters it is convenient to represent PLL circuit as on Fig.24. Parameters of this system are heavily affected by phase detector transfer characteristic. We used a detector with saturation areas and three output states. Theoretical and experimental investigations as well as computer simulation permit to draw the follow conclusions:

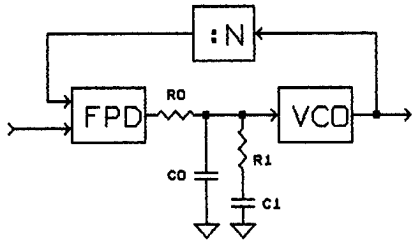
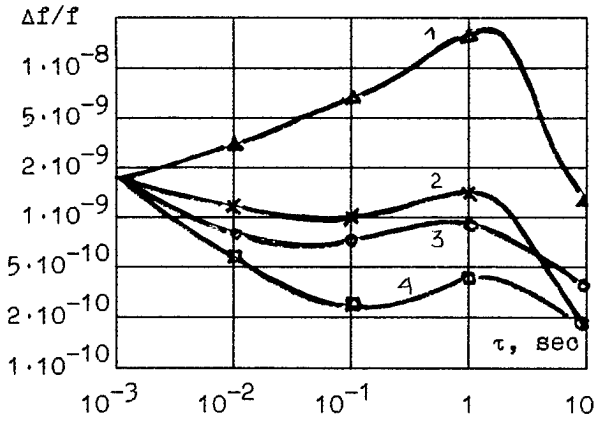


Fig.24. PLL system model.

- within linear system one cannot get good filtration of low frequency disturbance effects and small duration of transient process at the same time;
- form of transient curve is determined by total transfer constant of the closed system and depends on R and R relationship;
- duration of transient process is defined, in general, by time constant $\tau = (R_o+R_1) \cdot C_o$;
- cutt-off frequency of PLL amplitude - frequency curve, equivalent to LPF, for disturbance effect more depends on resistor R , but a degree of interference suppression beyond the bend depends on R_o/R_1

Use of non-linear filter can considerably improve dynamic parameters and maintain good filtration. In quasy-static mode its equivalent time constant increases, but in transient modes it decreases. Great attention was drawn to switching on process dynamics since among disturbance effects the initial switching on was the most strong. Good results were obtained by application a forcing circuit decreasing R (Fig.24) at initial time. Fig.25 show experimental results of Allan variance measuring at different LPF parameters.

Fig.26 show a process of DTCXO frequency setting (three realizations) in optimum case.



N	RO	CO	R1	C1
1	22K	1M	47K	22M
2	220K	1M	47K	22M
3	1.6M	1M	1.6M	22M
4*	220K	1M	100K	22M

*compensation system OFF

Fig.25. 100 MHz DTCXO output signal Allan variances vs. averaging time

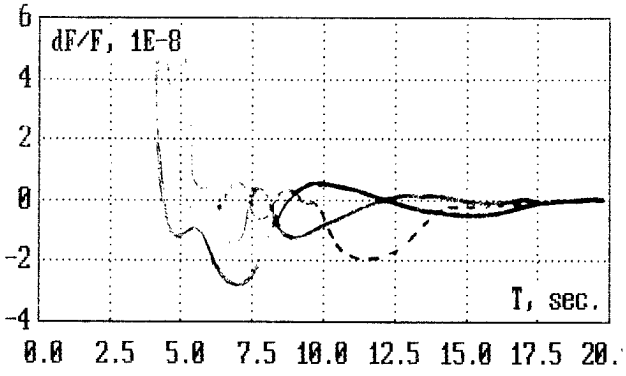


Fig.26. Frequency setting process. (3 realization)

4.4.3. Realization and experimental results.

By currently an intermediate version of 100 MHz DTCXO has been completely tested. It differs from end version by large overal dimensions. Two - frequency oscillator contains two hybrid microassemblies, and a digital part contains three half-customed LCI. Output VCO has 100 Mhz resonator exited at the first mechanical harmonic. Fig. 27 shows experimental F-T curve of the oscillator, and Fig.28 shows output oscillations spectrogram.

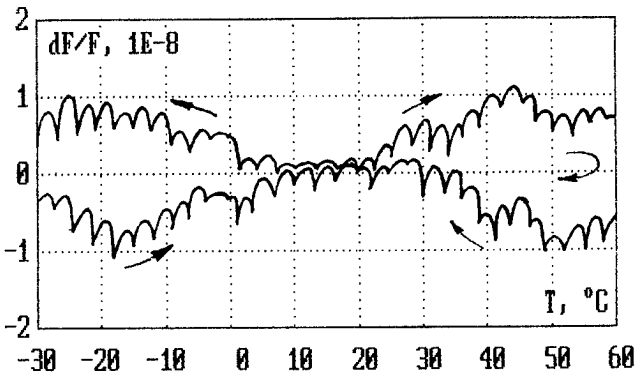


Fig.27. 100 MHz DTCXO F-T curve.

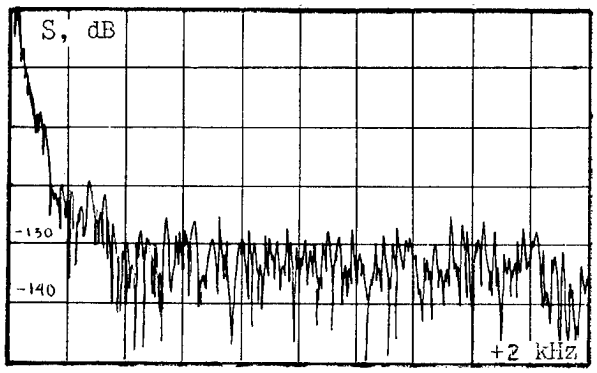


Fig.28. 100 MHz DTCXO signal spectrogramm

CONCLUSION

Analysis made in this paper permits thoroughly estimate new method of forming compensation function. The obtained results allow to consider it to be highly effective and its realization to be very simple. It is easy to notice that the circuit (Fig.15) partically coincides with circuits of many oscillators with stepped CF forming. Only synchronizer has a peculiarity. Its designing must be carried out with great care (identical at first sight circuits differing only by inverter on the counter input produce absolutely different results). In fact we only professionally used possibilities existing in process of measuring aliquant frequencies. However, the results of stability improvement are so essential (dones times) that it can be said about new approach.

The most features were evoked by DTCXO output oscillation spectrum since the control operation here is obtained in result of filtering two-lewel sequence with great drops. Nevertheless, both theoretical and experimental evidences have shown that do not present essential difference to the worst side from traditional DTCXO.

Extending named principles we have developed statistical parabolic interpolation methods possessing great accuracy [15]. But this is rather exotic. Just the method of statistical piecewuse-linear interpolation can be consider reliable "working horse". The method is useful both for designing the simplest and for precision DTCXO.

REFERENCES

1. Kosykh A., Ionov B., Vasiliev A. Temperature-dynamic model and temperature dynamic compensation of crystal oscillators. - Proc.46 A.F.C.S., 1992, p.425...433.

2. M.E.Frerking. Crystal oscillator design and temperature compensation. N.Y. ,Van Nostrand, 1978.

3. Curry G. Oscilloquartz D-TCXOs the case for defence. Electronic Engineering, November, 1988, p.41...47.

4. Benjaminson A.,Stalling S.C. A microcomputer compensated oscillator using a dual - mode resonator. Proc. 43 A.F.C.S., 1989.

5. Прак, Педьюто. Система температурной компенсации частоты на основе цифровых интегральных микросхем для задающих генераторов. Электроника, 1972, N 17, с. 63...66.

6. Invention certificate (Patent) SU1084938A (USSR) Термокомпенсированный кварцевый генератор /Багаев В.П., Косых А.В., Лепетаев А.Н. Опубл. Б.И. N 13, 1984.

7. Invention certificate (Patent) SU1206820A (USSR) Стохастический кусочно-линейный интерполятор /Косых А.В., Багаев В.П. Опубл. Б.И. N 3, 1986.

8. Schodowski S. Resonator Self-Temperature-Sensing using Dual-Harmonic-Mode oscillator. Proc. 43 A.F.C.S., 1989.

9. Invention certificate (Patent) N760398 (USSR) Кварцевый генератор / Марьяновский Л.С. Опубл. Б.И. N 32, 1980.

10. Abramson I. Two-mode quartz resonators for digital temperature compensated quartz oscillators. Proc.46 A.F.C.S., 1992.

11. Invention certificate (Patent) SU1145450 (USSR) Термокомпенсированный кварцевый генератор /Самойленко В.Ф. Опубл. Б.И. N 10, 1985.

12. Invention certificate (Patent) SU1573524 (USSR) Двухчастотный кварцевый генератор /Багаев В.П., Завьялов С.А. Опубл. Б.И. N 23, 1990.

13. Warwick G.A., Cosling W., Prescott A.J. A digital technique for temperature compensation of crystal oscillators. Proc. Conf. Radio Receivers and assoc. systems. Southampton, 1978, London, p.207...216.

14.Invention certificate (Patent) SU1659972A (USSR). Генератор импульсов /Багаев В.П., Косых А.В. Лепетаев А.Н., Ионов Б.П. Опубл. Б.И. N 24, 1991.

15.Invention certificate (Patent) SU1190455A (USSR) Устройство термокомпенсации кварцевого генератора. /Косых А.В., Опубл. Б.И. N 41, 1985.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

UNIVERSAL, COMPUTER FACILITATED, STEADY STATE OSCILLATOR, CLOSED LOOP ANALYSIS THEORY

ANALYSIS OF ALC AND SELF LIMITING CRYSTAL OSCILLATORS

Benjamin Parzen
consulting engineer
3634 Seventh Avenue
San Diego, CA 92103
(619) 291 0567

ABSTRACT

This paper derives and presents, in detail and with very high frequency resolution, the computer aided procedures for calculating the total performance of ALC and self limiting oscillators. It treats topics 1- the determination of the type of self limiting in effect in a given oscillator; 2- the use of commercially available computer aided analysis programs, and their attendant problems, in the analysis of ALC and self limiting oscillators; 3- the formulation of additional theory and procedures required to adapt the titled theory to self limiting oscillators; 4- the application of computer program BPT to topic 3 ; and 5- oscillator zeroing.

1. INTRODUCTION

Ref 1 presented the titled theory and some applications to UHF and microwave oscillators.

Ref 2 presented some applications of the titled theory to precision crystal oscillators.

While not explicitly stated it is evident that, since the DC environment was not considered, all the oscillators were provided with ALC systems, represented by dR.

This paper clarifies ALC oscillators and attempts to describe the analysis procedures for self limiting oscillators. The paper was made possible by new insight into oscillator non linear behavior gained from the titled theory. This has great importance as the great mass of crystal oscillators falls into the self limiting category. Note that some of the material, developed, is applicable only to harmonic oscillators.

The nomenclature and symbols are those in Ref 3 complemented by Ref 1. The reader's particular attention is directed to the List of Symbols beginning on Page 441 of Ref 3.

The literature is conspicuous in its absence of a satisfactory treatment of the self limiting oscillator. One exception is Ref 3 which contains a simplified presentation of self limiting in oscillators and the reader is referred to Sects 2.4, 2.5, and 2.7 of Chpt 2 and Chpt 6, which present the basic theory, and the sections on limiting in Chpts 7 to 11, which include applications. The reader is urged to acquire a working knowledge of the material therein as it is essential for understanding much of this paper.

The major function, of limiting is to fix the magnitude of the carrier amplitude, Funct A. A secondary function, Funct B, is that in performing Funct A, it does not deteriorate the oscillator frequency performance, i.e. the short and long term frequency stability.

1.1 Funct A Fundamentals

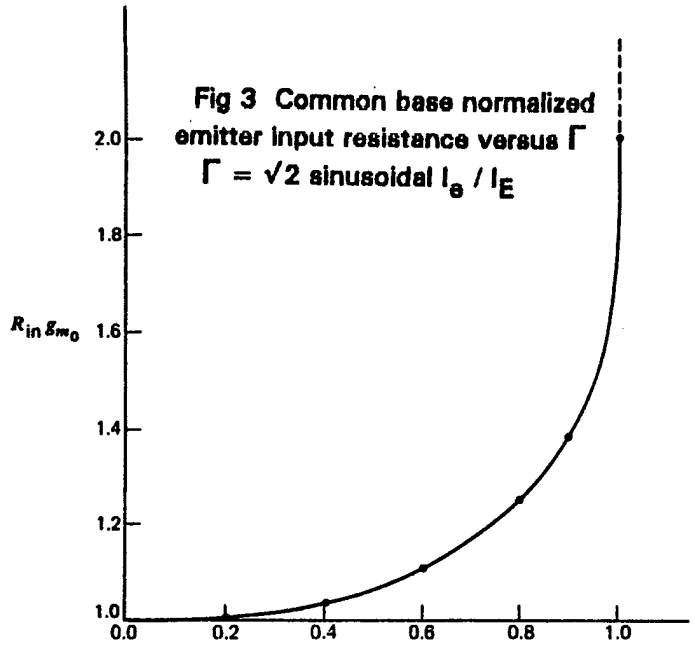
1.1.1 All oscillators

The oscillator carrier amplitude is indirectly controlled by the real parts of basic Eq 1 or 2, in Sect 2.1, which requires that the value of a parameter of a physical component, Pa, such as the gm of a transistor, is set so that the equation is satisfied. The system performing these operations is called the LIMITING system and is represented by dR (see Ref 1).

1.1.2 ALC oscillators (see Ref 1 and Chpt 14 of Ref 3)

In the linear oscillator, the equation is satisfied over

0-7803-0905-7/93 $3.00 © 1993 IEEE



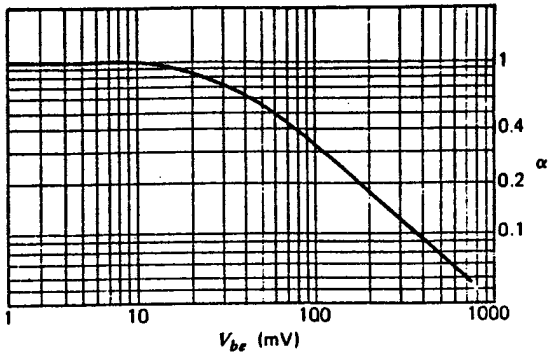


Fig 1 Common emitter α versus V_{be}

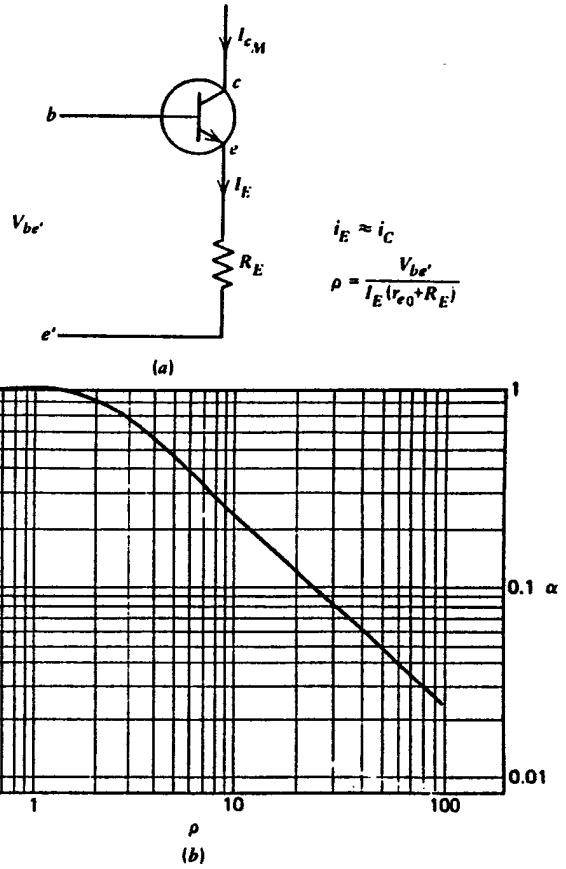


Fig 2 Common emitter with emitter degeneration
a. Model of circuit and definition of ρ
b. α versus ρ

BIPOLAR TRANSISTOR LARGE SIGNAL PARAMETERS

a wide range of oscillator amplitudes. The desired amplitude is set by an ALC (Automatic Level Control system) circuit including an output amplitude sensor, controlling a DC voltage generator, that produces V_O. V_O is subtracted from a reference voltage, V_{REF} to create V_E. V_E in turn controls Pa. Only 1 value of V_E can cause the equation to be satisfied. Therefore, to change the amplitude requires changing V_{REF} or V_O. The starting loop gain, ALo, is controlled by the ALC system to be larger than 1, since at starting V_O --> 0 and V_E --> V_{REF}.

1.1.3 Self limiting oscillators (see above mentioned refs)

The limiting is based on the non linear behavior of active devices such as transistors or passive devices auch as diodes. Typical important mechanisms are

1.1.3.1 Sinusoidal v_{BE} or $v_{BE'}$ (see Sect 6.2 of Ref 3)

a. "be", base emitter cutoff for bipolar transistors (gate source cutoff for fets) in which the large signal gain of the active device, A, is a function of the input signal amplitude and decreases as the input signal increases. This mechanism is operative when $v_C$min always exceeds $v_B$max. The degree of limiting is controlled by the input signal amplitude between base and emitter and the starting loop gain, ALo. ALo is approximately the starting A, Ao, divided by the steady state A.

The most important oscillator group with this limiting type is the Pierce oscillator family, which is probably also the most important oscillator family, and is fully described in Chpt 5 of Ref 3. In this group, A is proportional to the transistor gm so that Alo = gmo /gm where gmo is the starting and gm is the steady state gm. (Note that gm is used to denote both the generic and steady state gm).

Figs 1 and 2 shows the theoretically derived relationship between α = gm /gmo and the input voltage for silicon bipolar transistors. (Figs 1 and 2 are respectively Figs 2.12a and 2.13a in Ref 3)

b. Collector voltage limiting in which the non physical ac resistance, r_{ce}, is a function of v_C compared to v_B and decreases the more v_C swings below v_B. The degree of limiting is controlled by the base and collector bias voltages and the starting loop gain. The starting loop gain, ALo, is RT at starting, RTo, divided by the steady state RT where the change in RT is caused by the change in r_{ce}.

It is very desirable that the bias voltages be set so that a and b are not simultaneously operative.

"be" limiting is preferred over collector voltage limiting for the reasons in Sect 6.2.2 of Ref 3.

1.1.3.2 Sinusoidal i_E, where the crystal is in series with the emitter. (see Sect 6.3 of Ref 3)

a. Variation of the dynamic emitter resistance, Rin, as a function of I_E and I_e, which increases sharply as I_E /I_e increases. This mechanism is operative when $v_C$min exceeds $v_B$max. The degree of limiting is controlled by setting I_E and the starting loop gain. The starting loop gain, ALo, is RT at starting, RTo, divided by the steady state RT where RT approximates Rxtal + Rin. This type of limiting is also called Rin limiting. Examples of oscillators with this type of limiting are the Butler oscillator family described in Chpt 11 of Ref 3 and many of the oscillators in Chpt 13.
Fig 3 shows the theoretically derived relationship, for bipolar transistors, between Γ = SQRT(2) * I_e / I_E and Rin /Rino. (Note that Rino = 1 / gmo.) (Fig 3 is Fig 2.15a in Ref 3).

b. Collector voltage limiting. Identical to Sect 1.1.3.1b.

These types of limiting result in substantial degradation of the operating Q for high starting loop gain. They produce good noise performance at high Fourier frequencies since the crystal is effectively in series with the load.

1.1.4 Oscillators with diode limiters (see Sect 6.5 of Ref 3)

This is similar to Sect 1.1.3.1b except it permits much better control of the diode r_{ce} and the oscillator amplitude as the diode bias supplies are independent of the transistor's supplies.

1.2 Methods of Determining the Performance of Self Limiting Oscillators with Computer Aided Analysis Programs.

The reader is cautioned that in any computer aided analysis, it is extremely important that the oscillator entered into the computer truly represents the real oscillator. Most of the discrepancies between the

computer predictions and the experimental performance are often due to the disagreements between the computer and real oscillators' descriptions.

1.2.1 Carrier output, Funct A.

1.2.1.1 Fundamentals of computer aided analysis.

1.2.1.1.1 *Linear analysis programs*

Determining the carrier amplitude is a highly non linear problem. Therefore the customary linear analysis programs such as Touchstone and Supercompact are useless.

1.2.1.1.2 Non linear analysis programs

The transient analysis facility of nonlinear programs such as Spice, Libra, and Harmonica and similar programs can be used to simulate this problem.

Most of these programs include very good transistor models which are particularly useful at the higher carrier frequencies. However often the usefulness of the models is limited by the spread of the parameters of the real transistors. Obviously transistors having identical parameters as the model can be procured but at probably prohibitive cost. Also, whenever feasible, the transistors should be so stabilized that transistors with closely controlled parameters are not required.

Many of these programs, when applied to high operating Q crystal oscillators, are prohibitively slow. The programs with harmonic balance capability are considerably faster but still slow.
An acceptable speed can often be obtained, with an acceptable loss in accuracy, by replacing the simulated HI Q crystal with a much lower Q crystal but with the same R1.

Note that these programs also produce information on the wave shapes.

1.2.1.1.3 Program BPT

This program, although a linear program, together with the titled theory and the theory on limiting cited above will also solve this problem, with perhaps less accuracy than the non linear analysis programs but much faster and with greater understanding of oscillator operation.

For bipolar transistors, the loss of accuracy is due to the very simple transistor model, based upon theoretical considerations, which should be sufficient for oscillators with DC stabilized high Ft transistors. Also, the small shifts in the transistor DC biases, caused by the presence of the AC voltages, are neglected. Stabilization produced by a large amount of DC emitter degeneration is effective and improves the accuracy.

The problem is obviously highly non linear but the intuitive heuristic assumption is made that calculations, for high Q circuits, need consider only the fundamental component of the quantity, being calculated.

Of course, BPT produces no information on the wave shapes but does treat the oscillator phase noise.
It also requires the user to be more knowledgeable.

1.2.2 Oscillator phase noise

To the writer's knowledge, the only universal steady state oscillator simulation program that predicts phase noise is program BPT which is the only program to simulate the steady state operation in the frequency domain. This program is made possible by the titled theory.

2. THE ANALYSIS OF ALC and SELF LIMITING OSCILLATORS WITH PROGRAM BPT

IMPORTANT NOTE : On line auxiliary program *HELP* is always callable from within and outside BPT to help in performing the analysis. Items, in this paper, having names in *ITALIC* type are fully explained in *HELP*.

2.1 Oscillator Zeroing PROCEDURE

By oscillator zeroing is meant that the components in the Z configuration are being adjusted so that the equations below are satisfied at the specified frequency. (See Ref 1)

$$Zt = \Sigma Z = dR \quad (1)$$

or

$$Yt = \Sigma Y = dG \quad (2)$$

When Eq 1 or 2 is true then the configuration is ready to be converted into a working N configuration. (see Ref 1)

The zeroing could have been accomplished with only 1 computer command but is was felt that the user would better understand the procedures if each procedure was explicitly executed by the user.

Zeroing can also be performed on a Y configuration but the Z configuration is preferred because the usual resonator characteristics are more easily described and studied in the Z domain.

The zeroing procedure depends on the oscillator circuitry and the type of limiting. It is performed when specified in Sect 2.3.1e at the Fop of Sect 2.3.1d.

2.1.1 For the ALC oscillator
The zeroing procedure is that in section 2.1.2 or 2.1.3 as applicable.

2.1.2 For the Pierce oscillator family

$$RN \approx gm * X1 * X2$$

Therefore

2.1.2.1 The zeroing process is

a. Set the values of RV and XV to 0.

b. Set *FX* to Fop or close to Fop.

c. Execute *AG* so that Rt is less than 1e-6 * RT.
The default *AG* for that and much better accuracy can be obtained if so ordered by the user (coarse zeroing).

d. Execute *T*, for Xt = 0, on the oscillator frequency trimming component so that Eq 1 is satisfied to a max. Xt mag. of < 1E-6 * RT.

e. At this point, Eq 1 is almost satisfied but not well enough. The final step is to execute *AT* , after which |Xt| and (Rt - dR) should be less than 1e-4 * dR (fine zeroing).

Procedures *AG*, *T*, and *AT* are automatic and are described in *HELP*. The total zeroing process takes less than 5 s. in a 386 computer and several times that in a normal AT computer. All computers should of course, be equipped with a math coprocessor.

2.1.3 For the oscillator family wherein the resonator is in series with the emitter

$$RN \approx X1 * X2 / RT.$$

2.1.3.1 The zeroing procedure is the same as in Sect 2.1.2.1 except that step c is replaced by executing *T*, for Rt = 0, on X1 or X2 components so that Eq 1 is satisfied to a maximum Rt mag. of < 1E-6 * RT.

2.1.4 For Pierce family oscillators with "be" limiting the zeroing procedure is the same as in Sect 2.1.2

2.1.5 For the oscillator family with Rin limiting wherein the resonator is in series with the emitter

2.1.5.1. The zeroing procedure is the same as in Sect 2.1.2.1 except that step c is replaced by

c1 Add a resistor, labeled Rin1 in series with the transistor emitter and the resonator, labeled Rin1, if it is not already present.

c2 execute *T*, for Rt = 0, on Rin1.

2.1.6. For oscillators with collector voltage limiting

2.1.6.1 The zeroing procedure is the same as in Sect 2.1.2.1 except that step c is replaced by

c1 Add a resistor, labeled r_{cb}, between the transistor collector and base.

c2 execute *T*, for Rt = 0, on r_{ce}

See also: *HELP OSC KN AG AT T CKT CTR REF*

2.2 CRITERIA for the Determination of the Type of Self Limiting in Effect in a Given Oscillator (see Sect 2.3.1b5)

The method of analysis is strongly dependent upon the type of self limiting employed, as outlined in Sect. 1.1.3.

By now it should be evident that the type of limiting effective in a given oscillator is strongly dependent upon the values of the bias voltages and currents. It is very desirable that these be set so that 2 types are definitely not simultaneously operative.

Since all oscillators are candidates for collector voltage limiting it should be first established whether the limiting is of this type.

2.2.1 Collector voltage limiting oscillators
The oscillator is most likely a collector voltage limiting type if V_{CB} is less than 1.4 V_{cb} for all active devices where V_{CB} is the DC and V_{cb} is the RMS AC, voltage between base and collector or equivalent.

2.2.2 ALC oscillators
The oscillator is an ALC type if 1 of the following

conditions is met

The DC power supply voltages are not specified

V_{be} is less than 20 mV for bips and 50 mV for fets which ensures linear operation.

The oscillator components operate in a linear mode.

2.2.3 "be" limiting oscillators

The oscillator is of this limiting type if is a member of the Pierce oscillator group and it does not meet any of the Sects 2.2.1 and 2.2.2 criteria.

2.2.4 Sinusoidal Ie oscillators

The oscillator is of this limiting type if the crystal is in series with the emitter and it does not meet any of the Sects 2.2.1 and 2.2.2 criteria.

2.3 The Complete Analysis Procedure

2.3.1 The oscillator setup procedure

The procedure varies with the type of limiting and is as follows

a. Enter the complete ZN configuration circuit into the computer via a file or keyboard with *FF* or *KK*. Include components RV, XV, dR, {J}, test point sets for Vin and Vs, and Vn. (see Fig 2 in Ref 1). It is recommended that, for calculating the carrier voltage and current amplitudes, Vn be made a VW, a white noise source. Additional noise sources can be provided after Sect 2.3 is completed. Determine the oscillator Fop with *AF*.

If it is an ALC oscillator go to Step c else go to step b.

b1. Add a testpoint set, a TP component, labeled V_{CB}, across the transistor collector and base to measure V_{CB}

Execute *C* at *FX* = 0, (dc).

C calculates and indicates the voltages across the terminals of all components and their currents.

b2 For the oscillator with be limiting, note the value of the transistor gm in the parts list *L*, and label it gmo.

b3 For the oscillator with Rin limiting, note the value of the transistor I_E in the *C* display.

b4 Make a hard copy of the data for oscillators with collector limiting.

b5. Examine the circuit and the voltages to determine the probable type of limiting in accordance with the criteria of Sect 2.2.

c. Execute *KZ* which converts the ZN configuration into a Z configuration.

d. Execute *AF* which calculates the oscillator frequency, Fop.

At this point, the starting loop gain can be determined by executing *KN* followed by *AF*. Execute *KZ* before going to Step e.

e. Zero the oscillator at Fop in accordance with Sect. 2.1

f. Execute *KN*, which converts the Z configuration into a ZN configuration, and *C*.

2.3.2 Procedure for the determination of the value of dR in the ZN config.

This procedure depends on the oscillator circuitry and the type of limiting.

The value of dR is based upon 1 additional fact derived from the limiting and the *C* data.

2.3.2.1 For the ALC oscillator

The fact is obtained from the setting of dR to produce a desired carrier signal at a specified location. Typical signals are the crystal current and Vin in connection with the residual noise concept.

Therefore adjust dR until the desired signal amplitude, Ax, is obtained, using *C*.

It is not necessary to hunt for dR. Instead use the relationship:

If dR1 produces an amplitude A1, then Ax corresponds to a dRx = dR1 * A1 / Ax.

2.3.2.2 For Pierce family oscillators with be limiting

a. Note the value of transistor gm in the parts list with *L*.

b. Calculate α = gm / gmo. gmo was noted as per Sect 2.3.1.b2

c. Read V_{bex} from applicable Figs 1 or 2.

d. Adjust dR until $V_{be} = V_{bex}$

e. Now that the complete voltage performance is known, check, via Sect 2.2, that the correct type of limiting was chosen.

2.3.2.3 For Rin limiting oscillators

a. Note the value of Rin1 and the transistor gmo = gm in the parts list with *L*.

b. Calculate Rin = (Rin1 + 1 /gmo)

c. Calculate Rin * gmo and read Γ from Fig 3.

d. calculate $I_{ex} = \Gamma * I_E$ / SQRT 2. I_E was noted as per Sect 2.3.1.b3.

e. Adjust dR until $I_e = I_{ex}$

f. Now that the complete voltage performance is known, check, via Sect 2.1, that the correct type of limiting was chosen.

2.3.2.4 For oscillators with collector voltage limiting

a. Adjust dR until $I_e \approx 1.4 * I_E$ in the zeroed oscillator. I_E is obtainef from the data of Sect 2.3.1.b4. (see Eq 6.9 of Ref 3)

b. Now that the complete voltage performance is known, check, by confirming that 1.4 * Vcb ≈ VCB, that the correct type of limiting was chosen.

2.3.2.5 Save the ZN circuit, convert it into the corresponding Z configuration and save it.

The oscillator Z and Zn configurations are now ready for performance testing and for use as simulated oscillator versions of the real oscillator using all the facilities in BPT including noise calculation, CTR.

2.4 S PARAMETER ACTIVE DEVICE SELF LIMITING OSCILLATORS

It is assumed that the device's large signal s parameters and residual noise data were taken at the device's operating point when in the oscillator and that the operating point is in the limiting region. (see Ref 5) The analysis is then performed in the same manner as that of the ALC oscillator. The device characteristics should not be altered during zeroing.

3. CONCLUSIONS

The oscillator ALC and self limiting systems have been clarified to increase the users' comprehension of the mode of operation, design and analysis of these systems. In particular, the ALC system was viewed in terms of the much simpler, but more unfamiliar, oscillator, rather than control system parameters.

Algorithms, based upon the titled theory and the limiting theory in Ref 3, have been presented in detail to be used in the analysis of self limiting oscillators. The algorithms enable the calculation of the carrier steady state approximate magnitude with the aid of linear analysis programs. The algorithms are easily applied in program BPT and, with much greater difficulty, in some commercial analysis programs.
If the active device residual noise or noise sources data is available, then the noise performance of the oscillator can also be calculated.

Zeroing, which is an important part of the theory and algorithms has been intensively treated for all types of limiting.

Active devices, included in the limiting oscillators, are bipolar and FET transistors and devices characterized by S or Y parameters.

REFERENCES

[1] B. Parzen, "Universal, Computer Facilitated, Steady State Oscillator Analysis Theory and Some UHF and Microwave Applications", Proc. 45th Annual Frequency Control Symposium, pp. 368-383. May 1991.

[2] B. Parzen, "Universal, Computer Facilitated, Steady State Oscillator Analysis Theory, Some Applications to Precision Crystal Oscillators", Proc. 46th Annual Frequency Control Symposium, pp. 453-464, May 1992.

[3] B. Parzen, Design of Crystal and Other Harmonic Oscillators. New York: Wiley, 1983.

[4] W.P. Robins, Phase Noise in Signal Sources. London: Peter Peregrinus Ltd, 1982.

[5] G. Montrose et al., "Residual Noise Measurements of VHF, UHF, and Microwave Components", Proc. 43rd Annual Frequency Control Symposium, pp. 349-359, May 1989.

[6] B. Parzen, Oscillator and Stability Analysis in the Immittance Domain. In preparation.

APPENDIX A

Glossary of referenced BPT routines

A Add component

AF Calculate the Fop of Y, Z, and N configs.
Fop is defined as the frequency at which Xt or Yt = 0.
For a Z config., the linewidth ia also calculated.
For a Y or N config., the loop gain ia also calculated.

AG The automatic setting of the gain of an sctive device of a Z config. so that Rt = dR.
The gain parameter may be the gm of a transistor, the gain of an OP AMP, the mu of a tube or _21 of an S or Y parameter network.

AT The automatic fine tuning of a Z or Y config so that Zt = dR or Yt = dG.

C Calculation of all component voltages and currents.

CTR Calculates the CTR function of an oscillator (see Ref 1)

E Calculate the voltage as a function of any selected swept parameter of a component.

FF Loads a circuit file.

FX Sets the fixed frequency of calculation.

KK Loads a circuit, part by part, via the keyboard.

KN Converts a Z config. into a ZN config.

KZ Converts a ZN config. into a Z config.

L Displays the parts list.

T The automatic tuning of an immittance component parameter so that Rt = dR or Xt = 0 as selected.
The component may be a R, l, C, X, Y, Z, or transmisssion line component.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

DYNAMIC ANALYSIS OF MODULATED OSCILLATORS

R. L. Clark
McCoy Electronics
100 Watts Street
Mt. Holly Springs, PA 17065

D. T. Comer
Penn State University at Harrisburg
Middletown, PA 17057

Abstract: This paper provides a dynamic analysis method and model for analyzing oscillator circuits which may be modulated by nonlinear signals, or by linear signals which interact with resonator elements to produce spurious modes. This method proves to be more general than a traditional static analysis and more accurately predicts oscillator performance over a wide range of modulation conditions.

Background

It has been known for some time that the frequency modulation of oscillators at rates corresponding to resonator spurious mode offset frequencies can result in erratic oscillator performance [1]. In a more general sense, any modulation process, including noise effects and mechanical vibrations as well as intended modulation, may induce behavior somewhat different than a traditional analysis [2] would predict. A quantitative understanding of the type of oscillator behavior has been elusive, however.

The analysis of modulated or voltage controlled oscillators is traditionally based on either open-loop gain/phase analysis or negative resistance. In the former, the oscillation frequency is determined from the zero phase crossover point of the open loop transfer function. In the latter, the oscillation frequency is determined by finding the frequency at which the circuit impedance is precisely 180 degrees. A common method of modulating an oscillator is to change the bias voltage across a varactor. Voltage controlled oscillator analysis generally then proceeds by finding a series of discrete oscillation frequencies which correspond to certain oscillator component values. This type of analysis is generally referred to as a *quasi-static* or *static* analysis [3].

Of course, if the oscillator is being modulated sinusoidal, then sidebands will be generated. The amplitude of the sidebands are determined from a Bessel series expansion for a modulated sinusoidal function.

An implicit assumption in the preceding methods of analysis is that the oscillator frequency will instantaneously change with changes in component values. In reality, the modulated oscillator is a time varying system which cannot change frequency instantaneously. Although static analysis is extremely useful for predicting many attributes of oscillator performance, there are certain characteristics exhibited by real oscillators which are not revealed by static analysis. One method of overcoming the limitations of static methods is to consider the nonlinear element in the oscillator as a linear time varying element. This approach, taken in this paper, is described as a dynamic analysis as suggested by Smythe [3].

Dynamic Analysis

Consider an oscillator represented by the equivalent circuit of Fig. 1 where the energy to drive the load resistance R_L is provided by the negative resistance $-R$ and the resonator is comprised of the parallel L-C elements. The capacitance is assumed to be made up of a fixed, linear time invariant component C, and a nonlinear component C_1, such as might be created by a varactor diode.

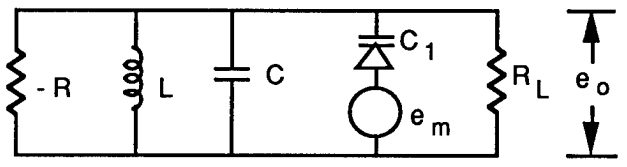


Fig. 1 Modulated oscillator representation

0-7803-0905-7/93 $3.00 © 1993 IEEE

The modulating signal e_m is assumed to be applied in such a way that it modulates the capacitor C_1 but is transparent to the primary oscillation signal e_o and can be either an externally supplied modulation signal, or intrinsic circuit noise caused by shot, thermal, or other noise effects.

Fig. 2 is an alternate representation of the capacitance C_1 where C_0 represents the linear, time invariant component of C_1 and $C(t)$ represents the time variant component, caused to vary by the modulating signal $e_m(t)$. The current $i(t)$ flowing into the equivalent capacitance may be expressed as [4]:

$$i(t) = i_0(t) + i_1(t) = \frac{d}{dt}[C_0 e(t) + C(t)e(t)] \qquad (1)$$

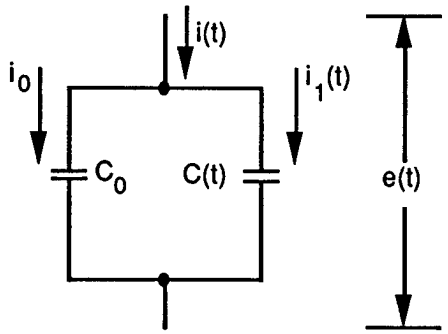


Fig. 2 Equivalent capacitance of oscillator circuit

Since C_0 is time invariant, the first term of eq. 1 is simply:

$$i_1(t) = C_0 \frac{d}{dt} e(t) = C_0 e'(t) \qquad (2)$$

The second term $i_1(t)$ may be expressed as:

$$i_1(t) = \frac{d}{dt}[C(t)\, e(t)] = C(t)e'(t) + e(t)C'(t) \qquad (3)$$

It can be shown that for most practical modulating signals, where the rate of change of the modulation and hence the rate of change of $C(t)$ is small compared to the rate of change of $e(t)$, that the magnitude of the second term of eq. 3 will be small relative to the first term and can be neglected [5]. Under these conditions, the current $i(t)$ may be expressed as

$$i_1(t) \approx C(t)\, e'(t) \qquad (4)$$

And the total current $i(t)$ is given by

$$i(t) = C_0 e'(t) + C(t)e'(t) \qquad (5)$$

Fig. 3 shows a possible equivalent circuit representation of eq. 5 where the first term is represented by the capacitor C_0 and the second term by the dependent current source $i_1(t)$.

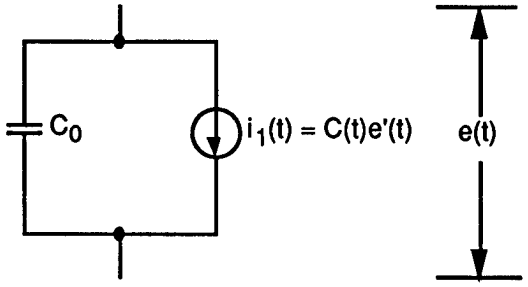


Fig. 3 Eq. 5 equivalent circuit

Fig. 4 shows an equivalent circuit of the oscillator of Fig. 1 in which the linear and modulated components of the capacitor have been replaced by the equivalent circuit of Fig. 3.

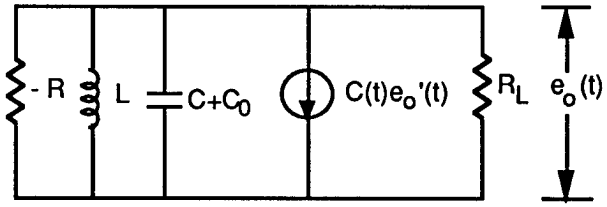


Fig. 4 Equivalent modulated oscillator

The advantage of this transformation is that the oscillator may be analyzed by standard circuit analysis or simulation techniques without simplifying assumptions and can accommodate additional circuit elements to account for spurious mode signals. Fig. 5 shows a modulated oscillator equivalent circuit which includes elements R_X, C_X, and L_X representing a higher order spurious mode in the resonator.

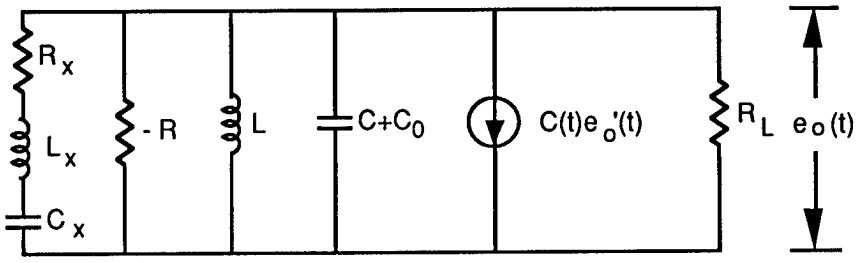


Fig. 5 Modulated oscillator with spurious mode

In the case of a simple harmonic modulating signal with a single pole oscillator topology, it is possible to show that the equivalent circuit of Fig. 4 will accurately predict the sidebands normally associated with such an FM system and gives results similar to a static analysis [5-6]. In the case of more complex modulating signals, or when a more complex oscillator network exists, such as spurious modes as represented in Fig. 5, the method inherently maintains extraneous signal components and is capable of simulating real oscillator behavior for any case that the modulating signal or spurious mode can be properly characterized.

It should be noted that even though the time varying capacitor characterization was developed and illustrated for a shunt/negative resistance oscillator it is in fact a general transformation and is valid for any oscillator in which the modulated capacitance can be identified.

Sinusoidal Modulation

If the modulation signal is sinusoidal, for example,

$$e_m = E_p \cos \omega_m t \tag{6}$$

then eq. 3 becomes

$$i_1(t) = C_1 \sin(\omega_m t) E_p \omega_0 (\sin \omega_0 t) + E_p \cos(\omega_0 t) \omega_m \cos(\omega_m t) \tag{7}$$

For $\omega_0 >> \omega_m$, eq. 7 may be simplified to

$$i_1(t) \approx C_1 E_p \omega_0 \sin(\omega_m t)(\sin \omega_0 t) \tag{8}$$

so that the time varying capacitance of Fig. 1 may be transformed to the circuit of Fig. 6.

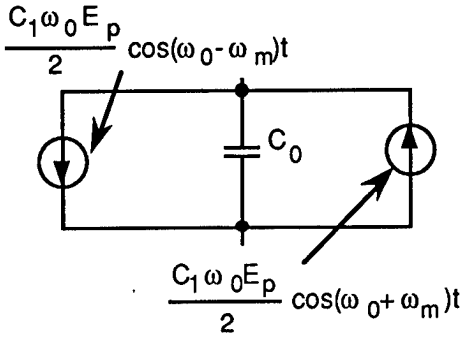


Fig. 6 Equivalent circuit of time varying capacitor with impressed signal

By use of this transform, conventional circuit analysis may now be conveniently applied to the modulated oscillator. Significant deviations from static analysis results were previously reported for the circuit of Fig. 7 where the additional crystal and resistor were added to emulate the effects of a spurious mode in the crystal [1].

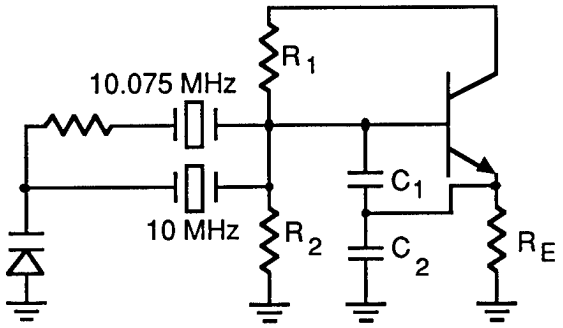
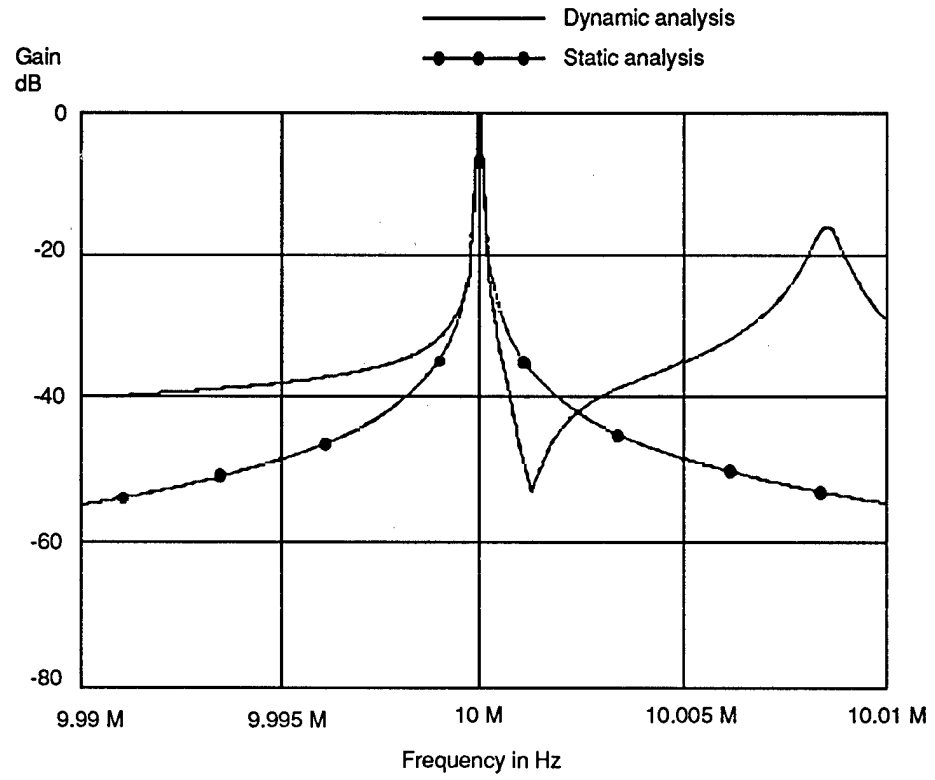
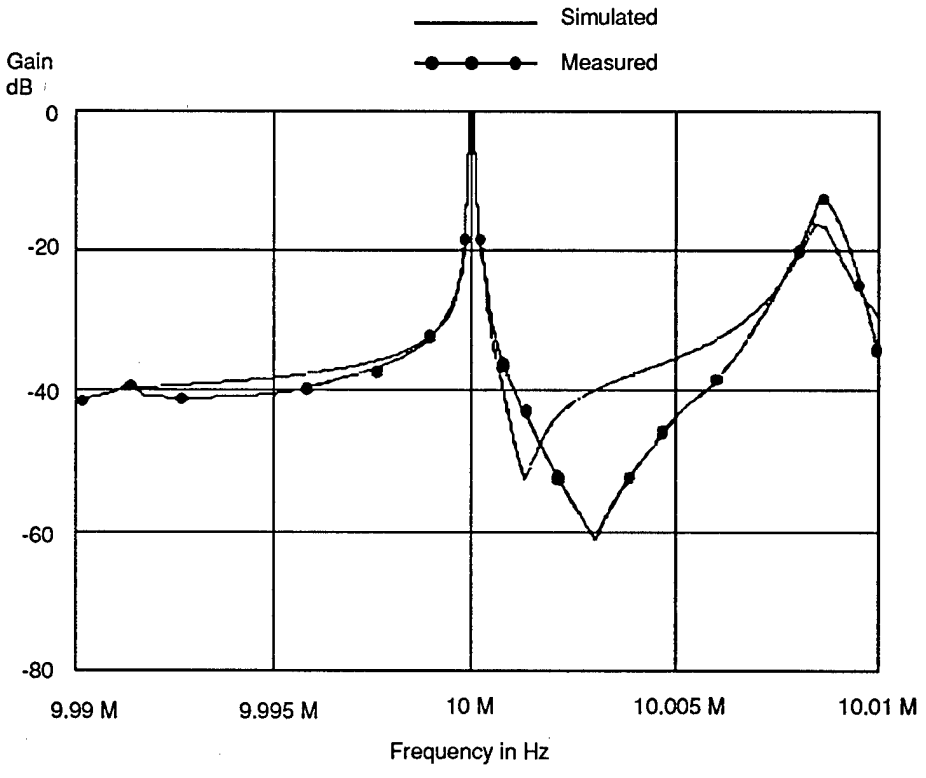


Fig. 7 Modulated Colpitts oscillator with spurious mode

Using the transformation of Fig. 5 and accounting for reductions in transistor transconductance due to limiting, simulations were performed on the modulated oscillator with results as shown in Fig. 8. The reduction in transistor gain effectively also sets the circuit 3 dB bandwidth, although this is not critical for higher modulation frequencies. For comparison purposes, the sideband levels predicted by static analysis are also shown.



(a)



(b)

Fig. 8 Simulated and measured sideband levels.

The precise location of the notch was extremely sensitive to the circuit values, so some departure between simulations and measurements is not surprising. Finally, it must be noted that since the transformation is general, this analysis suggests that sidebands are not symmetrical and may differ significantly from static analysis results if the feedback network has a response other than a single pole.

References

1. R. L. Clark, "Oscillator Insights Based on Circuit Q," Proceedings of the Forty-fifth Annual Symposium on Frequency Control, 1991.

2. D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," Proceedings of the IEEE 54, no. 2, pp. 329-330, Feb. 1966.

3. R. C. Smythe, "Acceleration Effects in Crystal Filters- A Tutorial," Proceedings of the Forty-fifth Annual Symposium on Frequency Control, 1991.

4. M. Ghausi, Electronic Circuits. New York: Van Nostrand Reinhold Co., 1971, ch. 1.

5. D. T. Comer, "Nonlinear Capacitor Characterization in Oscillator Circuits," Research Report No. SET-VLSI-108.0, Division of SET, Penn State Harrisburg, April 2, 1992.

6. M. Schwartz, Information Transmission, Modulation, and Noise. New York: McGraw-Hill, 1959.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE LEVER OSCILLATOR FOR USE IN HIGH RESISTANCE RESONATOR APPLICATIONS *

Kurt O. Wessendorf

Sandia National Laboratories Dept. 2254
Albuquerque New Mexico 87185-5800

Abstract

The Lever oscillator has been specifically designed for use with quartz resonator sensors. The use of quartz resonators as sensors is of particular interest and depending on the sensing environment, e.g., liquid, the oscillator design is both critical and difficult due to the wide dynamic range of resonator resistance possible due to damping of the resonator. Standard oscillator designs do not work well as sensor oscillators. An oscillator design will be presented that allows both frequency and loss (R_m) of the resonator to be determined over a wide dynamic range of resonator loss. The Lever oscillator uses negative feedback in a differential amplifier configuration to actively and variably divide (or leverage) the resonator impedance such that the oscillator can maintain the phase and gain of the loop over a wide range of resonator resistance.

Introduction

Standard oscillator designs, like the Pierce or Colpitts type, are good for frequency source oscillator designs which provide low phase noise and high over- all frequency stability. These designs do not necessarily work well for sensor applications because of their inherent phase and gain sensitivity to resonator loss, R_m. Operating the resonator at f_s is a good initial approach if one is to design an oscillator for use with a high loss resonator. At f_s the resonator is approximately a real impedance and is at or near the minimum impedance of the resonator. Figure 1 is the model for a bulk wave resonator, with f_s defined as the resonance of L_m and C_m. Since a high loss resonator is not likely to be highly inductive at any frequency, it would be undesirable to try to make the sensor operate at an inductive impedance as this will limit the dynamic range of the oscillator. It is also desirable to design this type of oscillator with one side of the resonator at a.c. ground. With one side of the resonator grounded the parasitic capacitances due to the fluid can be more easily controlled; this is especially true if only one side of the resonator is used for liquid exposure.

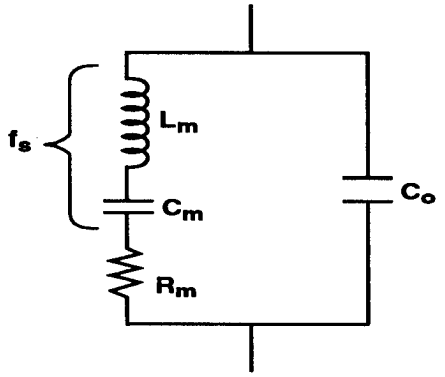


Figure 1. The quartz resonator model.

Design Goals

There are a few oscillator configurations which operate the resonator near f_s and have one side of the resonator grounded. None of these configurations meet the design goals of operating at f_s of the resonator and outputting a voltage proportional to the resonator loss R_m, by way of an automatic level control (ALC) circuit. For *in situ* resonator sensors the R_m of the resonator will have a wide dynamic range. This dynamic range depends upon the resonator and liquid properties but could be tens of ohms in air to several thousand ohms in a viscous liquid. The resonator loss is a function of the viscosity-density product (1). Since quartz resonator frequency shifts due to a given environment are

* This work performed at Sandia National Laboratories by the U.S. Department of Energy under contract DE-AC04-76DP00789

0-7803-0905-7/93 $3.00 © 1993 IEEE

determined by calculating the change in f_s of the resonator, it is convenient to design the oscillator to operate as close to f_s as possible such that the oscillator frequency mirrors the sensor-related frequency changes. The ability to operate at exactly f_s of a resonator is in itself a difficult task and the author has found that on an absolute basis it is impossible. In general if the impedance of C_0 is relatively large compared to R_m then f_s is close to the frequency at which the resonator impedance phase is zero. The R_m output voltage can also be used to determine the frequency error of the oscillator itself relative to true f_s if the shunt capacitance across the resonator is relatively constant throughout the measuring medium.

The Oscillator Design

Figure 2 is schematic of an oscillator that evolved into what is called the Lever oscillator. Understanding this circuit will help in understanding how the design goals were ultimately met.

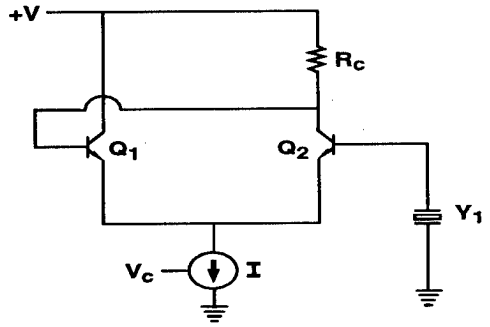


Figure 2. The non-inverting amplifier as an oscillator.

In this oscillator the collector of Q_2 is directly connected to base of Q_1 and the resonator is connected from the base of Q_2 to ground. The resonator will control (poorly) the gain and phase of the potential oscillation. The current source I is controlled by V_c and will be used in an ALC circuit to servo a given oscillation amplitude and thus will reflect the resonator loss. Transistor h is the reciprocal of transconductance g_m and is equal to (26 mV)/I_e where I_e is the individual transistor bias current. To determine the loop gain equations one could open the loop at the base of Q_1 and determine the loop gain equation for this amplifier, Figure 3. If the output impedance of the circuit is low relative to the input impedance the gain equation will be fairly representative of the actual oscillator. Equation (1) is the open loop gain equation for the circuit in Figure 3. This circuit, when V_{out} is connected to V_{in}, will oscillate when A_v is greater than or equal to one.

$$A_v \approx \frac{R_c}{2h + \frac{Z_{res}}{\beta}} \cdot \qquad (1)$$

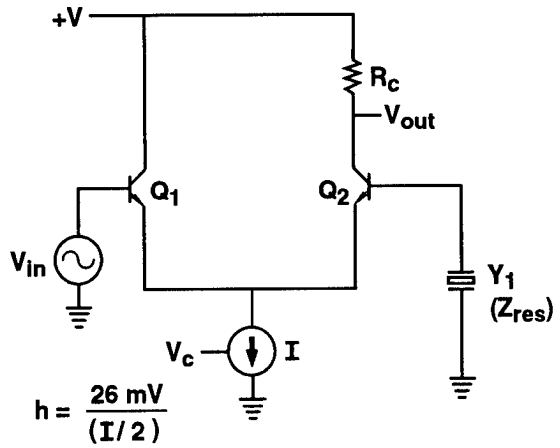


Figure 3. Open loop circuit for oscillator calculations.

If Q_1 and Q_2 are biased at reasonable current levels (0.1 < I < 10mA), the resonator will weakly control the oscillator because the resonator impedance is divided by the transistor β. Since the loop gain is more dependent on transistor h than the resonator impedance the oscillator sensitivity to resonator resistance is poor and the loop phase of the circuit will not be very sensitive to the resonator impedance. This circuit would work well for the application described if the resonator impedance was effectively divided by a smaller number (less than β) and was independent of transistor β. For example it would be desirable to have the resonator impedance divided by a large number when the resonator is a high impedance and divided by a smaller number when the resonator is a low impedance. This variable divider would help to compress the dynamic range of the oscillator transconductance over a wide dynamic range of resonator loss while still allowing enough sensitivity to determine the resonator loss via an automatic level control (ALC) circuit. This function can be obtained by modifying the oscillator of Figure 3 by placing a feedback resistor from the collector of Q_2 to the base of Q_1. The Lever oscillator design (open loop) can be seen in Figure 4. This schematic shows the feedback resistor, R_f, that is used to make the oscillator sensitivity to resonator impedance a function of the resonator impedance itself. This function can be made β independent and will allow the resonator impedance to be effectively divided (leveraged) as a function of the magnitude of the resonator impedance.

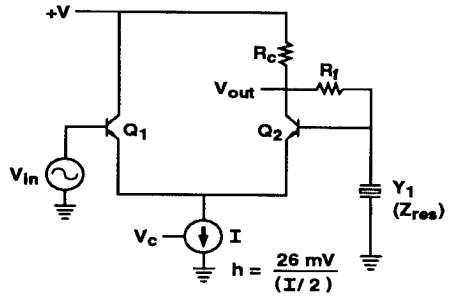


Figure 4. The Lever oscillator open loop circuit.

The Lever oscillator open loop gain, A_v if $R_c < R_f$:

$$A_v \approx \frac{R_c}{2h + \frac{Z_{res} \| [R_f + R_c]}{\beta} + \frac{R_c Z_{res}}{R_c + R_f + Z_{res}}}; \quad (2)$$

for oscillation A_v must be greater than or equal to one. If all the variables in equation (2) are real, then Z_{res} must also be real (have a zero impedance phase). If Z_{res} is real then the frequency of operation is close to f_s of the resonator and the resonator impedance can be approximated by R_m. Equation (2) can be simplified to:

$$A_v \approx \frac{R_c}{2h + \frac{R_c R_m}{R_c + R_f + R_m}} \geq 1 \quad (3)$$

where β is considered large enough to make the center term in the denominator of the loop gain equation (2) very small in comparison to the other two terms. The term

$$\frac{R_c R_m}{R_c + R_f + R_m} \quad (4)$$

is the "Lever" term, since it will determine how the resonator impedance is effectively divided (leveraged) by the feedback circuit elements, R_c and R_f. R_m is effectively divided by the function

$$\frac{R_c + R_f + R_m}{R_c} \equiv g(R_m) \quad (5)$$

which is small when the resonator impedance is small and large when the resonator impedance is large. The Lever term (4) is transistor β-independent and can be made to dominate the given gain equation (2) over a wide range of β. This variable leverage function allows the oscillator to operate over a wider dynamic range of resonator resistance while maintaining sensitivity at lower resonator resistances than if the divider effect was constant or dependent on β. Figure 5 shows a graph of the equation (5) vs R_m. This graph represents the variable divider effect the circuit has on R_m with R_c equal to 150 ohms and R_f equal to 1000 ohms. Due to the fact that the circuit is operating close to f_s of the resonator, Z_{res} is approximately R_m if C_o is relatively small.

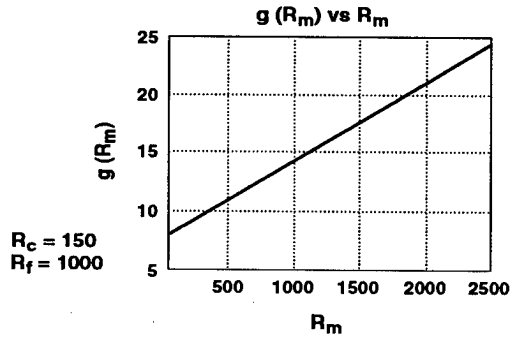


Figure 5. Graph of variable "Lever effect", $g(R_m)$ vs R_m.

For the values of R_c and R_f chosen, the resonator impedance R_m is effectively divided by 7 for small values of R_m and by approximately 24 for large values of R_m. By selecting the values of R_f and R_c such that $R_f/R_c << \beta$, the oscillator will be relatively independent of β. The choice of R_f and R_c in combination with the current source function (V_c) allow the designer to set the oscillator sensitivity to a desired range of R_m.

Lever Oscillator Practical Realization

The oscillator theory presented to this point has virtually ignored transistor capacitances and the practical matters of the design. Figure 6 is a schematic of a practical form of the Lever oscillator.

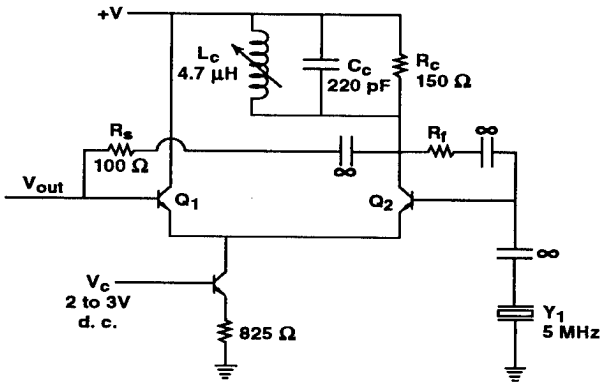


Figure 6. Lever oscillator schematic.

This schematic will be used to describe and derive the actual design performance of the oscillator function. In the theory stated above it was stated that if all the circuit components shown (Figure 4) were real then at the oscillation frequency the resonator would also be real. But in actuality, at frequencies greater than d.c., circuit capacitance will influence circuit performance. Capacitance from collector to ground of Q_2 combined with base to ground capacitance of Q_1 must be addressed. This capacitance, lumped together because of the circuit configuration, significantly degrades circuit performance: (1) this capacitance pulls the oscillator low in frequency (even for a small amount of capacitance) and (2) if the resonator becomes highly lossy, the circuit has a tendency to jump to a parasitic (capacitor controlled) frequency. These two problems were minimized by placing a parallel L-C (L_C and C_C) tank circuit across R_C. This tank is designed to be low Q (1 to 3) and is used to make the collector to ground impedance at Q_2 to be real (approximately R_C) at the desired frequency and to help constrain the oscillation frequency to the resonator and not to a parasitic frequency. This tank is adjusted such that with a resistor (low value, 100 ohms) in place of the resonator the tank circuit is adjusted to make the oscillator frequency approximately equal to the desired resonator frequency. With R_C real at the resonator frequency we have satisfied one of the assumptions given in the theory above.

The other assumption in the theory is that R_f is also real. Since R_f is a major component in the design, it is interesting to see the effect of capacitance across R_f due to the C_{bc} of the transistor. By using equation (3), substituting Z_{res} for R_m and Z_f for R_f one can solve for Z_{res} and determine the phase and magnitude of Z_{res} as a function of Z_f, R_C and h.

$$Z_{res} \approx \frac{[R_c - 2h][R_c + Z_f]}{2h} \tag{6}$$

and if $R_C < Z_f$

$$Z_{res} \approx \frac{[R_c - 2h][Z_f]}{2h} \tag{7}$$

where R_C and h are real. If these assumptions are good then it can be seen from equation (7) that the resonator impedance phase must be equal to the phase of Z_f. If operating close to zero impedance phase of the resonator is desired it is important to select R_f such that R_f is much less than the impedance of C_{bc} for the chosen transistor. If one was to design this oscillator with a transistor having a C_{bc} of approximately 3 pF and R_f of 1000 ohms then the impedance, Z_f for R_f in parallel with 3 pF at 6 MHz, will equal 994 ohms at -6.45°. For this case the resonator impedance phase at the oscillation frequency will also be approximately 6°. Although the equations derived throughout this paper are for small-signal low-excess loop-gain (one) conditions, the oscillator essentially behaves as equation (6) predicts; this is true for two reasons: (1) the oscillator will be used in an ALC circuit which is used to control the excess gain to a relatively low value so that R_m can be determined and (2) the differential amplifier limiting characteristics provide a symmetrical limiting action that makes the circuit behave more linearly than a simple single-transistor oscillator circuit. Figure 7 is a calculation of resonator impedance and phase versus transistor h using equation (6).

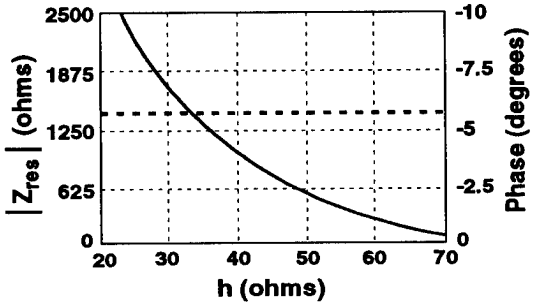


Figure 7. Calculated resonator magnitude and phase vs h calculations for the Lever oscillator.

These graphs demonstrate the wide dynamic range of resonator impedance possible with the oscillator controlling the crystal impedance phase to approximately the phase of the feedback impedance Z_f.

Lever Oscillator Performance Data

To indirectly measure resonator resistance an ALC loop was designed to control the oscillator gain (h) via the voltage controlled current source, Q_3, (Figure 8) of the oscillator.

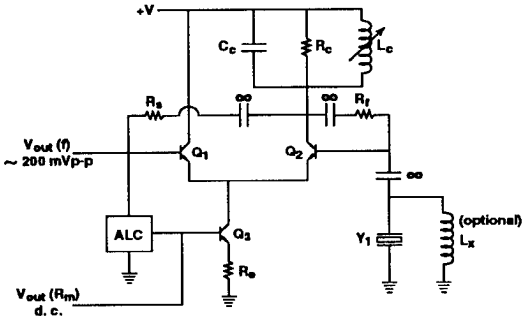


Figure 8. The Lever oscillator with ALC circuit.

The ALC was designed to control the oscillation amplitude at the base of Q_1 to approximately 200 mV peak-to-peak. Figure 8 also shows an inductor L_x which can be used to tune out static capacitance of the fixture and resonator . L_x can dramatically increase the upper resistance limit before the oscillator breaks into a parasitic oscillation. To evaluate the Lever oscillator for one set of circuit parameters, a 6 MHz test resonator was made using a standard low resistance device in series with added resistance. Figure 9 shows the test resonator used to evaluate the Lever oscillator.

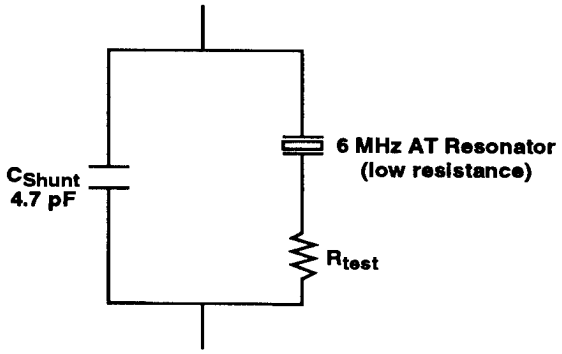


Figure 9. Test resonator with variable resistance.

C_{shunt} was included in the test resonator to more accurately simulate the C_0 of a standard resonator. This resonator was used to determine the actual resonator impedance phase at oscillation for a wide range of resistance, R_{test}. R_{test} was varied from 56 to 1700 ohms and the resonator impedance was measured on an impedance analyzer at the frequency of operation of the Lever oscillator for each tested resistance. Also the d.c. voltage of the ALC was recorded for all values of R_{test}. Data relating resonator impedance, phase, and d.c. output voltage as a function of R_{test} are presented in Figures 10 and 11 respectively.

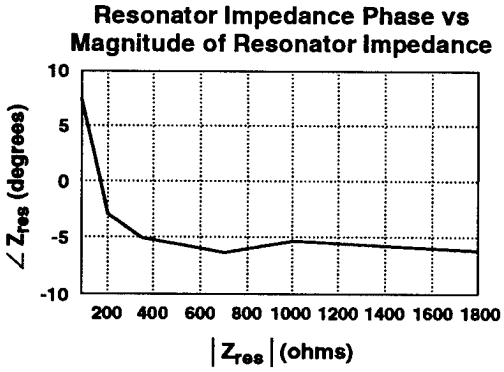


Figure 10. Measured resonator impedance phase vs magnitude.

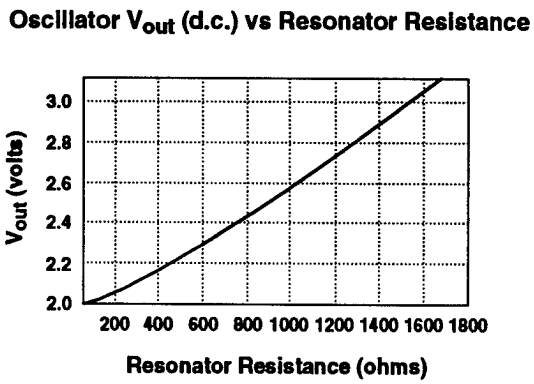


Figure 11. ALC voltage out vs resonator resistance.

These data demonstrate that the oscillator does indeed servo the resonator impedance to approximately the phase of Z_f, in this case approximately 6°, when the resonator is above 200 ohms. At resonator impedances below 200 ohms the resonator impedance phase increases. This phase shift at low resonator impedances is explained by the fact that the oscillator is less controlled by the resonator when the resonator impedance is low than when large and is therefore more sensitive to circuit parameters not used in the approximations presented here. To illustrate this, a more exact model of the oscillator was used which assumed the transistor h to be very slightly capacitive. When this capacitance is included in the calculations the resonator impedance phase increases with decreasing resonator impedance just as seen in the data, Figure 10. Figure 12 shows the calculated magnitude of Z_{res} versus the phase of Z_{res} with transistor h slightly capacitive.

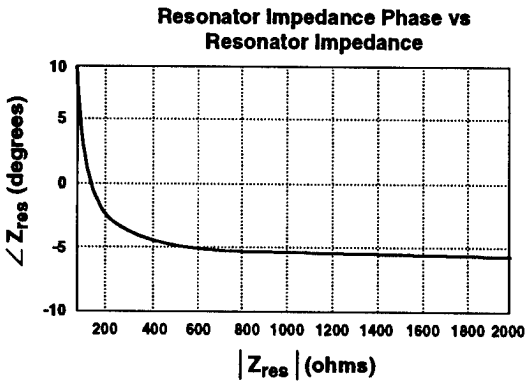


Figure 12. Calculation of resonator impedance phase vs magnitude for h slightly capacitive.

To demonstrate the β insensitivity of the Lever oscillator, a negative resistance model was used to calculate the input impedance (Z_{in}); the calculations are shown in Figure 13 as a function of h for transistor β's of 150 and 50.

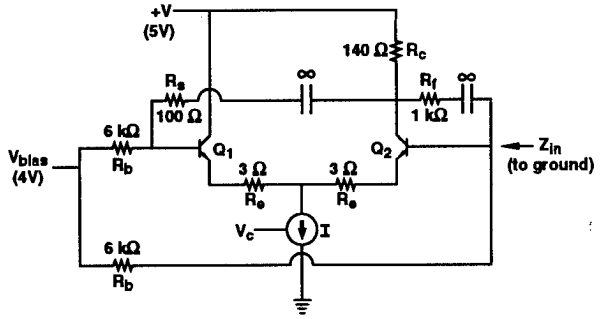


Figure 13. Lever oscillator circuit used in negative model calculations.

Figure 14 is a graph of the negative resistance function of the lever oscillator for β equal to 50 and 150. The oscillator β sensitivity is greatest at high resonator resistances where the negative feedback due to R_f and R_c is least effective. With the feedback used in this example the oscillator is relatively insensitive to transistor β for h between 20 and 65 ohms.

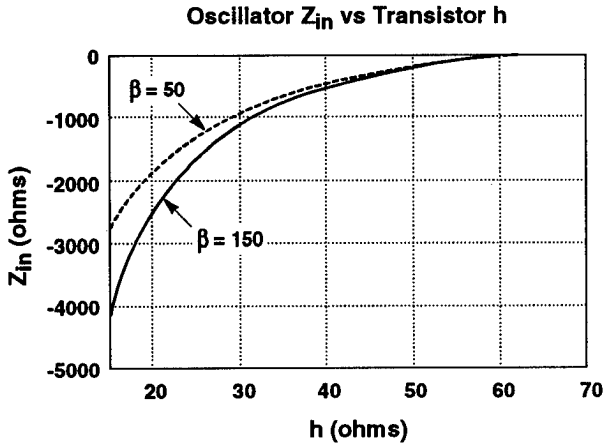


Figure 14. Calculation of β sensitivity vs h.

Oscillator Frequency Error

Now that it has been established that the Lever oscillator function can be made to act like an impedance phase servo it would be interesting to determine how this phase frequency relates to the true f_s of the resonator. If C_0 is zero then f_s is equal to the zero impedance phase frequency. But since C_0 is never zero then the oscillation frequency of the Lever oscillator will be offset from true f_s of the resonator by some amount depending on R_m, C_0, and the phase of the resonator impedance which the oscillator is servoing. Figure 15 is a graph showing the calculated difference between the frequency of zero impedance phase of the resonator and f_s of the resonator. This is done for two static capacitances of the resonator, 1 pF and 3 pF. The graph shows that, for C_0 equal to 1 pF, the frequency error with C_m equal to 25 fF and L_m equal to 31 mH is relatively small but increases with resonator resistance R_m.

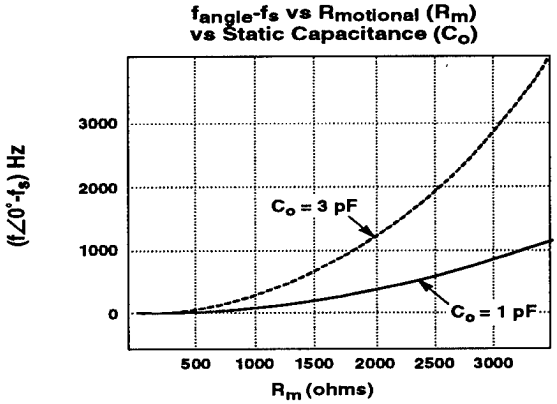


Figure 15. Frequency at 0° phase -f_s vs R_m.

With increasing C_0 the frequency error increases dramatically with R_m. Figure 16 is a graph of the same function but in this case a difference frequency for -6° of impedance phase and f_s is calculated and the resulting curve shows a negative error for small R_m, sweeping positive as the resistance increases.

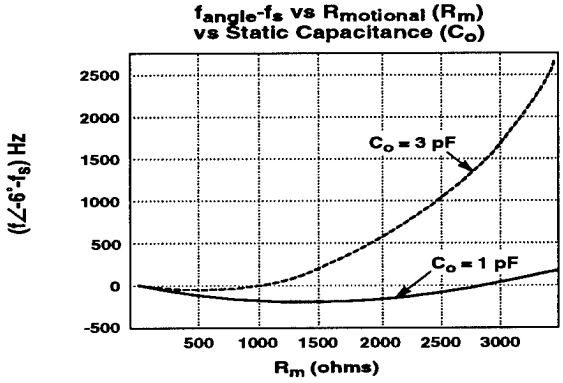


Figure 16. Frequency at -6° phase -f_s vs R_m.

This is to be expected since at f_s the resonator is essentially a capacitance (C_o) in parallel with resistance (R_m). Therefore, the resonator impedance phase at f_s is a negative quantity varying with R_m and/or C_o. To make a "true" f_s oscillator one would have to vary the phase that the oscillator servos as a function of R_m given a fixed C_o. Due to the fact that the Lever oscillator has a R_m-sensitive output voltage, one can use this output to correct for the frequency error calculated in Figures 15 and 16 if such accuracy is required. This correction could be implemented using digital analog techniques.

Conclusions

A series resonant oscillator design has been presented that operates over a wide dynamic range of resonator resistance. This resistance range is compatible with *in situ* resonator sensors used in liquid environments. The Lever oscillator is useful for these applications because of its ability to measure the f_s and R_m of the resonator. Also, the oscillator's variable leveraging effect has been shown to be useful for increasing the dynamic range of resonator resistance while maintaining sensitivity at lower resistances. The oscillator's ability to servo a resonator-impedance phase has been demonstrated and this phase can be made close to zero degrees by selection of transistor type and/or circuit resistance values. The relationship between the resonator frequency at a given phase with respect to true f_s was discussed along with correction methods. Currently this design is being implemented in a jet-fuel contamination sensor system and in a viscosity-density sensor using bulk wave resonators (2). Variations on this design have worked with resonators with R_m in the 3500 ohm range, at approximately 6 MHz, and direct coupled, integratable, versions of this oscillator design are also being developed for applications where small size is important.

Acknowlegements

The author would like to thank Allyn R. Anderson for his assistance in the development of the Lever ocillator. His fabrication, testing and evaluation of the design helped to make the Lever oscillator a robust, successful design. Thanks also to Steve J. Martin and Elmer A. Klavetter for giving the author the design challenge that precipitated the development of the Lever oscillator and for their technical support and encouragement.

References

(1) S. J. Martin, K. O. Wessendorf, C. T. Gebert, G. C. Frye, C. W. Cernosek, L. Casaus and M. A. Mitchell, "Measuring Lequid Properties With Smooth-And Textured-Surface Resonators," Proc. 47th Annual Frequency Control Symposium, June 1993.

(2) T. J. O'Hern, W. M. Trott, S. J. Martin and E. A. Klavetter, "Advanced Diagnostics for *in situ* Measurement of Particle formation and Deposition in Thermally Stressed Jet Fuels," AIAA 930363, Aerospace Science Meeting, Reno, Nevada, 1993.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

OVERMODED HIGH Q RESONATORS FOR MICROWAVE OSCILLATORS

G.R. Kline, K. M. Lakin, and K.T. McCarron

TFR Technologies, Inc.
701 SE Salmon
Redmond, Oregon 97756

Abstract

High overtone bulk acoustic wave resonators operating in the 1-6 GHz range have been fabricated on sapphire, lithium niobate, dilithium tetraborate, and quartz substrates in an investigation to obtain high Q feed back elements for use in low noise oscillators. C-axis normal aluminum nitride films were used for transduction for all the substrates yielding longitudinal modes. The resonators on Z-cut sapphire exhibited an unloaded fQ product in excess of 1×10^{14} at frequencies between 1.6 and 4.0 GHz.

I. Introduction

Considerable interest has been generated in thin film based acoustic devices fabricated by conventional microelectronic techniques. These devices form the basis for future fully integrated MIMIC having on chip frequency control elements or adjacent chips of comparable size to the MIMIC die. For applications where low noise oscillators are required, current materials processing limitations exclude full integration. However, high Q feedback elements of comparable size to the MIMIC die can be fabricated using piezoelectric thin films on highly polished and parallel low loss (high Q) substrates that need not be piezoelectric.

High Q feedback elements for on-frequency control of low phase noise oscillators can take the form of two-port narrow bandwidth filters[1], two-port microwave acoustic delay lines[2-3], or as one-port resonators. In the latter case the high Q resonators have to exhibit substantial inductive response to be effective in a microwave oscillator circuit. In order for the overmoded resonator to show an inductive response, the effective piezoelectric coupling coefficient of the resonator must be large enough to overcome finite Q and degradations due to high mode numbers. The maximum effective coupling coefficient, and hence the largest inductive response, occurs when the piezoelectric film is approximately one half wavelength thick.

Experimental results show that the basic AlN film has a coupling coefficient of approximately 7%. Equivalent circuit modeling of the high Q overmoded resonator suggests that the effective coupling coefficient of the resonator is given by the film coupling coefficient divided by the mode number of the resonator. Further analysis shows that the Q of the overmoded resonator departs from the material Q if the mode number is too large. The point of departure from the basic material Q is dependant on the effective piezoelectric coupling coefficient. This topic, relating to resonator design, is discussed for various substrates including sapphire and quartz in another paper[1].

A typical two-port resonator delay line configuration is shown in figure 1a. A ground plane metalization is deposited on both sides of a low loss highly polished and parallel substrate. This is followed by deposition of a piezoelectric film and patterning of the top electrode again on both sides of the substrate. Since the delay section does not have to be piezoelectric a great deal of latitude is enjoyed for selection of the substrate on the basis of acoustic loss and temperature coefficient properties. Multiple resonances occur at a repeat frequency which is the fundamental frequency of the entire structure. This is inversely proportional to the substrate thickness d. For a structure consisting of a 5 micron thick AlN film on a 1 mil AT-cut quartz substrate the response would repeat every 60 MHz either side of the center frequency of 1 GHz.

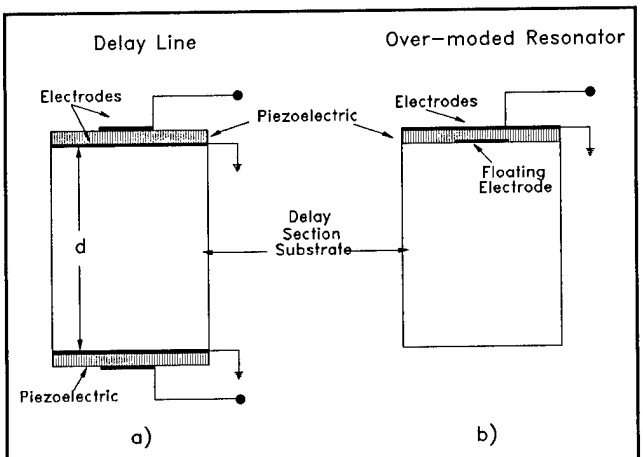


Fig. 1 One and two port overmoded resonator configurations a) A conventional two port resonator employing a low loss substrate as a delay line for overmoded response b) An overmoded resonator configuration where both electrodes are on the top surface via a series connected architecture implemented with a buried floating electrode.

0-7803-0905-7/93 $3.00 © 1993 IEEE

The approach used for fabrication of high Q overmoded resonators in this investigation is shown in figure 1b. A one port resonator was fabricated using a series connected electrode structure which places both the ground and the hot electrodes on top of the piezoelectric film. Using this approach there is no need to etch the piezoelectric to get to the ground plane as would be done in a conventional one-port device. One advantage of this approach over the delay line technique is that post-process trimming of the back side of the substrate can be done to adjust the mode spacing.

II. Experimental Results

C-axis normal AlN was used as the transducing thin film for all samples fabricated. The AlN film was deposited by DC magnetron reactive sputtering in a system specifically designed for high thickness uniformity. Thickness non-uniformity across a four inch wafer was measured at $<\pm 0.65\%$ by midband frequency measurements of fundamental mode ladder filters. All measurements were performed on an HP8753C network analyzer with a 6 GHz option using 50 ohm microwave co-planar wafer probes at room temperature. The Q values were calculated from a parabolic fit to the 1600 point raw S11 data. The same metalization mask set was used for all samples. The mask set had varying size electrodes to allow a wide range of frequencies to be measured. Generally speaking, the highest Q values measured were commensurate with an electrode area resulting in impedances near 50 ohms.

Overmoded resonators were fabricated on Z-cut sapphire, Z-cut lithium niobate, Z-cut dilithium tetraborate ($Li_2B_4O_7$), and AT-cut quartz at frequencies from 1.5 to 5.2 GHz. The sapphire crystals were 3" in diameter with thicknesses from 10 to 25 mils. They were processed as received. Most of the other crystals were polished on one side as received and so had to be further processed in house. They varied in size up to 3" in diameter with thicknesses from 8 to 20 mils. Table 1 shows a summary of the results. Also included for completeness in the table are the results from fabrication of overmoded ladder filters discussed elsewhere[1]. All resonators listed in the table exhibited sufficiently inductive response to be usable for oscillator control. The highest fQ product obtained was for Z-cut sapphire. Figure 2 shows an S11 plot for a single response for an overmoded resonator on sapphire with a series frequency of 1.60327 GHz and a parallel frequency of 1.60339 GHz. The maximum Q for this device was 68000 obtained near parallel resonance resulting in an fQ product of 1.1×10^{14}. The sapphire wafer was 25 mils thick resulting in a resonance spacing of 8.64 MHz and a mode number of 185.

As shown in Table 1 the fQ product holds fairly well for sapphire up to 4 GHz and then drops substantially at 5 GHz. This would seem to suggest that crystal parallelism or surface damage resulted in the lowering of Q at this frequency. Even at 5.2 GHz, however, the response is highly inductive as shown in figure 3. Lithium niobate and quartz also exhibit a diminishing fQ product as the frequency increases possibly as a result of surface damage. The Qs measured represent 50-80% of the intrinsic theoretical limits[4] for lithium niobate, sapphire, and quartz while none has been determined for dilithium tetraborate. The fQ product for dilithium tetraborate is relatively close to that measured by Ballato et. al.[5] at much lower frequencies but due to processing difficulties the current Q measurement may not reflect the intrinsic material Q. It is anticipated that refinements in wafer processing will raise the fQ product for this material.

The fQ product for lithium niobate was somewhat lower than for sapphire as expected. Figure 4 shows an S11 plot for lithium niobate at 1.6 GHz where a Q of approximately 54000 was measured. The resonance spacing was 11 MHz yielding a mode number of 147.

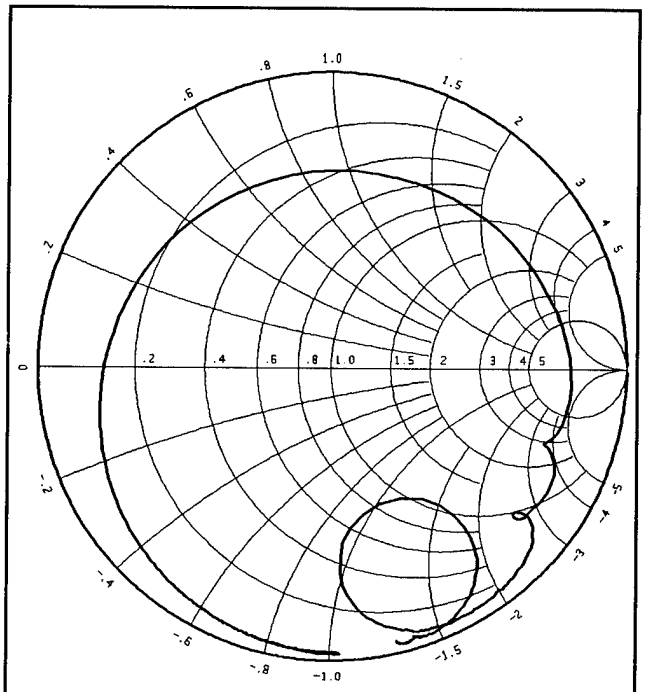


Fig. 2. Smith chart display of a single resonance of a high Q overmoded resonator on Z-cut sapphire. The center frequency is 1.603 GHz. A maximum Q of 68000 was obtained near parallel resonance for an fQ product of 1.1×10^{14}. The resonance spacing is 8.6 MHz yielding a mode number of 185.

Dilithium tetraborate and quartz were chosen for this investigation because they both have temperature compensating cuts. The TA-cut (for longitudinal waves) and TC-cut (for shear waves) in dilithium tetraborate and the AT-cut (for shear waves) in quartz are temperature compensating for bulk mode resonators. Figure 5 shows an S11 plot for a pair of resonances on AT-cut

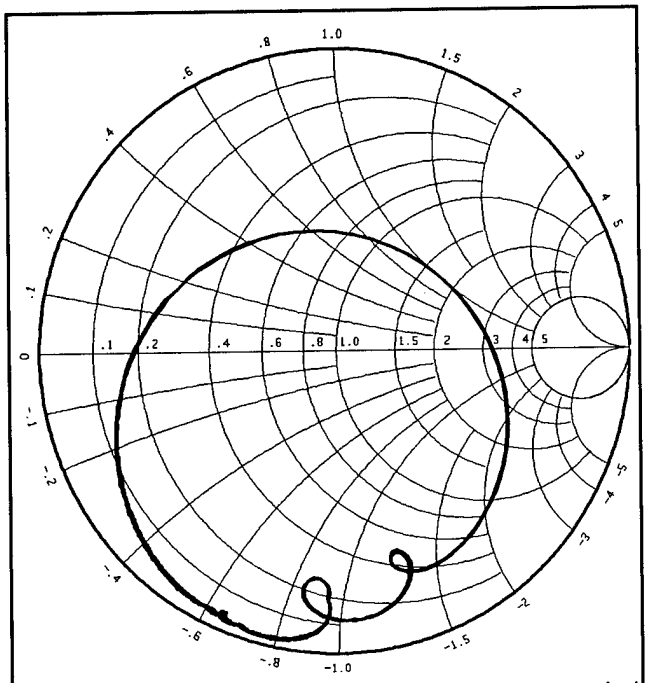


Fig. 3 Smith chart of a pair of resonances of an overmoded resonator on Z-cut sapphire. The center frequency is 5.233 GHz. A maximum Q of 14500 was obtained near parallel resonance for an fQ product of 7.6 x10^{13}. The resonance spacing is 21 MHz for a mode number of 247.

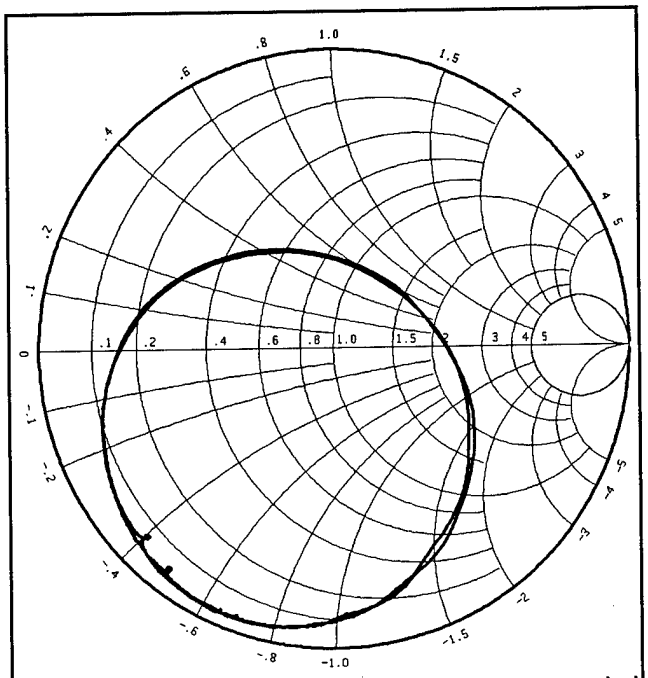


Fig. 5 Smith chart of a pair of resonances of an overmoded resonator on AT-cut quartz. The center frequency is 2.389 GHz. A maximum Q of 5300 was obtained near series resonance for an fQ product of 1.26 x10^{13}. The resonance spacing was 17 MHz for a mode number of 140.

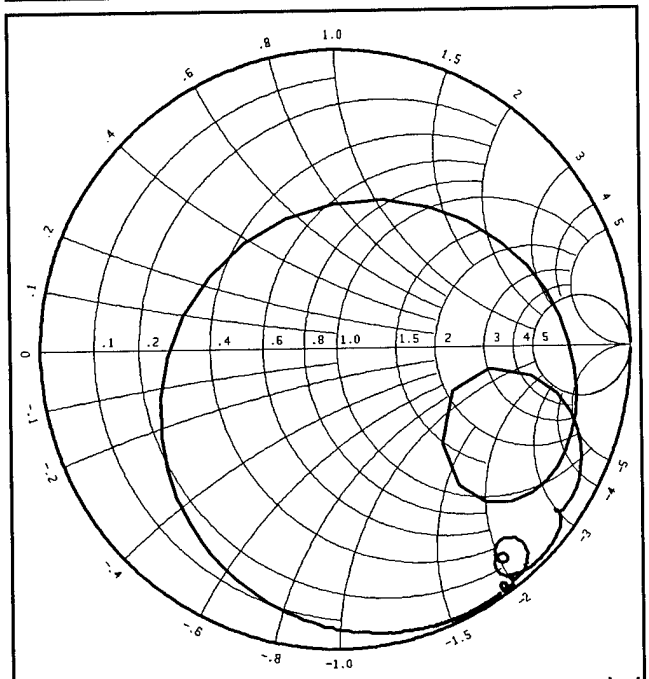


Fig. 4 Smith chart of one resonance of an overmoded resonator on Z-cut lithium niobate. The center frequency is 1.608 GHz. A maximum Q of 53700 was obtained near parallel resonance for an fQ product of 8.6 x10^{13}. The resonance spacing is 10.9 MHz yielding a mode number of 147.

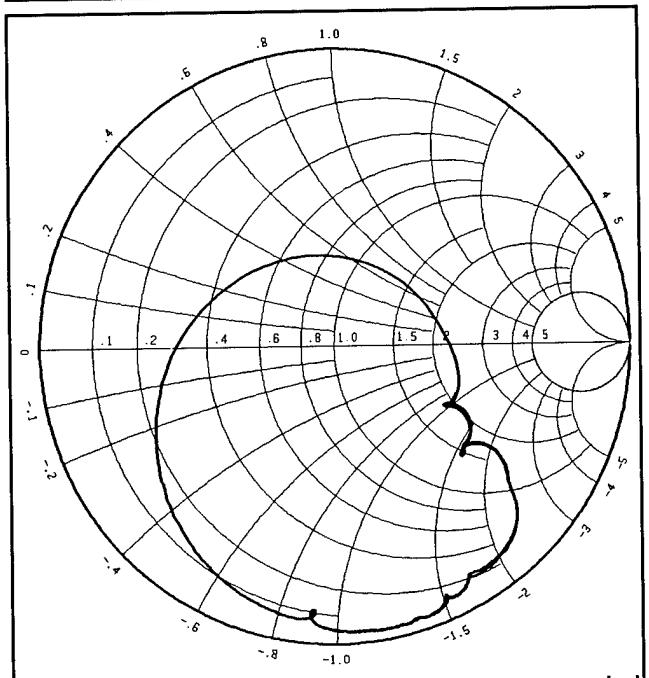


Fig. 6 Smith chart of a single resonance for an overmoded resonator on Z-cut dilithium tetraborate. The center frequency is 1.835 GHz. A maximum Q of 17500 was obtained near series resonance for an fQ product of 3.2 x10^{13}. The resonance spacing was 8.2 MHz yielding a mode number of 223.

SUBSTRATE	Q	f	fQ	MODE #	LOADED Q	IL	BANDWIDTH
Sapphire Z-cut							
"	68440	1.6034 GHz	1.10E+14	185			
"	44332	2.4640 GHz	1.09E+14	285			
"	25869	4.0138 GHz	1.04E+14	190			
"	14584	5.2326 GHz	7.60E+13	247			
"		1.6059 GHz	8.40E+13	186	30000-50000	4.97 dB	152 KHz
"		1.6092 GHz	8.05E+13	186	40000-50000	7.3 dB	76 KHz
Lithium Niobate Z-cut							
"	53695	1.6083 GHz	8.64E+13	147			
"	27600	2.4202 GHz	6.68E+13	240			
"		1.5013 GHz	7.80E+13	131	30000-50000	5.6 dB	126 KHz
Dilithium Tetraborate Z-cut	17472	1.8348 GHz	3.21E+13	223			
Quartz AT-cut							
"	9523	1.5209 GHz	1.45E+13	93			
"	5723	2.4059 GHz	1.38E+13	141			

TABLE 1

quartz at 2.4 GHz. The Q for this device was 5300 with a resonance spacing of 17 MHz resulting in a mode number of 140. Figure 6 is an S11 plot for a single resonance on Z-cut dilithium tetraborate at 1.8 GHz with a Q of 17500.

III Conclusions

High overtone bulk acoustic wave resonators have been fabricated on sapphire, lithium niobate, dilithium tetraborate, and quartz substrates at frequencies up to 6 GHz using AlN as the transducing material in a longitudinal mode. The resonators on Z-cut sapphire exhibited Qs as high as 68000 at 1.6 GHz with highly inductive response. Lithium niobate exhibited a Q of 54000 at 1.6 GHz. When used in combination with a mode-select filter these resonators would make good feedback elements for low phase noise oscillators where inherent high temperature stability is not necessary. For applications where temperature compensation is required dilithium tetraborate and AT-cut quartz could be used with some sacrifice in Q.

This work was supported by the U. S. Army Research Laboratory, Ft. Monmouth, NJ under a Phase II SBIR contract.

References

1. K. M. Lakin, G. R. Kline, and K. T. McCarron, "High Q Microwave Acoustic Resonators and Filters", IEEE MTT-S Int. Microwave Symp. Dig., June 1993.

2. M.M. Driscoll, R.A. Jelen, and N. Matthews, "Extremely Low Phase Noise UHF Oscillators Utilizing High-Overtone, Bulk-Acoustic Resonators", IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 39, No. 6, Nov. 1992 ,pp 774-779.

3. D.S. Bailey, M.M. Driscoll, R.A. Jelen, and B.R. McAvoy, "Frequency Stability of High-Overtone Bulk-Acoustic Resonators", IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 39, No. 6, Nov. 1992 ,pp 780-784.

4. R. A. Moore, J. T. Haynes, and B. R. McAvoy, "High Overtone Bulk Resonator Stabilized Microwave Sources", IEEE Ultrasonics Symposium Proceedings, 1981, pp. 414-424.

5. A. Ballato, J. Kosinski, S. Mallikarjun, and T. Lukaszek, "A Comparison of Predicted and Measured Properties of Doubly-Rotated Lithium Tetraborate Resonators", Proceedings of the 45th Annual Symposium on Frequency Control, 1991, pp. 217-221.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A STUDY OF A MSW DEVICE USING A LOW-LOSS Bi,Ga SUBSTITUTED YIG FILM HAVING A PROPER GROWTH INDUCED ANISOTROPY MAGNETIC FIELD CHARACTERIZED BY ULTRA-LOW FREQUENCY-TEMPERATURE DEPENDENCIES

Masayuki TANNO *, ** and Mitsuo NAKAZAWA **

* Shin-Etsu Chemical Co.,Ltd., 2-13-1 Annaka isobe Gunma , Japan
** Shinshu University , 500 Wakasato, Nagano City 380 , Japan

Abstract

A YIG film having a small frequency-temperature dependence with a low ΔH has been searched both theoretically and experimentally. It is found that a frequency-temperature characteristics of a low-loss Bi,Ga substituted YIG film epitaxially grown on a (111) oriented GGG substrate with a proper growth induced anisotropy magnetic field is better than that of a YIG shpere. A MSFVW chip resonator made from this new film is also demonstrated.

1. Introduction

A planar-type magnetostatic wave (MSW) device[10] made from a YIG film having a small frequency-temperature characteristics has a large possibility to become a counterpart of a YIG sphere device such as a microwave tunable filter or a oscillator which is widely used for a microwave measurement system.
In the past decade, some works were conducted to improve a frequency-temperature characteristic for YIG films[2,3,4,5]. H.L.Glass et al. reported that perpendicular resonance magnetic field at a constant frequency of a $La_{0.06}Y_{2.94}Fe_{4.13}Ga_{0.87}O_{12}$ film($4\pi Ms$ =410G) epitaxially grown on a (111) GGG substrate showed a parabolic temperature dependence due to a cubic crystal anisotropy and a stress induced anisotropy[2]. While P.Röshmann et al. pointed out that a La substitution and/or a large lattice mismatch caused a deterioration in resonance spectra. They used an Al substituted GGG (a=12.363Å) substrate to eliminate a La element in the film and made a mismatch free $Y_3Fe_{4.07}Ga_{0.93}O_{12}$ film ($4\pi Ms$= 380G) having almost the same temperature characteristics[3]. However, these materials still show a large frequency change of less than about 50MHz (-20°C~80 °C) compared with that of the YIG shpere. Further, as T.Ryuo et al. reported, the ΔH of La,Ga substituted YIG films grown on a yttrium substituted GGG substrate (a=12.367Å) rapidly increase when $4\pi Ms$ was less than 500 G [5]. Consequently, the former YIG films [2,3.5] have been troubled with an insufficient frequency-temperature characterisitc and a moderate ΔH.
In the present paper, in order to attain a more sophisticated YIG film with a low ΔH having as low frequency-temperature dependence as that of a YIG sphere, we searched a film composition and orientation both theoretically and experimentally and find that the temperature dependence due to the growth induced anisotropy magnetic field, the cubic crystal anisotropy magnetic field and the saturation magnetization of the YIG film may be canceled each other. Consequently, we obtain the following results. At first, a $Bi_{0.15}Y_{2.85}Fe_{4.34}Ga_{0.66}O_{12}$ film grown on a (111) GGG substrate with a growth induced anisotropy constant(Ku^{σ}) of 1.4×10^{4} erg/cm^3,a $4\pi Ms$ of 0.7$_3$ kG and a ΔH of 1.5 Oe at 9.2 GHz, shows a extremely low vertical resonance field change versus temperature. That is within 3(G) between -33 °C and 72 °C at a constant frequency of 9.2 GHz. The first and the 2nd order temperature coefficients of resonance field change α,β are α=-0.011(G/°C) and β=-0.210 $\times10^{-4}$(G/°C^2),respectively, when the resonance field deviation (δHres in G) versus temperature (T°C) expressed as the form of $\delta Hres=\alpha(T-25)+\beta/2\cdot(T-25)^2$.
Secondly , both the turning points of the curves and the temperature coefficients(α,β) of the films are controlled by using a proper annealing , and also the unwanted modes are suppressed.
Finally, a MSFVW chip resonator[9] made from the new material ($Bi_{0.15}Y_{2.85}Fe_{4.34}Ga_{0.66}O_{12}$,$4\pi Ms$=0.7$_3$ kG,Ku^{σ}=1.1×10^{4} erg/cm^3,ΔH =1.3 Oe) shows a high loaded Q of 2,000 at 3 GHz and a small frequency change versus temperaure within 10 MHz between 4 °C to 74 °C at a constant magnetic field. The frequency -temperature dependence of the MSFVW chip resonator, without any complex temperature compensation procedures, is less than that of a YIG sphere device.

0-7803-0905-7/93 $3.00 © 1993 IEEE

2. Material design

A Kittel's equation of a perpendicular resonance for a (111) oriented square YIG film is expressed as the formula(1) if we neglect a stress induced anisotropy magnetic field caused by a large mismatch between a film and a substrate[1,2,3]

$$f=\gamma\cdot\left(Hres-Nz'\cdot 4\pi Ms-\frac{4K1}{3Ms}+\frac{2Ku^{\varepsilon}}{Ms}\right) \quad (1)$$

where f is a frequency(MHz) , γ is a gyromagnetic ratio (2.8MHz/Oe), Hres is a perpendicular resonance magnetic field, Nz' is an effective demagnetizing factor for a square chip (Nz'=(Nz-Nx),Nx+Ny+Nz=1, Nx=Ny),4πMs is a saturation magnetization ,K1 is a first order constant of a cubic anisotropy(erg/cm^3), Ku^{ε} is a growth induced anisotropy constant(erg/cm^3). In this work, firstly , in order to calculate a temperature coefficient of a purpendicular resonance magnetic field at a constant frequency , the Nz' and the Ku^{ε} is derived experimentally (fig.1,2), where we use a calculated value of a temperature dependence for a 4πMs by the molecular field theory [7] and the previously reported value of the K1[8]. In the fig.1 , the Nz' values for the 1mm square YIG chips having various film thickness on (111) GGG substrates are experimentally derived from a purpendicular resonance field at a constant freuency of 9.2 GHz as a function of the aspect ratio(h/l), where h is a film thickness and l is a lateral length(1mm). In the ferrimagnetic resonance (FMR) measurement, we used a JES-FE2XG made by JEOL Co.. The experimental relation between Nz' and the aspect -ratio is expressed as Nz'=1-1.76·h/l. Joseph's first order theory for a center portion of a square chip[6] is plotted as a reference.

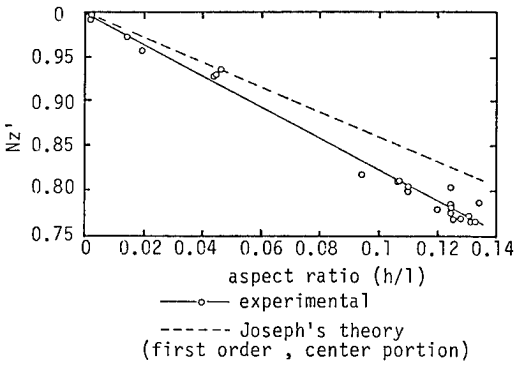


Fig.1 An effective demagnetizing factor (Nz') vs. aspect ratio of YIG square chips(1.0×1.0mm). {h:film thickness, l:lateral length(1mm)}

The fig.2 shows an example of temperature dependencies of a Ku^{ε} calculated from the FMR measurement by using the equation(1). The sample is a $Bi_{0.15}Y_{2.85}Fe_{4.34}Ga_{0.66}O_{12}$ chip (1mm square) grown on a (111) GGG substrate. The Ku^{ε} caused by a Bi sub-stitution decrease with a increase of an annealing hour(0 hr.,18hrs.,81hrs.,148hrs.,213hrs.,365hrs., at 950 °C in air). This indicates that the different Ku^{ε} are allowed in the same film composition.

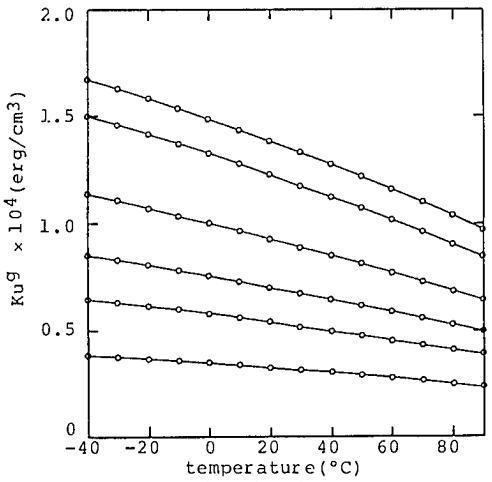


Fig.2 Measured growth induced anisotropy constants (Ku^{ε}) of a $Bi_{0.15}Y_{2.85}Fe_{4.34}Ga_{0.66}O_{12}$ film as a function of temperature.

And the next , we define the change of the perpendicular resonance field (δHres) at a constant frequency vs. temperature as the following equation.

$$\delta Hres(T)= Nz'\cdot 4\pi Ms+\frac{4K1}{3Ms}-\frac{2Ku^{\varepsilon}}{Ms}-Hresc$$
$$=\alpha\cdot(T-25)+\beta/2\cdot(T-25)^2 \quad (2)$$

where T is temperature(°C), Hresc is a temperature independent value, α(G/°C),β(G/°C^2) is a 1st and a 2nd temperature coefficients of a resonance field change ,respectively.

The α and the β for (111) oriented Ga substituted YIG chips(1mm square,h=50μm) having Ku^{ε}=0, 1.35, 2.0×10^4 erg/cm^3 (at 25°C) are calculated as a function of 4πMs(Fig.3,4). The temperature dependence for Ku^{ε}=2.0×10^4 erg/cm^3 is simply estimated from the curve for Ku^{ε}=1.35×10^4 erg/cm^3. As for Ku^{ε}=0 , the former material[2,3,5], α=0 is obtained at 4πMs of near 0.43 kG.When Ku^{ε}=1.35 or 2.0×10^{4}erg/cm^3,the 4πMs which gives α=0 is increased to near 0.7kG and the β becomes quite smaller than that for Ku^{ε}=0. On the contrary ,for the other cuts , for example (100) orientation ,temperature compensation does not occur due to a opposite sign of the cubic anisotropy

field term. Thus, it is predicted that a lattice mismatch free Bi,Ga substituted YIG film ($4\pi Ms$= ~0.7kG) grown on a (111) oriented substrate with proper growth induced anisotropy (Ku^{g}=1-2×10^{4} erg/cm^{3}) show a good temperature-resonance field characteristics and a small ΔH[5].

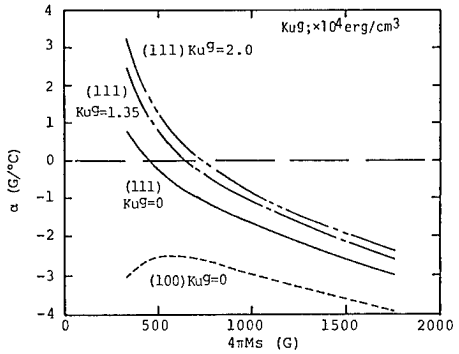


Fig.3 First order temperature coefficients(α) of a resonance magnetic field change for Ga substituted YIG films (calculation).

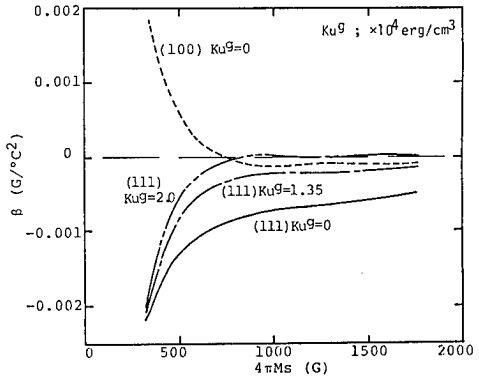


Fig.4 2nd order temperature coefficients(β) of a resonance magnetic field change for Ga substituted YIG films (calculation).

3. Experiment

3-1. Material preparation

In order to confirm the prediction in the section 2, Bi,Ga substituted YIG films were grown on a (111) oriented GGG (a=12.383Å) and a Y substituted GGG (a=12.367Å) substrate using a conventional liquid phase epitaxial(LPE) method. The oxide materials, Y_2O_3 ,Bi_2O_3, Ga_2O_3, Fe_2O_3, B_2O_3 and PbO were mixed in a platinum crucible and heated up to 1,100°C in a kanthal wound furnace. The Bi,Ga substituted YIG films were obtained by immersing the substrates into the melt. The growth temperature was typically 900°C. The substrate was rotated with a rotation rate of 50rpm at the reversal period of 2 sec.. The typical film thickness was 50μm.

3-2. Temperature-resonance field dependencies

Based on the analysis in the section 2 ,the temper -ature compensated materials were searched experi -mentally (the table.1). In the table , measured or estimated important characters , film composition, film thickness, $4\pi Ms$, lattice mismatch (Δa), substrate, ΔH at 9.2GHz, Ku^{r}, K1, turnover temperature , 3.0(G) temperature range ,α and β are listed. Where the film composition was determined by a X ray fluorescence method. The film thickness was measured by a spectrophotometer. The $4\pi Ms$ was measured by a vibrating sample method (VSM) and a saturation magnetic field for a Faraday rotation. The Δa was measured by a X ray double crystal method. The ΔH and the temperature dependence of the perpendicular resonance field at 9.2GHz were measured by the FMR. The k1 was estimated from ref.[8] which correspond to the measured Ga content. The Ku^{r} was calculated

Table.1 Temperature compensated materials for MSFVW devices

No.	film composition $Bi_xY_{3-x}Fe_{5-y}Ga_yO_{12}$		film thickness	$4\pi Ms$	Δa	substrate	ΔH	Ku^{r}	K1	turnover temperature	3.0 gauss temperature range	α	β
	x	y	(μm)	(kG)	(Å)		at 9.2GHz (Oe)	×10^{4} erg/cm^{3}	×10^{4} erg/cm^{3}	(°C)	(°C)	(G/°C)	×10^{-2} (G/°C^{2})
1	0.15	0.66	52.6	0.73	+0.001	(111)GGG	1.5	1.4	-0.31	20	105	-0.011	-0.210
2	0.15	0.66	53.9	0.73	+0.001	(111)GGG	1.3	1.1	-0.31	30	80	-0.002	-0.308
3	0.15	0.67	53.0	0.70	+0.001	(111)GGG	1.5	1.2	-0.30	60	80	+0.107	-0.277
4	0.14	0.66	49.1	0.71	+0.001	(111)GGG	1.8	1.0	-0.31	0	100	-0.069	-0.294
5	0.12	0.74	46.8	0.62	+0.019	(111)Y:GGG	2.5	0.6	-0.27	10	80	-0.032	-0.412
6	0.06	0.92	50.2	0.38	+0.002	(111)Y:GGG	4.3	0.0	-0.19	30	45	+0.022	-1.270

from the equation (1). All the samples for the FMR measurement in the table.1 were 1mm square shape. The α and the β obtained experimentally are compared with the theoretical calculation (Fig.5 ,Fig.6). As expected from the analysis , the sample No.1 to No.5 having Ku^{ε} of 0.6 to 1.4 $\times 10^4$ erg/cm^3, 4πMs of 0.6_2 to 0.7_3kG have a large 4πMs for α=0 and a quite small β compared with the sample No.6 ($Bi_{0.06}Y_{2.94}Fe_{4.08}Ga_{0.92}O_{12}$, 4πMs=0.3_8 kG, Ku^{ε}=0) which is almost the same material as that of the refs.[2,3,5].

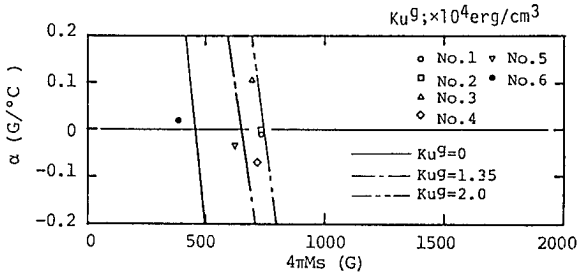


Fig.5 First order temperature coefficients (α) of Bi,Ga substituted YIG films epitaxially grown on (111) GGG and Y:GGG substrates(experiment).

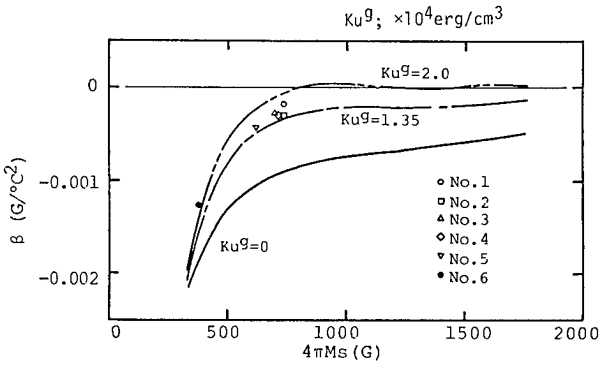


Fig.6 2nd order temperature coefficients (β) of Bi,Ga substituted YIG films epitaxially grown on (111) GGG and Y:GGG substrates(experiment)

In the fig.7, we compare the change of the resonance magnetic field versus temperature at the constant frequency of 9.2GHz for the samples No.1 and No.6. As for the sample No.1, the turnover temperature is 20 °C and the change of the resonance field is within 3 gauss from -33 °C to 72 °C. The 3 gauss temperature renge for the sample No.6 is 5 °C to 50 °C. That is quite narrower than that of the No.1.

We show a typical FMR spectrum of the sample No.2 (1.5mm ×2.5mm) having a ΔH of 1.30e which is as close value as that for a non doped YIG film(fig.8).

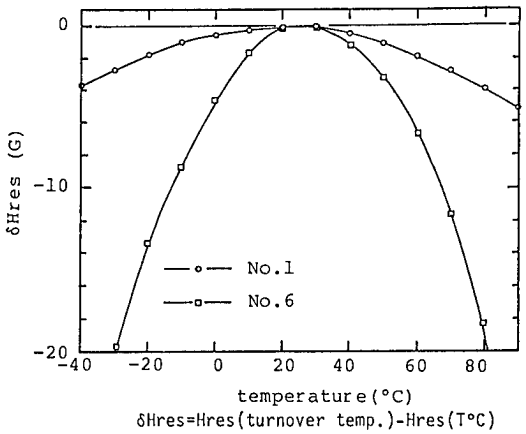


Fig.7 Temperature dependencies of the δHres for the samples No.1 and No.6.

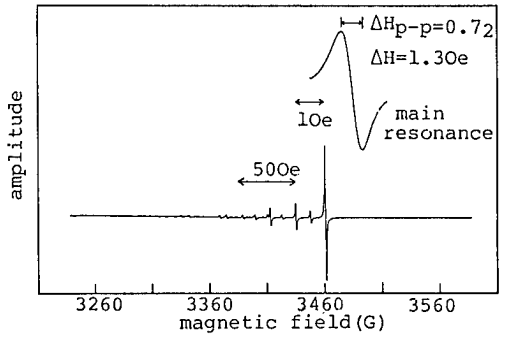


Fig.8 A FMR spectrum of a $Bi_{0.15}Y_{2.85}Fe_{4.34}Ga_{0.66}O_{12}$ chip(sample No.2,1.5×2.5mm).

3-3. Control of a resonance by an annealing

We find that controlling the turnover temperature only by a growth temperature is quite difficult. Actually , a growth temperature range of 0.2°C at about 900 °C was required to control the turnover temperature from 20 °C to 40 °C. An annealing may be the most suitable method to control the turnover temperature precisely ,since the ku^{ε} may be controlled by a proper annealing.

In the fig.9, the turnover temperature , the α and the β for a $Bi_{0.15}Y_{2.85}Fe_{4.34}Ga_{0.66}O_{12}$chip ($4\pi$Ms= 0.7_3 kG, Ku^{ε}=1.3×10^4 erg/cm^3) were examined with decrease of the Ku^{ε} by the annealing at 950°C in air. After about 500 hours of annealing, the Ku^{ε} was absolutely eliminated. The turnover temperatures were controlled from 40 °C to -75 °C , however the temperature coefficients (α、β) increased with a decrease of the Ku^{ε}.

Another effects,a reduction of unwanted modes, on the FMR spectrum by annealing were observed (Fig.10

(a,b)). Spurious modes which sometimes occured for an as grown $Bi_{0.15}Y_{2.85}Fe_{4.34}Ga_{0.66}O_{12}$ chip (Fig.10 (a)) were suppressed after the annealing (Fig.10(b)) (950°C 18 hrs.,in air). An inhomogeneity of a Bi content or a Ku^5 for an as grown film might be improved by an anealing.

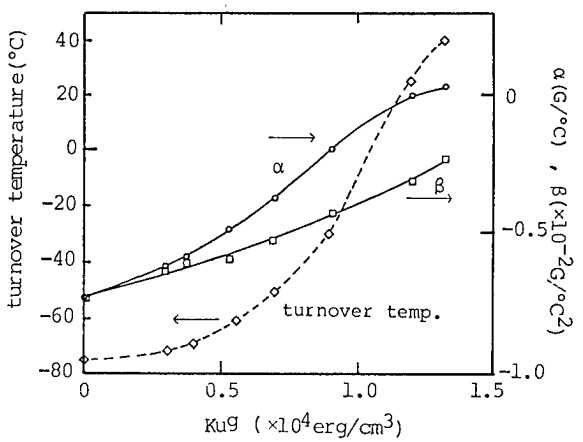
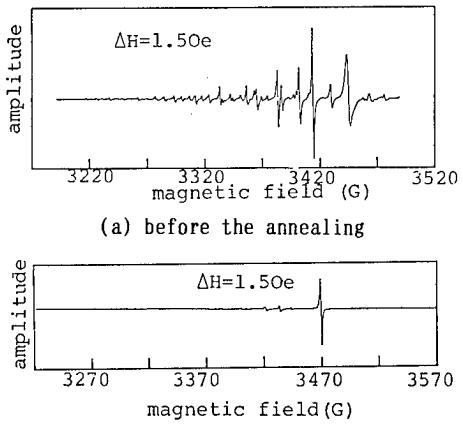


Fig.9 Temperature coefficients of a resonance field for a $Bi_{0.15}Y_{2.85}Fe_{4.34}Ga_{0.66}O_{12}$ film as a function of the Ku^5. The Ku^5 was controlled by an annealing.



(a) before the annealing

(b) after the annealing (950 °C, 18 hrs.)

Fig.10 An effect of an annealing on resonance spectra for a $Bi_{0.15}Y_{2.85}Fe_{4.34}Ga_{0.66}O_{12}$ chip.

4. A MSFVW resonator made from the new material

In order to examine a frequency-temperature characterisitics of the presented materials ,a magnetostatic forward volume wave(MSFVW) chip resonator [9] has been constructed. The chip size was 1.7×3.4mm. The resonator having 4 electrodes and 2 bonding pads was made by a conventional photolithography. The chip resonator was mounted on an electro-magnet and a magnetic field was applied perpendicularly to the chip resonator. The electro-magnet was designed to have an uniform magnetic field in a 5mmϕ region by using a finite element method(FEM). A typical $|S21|$ response near 3GHz for the resonator made from the sample NO.2 is shown in fig.11. The loaded Q defined as $Q=f/\Delta f$ (f:frequency ,Δf:3dB band width of the dip) of the device was about 2,000. In the measurement , we used a HP 8757A(a scalar network analyzer) and a HP 8340B (a synthesized sweeper). A temperature dependence of this resonator at a constant magnetic field was measured (fig.12). A small resonance frequency change versus ambient temperatures was obtained. The frequency change was within 10 MHz between 4 °C to 74 °C. The dotted line indicates the calculated Δf using the temperature coefficients in the table.1 (sample No.2), whrere $\Delta f=2.8\cdot(\alpha(T-25)+\beta/2\cdot(T-25)^2)$(MHz), α=-0.002(G/°C) ,β=-.308(G/°C^2).The frequency-temperature characteristics well reflected the temperature-resonance field characteristics as examined in the section 3. Finally, the MSFVW chip resonator using the present material , without any temperature compensation procedures, shows a ultra-low frequency-temperature dependence, which was less than that of a YIG sphere device(fig.12)[9].

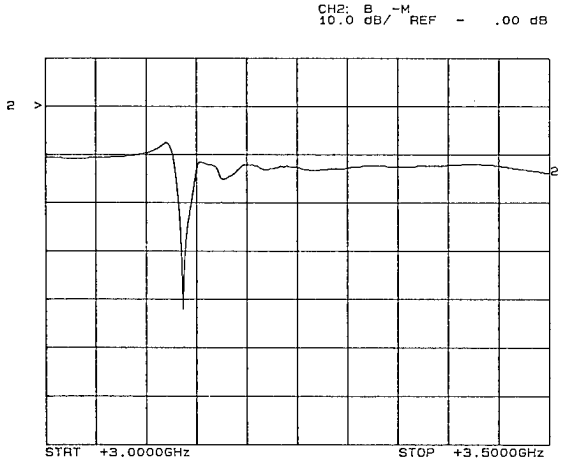


Fig.11 A typical $|S21|$ resoponse of a MSFVW chip resonator[9] made from the present material (sample No.2).

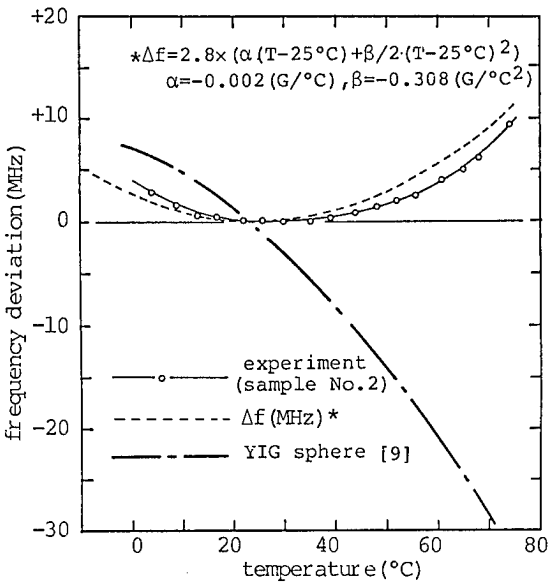


Fig.12 Frequency deviations versus temperature for the MSFVW resonator using the present material (sample No.2) in comparison with a YIG sphere device[9].

5. Conclusion

A YIG film having a small frequency-temperature dependence with low ΔH has been searched both theoretically and experimentally. It is found that frequency-temperature characteristics of a low-loss Bi,Ga substituted YIG film epitaxially grown on a (111) oriented GGG substrate with proper growth induced anisotropy magnetic field is better than that of a YIG shpere. We convince that a planar-type MSW device using the present material will replace the YIG sphere device.

References

[1] C.Kittel,"On the Theory of Ferromagnetic Resonance Absorption",Phys.Rev.71,pp.270,1947.

[2] H.L.Glass,J.H.W.Liaw and M.T.Eliott,"Temperature Stablization of Ferrimagnetic Resonance Field in Epitaxial YIG by Ga,La substitution.",Mat.Res .Bull.Voll.12,pp.735-740. 1977.

[3] P.Röshmann and T.Tolksdolf,"Epitaxial Growth and Annealing Control of Thick Homogeneous Ga Substituted Yttrium Iron Garnet Films" ,Mat.Res. Bull.,Vol.18,pp.449-459,1983.

[4] H.Le Gall,J.P.Castera,P.Hartmann and D.Mahasoro, "Influence of Ga-substitution the Magnetostatic Waves Properties in Garnet Films",IEEE Trans.on MAG.,Vol.MAG-22,No.5,Sep.,1986.

[5] T.Ryuo and M.Tanno,"High Quality YIG Film for MSW Devices",Proc. of IEEE Ultrasonics Symp., pp.237-240,1988.

[6] R.I.Joseph and E.Sclomann,"Demagnetizing Field in Nonellipsoidal Bodies", J.A.P., Vol.36,No.5 ,May,pp.1579-1593.

[7] P.Hansen , P.Röshmann and T.Tolksdolf, "Saturation Magnetization of Gallium-substituted Yttrium Iron Garnet",J.A.P.,Vol.45,No.6,pp.2728-2732,1974.

[8] P.Hansen, "Anisotropy and Magnetostriction on Gallium-substituted Yttrium Iron Garnet" ,J.A.P.,Vol.45,No.8,pp.3638-2642,1974

[9] K.Ito et al.,"Voltage Controlled Oscillators with Planar Resonators",Proc.of IEEE Ultrasonics Symp.,pp.323,1989.

[10]J.P.Castera,"State of the Art in Design and Technology of MSW Devices (invited)",J.A.P. ,55(6),pp.2506,15 Mar.,1984.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

EXTREMELY HIGH-LOOP POWER GHz RANGE SURFACE TRANSVERSE WAVE OSCILLATORS USING AB-CLASS AMPLIFIERS

IVAN D. AVRAMOV

Institute of Solid State Physics
72 Tzarigradsko Chaussee Boulevard
1784 Sofia
Republic of Bulgaria

Abstract

This paper describes novel microwave feedback power oscillators, stabilized with 1 GHz shear horizontal (SH) surface acoustic wave low-loss resonant devices, dissipating a radio frequency (RF) power of more than 1 W in the loop. The loop power of 35 dBm is generated by matched, highly efficient AB-class power amplifiers. A record value of -194 dBc/Hz for the thermal phase noise floor is achieved with a voltage controlled oscillator (VCO) featuring a tuning range of 1.2 MHz and an efficiency of 19%. When operating as linear, directly frequency modulated (FM) microwave power transmitters, such VCO are capable of transmitting data rates exceeding 160 Kbits/s and 300 Kbits/s in FM bandwidths of 150 KHz and 600 KHz respectively.

Introduction

Surface acoustic wave (SAW) oscillators have established themselves as stable frequency sources, featuring excellent phase noise performance in the lower microwave range [1], [2]. Johnson phase noise floors of -184 dBc/Hz have been demonstrated at 500 MHz with SAW two-port resonator stabilized oscillators [1]. As shown in [1] the single-sideband noise-to-carrier ratio $\mathcal{L}(f)$ in dBc/Hz for the Johnson phase noise, also called thermal noise floor, in a feedback oscillator can be obtained by following equation:

$$\mathcal{L}(f) = -174 + G + F - P_o \qquad (1)$$

where G/dB is the compressed power gain of the loop amplifier, also equal to the loss in the feedback loop, F/dB is the amplifier noise figure and P_o/dBm is the oscillator carrier output power, also equal to the loop power level in the feedback oscillator. From (1) it is evident that, in order to minimize the oscillator thermal noise floor $\mathcal{L}(f)$, one should decrease G and F and increase P_o. SAW resonators usually have insertion losses in the 3-6 dB range. Somewhat lower values can be achieved only at the expense of reduced resonator loaded Q. Not much improvement can be achieved by reducing the noise figure F either since bipolar transistor amplifiers have a very low noise factor already which is typically in the 2-6 dB range. Obviously the only real improvement of $\mathcal{L}(f)$ can be achieved by increasing the oscillator loop power since the only limit to this is the power handling ability of the acoustic frequency stabilizing element. It is well known that SAW resonators with durable Cu- or Ti-doped Al electrodes experience substantial aging even at relatively low incident power levels. For example a 640 MHz resonator, operated at a power of 24 dBm, exhibited a decrease of its resonant frequency by more than 260 ppm after only a few days of operation [3]. Dual-track SAW resonator configurations can stand a few dB more of incident power but they require a complex laser trimming procedure [4].

So far the utilization of shear horizontal surface acoustic waves seems to be the best solution of the power handling problem in resonant devices. One of their most important advantages is the ability to handle extremely high radio frequency power levels on a small substrate area without deteriorating the electrical or aging device performance. As shown in [3] and [5] the maximum RF power, P_{max}, a SH-wave resonator can dissipate, can be calculated using the following equation:

$$P_{max} = \left(\frac{|T_{12}|_p}{2050} \right)^2 \frac{L_e D W}{Q_U \lambda} \qquad (1)$$

where $|T_{12}|_p$ is the highest component of the shear stress that can occur in the SH-wave device and here it is assumed to be as high as the degradation peak stress value of $2.8*10^8$ N/m^2 for SAW resonators with Al-Cu metallization, Q_U is the device unloaded quality factor, λ is the acoustic wave length, D is the penetration depth into the bulk, which is assumed to be 0.75λ [3], W is the acoustic aperture and L_e is the effective resonant cavity length. According to (1), and as confirmed experimentally in [3], [5] and [6], SH-wave based high-Q and low-Q resonators as well as combined mode resonator filters (CMRF) were found to be able to handle RF power levels of more than 2 W while occupying a substrate area of only about 3 mm^2 on temperature stable rotated Y-cuts of quartz. If such devices are used to stabilize the frequency of feedback oscillators extremely low thermal noise floor values can be achieved. Such state-of-the-art SH-wave based power oscillators with Johnson noise values as low as -194 dBc/Hz are subject of this paper.

0-7803-0905-7/93 $3.00 © 1993 IEEE

Operation Principle of a SH-wave Based Feedback Oscillator Using an AB-class Power Amplifier

The operation principle of a SH-wave oscillator, using an AB-class power amplifier, is shown in Fig. 1. The circuit consists of a SH-wave frequency stabilizing device, usually a surface transverse wave (STW) resonator or CMRF, a loop adjust line for setting the phase condition of oscillation, a signal sampling circuit for delivering part of the RF power to the load and an AB-class power amplifier, satisfying the amplitude condition of oscillation and generating the loop power, responsible for the oscillator Johnson phase noise. The circuit contains also matching networks at the input and output of the AB-class amplifier. They are necessary for proper B-class operation. For electronic frequency tuning and modulation a voltage controlled variable phase shifter (VPS) can also be included in the loop (see below).

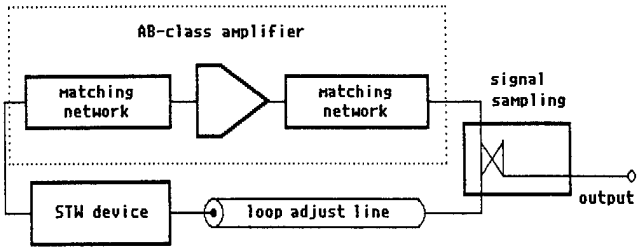


Fig.1. Schematic of a SH-wave feedback oscillator using an AB-class power amplifier.

Except for achieving the desired high values of the loop power, the oscillator has to be selfstarting. This means that oscillation has to reliably build up when the supply voltage is turned on. The key to achieving this is in the nonlinear function of the AB-class amplifier. Its transistors have to be biased in such manner that for small input drive power levels it behaves as a linear A-class amplifier, providing enough gain for the oscillation to build up. When, after a few loops, the input power P_{in} achieves a certain threshold value P_{TB} the amplifier starts operating in the switched B-class mode, generating a compressed loop power which can exceed several Watts. The typical performance of an AB-class amplifier is illustrated in Fig. 2 which shows the amplifier gain as a function of frequency for P_{in} values changing from -36 dBm to +14 dBm. The gain is about 13 dB at 1 GHz and increases linearly with input drive levels of -36 dBm to +4 dBm. This causes a parallel shift of the lowest five curves in vertical direction. With an input drive level of +8 dBm the gain curve changes its shape abruptly and, with an input power level of +14 dBm the output power achieves the value of +27 dBm. Dependent on the transistors used the output power level can achieve much larger values if proper matching conditions to the source and load impedances are created. In this study, for example, an output power of +35 dBm was achieved at 1 GHz using a highly efficient bipolar transistor with a transition frequency of 5 GHz and a total direct current (dc) power dissipation of 1 W.

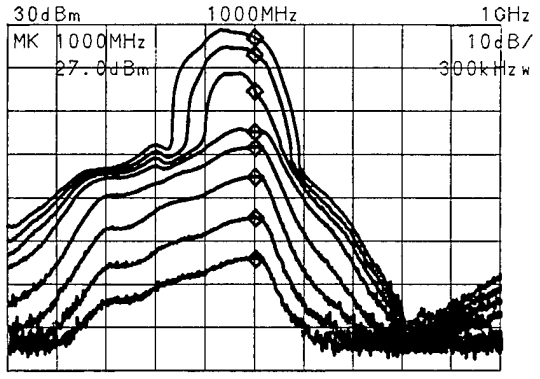


Fig. 2. Gain versus frequency of an AB-class power amplifier with the input power P_{in} as a parameter: P_{in}=-36 dBm (lowest curve), -26, -16, -6, +4, +8, +11, +14 dBm (highest curve). horizontal: 100 MHz/; vertical: 10 dB/.

Obviously, for proper oscillator operation, following three conditions have to be satisfied:

1. **Phase condition:**

$$\phi_a + \phi_e = 2k\pi \qquad (3)$$

2. **Amplitude condition:**

$$G_s + L_l \geq 0 \qquad (4)$$

3. **Power condition:**

$$P_{in} \geq P_{TB} \qquad (5)$$

Here ϕ_a is the phase shift of the acoustic resonator, ϕ_e is the electrical phase shift, associated with the amplifier, matching networks, loop adjust lines, phase shifters, power sampling circuits etc., k is a positive integer, G_s/dB is the small signal gain of the amplifier and L_l/dB is the overall loop loss associated with all passive components in the circuit.

If only conditions 1. and 2. are satisfied then oscillation will occur but the carrier loop power will be very low, typically a few dBm or less. This will be the case if the AB-class amplifier is poorly designed and goes into saturation without satisfying condition 3 or if matching conditions at the input and output are bad. From Fig. 1 it is evident that the higher the output power and the lower the loop loss the easier condition 3. is satisfied.

Oscillator Realization

All oscillators, considered here, were built with the circuit shown in Fig. 3. Its loop components are an AB-class amplifier, a Wilkinson power divider, a loop adjust line, a VPS and an A-class amplifier/limiter. The AB-class amplifier

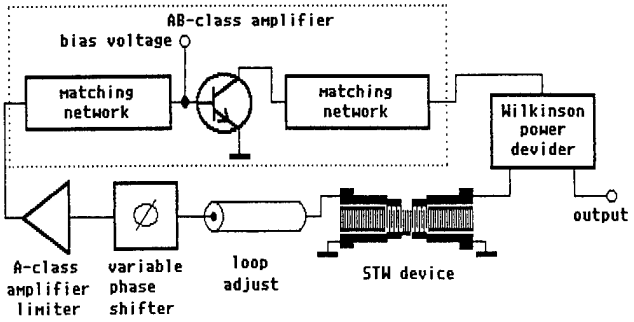


Fig. 3. SH-wave based power VCO circuit.

is realized with a single 1 W bipolar transistor, carefully matched at its input and output for proper B-class operation [7]. A bias voltage is applied to the base of the transistor to guarantee small signal A-class operation for selfstarting. This power transistor stage is driven by an A-class amplifier which has a complex and very important function. On one hand it provides sufficient gain for satisfying the amplitude condition and isolates the VPS from the AB-class stage. This isolation is necessary because the VPS output impedance change with tuning would strongly deteriorate the matching conditions at the AB-class amplifier input. On the other hand the A-class amplifier satisfies the power condition for oscillation and performs the limiting function in the loop. Without limiting the input power P_{in} (see Fig. 1) would increase too much and this would thermally overload the AB-class power transistor stage. The VPS is a widely used high-pass CLC-type varactor tuned circuit [8] which was found to perform well even at +27 dBm of incident power. Its insertion loss at 1 GHz was typically below 1 dB. The Wilkinson power divider [8] is used to split the carrier power into two equal parts. One of them is incident on the SH-wave device and the second one is delivered to the 50 Ohm load at the oscillator output. In microstrip technology this power divider was found to have an insertion loss of 3.8 dB in each branch.

Results

Two types of SH-wave resonant devices were implemented in the oscillator designs: high Q STW resonators with an insertion loss of 7 dB and a loaded Q of about 4000 [5], and wide band CMRF with an insertion loss of 4-5 dB and a group delay below 150 ns for tuning ranges exceeding 0.1% of the resonant frequency and FM rates of up to 300 Kbits/s [6]. Both types of devices were found to conveniently handle RF-power levels of 2W. In the CMRF based oscillator the G value was 9.6 dB (4.8 dB CMRF insertion loss, 3.8 dB loss in the Wilkinson power divider and 1 dB loss in the VPS and the loop adjust line). The noise figure F of the AB-class amplifier was measured as 5 dB and the loop power was 34.8 dBm with an incident power level of 31 dBm at the SH-wave device input transducer. According to (1) this yields a $\mathcal{L}(f)$ value of 194.2 dBc/Hz. This result is confirmed experimentally in Fig. 4 which is the phase noise performance of this oscillator. The 194 dBc/Hz value is achieved at an offset of 6 MHz away from carrier. This value is by 10 dB lower than the lowest phase noise floor reported to date [1] and, to the best of the author's knowledge, is the lowest Johnson noise achieved with a microwave oscillator, stabilized with an acoustic wave resonator, to date.

The fixed frequency 1 GHz oscillators were found to have a Johnson phase noise floor of -189 dBc/Hz and an output power of 29 dBm in a 50 Ohm load. Compared to the CMRF based power oscillators these values are lower because of the higher insertion loss of the high-Q devices.

Fig. 5 is the harmonic content of the CMRF based power VCO without additional filtering. It is evident that the harmonics, generated by the amplifier circuits in the loop, are suppressed by more than 25 dBc which is due to the low small

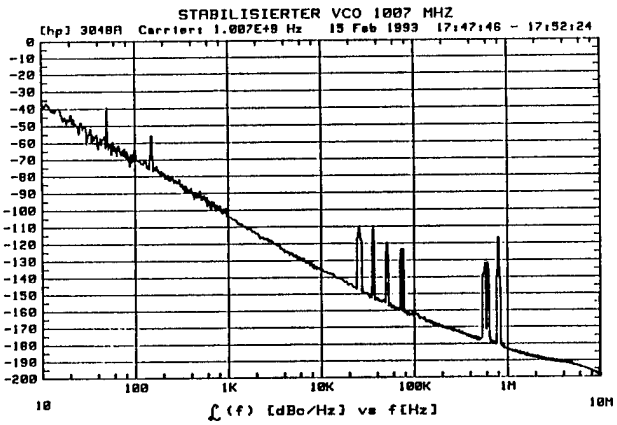


Fig. 4. Phase noise performance of a 1007 MHz CMRF based VCO with a tuning range of 1.2 MHz and ability to transmit FM-data at a rate of 300 Kbits/s in a 600 KHz FM-bandwidth.

signal gain compression this oscillator can stand. In this particular case its value was only 2.4 dB and nevertheless no difficulties with selfstarting or instabilities with tuning were observed.

Fig. 6 is part of the tuning characteristic of a 1 GHz power VCO stabilized with a low-Q SH-wave resonator similar to the one described in [5]. The output power variation over the whole tuning range did not exceed ± 0.5 dB. At an operating point of about 12 V the curve in Fig. 6 has a linear part where extremely linear FM is possible. This is illustrated in Fig. 7a) and 7b) which show wideband FM spectra of the VCO for modulating sinusoidal frequencies of 2 and 68 KHz respectively. The modulation bandwidth was 150 KHz. The symmetry of

both spectra indicates a very low nonlinear distortion. The maximum modulating frequency at which the modulation index became equal to 1 was 80 KHz which corresponds to a data rate of 160

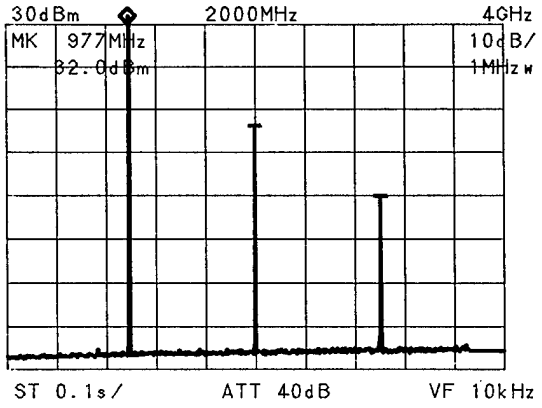


Fig. 5. Harmonic content of a 1 GHz CMRF based power VCO.
Frequency span: 4 GHz
Marker value at the fundamental frequency: +32 dBm/977 MHz.

Kbits/s. Similar experiments were performed with CMRF based power VCO. At a carrier frequency at which the CMRF group delay was 150 ns the maximum data rate achieved was 300 Kbits/s in a FM bandwidth of 600 KHz.

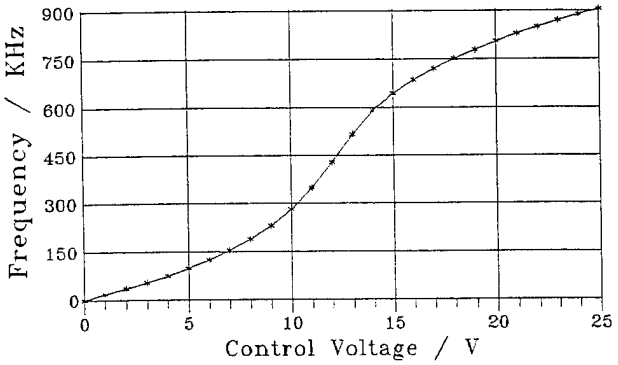
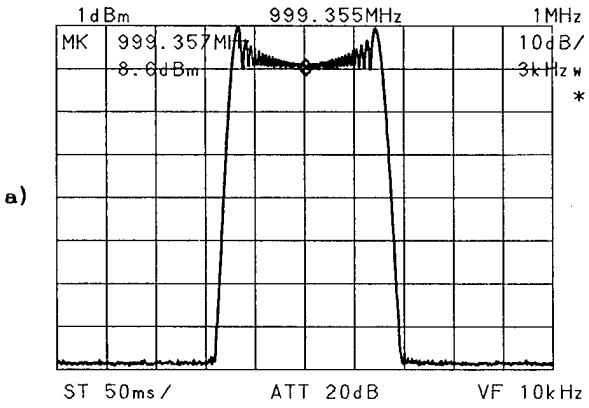


Fig. 6. Tuning characteristic of a power VCO stabilized with a 1 GHz SH-wave low-Q resonator with a group delay of 500 ns.

Efficiency Considerations

The RF/dc efficiency can be defined as the ratio between the carrier RF power in the load and the total d. c. power consumption of the oscillator. Since a Wilkinson power divider is used in the circuit shown in Fig. 3 only half of the carrier power is delivered to the load. Thus the RF/dc efficiency in this case can never exceed 50%, even if the efficiency of the active circuitry in the loop would be 100%. The CMRF based oscillator in Fig. 3 demonstrates a RF/dc efficiency of 19%. while the oscillators, stabilized with the high-Q devices, had an even lower efficiency of about 15%. In both cases the main limiting factor was the A-class amplifier. Still the RF/dc efficiency of these power oscillators is 5-20 times better than that of SAW oscillators using A-class amplifiers.



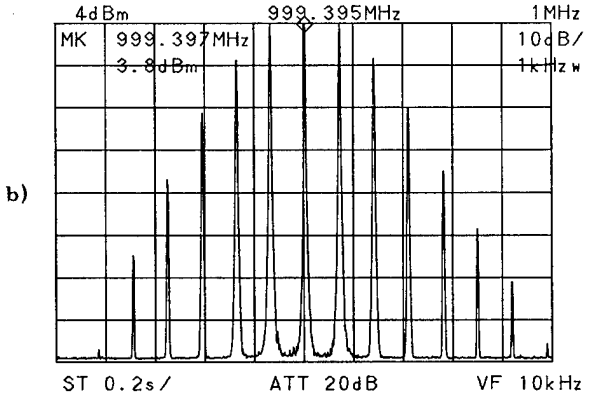


Fig. 7. FM-spectra of low-Q SH-wave resonator stabilized power VCO for modulating sinusoidal frequencies of:
a) 2 KHz; b) 68 KHz.

There are two ways of improving the RF/dc efficiency of the power oscillators, considered here:

1. By designing highly efficient AB-class amplifier circuits. B-class amplifiers, using bipolar transistors can achieve efficiency values of 70-80%. There is no reason why they should not work well in the AB-class mode. An AB-class amplifier with a 80% efficiency would increase the oscillator efficiency to 40%.

2. By replacing the Wilkinson power divider with a directional coupler, featuring a very low insertion loss in the load branch and an increased insertion loss in the loop branch. In this case the oscillator efficiency would approach 80% but at the expense of a higher noise floor which,

according to (1), would increase due to the higher loop loss. Reversing both branches of the directional coupler would have the opposite effect: The noise floor would be improved by almost 3 dB at the expense of reduced output power and lower RF/dc efficiency accordingly.

Thermal Stress Behaviour of the SH-wave Resonator

If the incident RF power on the SH-wave device exceeds 1 W it is quite obvious that the resonator will experience a substantial thermal shock immediately after the oscillator is turned on. In fact the frequency shift in the first few minutes of the oscillator operation will depend on the RF power, dissipated in the acoustic device, on its turn-over temperature and its thermal contact with the case and the environment. In this study the device turn-over temperature was about 35 deg. C and the resonators were attached with silicon adhesive to the oscillator circuit. The incident RF power was 30 dBm. In most cases the frequency down shift in the first 1 minute, after the oscillator was turned on, was about 100 ppm and increased by another 100 ppm after 10 minutes of operation. These 200 ppm of initial frequency shift correspond to a temperature of about 90 deg. C on the surface of the substrate (see Fig. 31, 32 and 33 in Reference [5]). All oscillators were found to conveniently stand this thermal and RF power stress after several weeks of operation. No evidence of performance degradation or aging was observed.

Conclusions

This paper has demonstrated SH-wave based microwave fixed frequency and voltage controlled power oscillators with state-of-the-art white phase noise floor performance. The combination of a highly efficient AB-class power amplifier with a RF-power durable SH-wave resonant device as an acoustic frequency stabilizing element in the loop of a feedback oscillator was found to be an excellent technical solution for achieving Johnson noise values of -194 dBc/Hz. This result was obtained with a carrier loop power of +35 dBm and an incident power level of up to +31 dBm on the acoustic device. The oscillators were found to operate at a RF/dc efficiency of 19% and deliver a power of up to +32 dBm in a 50 Ohm load. Excellent tuning of up to 0.12 % of the carrier frequency and direct wideband FM at a rate of 300 Kbits/s were demonstrated with CMRF based VCO.

In the author's opinion further optimization of the power oscillator circuit can reduce the thermal noise floor below the -200 dBc/Hz value. This can be achieved by:

- increasing the loop power by 2 more dB. The SH-wave devices can conveniently stand a continuous power level of +33 dBm,

- reducing the loss in the loop branch of the power sampling circuit by 3 dB

- reducing the loss in the acoustic device by 1.5-2 dB (see Reference [6]),

-reducing the noise figure of the A-class amplifier by 2-3 dB using appropriate noise matching.

It is the author's belief that the proposed technique would be very competitive in extremely low-noise VCO for phase-locked-loop applications and in FM microwave power transmitters for local area radio networks and portable communications.

Acknowledgements:

The author wishes to gratefully acknowledge Drs. I. B. Yakovkin, A. S. Kozlov and A. Aseev from the Institute of Semiconductor Physics in Novossibirsk, Russia for the opportunity to have the SH-wave devices fabricated at those laboratories. I also thank Dipl.-Ing. T. Felgentreff from the Technical University in Munich, Germany for his help in performing the phase noise measurements and Dr. T. E. Parker from Raytheon Research Division in Lexington, Massachusetts for his help with evaluation of the phase noise data and for the fruitfull discussions.

This work was supported by the National Research Fund of Bulgaria.

References

[1]. *T. E. Parker and G. K. Montress, "Precision Surface Acoustic Wave (SAW) Oscillators", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, May 1988, Vol. UFFC-35, No. 3, pp. 342-364.*

[2]. *G. K. Montress and T. E. Parker, "State-of-the-Art SAW Oscillators: Design and Performance, Proc. 7-th European Frequency and Time Forum, EFTF-93, March 16-18, 1993, Neuchatel, Switzerland (in press).*

[3]. *T. L. Bagwell and R. C. Bray, "Novel Surface Transverse Wave Resonators with Low Loss and High Q", IEEE 1987 Proc. Ultrasonics Symposium, pp. 319-324.*

[4]. *J. A. Greer, G. K. Montress, T. E. Parker, "Applications of Laser-Trimming Techniques for All Quartz Package, Surface Acoustic Wave Devices, Proc. 1989 IEEE Ultrasonics Symposium, pp. 179-184.*

[5]. *I. D. Avramov, "Microwave Oscillators Stabilized with Surface Transverse Wave Based Resonant Devices", Proceedings of the 1992 IEEE Frequency Control Symposium, May 29-31, 1992, Hershey, PA, pp. 391-408.*

[6]. *I. D. Avramov, "1 GHz Low Loss Coupled Resonator Filter Using Surface Skimming Bulk Waves and Bleustein-Gulyaev Waves", Electronics Letters, February 1991, Vol. 27, No. 5, pp. 414-415.*

[7]. *Ю. В. Завражнов, И. И. Каганова, Е. З. Мазель, "Мощные высокочастотные транзисторы", Москва, "Радио и связь", 1985.*

[8]. *G. K. Montress, T. E. Parker, "High Stability SAW Oscillators", Short Courses Presented at the 1991 IEEE Ultrasonics Symposium, December, 1991.*

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

Phase Noise and Frequency Stability of Ka-Band Harmonic Dielectric Resonator Oscillators.

Muhammad Mizan,Thomas Higgins and Dana Sturzebecher

Army Research Laboratory, E&PS Directorate, Fort Monmouth, N.J. 07703

ABSTRACT

Harmonic Dielectric Resonator Oscillators (HDROs) designed for 25 and 27 GHz were fabricated using a two stage GaAs MESFET amplifier and a passive two port Resonant Structure (RS). The RS consists of a metal cavity, an X-band ceramic dielectric resonator, 50Ω microstrip transmission lines and a support for elevating the dielectric resonator inside the cavity. The electromagnetic behavior of the RS was analyzed using Hewlett-Packard's High Frequency Structure Simulator (HFSS). The simulator was used to calculate the EM fields and S-parameters of the passive RS. Solutions were obtained for the dominant mode. The predicted computer results were then compared to measured data. The material properties of dielectric resonators were obtained from two different vendors and evaluated for low noise microwave oscillator applications. The important material properties for oscillator applications are the unloaded Q, temperature coefficient, and dielectric constant. Measured data for resonators from both vendors are presented. The measured phase noise levels at 27 GHz are a 15-20 dBc/Hz improvement over previously published results for DROs. The oscillator also demonstrated a superior frequency vs. temperature characteristic, over the temperature range ±55 °C, of less than 40 ppm.

INTRODUCTION

The absolute phase noise of a dielectric resonator oscillator (DRO), in a parallel feedback configuration, is governed mainly by the noise of the oscillator's loop amplifier and the bandwidth of the resonant circuit. The resonator bandwidth is defined as $BW = f_0/Q_L$ where f_0 is the center frequency, and Q_L is the loaded Q of the resonator. To enhance an oscillator's absolute phase noise, one must either lower the amplifier noise, or increase the resonator Q_L. The task for designing low noise frequency sources at L and S-band is simplified due to the availability of extremely low 1/f noise BJT amplifiers, and high Q dielectric resonators. DROs at X-band, have been designed and tested to show phase noise levels of -65 dBc/Hz and -93 dBc/Hz, at 100 Hz and 1 KHz offset frequencies [1]. To obtain similar phase noise levels at Ka-band, one would need very high Q values, as calculated by:

$$\mathcal{L}_c(f) = \mathcal{L}_o(f) - 20Log(f) + 20Log\left[\frac{F_0}{2Q_L}\right]$$

where $L_C(f)$ is the single-sideband phase noise in dBc/Hz, $L_O(f)$ is the open-loop amplifier 1/f (or residual) noise, f is the carrier offset or noise frequency in Hz, F_0 is the carrier frequency and Q_L is the loaded Q of the DR [2]. One method of achieving low phase noise at higher frequencies is to multiply a low noise, low frequency source and accept a noise degradation of 20•LOG (n), where n is the harmonic number. While this approach may have a good phase noise characteristic close-in to the carrier, far from the carrier the performance is only marginal. Disadvantages of this method are large size, low efficiency, circuit complexity and the need for a low noise multiplier circuit. A source operating at a fundamental frequency, offers a good phase noise characteristic far from the carrier, but poor close-in noise performance. A 27 GHz dielectric resonator loaded cavity can be designed to provide Q values of 3,000 to 4,000. Typical amplifier 1/f noise levels at this frequency are approximately -115 dBc/Hz, at a 100 Hz offset, with a -10 dB/decade slope. Theoretical phase noise levels at 27 GHz, were calculated from this data and compared to 2nd harmonic of Ku-band DRO and 3rd harmonic of X-band DRO. The results are shown in Table 1.

U.S. Government work not protected by U.S. copyright

Oscillator Type	Loaded Q	Phase Noise dBc/Hz @ offset frequencies			
		100 Hz	1 KHz	10 KHz	100 KHz
Fundamental 27 GHz DRO	4,000	-24	-54	-84	-114
2nd Harmonic of Ku-Band DRO	5,800 at Ku-Band	-32	-62	-82	-104
3rd Harmonic of X-Band DRO	12,700 at X-Band	-43	-71	-98	-126

Table 1. Phase of different DRO types noise at Ka-Band. The 3rd harmonic DRO is actual measured data.

The theoretical phase noise of a 2nd harmonic oscillator was estimated from the measured absolute phase noise at the Ku-band fundamental operating frequency, while assuming noiseless multiplication except for the 20•Log(n) degradation. The phase noise of the 3rd harmonic oscillator is the actual measured absolute phase noise at Ka-Band.

FINITE ELEMENT ANALYSIS OF THE RESONANT STRUCTURE

The purpose of this analysis is to determine the resonant frequencies of the high Q structure. The model created for analysis on HFSS is shown in Figure 1. The 50 ohm microstrip transmission lines are on 0.0254 cm thick duroid board with a dielectric constant of 2.2. The separation between excitation and receive port was optimized for coupling energy into and out of the resonator, while maximizing isolation between the ports. The DR support post was glued to the brass plate through a cutout in the dielectric. The DR was mounted on the low loss and low dielectric constant support post. Figure 1 also shows the mesh that was generated by the software. For this application we have found that the HFSS demands a high speed computer with large memory, however is the only commercially available EM field solver that is capable of handling arbitrary 3D geometries. The HFSS version A.02.06, running on a HP Apollo Series 735 workstation with 145 Mbytes of RAM and 500 Mbytes of usable disk space, solved for

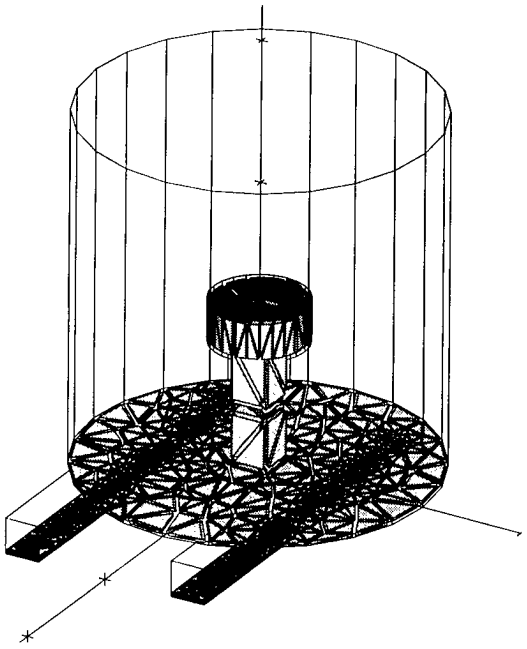


Figure 1. 3-D geometry created within HFSS.

forty frequency points between 8 and 9.5 GHz in 60 hours. The simulated and measured data are shown in Figure 2. It has not been determined if the differences in the data, shown in Figure 2, are due to inaccuracies in the measurement of the DR and cavity dimensions or approximations of the geometries within the software. Using the post processing software, the magnitude of E and H fields at various cut planes could be displayed, an example of which is shown in Figure 3.

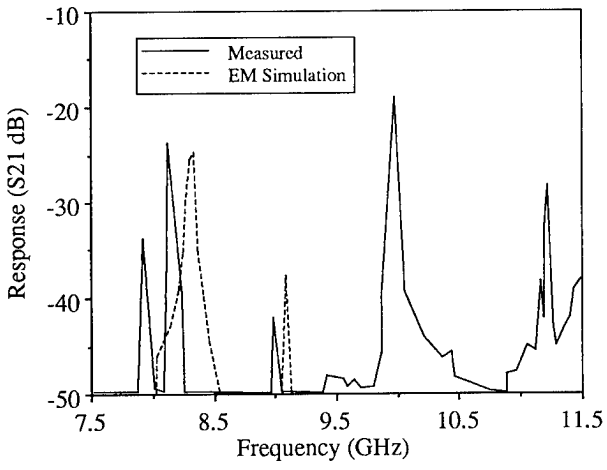


Figure 2. Measured and EM simulated response of X-band RS. The simulated data is 8-9.5 GHz only.

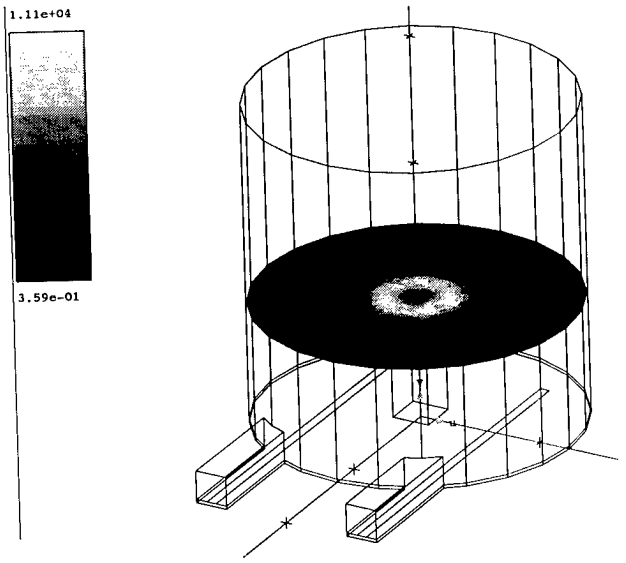


Figure 3. Magnitude of the E-field in a plane thru the DR.

RESONATOR CHARACTERIZATION

Trans-Tech and Murata Erie's dielectric resonators were evaluated for low noise oscillator applications. The resonators were tested for insertion loss (IL) and Q_L in an X-band RS configuration under TE_{018} mode operation. The results of this investigation are presented in Table 2.

Vendors	Frequency (GHz)	IL(dB)	Q_L	Epsilon
Murata Erie E-Series	9.857	4.5	3,457	24.2
		5.1	8,099	
		7.8	12,047	
Murata Erie S-Series	10.0876	7.5	3,435	28.6
		8.3	6,029	
		12.3	8,428	
Trans-Tech 8700 Series	10.3727	15	1,783	29.4
		17	3,345	
		21.1	4,378	

Table 2. Measured insertion loss and Loaded Q values for different resonator materials.

It can be seen from the table that Murata Erie's E-series material has the highest loaded Q with minimum insertion loss, and is therefore the optimum choice for low noise oscillators. The unloaded Q specified by the manufacturer for this material at X-band is about 25,000

OSCILLATOR DESIGN AND PERFORMANCE

The two X-band DROs were designed so the third harmonic (Ka-band) could be extracted. A block diagram of the oscillators is shown in Figure 4. The resonant circuit had a Q_L of 13,000 at X-band, and consisted of a cavity loaded with a Murata Erie 0.0 ppm/°C dielectric resonator, a support post, and microstrip transmission lines for input and output ports.

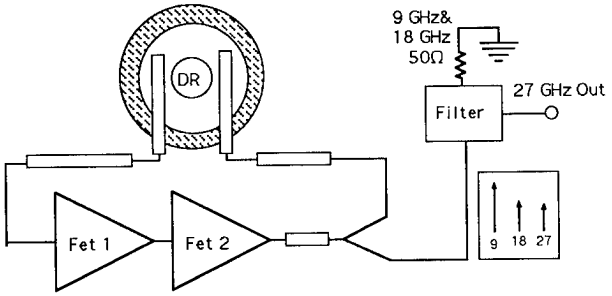


Figure 4. Diagram of the oscillator configuration.

The design incorporated a two-stage MESFET amplifier, which was required to overcome the insertion loss of the resonant circuit to satisfy oscillation criteria. The MESFETs used, were chosen because they had the low 1/f noise levels of -140 dBc/Hz and -150 dBc/Hz at 100 Hz and 1 KHz offset frequencies, and is shown in Figure 5. Gate and drain voltages were adjusted for minimum noise. Other devices were investigated for this application such as HEMT and HBT amplifiers. The effect of different bias conditions on 1/f noise in a Fujitsu HEMT amplifier is shown in Figures 6 & 7. An HBT amplifier was also tested for its 1/f noise characteristic and is shown in Figure 8.

The HDROs produced approximately -16 dBm of output power at Ka-band. The absolute phase noise of the HDRO at Ka-band, was measured with a HP 3048A

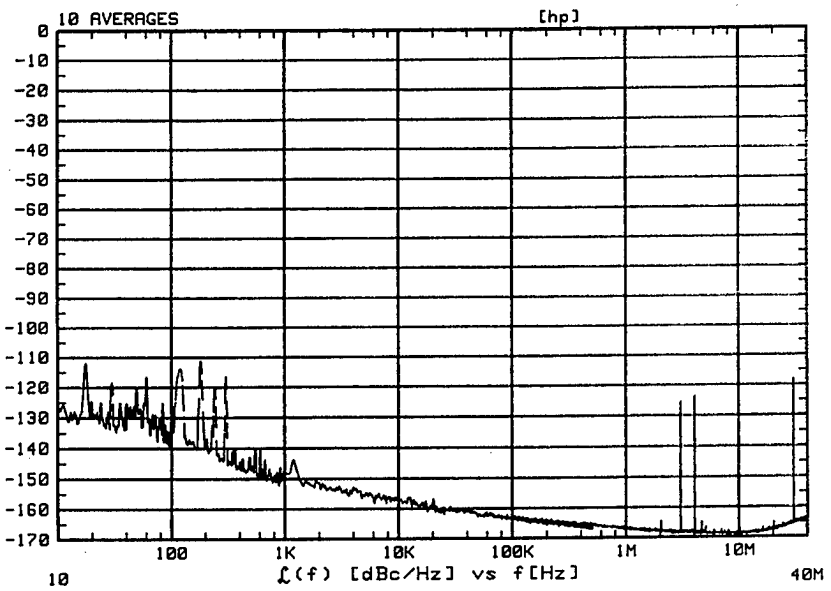


Figure 5. 1/f noise of MESFET amplifier.

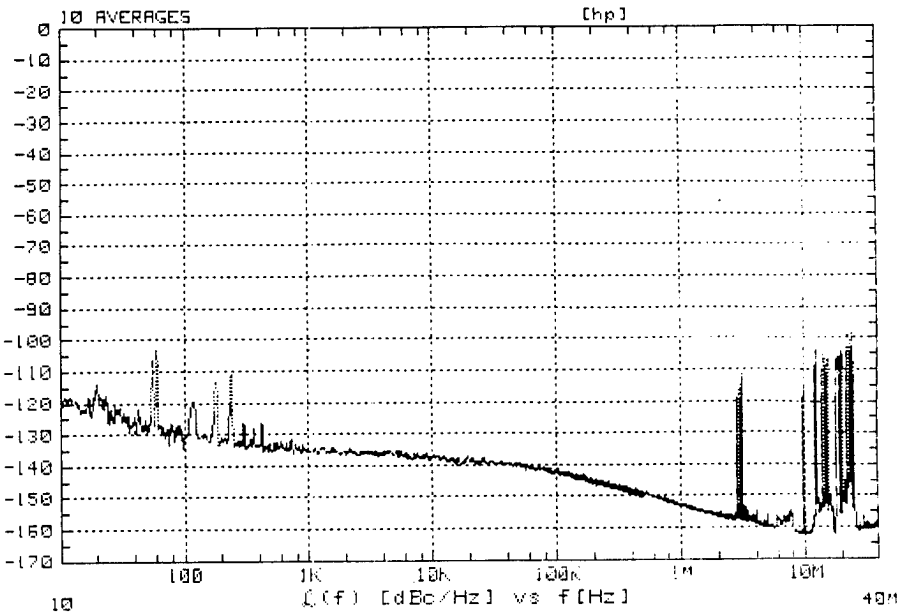


Figure 6. 1/f noise of HEMT amplifier under the bias conditions Vd=2.0 and Id=10.0 mA.

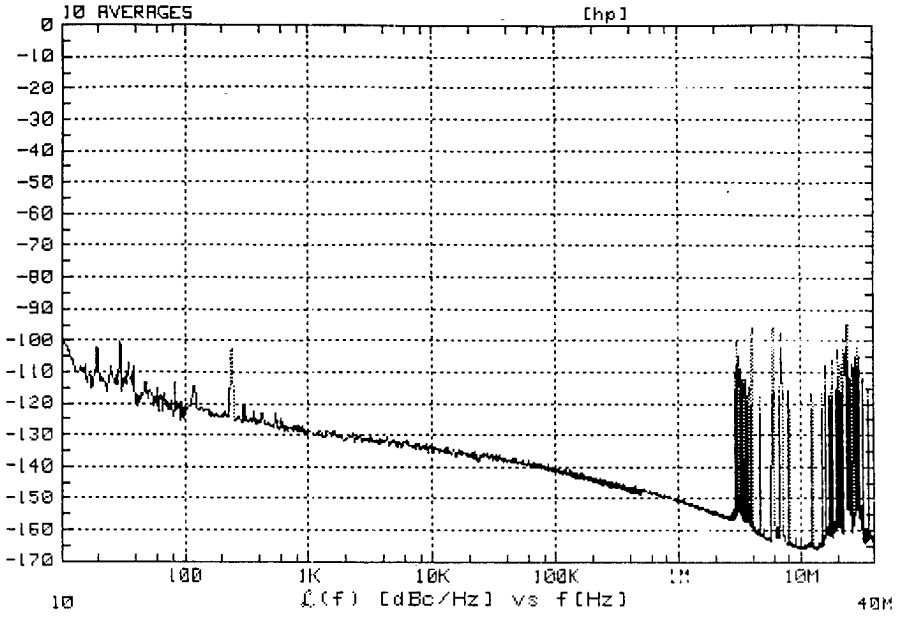


Figure 7. 1/f noise of HEMT amplifier under the bias condition of Vd=3.0 and Id=20.0 mA.

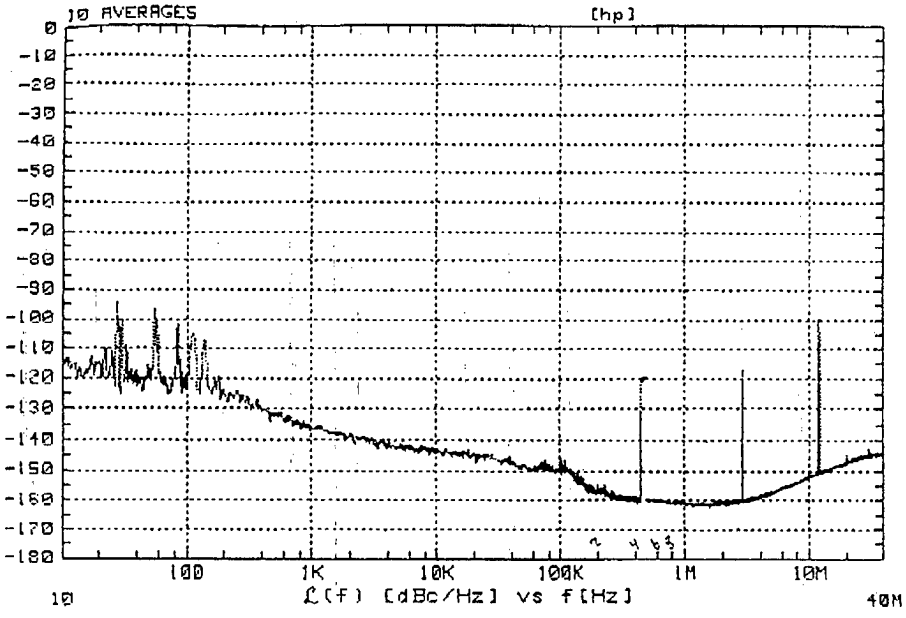


Figure 8. 1/f noise of HBT amplifier.

measurement system is shown in Figure 9. Three port filters were designed to terminate the fundamental frequency in 50 ohms while passing the 3rd harmonic. The 27 and 25 GHz HDROs were used to generate an intermediate frequency (IF) of 2.15 GHz, which was downconverted to a second IF of 650 MHz by using a 1.5 GHz extremely low noise local oscillator (LO).

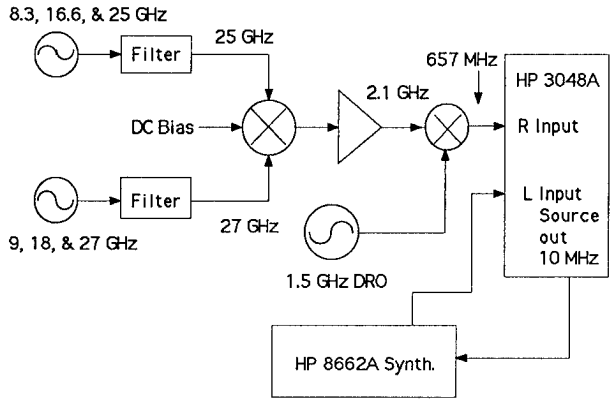


Figure 9. Test setup used for Phase noise measurement.

The resulting measurement provided the phase noise of both HDROs and the system noise. A 3 dB correction must be subtracted from the data to derive the phase noise of the individual HDROs. The measured absolute phase noise data is shown in Figure 10, and is 15-20 dBc/Hz lower than the data reported in references [3-5]. The frequency stability of the 27 GHz oscillator was measured over the temperature range of -55 °C to +55 °C, and is shown in Figure 11.

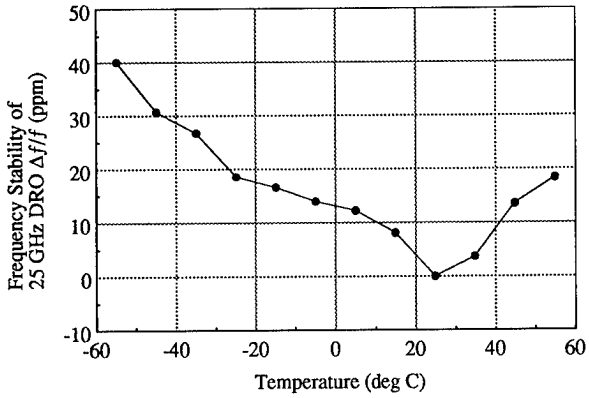


Figure 11. Frequency Stability vs. Temperature for the 27 GHz HDRO.

The largest shift in frequency is 40 ppm. This is an improvement over commercially available DROs, by a factor of three. It should be noted that no temperature compensation techniques were used to achieve this performance. The output power spectrum at 27 GHz with a 500 KHz span is shown in Figure 12.

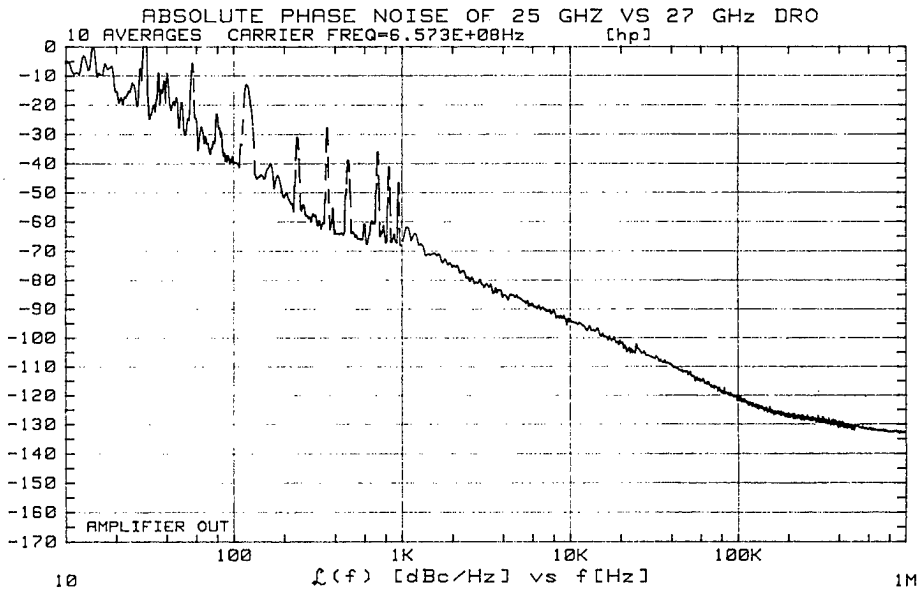


Figure 10. Absolute phase noise of HDRO at 27 GHz.

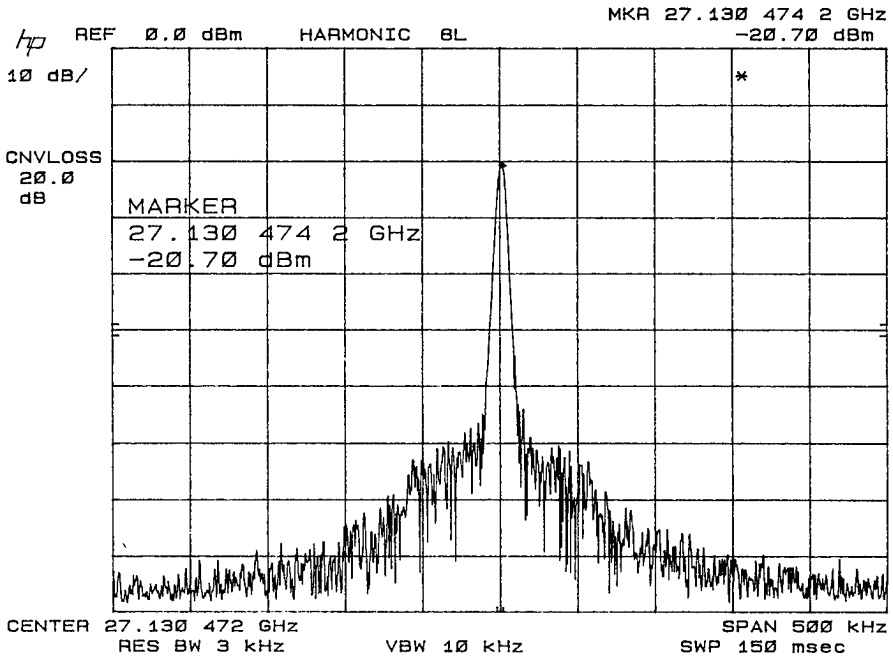


Figure 12. Output power spectrum of HDRO (third harmonic of X-band DRO).

CONCLUSION

Superior phase noise and frequency stability has been achieved for non-temperature compensated free running 25 GHz and 27 GHz HDROs. The design consisted of a high Q dielectric resonator, and a low 1/f noise, two-stage MESFET amplifier. Phase noise levels reported in this paper are 15 dBc/Hz better than any previously reported data. The frequency stability measurements demonstrated an improvement, by a factor of three, over commercially available DROs. It is believed that this is the first Ka-band Harmonic DRO that has been characterized in terms of phase noise from 10 Hz to 100 KHz offset frequencies.

REFERENCES

[1] M. Mizan, T. Higgins, and D. Sturzebecher, "Ultra-Stable, Low Phase Noise Dielectric Resonator Stabilized Oscillators for Military and Commercial Systems", To be published in NASA's Technology 2002 Conference digest of papers, Baltimore, MD, December 1-3, 1992.

[2] M. J. Loboda, T. E. Parker and G. K.Montress, "Frequency Stability of L-Band, Two Port Dielectric Resonator Oscillator," IEEE Transactions on MTT, Vol. MTT35, No. 12, Dec. 1987, pp. 1334-1339.

[3] A.P.S. Khanna and E. Topacio, "A Highly-Stable 36 GHz GaAs FET DRO with Phase-Lock Capability", Microwave Journal, July 1989, pp. 117-122.

[4] Koichi Uzawa and Kazutoshi Matsumoto, "Low Noise Microwave Oscillator Using Ultra High Q Dielectric Resonator", 1991 IEEE MTT-S Digest, pp. 835-838. IEEE Catalog No. CH2870-4/91.

[5] A.P.S. Khanna, Carl Creamer Ed Topacio, and Edmar Camargo, "A 45 GHz Highly Stable +15 dBm Low Noise GaAs MESFET Source Using a GaAs PHEMT as Frequency Doubler", Microwave Journal, November 1991, pp. 117-127.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A TEMPERATURE COMPENSATED 1 GHz STW BASED MULTIFREQUENCY OSCILLATOR

Marian Angelov Taslakov,

Institute of Electronics,
72 Tzarigradsko Chaussee, 1784 Sofia, Bulgaria

Abstract

Temperature compensation of a 1 GHz VCO, stabilized with a SSBW/STW combined mode resonator filter has been performed at 128 frequencies. Analogue and digital compensation techniques have been used. The analogue compensation is common for all frequencies and the digital one is individual. The overall instability for each of the frequencies is ± 2 ppm over the temperature range of -30...+60 C.

Introduction

Temperature compensation (TC) techniques using analogue and digital methods, have been widely used in crystal and surface acoustic wave oscillators to improve their temperature stability /1/,/2/. In 1991 we proposed a novel combined TC method that was successfully applied to a 1 GHz oscillator stabilized with a surface skimming bulk wave / surface transverse wave (SSBW/STW) combined mode resonator filter (CMRF)/3/, /4/, /5/. According to this method the temperature dependence is partly compensated for by an analogue circuit. Precision compensation is performed by a digital circuit. The overall instability achieved was ±1.4 ppm in the temperature range (-45...+75) deg C. Now we use a similar method for combined TC of a voltage controlled oscillator (VCO), stabilized with a CMRF. The VCO has a tuning range of a 1.2MHz which makes TC at 128 discrete frequencies with a step of 6.25kHz within this tuning range possible.

Circuit Description and compensation principle

The block diagram of the uncompensated, CMRF stabilized VCO is shown in Fig. 1. It consists of two amplifiers A1 and A2 with a phase shifter between them and a CMRF in the feedback loop. A3 is a buffer amplifier. The oscillator has a tuning range of 1.2 MHz at room temperature for a control voltage range from 0.4 to 12.4 V and an output power of 0 dBm. The CMRF is mounted on a precise Pt resistive temperature sensor which senses the crystal temperature and converts it into a temperature proportional voltage used by the compensation circuit.

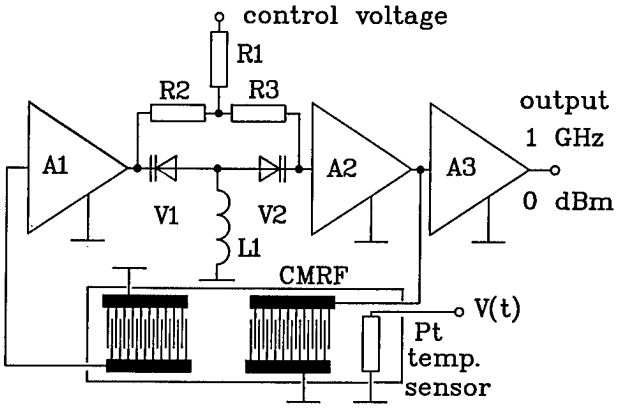


Fig.1 Block diagram of the uncompensated CMRF based VCO with a tuning range of 1200ppm.

Fig. 2 shows the temperature dependence of the VCO frequency for 33 values of the control voltage with a step of 0.4 V. The curves are all different because of nonlinearities in the CMRF phase response, the VCO tuning characteristic and their temperature dependences, and have a close to the parabolic shape.

0-7803-0905-7/93 $3.00 © 1993 IEEE

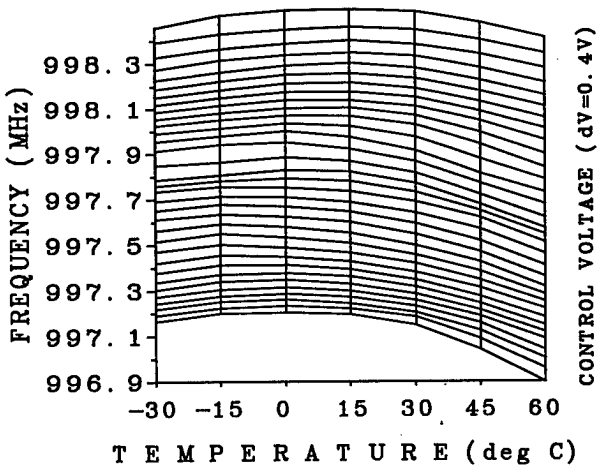


Fig.2 Frequency dependence of the VCO on temperature for different control voltages with a step of 0.4 V.

Fig. 3 shows the frequency as a function of two arguments: temperature and control voltage. When cutting this surface with horizontal planes, corresponding to constant frequencies, curves, giving the dependencies between the control voltage and the temperature, necessary for achieving a temperature independent frequency, are obtained. They show how the control voltage in a function of temperature must change to keep the frequency constant. This dependences are shown in Fig. 4 for 11 frequencies with a step of 100 kHz. These dependences are synthesized by the TC circuit in Fig. 5. The temperature proportional voltage from the Pt

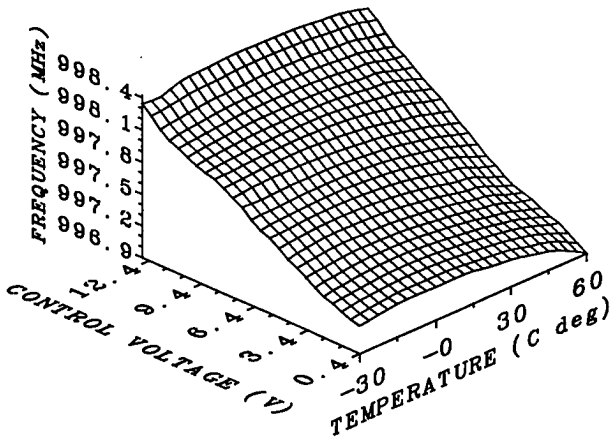


Fig.3 Frequency as a function of two arguments-temperature and control voltage.

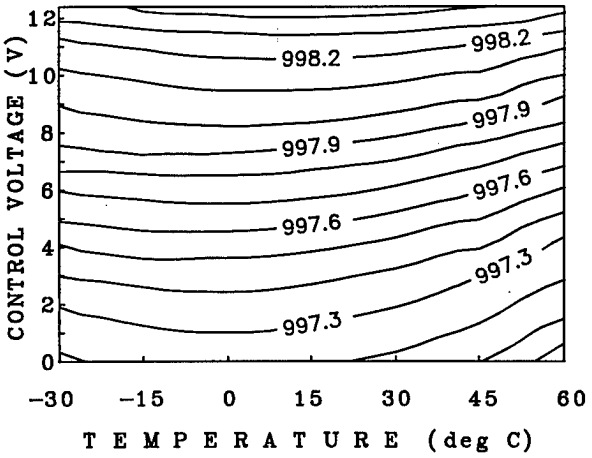


Fig.4 These control voltage / temperature dependencies have to be synthesized by the TC circuit in order to obtain temperature independent discrete frequencies.

temperature sensor controls the analog compensation circuit (ACC), performing a rough analog compensation, common for all desired frequencies. ACC is performed by an operation amplifier based functional converter having a trapezoidal converting characteristic which approximates the medium of the curves shown in Fig.4. Thus the overall temperature instability is reduced by a factor of 4 to 7, dependent on the individual frequency. Also the slope of the temperature - frequency dependence is reduced, which is important for precise compensation. The above mentioned voltage is also used to control the digital compensation circuit (DCC) which performs a precise individual compensation at each of the desired frequencies. The DCC consists of an analogue to digital convertor (ADC) which converts the temperature proportional voltage into 256 binary codes, each corresponding to one temperature point. They are used to address the EPROM. The DCC serves also as a frequency selection unit (FSU) which allows frequency hopping to any of the desired frequency on a command from the control unit (CU). This information is also used to address the EPROM. Thus the address inputs of the EPROM receive 256 possible values, carrying information on the current crystal temperature for one of 128 selected frequencies. The information for each pair of temperature and selected frequency (Ti,Fj, i=0 to 255, j=0 to 127) is written in one byte of the EPROM. The capacity of the EPROM is 32 KBytes = 256x128. The binary number, which is written in each byte of the EPROM,

contains the information for the VCO tuning voltage so that the selected frequency is independent from temperature. This binary number is converted by the digital to analogue converter (DAC) into a tuning voltage. The ACC and DCC voltages, containing the

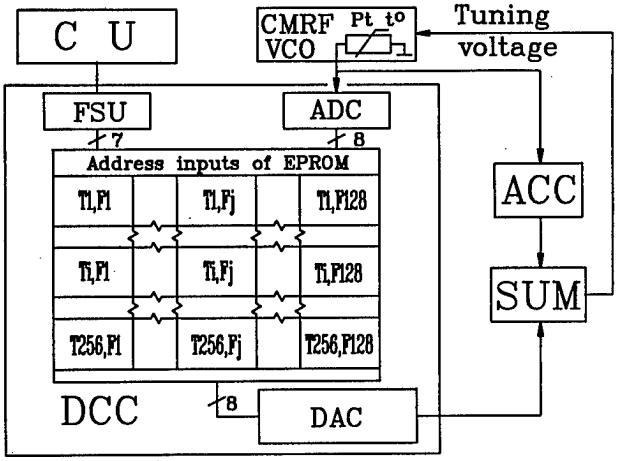


Fig.5 Block diagram of the TC circuit.

information on the desired frequency and the temperature compensation, are summed by a summer (SUM) and applied to the control input of the VCO.

Fig. 6. illustrates the temperature instabilities of the VCO for part of the frequencies after the common analog compensation. These temperature dependences also differ from each other but the overall instability after this rough compensation is significantly reduced compared to the data from Fig. 2. The differences between the curves in Fig. 6 make a precise individual compensation at each of the desired goal frequencies necessary. This is performed by the DCC. The compensation principle for precise TC is as follows: The VCO and TC circuit are placed into a thermostat and the temperature is swept in the desired operating temperature range, in our case (-30 ...60) deg C. The control voltage is changed in such manner that one of the desired frequencies is kept constant. This change is performed by the DAC controlled by a computer. For this reason the connection between the EPROM and the DAC is interrupted. The compensation principle accounts automatically for the temperature dependences and nonlinearities of all elements of the VCO and TC circuits as phase shifter, temperature sensor, ADC and DAC. The data, obtained during the temperature cycle, is written into the EPROM in the bytes corresponding to the

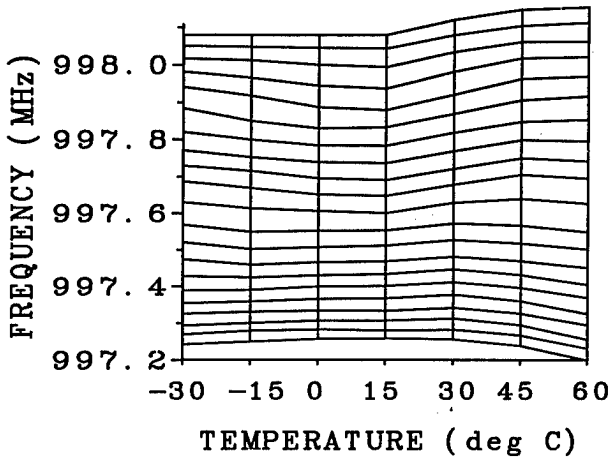


Fig.6 Frequency temperature dependence after the rough TC.

desired frequency. After that the connection between the EPROM and the DAC is recovered.

The most precise way for the individual compensation would require digital compensation at each of the 128 frequencies. In this case a tedious and time consuming calibration procedure would be necessary. To simplify the compensation cycle we decided to make a precise digital compensation only at 8 equally spaced mother frequencies. The compensation data for all intermediate frequencies was calculated by interpolation from the data of two adjacent mother frequencies.

Results

Fig. 7 is the experimental data from the overall temperature instability at five VCO goal frequencies after the compensation. The upper and the lower curves correspond to two mother frequencies while the medium curves correspond to three goal intermediate frequencies. It is seen that the overall instability for all five curves does not exceed ± 2 ppm. The temperature hysteresis of the temperature frequency dependence is about 3 ppm under the following conditions: time of temperature increase: 4 hours, hold at 60 deg C: 20 minutes and time of temperature decrease: 4 hours.

Discussion

The proposed method for combined multifrequency analog and digital compensation is simple and can be applied to different types of VCO with temperature dependencies of any kind.

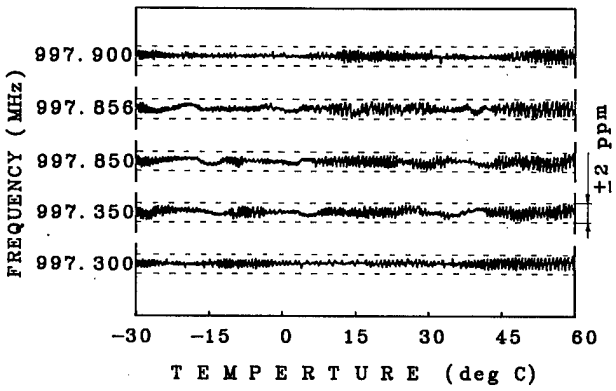


Fig. 7. Overall temperature instability after the individual compensation for five frequencies.

Although the stability of ± 2 ppm achieved is comparable with that of best crystal TCO, in our opinion this is not the limit. Further decrease of the temperature instability can be obtained by increasing the number of bits of the ADC and DAC and the capacity of the EPROM. A better method would be to approximate the curves of Fig. 4 with parts of straight lines. Now we approximate these curves with stepped curves. This causes jumps in the frequency temperature dependence (see Fig. 7). Better approximations can be realized by improved digital - analog compensation circuits, which permit also an automatic cycle of the temperature calibration and removes the frequency jumps in the temperature frequency characteristic.

Conclusions

We demonstrated a simple method for precise temperature compensation of a stable SSBW/STW based VCO for synthesizer applications. The compensation circuitry guarantees zero warm-up time of the oscillator, is readily realized in CMOS technology and has d.c. power consumption of 16 mW from a 5 V source. We feel that the circuit would be competitive in portable communications.

Acknowledgements

The author wishes to thank Dr. Ivan D. Avramov from the Institute of Solid State Physics in Sofia for developing the SSBW/STW VCO used in this study and for the useful discussions.

References

/1/ Kinsman, R.G., "Temperature compensation of crystals with parabolic temperature coefficients", Proc. 36-th Ann. Symp. on Frequency Control, 1982, pp.102-107.

/2/ Cowan, V.D., Slobodnik, A. J., Roberts, J. A., and Silva, J. H.,"A 300 MHz digitally compensated SAW oscillator", IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control, Vol 35, No 3, May 1988, pp. 380-385.

/3/ Avramov, I. D.,"1 GHz low loss coupled resonator filter using surface skimming bulk waves and Blewstein-Gulyaev waves", Electron. Lett., 1991, 5, pp.414-415.

/4/ Avramov, I. D., and Kozlov, A. S., "Design of combined mode resonator filters based on surface transverse and surface skimming bulk waves", Proc. ofthe 10-th European Conference on Circuit Theory and Design ECCTD'91, Lyngby, Denmark, 1991, pp. 729-738.

/5/ M.A. Taslakov, I.D. Avramov, "A Temperature Compensated 1GHz STW Oscillator", Proc. of the 1991 IEEE Ultrasonics Symposium, Lake Buesta Vista, Florida, USA, 1991, pp.275-278.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

HIGH-OVERTONE, BULK ACOUSTIC RESONATOR FREQUENCY STABILITY IMPROVEMENTS

S. P. Caldwell, M. M. Driscoll, S. D. Stansberry
D. S. Bailey, H. L. Salvo**

*Design and Producibility Engineering Division
**Advanced Technology Division
Westinghouse Electronic Systems Group, Baltimore, MD

Abstract

In an effort to counter the effects of both moderate-term and long-term frequency instability associated with (-30PPM/°C) frequency-temperature coefficient in High Overtone, Bulk Acoustic Resonators (HBAR) AFC loop stabilization of low noise, 640MHz HBAR oscillators has been demonstrated. A thermoelectric (TE) cooler is used as the resonator frequency tuning element. The TE cooler approach provides a means of obtaining near-equal device heating and cooling time-response characteristics, allowing stable loop operation. Using this technique, long-term, fractional frequency stability improvement to 5 parts in 10^7 has been demonstrated. A frequency-lock (rather than phase-lock) loop was mechanized in order to more easily cope with the relatively long thermal time constants without resorting to ultra-narrow loop bandwidth.

With regard to improvements in HBAR short-term frequency stability, prototype resonators have been fabricated on thicker (700nsec delay) YAG crystals and evaluated in oscillators operating at both 320MHz and 640MHz that exhibit flicker-of-frequency noise comparable, on a Sy(f) basis, to that reported for ultra-low noise, multiple crystal oscillators operating in the 80-100MHz range [1]. The HBAR oscillator design is somewhat novel in that both HBAR and SAW technology is used, with a conventional quartz SAW resonator included in the sustaining stage circuitry for suppression of adjacent (700KHz spacing) HBAR resonant responses.

Introduction

The Westinghouse High Overtone, Bulk Acoustic Resonator (HBAR) provides the highest loaded Q at microwave frequency of any known acoustic resonator. The HBAR is similar to the traditional quartz crystal resonator, except for the method used for transducing electrical signals into and out of the resonating crystal. As shown in figure 1, thin sputtered ZnO transducers and metal electrodes are deposited on a highly polished, low acoustic loss crystal. The unique nature of this configuration is that the resonant crystal itself can be made from materials or crystal orientations that need not exhibit piezoelectricity. This allows consideration of many crystals for which obtainable resonant Qs are an order of magnitude greater than those of quartz. Included are yttrium aluminum garnet (YAG), lithium niobate and tantalate, sapphire, and others. HBARs designed for longitudinal mode propagation in YAG have exhibited best performance achieved to date in terms of high operating frequency (640MHz), high loaded Q (110,000), high drive level capability (16dBm), low vibration sensitivity (10^{-9} to 10^{-10} per g), and good short-term frequency stability ($Sy(f=100Hz)=5X10^{-26}$) [2,3].

The single most significant drawback associated with the use of these devices is the relatively high-value (-30PPM/°C) frequency-temperature coefficient. In addition, the resonator topology is such that the film transducers tend to isolate the crystal (in an electrical equivalent circuit sense) from the input and output ports. As a result, use of external, variable reactance-frequency tuning is not feasible. Unless used in an extremely temperature-stable environment, the relatively high value of YAG HBAR frequency -temperature coefficient has an adverse effect on resultant HBAR -stabilized oscillator signal moderate-term and long-term frequency stability.

With regard to short-term frequency stability, HBAR loaded Q values are such that the HBAR itself (rather than the oscillator sustaining stage amplifier) is the dominant contributor to oscillator signal flicker-of-frequency noise. This implies that significant additional reductions in HBAR oscillator signal (flicker-of-frequency) noise are possible via attainment of corresponding reductions in resonator short-term frequency instability. Measurement of monolithic film resonator short-term frequency stability suggests the dominant source of flicker-of-frequency noise in the HBAR is associated with the sputtered ZnO transducers [4].

0-7803-0905-7/93 $3.00 © 1993 IEEE

HBAR Oscillator Frequency Stabilization

Since the HBAR is not amenable to reactance-frequency tuning, but has a relatively linear, large-value frequency-temperature coefficient, temperature can be used as a method of controlling HBAR oscillator signal frequency drift. Since the HBAR crystal itself is mounted in an hermetic enclosure along 75% of the entire crystal perimeter, the crystal-to-case thermal resistance and time constant are smaller than those associated with temperature control of conventional (i.e., quartz crystal) bulk wave resonators.

Figure 2 shows a simplified block diagram for AFC loop stabilization of an HBAR oscillator using a thermoelectric cooler as the frequency tuning element. A more detailed diagram for the AFC loop components is shown in figure 3. As shown in the figure, the loop frequency detector is a pulse-count type discriminator that employs a custom, I/Q detection integrated circuit. The time response characteristics of the HBAR/TE cooler can be characterized by a cascade of two effective thermal time constants: one representing the TE cooler-to-HBAR package interface (τ =3 to 5 seconds) and one representing the HBAR package-to-YAG crystal interface (τ =10 to 15 seconds). The loop integrator feedback network "corner" frequency was selected to match the longer thermal time constant, and overall loop gain adjusted for a nominal AFC loop bandwidth on the order of .01Hz. A 5 watt, stereo power amplifier IC was used to drive the 3.5 watt, TE cooler module cemented to the sealed HBAR package cover. A block diagram of the oscillator sustaining stage is shown in figure 4.

Using the arrangement of figure 2, HBAR oscillator frequency stability was improved to $5X10^{-7}$ in the operational (standard cabinet rack-mount in engineering laboratory) environment.

Figure 5 shows the phase noise measurement results obtained for two, phaselocked, AFC-stabilized HBAR oscillators operating at 640MHz. In order to make the measurement, the AFC feedback loop in one of the oscillators was temporarily disabled, with oscillator fine frequency tune capability (for phaselock) obtained by including a voltage controlled phase shifter in the oscillator sustaining stage circuit. The data is consistent with that previously obtained for non-stabilized HBAR oscillators and indicates no spectral degradation due to inclusion of the AFC feedback control loop.

Long Delay HBAR Short-Term Frequency Stability

Oscillator resonator loaded Q value may be considered a measure of the degree to which the conversion of sustaining stage circuit open loop, phase noise to oscillator closed loop, frequency noise is minimized. The extremely high values of loaded Q exhibited by the HBAR are such that oscillator signal flicker-of-frequency noise is almost entirely due to short-term frequency instability in the HBAR itself. Therefore, additional reductions in HBAR flicker-of-frequency instability would result in corresponding spectral improvements in the HBAR oscillator output signal. As mentioned earlier, there is reason to believe that the ZnO film transducers are the dominant contributors to HBAR self-noise.

With this in mind, a prototype HBAR was fabricated using a YAG crystal 3.5 times thicker than that previously utilized. In the longer delay HBARs, the influence of the films on HBAR resonant frequency is substantially reduced. An existing ZnO transducer mask set, optimized for use with shorter delay HBAR crystals, was used for the experiment. Use of this mask with the longer delay crystal resulted in an acoustic mismatch and correspondingly high (25dB) device insertion loss. Mask redesign will lower insertion loss to acceptably low values in the range 10 to 15dB.

As shown in figure 6, loaded Q values of 175,000 and 225,000 were obtained for the prototype resonator at 640 and 320MHz, respectively. The resonator was evaluated in oscillators operating at 320 and 640MHz. Two 50-ohm modular amplifiers were used to provide requisite sustaining stage gain. Conventional, L-C tuned circuits could not provide adequate circuit selectivity in the oscillator for rejection of adjacent HBAR resonances, occurring every 700KHz. Therefore, a surface acoustic wave resonator (SAWR) was used for this purpose. The SAWR was highly loaded in order to insure that the effect of short-term frequency instability in the SAWR itself would be negligible. The oscillator block diagram is shown in figure 7.

Measurements of HBAR oscillator output signal phase noise at 320 and 640MHz were made via comparison to a low noise, reference signal generator employing a four-crystal, 80MHz oscillator and cascaded frequency doublers, including 320MHz SAWR bandpass filters [1]. The results of phase noise measurements made for a pair of near-identical, reference generators are shown in figure 8. As shown in the figure, the flicker-of-frequency noise for the two reference generators is characterized by a phase noise sideband level at 640MHz of:

$$£(100Hz) = -120dBc/Hz \quad (1)$$

At 320MHz then:

$$£(100Hz) = -126dBc/Hz \quad (2)$$

The results of phase noise measurements made at 640MHz and 320MHz by phase locking a reference generator to the (long delay) HBAR oscillator, are shown in figures 9 and 10. As shown in figure 9, a flicker-of-frequency noise level of £(100Hz)= -120dBc/Hz was measured. This indicates that HBAR

oscillator flicker-of-frequency performance is near-identical to that of a single reference generator and is on the order of £(100Hz)=-123dBc/Hz. This represents a 3-6dB improvement, compared to the best average results previously obtained with shorter delay HBARs. Since the HBAR oscillator operates directly at 640MHz, the implication of this data is that, compared to the 80MHz oscillator-based source, use of the long delay HBARs can provide improved performance in the region fm>1KHz. In this region, the HBAR oscillator phase noise sideband characteristic is still primarily flicker-of-frequency (30dB/decade), while that of the frequency-multiplied crystal oscillator-based source is "bottoming out."

The results of phase noise measurements conducted for HBAR oscillator operation at 320MHz (figure 10) indicate slightly better flicker-of-frequency noise performance, characterized by £(100Hz)= -127dBc/Hz. It must be appreciated that the data in figures 9 and 10 is based on measurement of a single, available device and does not necessarily represent best attainable (lowest noise) performance. More definitive determination of long delay HBAR performance is pending in-progress fabrication of larger numbers of devices, including use of a new transducer design for lower resonator insertion loss.

Conclusions

AFC loop stabilization of HBAR oscillators to stabilities on the order of $5X10^{-7}$ is possible, using temperature-frequency tuning.

Use of a thermoelectric cooler provides a mechanism for achieving near-equal heating and cooling thermal time constant, aiding in loop stability margin.

State-of-the-art flicker-of-frequency noise performance has been demonstrated using long delay HBARs in an oscillator circuit incorporating dual use of HBAR and SAWR technology.

Use of long-delay HBARs shows promise as a means of achieving further reduction in device, and hence, oscillator signal flicker-of-frequency noise.

References

[1] M. M. Driscoll, "Reduction of Quartz Crystal Oscillator Flicker-of-Frequency and White Phase Noise (Floor) Levels and Acceleration Sensitivity Via Use of Multiple Resonators," Proc. 1992 IEEE Freq. Contr. Symp., May 1992, pp. 334-339.

[2] M. M. Driscoll, et al., "Extremely Low Phase Noise UHF Oscillators Utilizing High-Overtone, Bulk-Acoustic Resonators," IEEE Trans. UFFC, Vol. 39, No. 6, Nov. 1992, pp. 774-779.

[3] D. S. Bailey, et al., "Frequency Stability of High-Overtone, Bulk-Acoustic Resonators," IEEE Trans. UFFC, Vol. 39, No. 6, Nov. 1992, pp. 780-784.

[4] M. M. Driscoll, et al., "Design and Evaluation of UHF Monolithic Film Resonator-Stabilized Oscillators and Bandpass Filters," Proc. 1987 IEEE MTT-S Symp., June 1987, Vol. II, No. V-4, pp. 801-804.

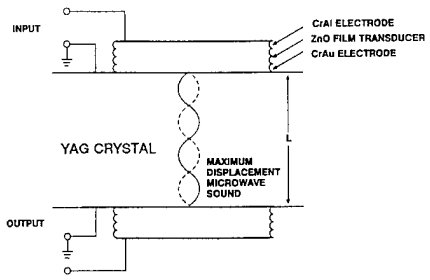


Figure 1. High Overtone, Bulk-Acoustic Resonator

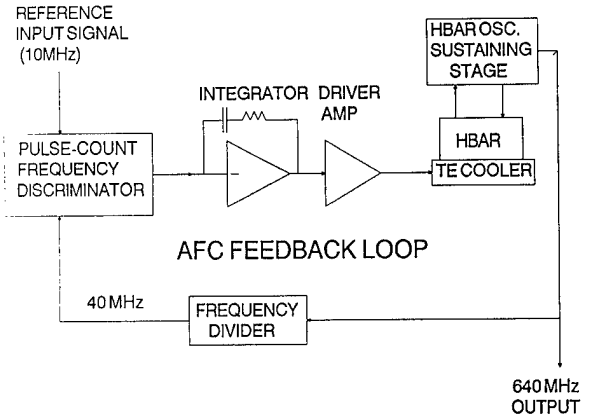


Figure 2. Simplified Block Diagram of AFC-Stabilized HBAR Oscillator

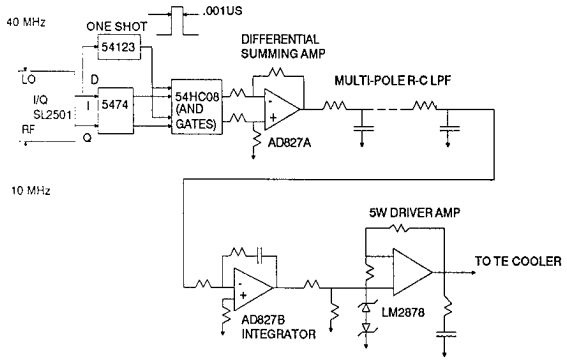


Figure 3. AFC Loop Component Block Diagram

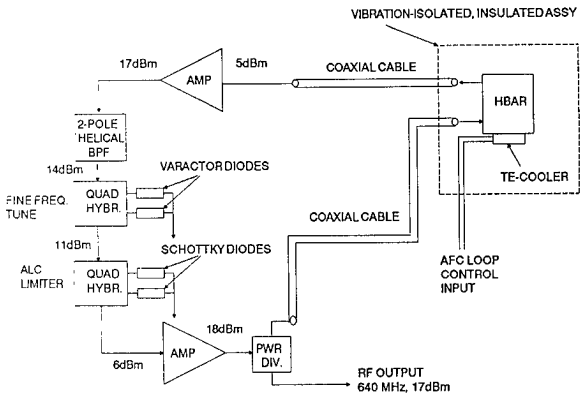


Figure 4. TE-Cooled HBAR Oscillator Block Diagram

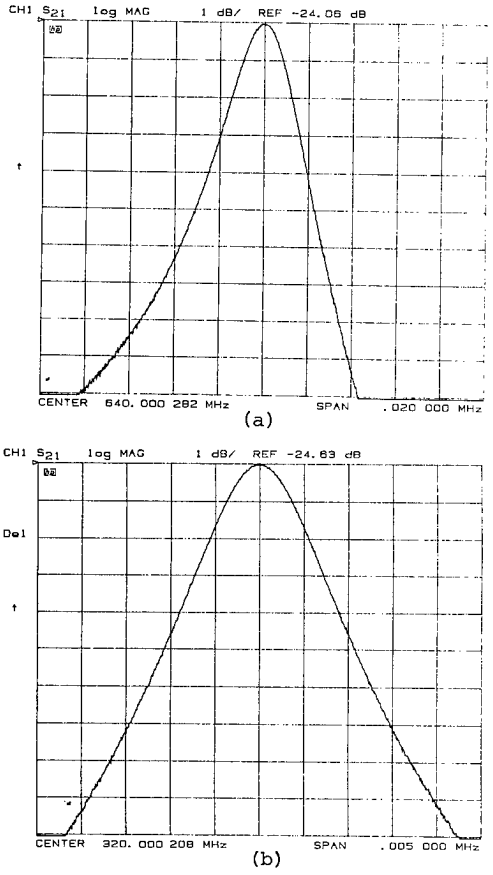


Figure 6. Measured Transmission Response for Long Delay HBAR
(a) 640MHz (b) 320MHz

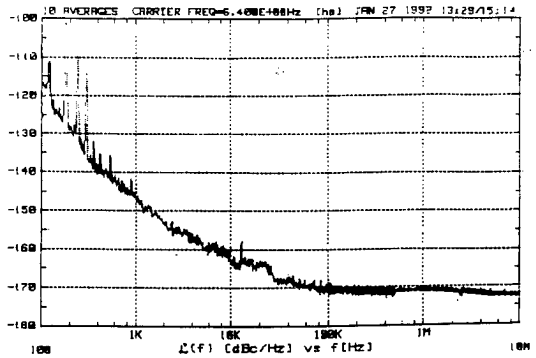


Figure 5. Measured Phase Noise for Two, AFC-Stabilized HBAR Oscillators

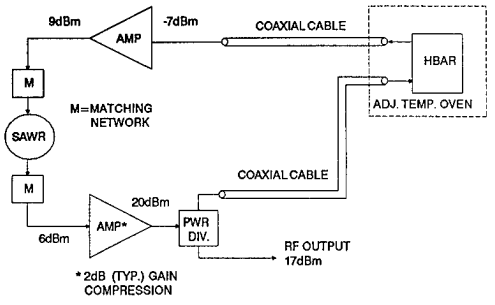


Figure 7. Long Delay HBAR Oscillator Block Diagram

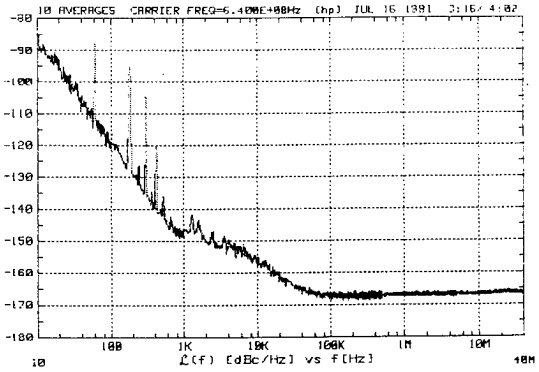


Figure 8. Measured Phase Noise at 640MHz for Two, 80MHz Crystal Oscillator-Based, Reference Signal Generators

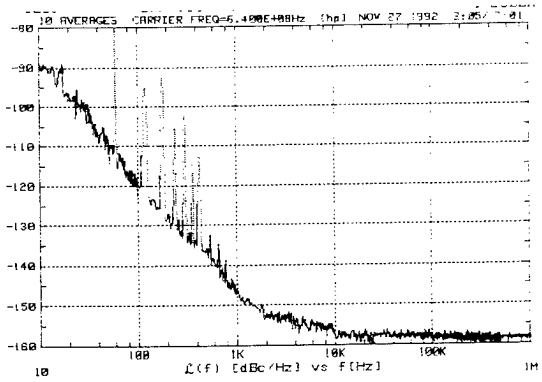


Figure 9. Measured Phase Noise for Reference Signal Generator Locked to Long Delay HBAR Oscillator at 640MHz

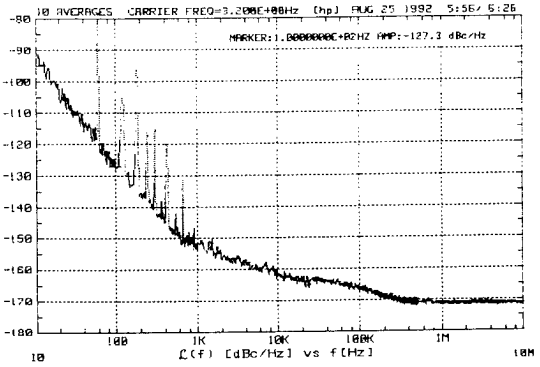


Figure 10. Measured Phase Noise for Reference Signal Generator Locked to Long Delay HBAR Oscillator at 320MHz

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

HIGH-Q TE STABILIZED SAPPHIRE MICROWAVE RESONATORS FOR LOW NOISE APPLICATIONS

*M.E. Tobar, *A.J. Giles, †S. Edwards and †J. Searls

*Department of Physics, University of Western Australia, Nedlands, WA, 6009, Australia.
†Poseiden Scientific Instruments Pty. Ltd., 14 Mouat St. Fremantle, 6160, WA, 6160, Australia.

Abstract: Two low noise high-Q Sapphire Loaded Cavity (SLC) resonators, with unloaded Q values of 2×10^5 and very low densities of spurious modes have been constructed. They were designed to operate at 0 degrees Celsius with a centre frequency of 10.000000 GHz. The cavity was cooled with a Thermo-Electric (TE) Peltier element and in practice achieved the required centre frequency near 1 degree Celsius. The resonator has a measured frequency-temperature coefficient of -0.7MHz/K, and a Q-factor which is measured to be proportional to $T^{-2.5}$. An upper limit to the SLC residual phase noise of $\mathcal{L}$(100Hz)=-147.5 dBc/Hz, $\mathcal{L}$(1kHz)=-155 dBc/Hz and $\mathcal{L}$(10kHz)=-160 dBc/Hz has been measured. Also we have created a free running loop oscillator based on the SLC resonator and measured its frequency noise using the other as a discriminator.

Introduction

The performance of modern communication systems is determined by the spectral purity of the local oscillators in the microwave transmitters and receivers. To construct state of the art oscillators it is necessary to develop high-Q resonators for use as stabilizing or filtering elements. Besides a high-Q, a resonator must also have a low value of residual self noise.

The dielectric material with the lowest known loss tangent at microwave frequencies is monocrystalline Al_2O_3 (sapphire). Over the last ten years various groups have developed sapphire dielectric resonators with extremely high Q values at room and cryogenic temperatures [1-9]. This paper describes the properties of two low noise high-Q SLC resonators which have been jointly designed by the University of Western Australia and Poseiden Scientific Instruments Pty. Ltd.

By choosing the appropriate mode the SLC resonators have been designed to have the maximum spurious mode separation possible without the cavity walls degrading the Q values below the limit imposed by the sapphire dielectric loss tangent [10]. High Q values of 2×10^5 are achieved without the need for cryogenics making these cavities very portable and robust, and allowing them to be fitted into any present scheme including mobile applications. The reproducibility of our design method is highlighted by our ability to build two identical cavities with prespecified and matched resonant frequencies.

Temperature Effects

The SLC resonators were both constructed to operate with a centre frequency of 10.00-0 GHz to 6 decimal places. By choosing the appropriate sapphire and cavity dimensions they have been designed to operate at the prescribed centre frequency at 0 degrees Celsius. In practice our cavities have obtained this centre frequency near 1 degree Celsius. The frequency coefficient with respect to temperature of these cavities is -70 ppm/K, and hence a temperature stability of about 1 mK is achieved to maintain the accuracy of the centre frequency. External control of the centre frequency is possible over a range of ±100 kHz via a voltage control input to the temperature controller. This control allows the resonator's centre frequency to be controlled for the purpose of long term frequency stability as required in such applications as multiple filters and stable oscillators.

SLC Q-factor and Frequency Dependence

The Q-factor and frequency dependence on temperature were measured via the circuit shown in figure 1.

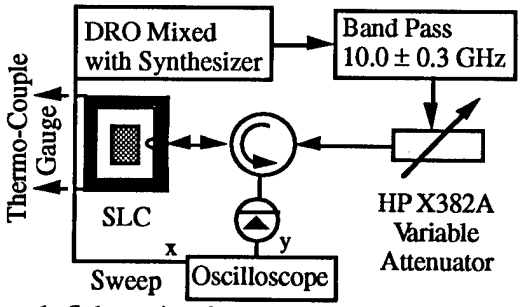


Figure 1. Schematic of the experimental set up used to measure the resonant frequency, loaded Q-factor and electrical coupling as a function of temperature.

0-7803-0905-7/93 $3.00 © 1993 IEEE

For these measurements the SLC resonator was configured as a one port device. Two thermo-couples were placed in good thermal contact with the SLC resonator which was placed inside a refrigerator capable of cooling to -17 degrees Celsius. The temperature was monitored by a thermo-couple gauge. To measure the coupling, Q-factor and resonant frequency we used the slow sweep method and traced out on the oscilloscope the magnitude squared of the reflection coefficient as a function of frequency. If the incident source is tuned off resonance no power will enter the resonator, and the maximum possible power (P_I) will be reflected. To measure the resonant frequency (f_0), the frequency of the minimum point of reflection (maximum absorption) was measured. From the reflected power on (P_R) and off resonance (P_I) the coupling can be calculated by;

$$\beta = \frac{1-\sqrt{P_R/P_I}}{1+\sqrt{P_R/P_I}} \qquad (1).$$

To measure the loaded bandwidth the frequency span (Δf_L) where the reflected power that was less than $(P_I+P_R)/2$ was measured. The unloaded Q-factor can then be calculated by;

$$Q_0 = \frac{f_0(1+\beta)}{\Delta f_L} \qquad (2).$$

To obtain accurate measurements for the unloaded Q-factor we must make accurate measurements of the coupling. A HP X382A precision attenuator was placed in front of the swept frequency source. To obtain an accurate measurement of the ratio P_R/P_I the level of the off resonance reflected power was reduced by increasing the attenuation until it was equal to the on resonance reflected power. This enabled us to measure P_R/P_I by directly reading the amount of attenuation added. Frequency and unloaded-Q factor measurements as a function of temperature are shown below. Throughout these measurements the ratio of β/Qo remained constant as a function of temperature and was equal to about 2×10^{-6}.

Braginsky et al [11] measured the loss tangent of sapphire as a function of temperature at cryogenic temperatures. They found that in the temperature range below 200 K the temperature dependence of the Q factor could be written as;

$$Q_0 \propto T^{-4.75} \qquad (3).$$

They did not show any clear results above this temperature. By fitting a power law curve to the data (given as the continuous line in fig. 3) we determined that temperature dependence of the Q-factor can be written as;

$$Q_0 = 3.2\times10^{11}\ T^{-2.5} \qquad (4).$$

Thus the power law dependence at room temperature is significantly smaller than the power law dependence at cryogenic temperatures.

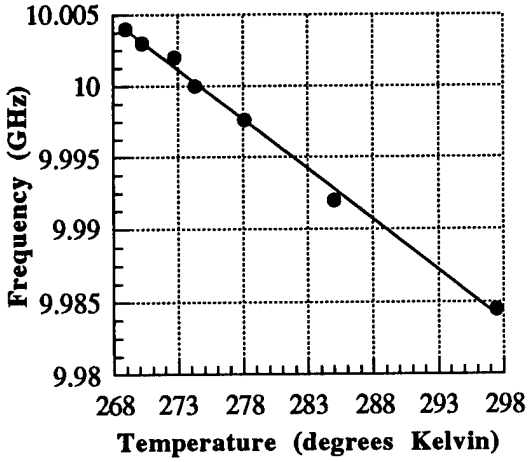


Figure 2. Measured frequency versus temperature for the SLC resonator. The frequency dependence is due to the dielectric constant changing as a function of temperature. The line of best fit gives a frequency-temperature coefficient of -0.7 MHz/K

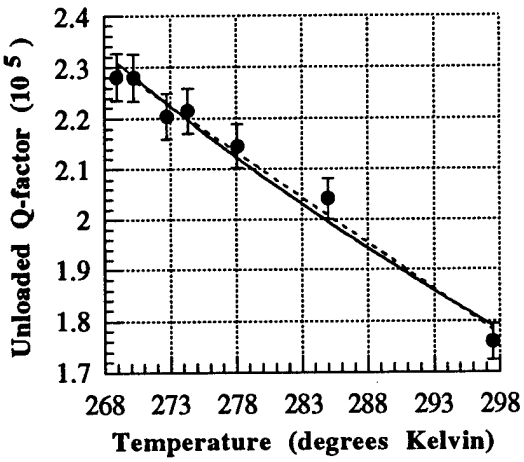


Figure 3. Unloaded Q-factor versus temperature. The dashed line is the best linear fit to the data and the continuous line is the best power law fit to the data. The line of best fit gives a Q-factor temperature dependence of $Q_0 = 7.16\times10^5 - 1.81\times10^3$ T.

Spurious Mode Suppression

To obtain a high Q in a sapphire loaded cavity, "whispering gallery-like" modes in the sapphire crystal must be used to limit the degradation of the Q by the cavity walls. This is necessary due to the relatively low

dielectric constant of about ten for sapphire. However when operating in this regime the density and coupling to spurious modes will be increased. This is undesirable since a spurious mode can degrade the resonators performance by providing off resonance feed through as well as translating their undesired frequency sensitivity and Q values to the operational mode [12,13].

Methods for suppressing undesired modes for conventional cavity resonators for high power applications have been presented previously [14]. They placed microwave absorbers in the cavity walls that intentionally did not affect the operational mode. Our method was achieved without the need for microwave absorbers. The cavity was made non-radiating with respect to the operational mode and radiating with respect to most spurious modes. This modified cavity in which the sapphire sits reduces the Q and coupling values of the spurious modes dramatically. Figs. 4-6 compare the mode density over X-band when the sapphire crystal was in an open environment, in a normal closed cavity and in the special spurious mode suppression cavity respectively.

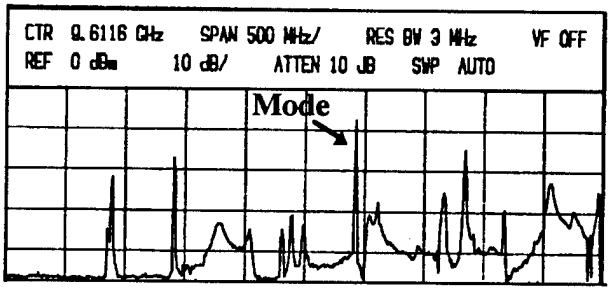


Figure 4. Transmission properties for the open sapphire resonator from 7 to 12 GHz.

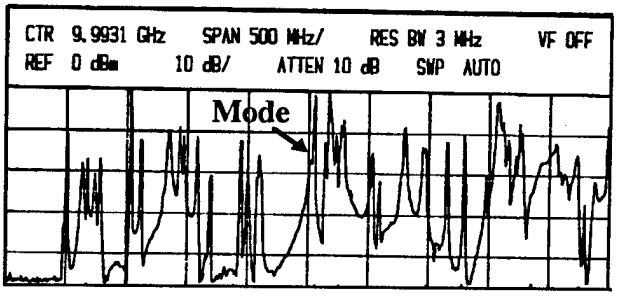


Figure 5. Transmission properties for the SLC resonator placed in a conventional cavity, from 7.4 to 12.4 GHz.

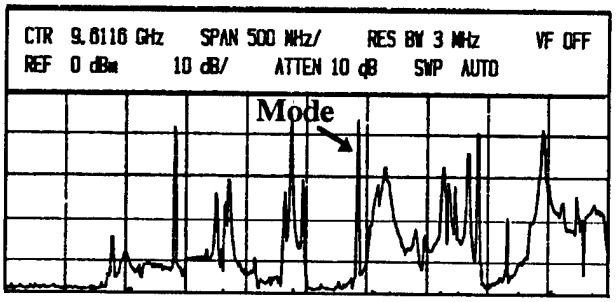


Figure 6. Transmission properties for the SLC resonator placed in the spurious mode suppression cavity, from 7 to 12 GHz.

Figure 5 shows that this resonator would be very difficult to operate with the sapphire placed in a conventional cavity. A spurious mode was seen to be interacting directly with the operational mode, severely affecting its Q value and the level of off resonance feed through. In the special spurious mode suppression cavity, the spurious mode interaction is eliminated, with the off resonance feed through reduced to 40 dB below resonance without affecting the Q value of the resonator. The spurious mode density was not as good as the open dielectric case, however we gain the benefit of eliminating radiation losses and gaining 20% more Q. The cavity is also necessary so we can design the right resonator frequency to the pre-specified accuracy when cooled to 0 Celsius.

Cavity Residual Noise Measurements

Residual phase noise measurements of the SLC resonators were performed using a phase sensitive bridge circuit as shown in fig. 7. Results are shown in figs. 8 and 9. This type of phase noise measurement is limited by the driving oscillator and mixer phase noise. In fact if the oscillator phase noise is large enough, this same type of system may be used to measure the frequency or phase noise of the driving oscillator. This is possible because a resonator is a dispersive element and will decorrelate some of the oscillator's frequency noise in its arm relative to the other arm of the bridge circuit. The decorrelated component will not cancel in the mixer and will be sensed by the measuring system.

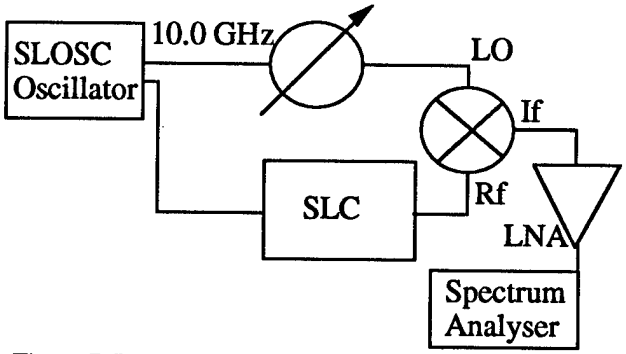


Figure 7. Resonator residual phase noise measurement.

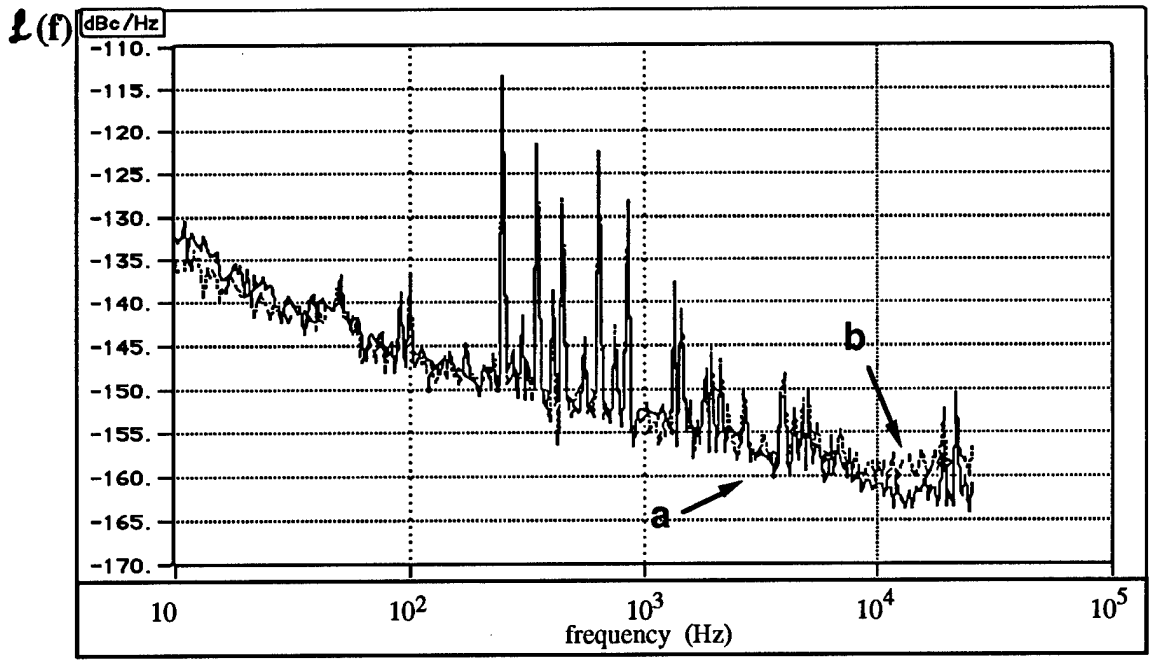


Figure 8. Residual phase noise measurement of SLC1. Measurement **a** (bold line) is the mixer noise floor and **b** is the residual noise measurement, there is little difference between the two indicating this is an upper limit to the resonator noise. Above 10 kHz the residual measurement is above the noise floor, due to dispersion of the HP-8662A synthesizer phase noise. The level here is consistent with its measured phase noise of about -140 dBc/Hz at 10 kHz.

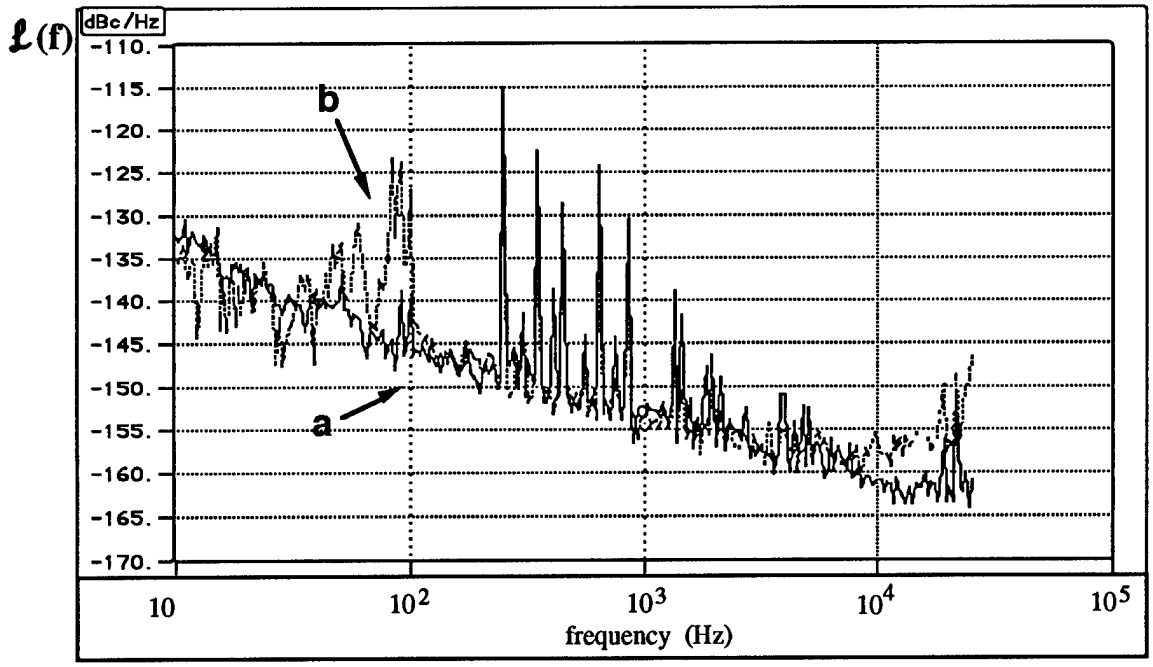


Figure 9. Residual phase noise measurement of SLC2. Measurement **a** (bold line) is the mixer noise floor and **b** is the residual noise measurement, there is little difference between the two indicating this is an upper limit to the resonator noise. Between 50 to 100 Hz some excess noise due to an insecure cavity probe existed and was sensitive to vibrations. Above 10 kHz the residual measurement is above the noise floor, due to dispersion of the HP-8662A synthesizer phase noise. The level here is consistent with its measured phase noise of about -140 dBc/Hz at 10 kHz.

To reduce the dependency on the oscillator noise a second matched cavity may be included in the other arm so near equal dispersion will occur in both arms. Thus the oscillator noise will be nearly correlated at the LO and Rf ports of the double balanced mixer, and will be suppressed. A measurement of this type was recently reported when measuring the residual noise in High-overtone Bulk-Acoustic Resonators (HBAR) [15]. This measurement detects the combined uncorrelated residual phase noise of the two resonators, and will not detect any correlated noise between the two resonators. Also it relies on a good matching of Q values. A 10% match will create a 20 dB suppression of oscillator noise. We originally developed two SLC resonators with the intention of implementing this measurement if necessary.

Due to the development of the low phase noise Sapphire Loaded Superconducting Cavity (SLOSC) oscillator, it was not necessary to implement the two resonator measurement. The phase noise of a 9.7 GHz SLOSC oscillator has been measured to be $\mathcal{L}(f_m)$=-140 dBc/Hz at 1 kHz [7]. To drive the residual noise measurement a HP-8662A 0-1.2 GHz synthesizer was mixed with the SLOSC oscillator, to create a 10 GHz signal that could be tuned, swept and modulated with a calibration signal.

The spectral density of residual phase fluctuations in the SLC resonator was determined by;

$$S_{\phi r}(f_m) = \frac{S_V(f_m)}{k_\theta^2\, G} \tag{5}$$

Here $S_V(f_m)$ [V^2/Hz] is the measured voltage spectral density, k_θ [V/rad] is the mixer phase sensitivity and G is the power gain of the low noise amplifier after the mixer. If the oscillator driving the measurement is the limiting component, then the equivalent detected oscillator spectral density of fractional frequency fluctuations is given by;

$$S_y(f_m) = \frac{S_V(f_m)}{f_0^2\, k_f^2\, G} \tag{6}$$

Here k_f [V/Hz] is the mixer frequency sensitivity and f_0 is the oscillator frequency. If the oscillator is incident on the cavity resonance then the rate of change of frequency with respect to phase is given by;

$$\frac{df}{d\theta} = \frac{f_0}{2\,Q_L} \tag{7}$$

Thus phase and frequency sensitivity may be related by;

$$k_\theta = \frac{f_0}{2\,Q_L}\, k_f \tag{8}$$

Thus combining (5), (6) and (8) the detected noise with respect to the cavity residual phase noise may be expressed in terms of the fractional frequency noise with respect to the driving oscillator;

$$S_{\phi r}(f_m) = 4\,Q_L^2\, S_y(f_m) \tag{9}$$

The double sideband oscillator phase noise is related to the fractional frequency noise by;

$$S_\phi(f_m) = (f_0/f_m)^2\, S_y(f) \tag{10}$$

Therefore the phase noise of the oscillator may be referred to the residual phase noise of the cavity by;

$$S_\phi(f_m) = \left(\frac{f_0}{2\,Q_L\, f_m}\right)^2 S_{\phi r}(f_m) \tag{11}$$

Thus if the phase noise of the driving oscillator is known it may be determined whether it is the limiting factor in the phase noise measurement.

The residual noise measurements shown in figs. 8 and 9 determine an upper limit to the residual phase noise of the SLC resonator. Between 10 to 10 kHz the measurement was dominated by the mixer noise floor. Above 10 kHz the phase noise of the HP-8662A synthesizer was measured. In both cases the magnitude of the synthesizer can be determined from (11) to be about -140 dBc/Hz which is close to the measured value [8].

The probes in SLC2 were not set rigidly in the cavity. This lead to this resonator exhibiting some excess residual phase noise between 50 and 100 Hz.

Oscillator Phase and Frequency Noise Measurements

High-Q resonators may be used as frequency discriminators to measure oscillator frequency noise. In this section we report on the measured phase noise of a commercial phased locked, multiplied cavity oscillator and a free running SLC loop oscillator.

A MITEQ PLM-5-9040-0-15P oscillator multiplied to 9040 MHz, was mixed with a HP-8662A synthesizer to create a 10 GHz signal. Then in the same way as shown in figure 7 we measured the voltage spectrum at the

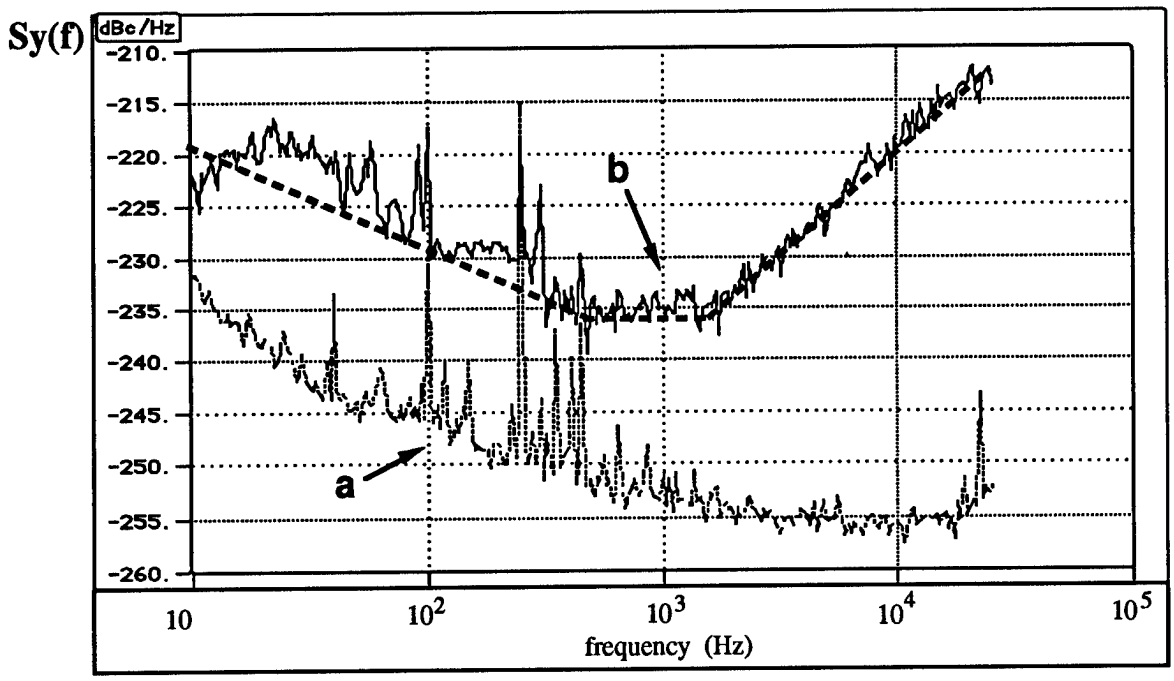


Figure 10. Frequency noise of a MITEQ phase locked cavity oscillator using SLC1 ($Q_L = 10^5$) as a discriminator, shown by measurement **b**. Measurement **a** shows the noise floor.

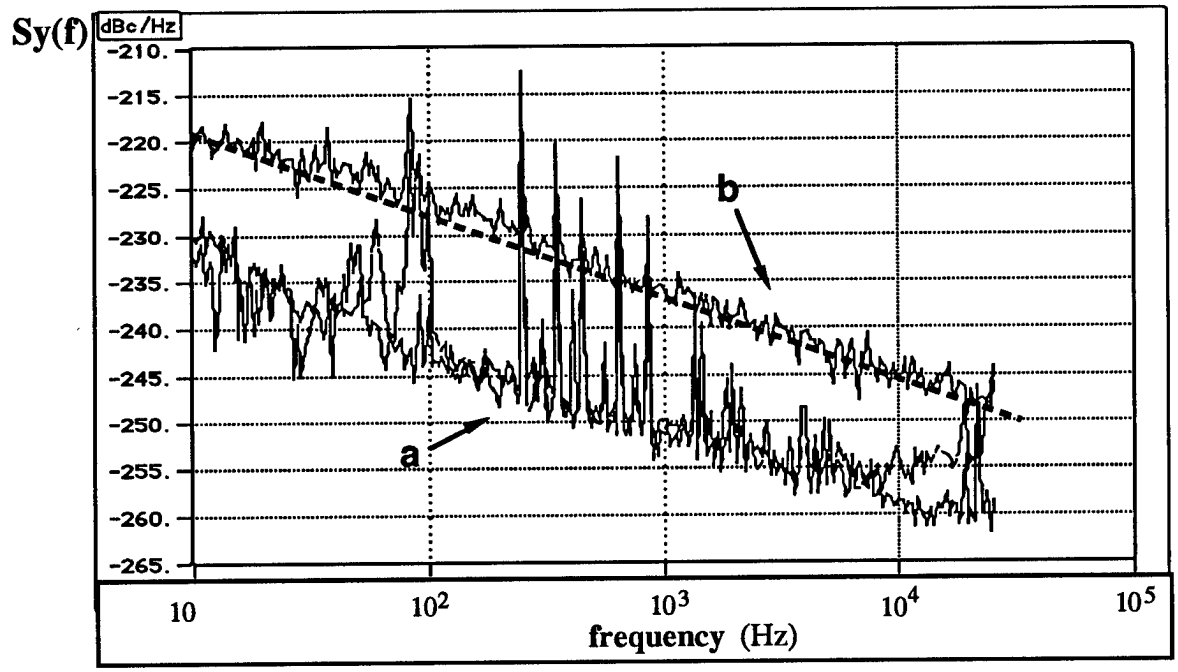


Figure 11. Frequency noise of the free running loop oscillator shown in figure 12, using SLC2 ($Q_L = 7.5\ 10^4$) as a discriminator, shown by measurement **b**. Measurement **a** shows the noise floor as measured in fig. 9 and referred from the residual phase noise noise measurement to oscillator frequency noise via equation (8). Generally the measurement is 15 dB above the noise floor. Below 100 Hz the probe vibrational effects of SLC2 can be seen in the noise spectrum.

output of the low noise amplifier. Fig. 10 shows the measured fractional frequency spectral density of the oscillator. We have determined that this oscillator has a value of $S_y(f)$=-235 dBc/Hz at 1 kHz. This corresponds to an $\mathcal{L}(f)$=-98 dBc/Hz at 1 kHz.

A simple free running oscillator was created using SLC1, a 9.65 to 10.25 GHz band pass filter and a T-12034 Avantek amplifier. A schematic is shown in fig. 12.

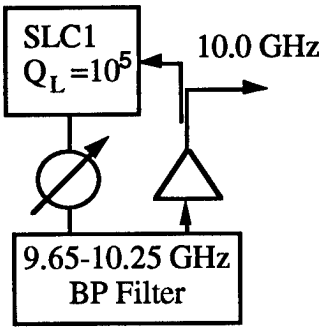


Figure 12. Free running loop oscillator using SLC1 as the frequency determining element

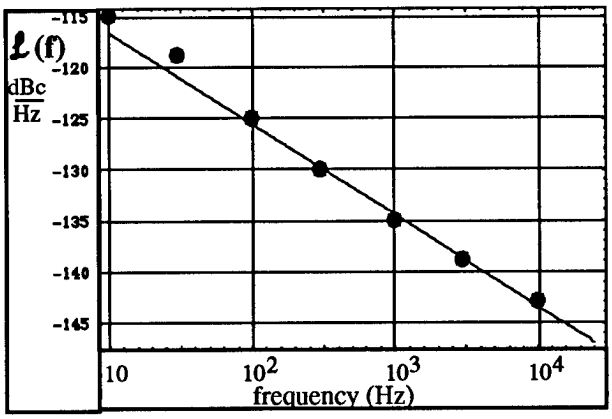


Figure 13. The continuous line is the inferred amplifier noise of the Avantek T-12034 from the SLC oscillator measurements. The points are independently measured values of the amplifier noise [16].

Because of the low spurious mode density in the SLC resonator, the 600 MHz bandwidth of the band pass filter was adequately narrow to operate this oscillator only on the desired mode. The frequency noise of this oscillator was measured using SLC2 as a frequency discriminator. The frequency of SLC2 was matched to the oscillator by varying its operation temperature. To finally match the oscillator frequency exactly to the resonance frequency of SLC2, the phase shifter in the free running oscillator was adjusted. The resultant fractional frequency spectral density measured is shown in fig. 11. This oscillator exhibits flicker of frequency noise with a value of about $S_y(f)$=-237 dBc/Hz at 1 kHz. This corresponds to an $\mathcal{L}(f)$=-100 dBc/Hz at 1 kHz.

From the oscillator noise the amplifier noise may be calculated using Leeson's model [17]. Fig. 13 shows the inferred amplifier noise from the oscillator noise presented in fig. 11, compared with previous direct phase noise measurements of the same amplifier [16]. From these measurements it is evident that the oscillator is limited by the amplifier noise.

Conclusions

We have demonstrated the ability to construct a sapphire resonator to a prespecified frequency, in this case 10 GHz ±1 kHz. The fine tuning and stabilization is achieved by thermo-electric temperature control. At 273K the unloaded Q was determined to be 2.2×10^5 and from 268K to 298K the Q was proportional to $T^{-2.5}$. By using a special cavity design, the coupling to undesired modes is considerably reduced, resulting in an apparent lower mode density. An upper limit to the residual noise of the cavity of $\mathcal{L}(f_m) \sim -125/f_m$ (dBc/Hz) was measured over the range 10 Hz to 10 kHz, which was limited by the mixer noise floor. No 'excess' noise was observed in these measurements. A feedback oscillator constructed around one of the resonators displayed phase noise consistent with (amplifier) device noise of $\mathcal{L}(f_m) \sim -105/f_m$ dBc/Hz.

Acknowledgments

The authors would like to thank Eugene Ivanov, Tony Mann, Marco Costa and Andre Luiten for their general assistance. They would also like to thank David Blair for helping to make this collaboration possible. This research was supported by the Australian Research Council.

References

[1] V.B. Braginskii, V.I. Panov, "Superconducting resonators on sapphire," IEEE Trans. on Magnetics, vol. 15, no. 1, pp 30-32, Jan. 1979.

[2] D.G. Blair and S.K. Jones, "High quality factors for a sapphire loaded superconducting cavity resonator," Electron. Lett., vol. 23, pp. 817-818, 1987.

[3] A.J. Giles, S.K. Jones, D.G. Blair and M.J. Buckingham, "A high stability microwave oscillator based on a sapphire loaded superconducting cavity," in Proc. IEEE 43rd Annu. Symp. Freq. Contr., 1989, pp 89-93.

[4] G.J. Dick and J Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Dielectric Ring Resonator for Ultra-Low Noise", IEEE Trans.Ultrason. Ferroelec. Freq. Contr., vol. 37, pp 339-346, Sept. 1990.

[5] G.J. Dick and R.T. Wang, "Ultra-stable performance of the superconducting maser," IEEE Trans. on Instrum. and Meas., vol 40, no. 12, pp. 174-177, Dec. 1991.

[6] M.M. Driscoll et al., "Cooled ultra-high Q, sapphire dielectric resonators for low noise microwave signal generation," IEEE Trans. Ultrason. Ferroelec. and Freq. Contr., vol. 39, pp 405-411, May 1992.

[7] G.J. Dick and D.G. Santiago, "Microwave frequency discriminator with a cryogenic sapphire resonator for ultra-low phase noise," in Proc. of 6th European Frequency and Time Forum, 1992.

[8] M.E. Tobar and D.G. Blair, "Phase noise analysis of the sapphire loaded superconducting niobium cavity oscillator," to be published in IEEE Trans. MTT., vol. 42, no. 2, Feb. 1994. (in press)

[9] A.N. Luiten, A.G. Mann and D.G. Blair, "Ultra high Q-factor cryogenic sapphire resonator," to be published in Electron. Lett. 1993.

[10] E.N. Ivanov, D.G. Blair and V.I. Kalinichev, "Approximate approach to the design of shielded dielectric disk resonators with whispering gallery modes," IEEE Trans. MTT., vol. 41, no. 4, April 1993.

[11] V.B. Braginsky, V.S. Ilchenko and Kh.S. Bagdassarov, "Experimental observation of fundamental microwave absorption in high-quality dielecric crystals," Phys. Lett. A., vol. 120, no. 6, pp 300-305, March 1987.

[12] M.E. Tobar, "Effects of spurious modes in resonant cavities," submitted J. Phys. D., 1993.

[13] M.E. Tobar and D.G. Blair, "A generalized equivalent circuit applied to a tunable sapphire loaded superconducting cavity," IEEE Trans. MTT., vol. 39. no. 9, pp. 1582-1594, Sept. 1992.

[14] S. Nogi, K. Fukui, M. Kojima and S. Tanaka, "Effective methods for suppression of undesired cavity modes using microwave absorbers," Electronics and Communications in Japan, part 2, vol. 74, no. 9, 1991.

[15] D.S. Bailey, M.M. Driscoll, R.A. Jelen, B.R. McAvoy, "Frequency stability of high-overtone bulk-acoustic resonators," IEEE Trans. Ultrason. Ferroelec. Freq. Contr., vol. 39, no. 6, pp 780-784, Nov. 1992.

[16] A.G. Mann, unpublished data, Department of Physics the University of Western Australia.

[17] D.B. Leeson, "A simple model of feedback oscillator noise spectrum," in Proc. IEEE, vol. 54, no. 2, pp. 329-330, Feb. 1966.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

IMPROVED SAPPHIRE DIELECTRIC RESONATORS FOR ULTRASTABLE OSCILLATORS

A. N. Luiten, A. G. Mann and D. G. Blair,

Department of Physics, University of Western Australia, Nedlands, W. A. 6009, Australia.

Abstract

Microwave oscillators based on cryogenic sapphire dielectric resonators have been shown to achieve Allan deviations below 10^{-14} for integration times from 1 second to over 100 seconds. These results were achieved with resonators having Q-factors ~10^8 at 9.7 GHz and using a synthesis of loop oscillator operation with active Pound frequency stabilization [1]. Almost all instability in that system could be attributed to the external circuit. Higher Q-factor resonators should enable proportionally improved oscillator performance. In this paper we report measurements on "whispering gallery" modes in new, ultra-high purity resonators. Below 10 K these resonators exhibit a dielectric loss tangent decreasing approximately linearly with temperature to at least 1.6 K where the unloaded Q-factor is ~8 x 10^9. We exploit the existence of paramagnetic ions with very high frequency electron spin resonance to obtain frequency-independent mode frequency-temperature compensation. A pair of oscillators based on these resonators, operating initially at 11.9 GHz and with Q-factors exceeding 10^9, has been investigated. Based on previous measured oscillator parameters Allan deviations around 10^{-15} may be anticipated for 6 K operation.

Introduction

With the development of laser-cooled atomic frequency standards a requirement exists for new "flywheel" oscillators at least an order of magnitude more stable than state of the art quartz oscillators. Excellent short term stability can be provided by cryogenic oscillators where resonator quality factors are in excess of 10^9. High purity monocrystalline sapphire at cryogenic temperatures is an excellent dielectric for such resonator applications owing to its high mechanical rigidity, low thermal expansion and very low dielectric loss. Ultra-stable microwave oscillators such as our Sapphire-Loaded Superconducting Cavity Oscillator (SLSCO) [1, 2] and JPL's Superconducting Cavity Maser (SCM) [3] both employ as the frequency determining element a very high Q sapphire resonator. Whereas the SCM operates in a low order mode the SLSCO utilizes a high order mode resulting in much smaller fields at the shield walls and employs [1] a synthesis of loop oscillator and the form of Pound frequency stabilisation developed to a high degree by Stein and Turneaure [4]. A vital property of both these oscillators is the existence of a maximum of mode frequency as a function of temperature and power which allows fractional frequency stabilities of 10^{-14} or better with quite modest temperature and power regulation. A fractional frequency stability of 9 x 10^{-15}, flat from 1 to 300 seconds integration time, has been reported with a 3 cm diameter sapphire resonator [1]. The frequency-temperature maximum of this resonator is due to the compensating effects of the Curie law temperature dependence of the ac susceptibility of residual paramagnetic impurities in the sapphire and the temperature-dependent surface reactance of the nearby superconducting walls [5, 6]. However the walls significantly limited the Q. To obtain higher Q factors and substantially increase the number of usable modes we have investigated more highly confined modes at X-band in larger, 5 cm diameter resonators [2, 6, 7]. These resonators also possess a frequency-temperature maximum, but now the temperature dependence of the dielectric constant instead of the superconducting penetration depth compensates the effect of the paramagnetic impurities.

The Resonator

The modes in a cylindrical dielectric resonator may be designated H-type (quasi-TE) or E-type (quasi-TM). Each mode is described by three mode numbers m, n and p characterizing the field variations in the azimuthal, radial and axial directions respectively. We have restricted our study to those modes with the greatest field confinement to the dielectric, hence highest potential Q factor, namely those which have large azimuthal and small axial and

0-7803-0905-7/93 $3.00 © 1993 IEEE

radial mode numbers; these are the H_{m11} ($TE_{m1\delta}$) and E_{m11} ($TM_{m1\delta}$) modes having m azimuthal field periods and only one for the radial and axial mode numbers, also known as "whispering gallery" modes.

We have focused on resonators cut from Crystal Systems (CS) "HEMEX" grade 1991 material, since this seems to be the lowest loss material we have tested so far. We have investigated two resonators with the geometry shown in figure 1. Both are single crystals with their c-axis oriented along the geometric axis (to within 1 degree), optically polished surfaces and have integral (1.5 cm diameter) mounting spindles at each end. The resonators were cut from different boules. Resonator 1 was from a typical boule while resonator 2 was from atypical material, known to be slightly contaminated with titanium.

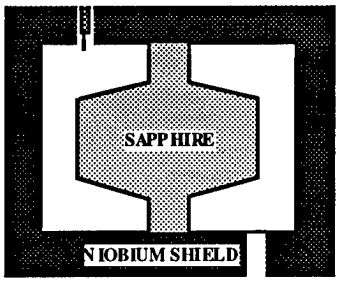


Figure 1 The sapphire resonator

The sapphire resonator is mounted inside a cylindrical niobium shield (inside diameter 8 cm and height 5 cm). The niobium shield was constructed with knife edges between the central cylinder and end plates to preserve the Q of cavity-like modes with currents crossing the walls thereby preventing Q degradation of nearby sapphire H or E modes. Coupling to the mode fields was via two fixed probes to enable loop oscillator operation on H_{m11} modes. One of these was a magnetic field loop probe in the end wall and the other a radial electric field probe in the center of the cylindrical wall. The sapphire was ultrasonically cleaned in a bath of nitric and hydrofluoric acid. It was given a final methanol rinse before assembly. The niobium shield was cleaned in nitric acid. The assembled resonator was placed in a vacuum pot which was then evacuated and permanently sealed. The pot was attached to a cryostat for cooling to liquid nitrogen or helium temperatures. Temperature control was provided by a carbon glass thermometer and a heater mounted in thermal contact with the shield.

Q factors

The unloaded electrical quality factor of a shielded dielectric resonator is given by:

$$Q_0^{-1} = F_C (\tan\delta + \chi'') + R_S \Gamma^{-1} \qquad (1)$$

where F_C is the confinement factor, determined by the ratio of the mode's energy stored in the dielectric to the total stored energy, $\tan\delta$ is the loss tangent of the dielectric and χ'', the imaginary part of the ac susceptibility, represents losses due to paramagnetic impurities in the dielectric. The third term represents losses at the walls: R_S is the surface resistance and Γ is a geometry factor describing the magnetic field at the walls.

The Q as a function of temperature around 12 GHz for resonator 1 has already been reported [7] and is shown in Figure 2. The Q factor shows a steady improvement on cooling; from 2.8×10^9 at 6 K, to 4.5×10^9 at 4.2 K, to 8.3×10^9 at 1.55 K.

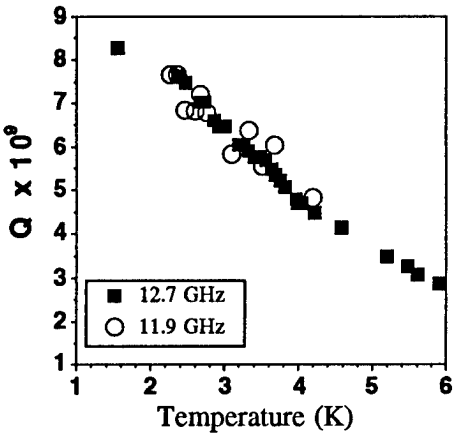


Figure 2 Unloaded Q as a function of temperature at 12.7 GHz ($E_{18\,1\,1}$) and 11.9 GHz ($H_{14\,1\,1}$) for resonator 1.

In Figure 3 we display the temperature dependence of Q for the $H_{14\,1\,1}$ mode in resonator 2. The Q is again a monotonically decreasing function of temperature. Below 6 K it is somewhat lower than in resonator 1, which is possibly a consequence of the relatively large concentration of titanium. Both Crystal Systems HEMEX resonators exhibit chromium and iron impurity levels of the order of parts per billion [2], as determined by measurement of the mode frequency shift (χ') as the electron spin resonances (ESRs) at 11.45 GHz (Cr^{3+}) and 12.05 GHz (Fe^{3+}) are alternately saturated [6]. Thus paramagnetic losses from the Cr^{3+} and Fe^{3+} ESRs, are expected to be negligible. The effect of the niobium walls is very small for the $H_{14\,1\,1}$ mode as is evident from the small change in Q across the niobium superconducting transition temperature, at 9.25 K. Since the BCS surface resistance at X band drops by at least two orders of magnitude from 9.25 K to 6 K [8] we conclude that below 6 K the observed Q values are determined by

the intrinsic loss-tangent of the sapphire dielectric.

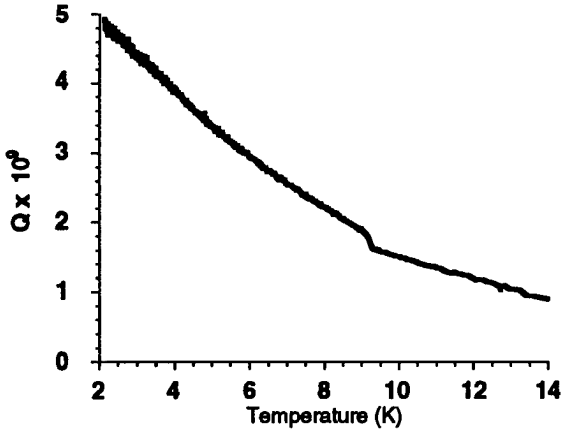


Figure 3 Unloaded Q as a function of temperature at 11.9 GHz ($H_{14\,1\,1}$) for resonator 2.

At 4.2 K previous measurements in similar size resonators fabricated from Union Carbide (1989) sapphire with chromium and iron impurities at the ppm level have shown maximum Qs of only 1.5×10^9 [2, 6], which is similar to the results of other workers [3, 9]. Braginsky [9] has shown that below 50 K the dielectric loss is strongly dependent on the level of crystal imperfection, as determined by the rate of crystal growth. We speculate that the high Qs in HEMEX resonators may be attributable to the lower defect density expected from the Heat-Exchange Method of crystal growth as compared to the conventional Czochralski process employed by Union Carbide.

Frequency-Temperature Maximum

The temperature dependence of a given mode frequency, f, when the metallic shield has negligible influence, may be expressed as:

$$(f_0 - f) / f_0 = AT^4 + \sum C\,D(T) \qquad (2)$$

where f_0 would be the frequency at absolute zero if no paramagnetic impurities were present. The first term represents the combined temperature dependence of the dielectric constant and the thermal expansion coefficient of the sapphire. A is somewhat mode dependent, being largest for E_{m11} modes and smallest for H_{m11} modes, the ratio of the two extremes being about 1.5. The second term represents the sum of the real part of the ac susceptibility, χ', of all paramagnetic ions. The coefficient C is strongly mode dependent (through a mode filling factor) and proportional to ion concentration and the square of its magnetic moment (or g factor). Below the ESR of the dominant paramagnetic species mode frequency-temperature "self-compensation" can occur. That is, for all modes the frequency passes through a maximum at some temperature, T_m. T_m is a very slow function of ion concentration: in the Curie law regime (D(T) = 1/T) it is proportional to $C^{1/5}$ while the second derivative of frequency with respect to temperature is proportional to $C^{2/5}$ [6]. Operation near the frequency maximum significantly relaxes the temperature control requirements. Typical curvatures in H modes at the maximum, in fractional frequency terms, are about 10^{-9} K^{-2}. This would imply that, with readily achievable 0.1 mK regulation, a temperature offset of 10 mK from the peak would limit the (long term) fractional frequency fluctuations to ~ 1×10^{-15}.

Mo^{3+} appears to be the dominant paramagnetic ion in our resonators, which is not unexpected since the sapphire is grown almost totally enclosed in a molybdenum crucible. The lowest ESR of Mo^{3+} is at 165 GHz [10], equivalent to a spin-state (Kramer's doublet) energy level spacing, $\Delta E/k$, of some 7.9 K. Since the resonator mode frequencies of interest are well below the ESR one obtains frequency-temperature self-compensation for all microwave modes up to the limit of our test equipment (22 GHz), allowing considerable freedom of choice of operating frequency. T_m then only depends on the mode type. For E_{m11} modes between 4.7 and 22 GHz in resonators 1 and 2 the frequency maximum is observed around 9-10 K. The behaviour of the H_{m11} modes is illustrated in Figure 4 by $H_{14\,1\,1}$.

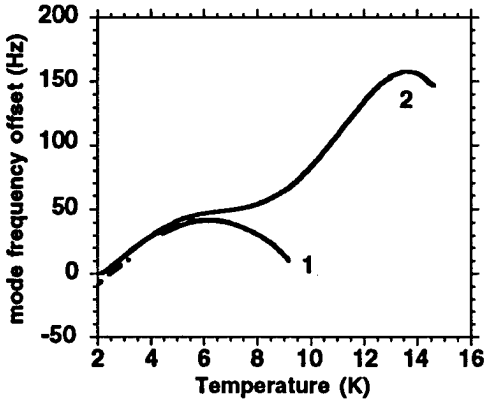


Figure 4 Mode frequency temperature dependence of $H_{14\,1\,1}$ mode for resonators 1 and 2.

For H_{m11} modes in resonator 1 the temperature maximum is at 6 K, as expected for a filling factor 20 times smaller than in E modes. However the H_{m11} modes in resonator 2 exhibit an unusual frequency-

temperature dependence, which closely follows that of resonator 1 up to 6 K, but rises to a frequency maximum at 13.5 K. This may be explained by the presence of a second paramagnetic species in resonator 2, which couples well to the H modes and negligibly to the E modes and whose spin resonance becomes activated only above several K. A strong candidate is Ti^{3+}, one, if not the only ion, which is highly anisotropic in the corundum lattice (parallel g factor 1.07, perpendicular g factor 0.1) and has a lowest electron spin resonance at about 1 THz [11] ($\Delta E/k$ ~54 K). The curves in Figure 4 can be fitted quite well with D(T) being of the form, $\tanh[-\Delta E/2kT]$. Further details of this aspect of our work will be published elsewhere.

Spectroscopy: In an attempt to identify the dominant paramagnetic species we have performed conventional X-band ESR spectroscopy from room temperature down to 100 K on mm size samples from a boule similar to resonator 1 and the same boule as resonator 2. A typical magnetic field scan is shown in Figure 5 for two orientations of the applied magnetic field to the crystal c axis, 90° (perpendicular) and 45°. The broad resonance at g = 2 is unidentified and may be an artefact of the sample chamber. The sharp resonance, whose g factor varies as the crystal is rotated, from a maximum of 4 at 90° to a minimum of 2 at 0°, is readily identified as Mo^{3+} [12].

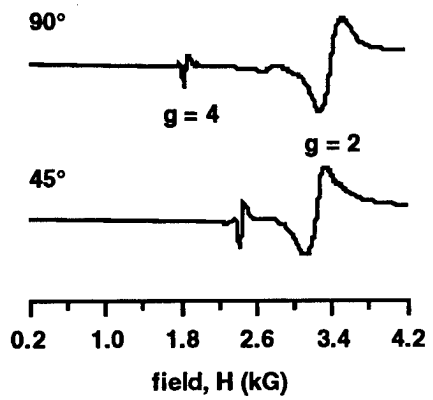


Figure 5 Typical magnetic field scans at 100 K, of a CS sample using X-band ESR spectroscopy, for 90 degree and 45 degree orientations.

Further confirmation is provided by a narrow field scan (Figure 6) which reveals its two isotopic satellite peaks. The concentration, estimated by calibrating the spectrometer with a ruby sample, is about 0.3 ppm with a variation of some 50% between boules, in agreement with the ratio of C coefficients for E modes in the two resonators. Unfortunately the linewidth of Ti^{3+} is so broad that it is not revealed by this technique until temperatures well below 77 K. Previous labelling of the dominant paramagnetic species in CS HEMEX as Ti^{3+} [2, 13] is therefore incorrect.

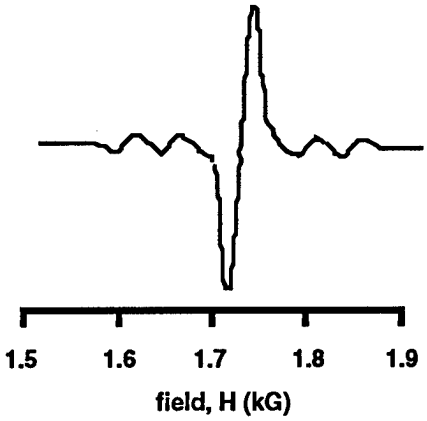


Figure 6 An ESR magnetic field scan at 90 degrees orientation and 100 K, revealing primary and two (isotopic) satellite resonances of Mo^{3+} in a CS HEMEX sample.

Frequency-Power Dependence

By monitoring the change in heater power required to maintain the resonator at constant temperature we have been able to measure the dependence of mode frequency on microwave power dissipated in the resonator, at least for modes with high coupling coefficients, principally H_{m11}. At high power (~10 mW) the measured slope of fractional frequency shift with power of the $H_{14\,1\,1}$ mode in both resonators (~ 10^{-8} W^{-1}) could be consistent with the static radiation pressure-induced permittivity change quoted by Braginsky [14], which is strictly proportional to stored energy. At low power in resonator 1 we see a significant deviation from linearity in mode frequency as a function of power (Figure 7). The $H_{13\,1\,1}$ (11.25 GHz) mode exhibits a positive initial slope and an extremum, while the $H_{14\,1\,1}$ (11.9 GHz) mode has a high initial (negative) slope. Such behaviour is characteristic of saturation of the chromium 11.45 GHz ESR by microwave power at the mode frequency, first seen in the original 3cm resonator and which changes sign above the resonance [15]. The sensitivity to chromium is magnified by the proximity of the $H_{13\,1\,1}$ and $H_{14\,1\,1}$ mode frequencies to 11.45 GHz. Due to a leak in the cryostat we have not investigated the low power regime of resonator 2. Thus if one cannot take advantage of a frequency-power extremum the limiting frequency stability will be about 10^{-15} for 0.1 mW dissipated in the resonator, assuming power fluctuations of 0.1 %, which are typical of microwave amplifiers used in our loop oscillator configuration.

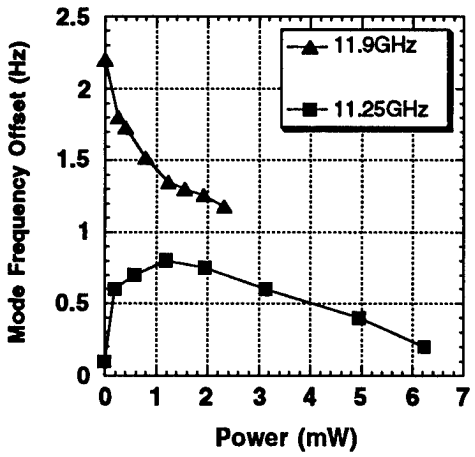


Figure 7 The dependence of mode frequency on power dissipated in resonator 1 for the $H_{13\,1\,1}$ (11.25 GHz) and $H_{14\,1\,1}$ (11.9 GHz) modes.

Conclusions

We have shown that in 5 cm Crystal Systems "HEMEX" resonators unloaded Q values, not limited by wall losses, of 5-8 x 10^9 are possible at 2 K and 3 x 10^9 at 6 K. Frequency-temperature self-compensation, in which the temperature dependence of the real part of the ac susceptibility of paramagnetic impurities balances that due to the dielectric constant of the sapphire, can be achieved in H modes at convenient operating temperatures. Because chromium and iron concentrations are very small, ions with very high frequency ESRs provide essentially frequency-independent self-compensation, allowing freedom of choice of operating frequency. The dominant paramagnetic ion has been identified as Mo^{3+}. Since the Mo^{3+} concentration is essentially fixed by the crystal growth process and we desire both the lowest possible frequency-temperature coefficient and dielectric loss, the H_{m11} modes are the best candidates for high stability HEMEX resonators. The observed power dependence of mode frequency, ~ 10^{-8} W^{-1}, will require that resonator power fluctuations are maintained at less than 100 nW if frequency stability of 1 x 10^{-15} is to be achieved.

Acknowledgments

The authors would like to thank Adrian Giles, Marco Costa and Mike Tobar for their general assistance. This research was supported by the Australian Research Council.

References

[1] A.J. Giles, S.K. Jones, D.G. Blair and M.J. Buckingham, "A high stability microwave oscillator based on a sapphire loaded superconducting cavity," in Proc. IEEE 43rd Annual Frequency Control Symposium, 1989, pp 89-93.

[2] A.G. Mann, A.N. Luiten, D.G. Blair and M.J. Buckingham, "Ultra-Stable Cryogenic Sapphire Dielectric Microwave Resonators", Proc. IEEE 46th Annual Frequency Control Symposium, 1992, pp 167-171.

[3] G.J. Dick and J Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Dielectric Ring Resonator for Ultra-Low Noise", IEEE Trans.Ultrason. Ferroelec. Freq. Contr., vol. 37, pp 339-346, Sept. 1990.

[4] S. R. Stein and J. P. Turneaure, "The Development of the Superconducting Cavity Stabilized Oscillator", Proc. 27th Annual Frequency Control Symposium, 1973, p414 .

[5] S.K. Jones, D.G. Blair and M.J. Buckingham, "Effect of paramagnetic impurities on frequency of sapphire-loaded superconducting cavity resonator", Electron Lett. vol 24 , pp 346-347, 1988.

[6] A G Mann, A J Giles, D G Blair and M J Buckingham, "Ultra-stable cryogenic sapphire dielectric microwave resonators: mode frequency temperature compensation by residual paramagnetic impurities", J. Phys. D., vol 25, pp 1105-1109, 1992.

[7] A N Luiten, A G Mann and D G Blair, "Ultra High Q Factor Cryogenic Sapphire Resonator", Electron Lett., vol 29, 1993, in press.

[8] C.M. Lyneis, Experimental Studies of the Microwave Properties of 8.6 GHz Superconducting Niobium, Niobium-Tantalum and Tantalum Cavities, Ph.D. Thesis, Stanford University, Stanford, CA, 1974, unpublished, pp 39-42.

[9] V.B. Braginsky, V.S. Ilchenko and Kh. S. Bagdassarov, "Experimental Observation of Fundamental Microwave Absorption in High Quality Dielectric Crystals", Phys Lett A, vol 120, pp. 300-305, 1987.

[10] A.A. Mirzakhanyan and K.N. Kocharyan, "Temperature Dependence of the Interdoublet ESR Spectrum of Mo^{3+} Ions in Corundum at Submillimeter Wavelengths", Sov. Phys. Solid State, vol 23, 1981, pp 49-53.

[11] R.R. Joyce and P.L. Richards, "Far-Infrared Spectra of Al_2O_3 Doped with Ti, V, and Cr", Phys. Rev., vol 179, 1969, pp 375-380.

[12] E.G. Sharoyan, O.S. Torosyan, E.A. Markosyan and V.T. Gabrielyan, "EPR and Spin-Lattice Relaxation of Mo^{3+} Ions in Corundum", Phys. Stat. Sol. (b), vol 65, 1974, pp 773-778.

[13] A.N. Luiten, A.G. Mann, A.J. Giles and D.G. Blair, "Ultra-Stable Sapphire Resonator-Oscillator", IEEE Trans. Instrumentation and Measurement, vol 42, 1993, in press.

[14] V.B. Braginskii, V.P. Mitrafanov and V.I. Panov Systems with Small Dissipation. Chicago: University of Chicago, 1985, p 78.

[15] A.J. Giles, Electron Spin Resonance Phenomena in a Sapphire Loaded Superconducting Cavity Oscillator, Ph.D. Thesis, Physics Dept., University of Western Australia, 1993, unpublished.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

MICROWAVE OSCILLATORS INCORPORATING CRYOGENIC SAPPHIRE DIELECTRIC RESONATORS

C. A. Flory and R. C. Taber

Hewlett-Packard Laboratories
3500 Deer Creek Road - Bld. 26M-3 Palo Alto, CA 94303-1392

Abstract

Progress is reported on efforts to develop a commercially-viable high purity X-band signal source incorporating a cryogenic sapphire dielectric resonator. The resonator design is of the whispering gallery type to take advantage of the excellent electromagnetic field confinement offered by this geometry. Complications resulting from the high spurious mode density of this type of resonator have been eliminated by developing a very accurate and complete mode analysis program which fully incorporates the dielectric anisotropies of the sapphire ring. This program allows the design of a window in the frequency domain where no unwanted modes exist, with accurate placement of the desired mode at the center of this region.

Evaluation of the phase noise properties of simple oscillators incorporating these resonators has been done. Two-oscillator phase noise comparison measurements have yielded extremely encouraging results.

Introduction

Ideally, a high purity X band signal source should supply state of the art phase noise performance over the broadest possible range of offset frequencies. In present day systems, this is accomplished by patching together the performance of several components. The reasoning that leads to this circumstance is widely understood. [1] For low offset frequencies ($f \approx 1$ Hz) the best performance available is obtained by producing high harmonics of relatively low frequency sources ($\nu_0 \approx 10$ MHz). This choice has been dictated by the availability of high resonator Q, low resonator frequency flicker and low amplifier phase noise at frequencies near 10 MHz compared to the corresponding parameters available at microwave frequencies. For larger offset frequencies, best performance is provided by oscillators operating directly at microwave frequencies. This is due to the fact that the far-from-carrier phase noise of well designed oscillators is only weakly dependent on the frequency at which the oscillator operates. However, the process of frequency multiplication inevitably produces an increase in phase noise by a factor of N^2, where N is the frequency multiplication factor, so many X band signal sources are configured in such a way that the behavior close to the carrier is dominated by a frequency-multiplied RF source while the behavior far from the carrier is governed by an X band resonator. Unfortunately, the process of merging these two regimes involves several compromises, and the results remain less than ideal.

It has been evident for some time that improved performance over a broad range of frequency offsets should be obtainable by means of cryogenic sapphire dielectric resonators without relying on an RF oscillator to control any part of the behavior. [2,3,4] As sapphire is cooled, its loss tangent varies in proportion to $1/T^5$. [5] As a result, it is possible to obtain resonator Q values at 77K that are high enough to adequately compensate for the inferior phase noise of existing X band amplifiers. Unfortunately, dielectric resonators generally display strong vibrational sensitivity, and the temperature sensitivity of the dielectric constant of sapphire at 77K is relatively severe with a value of $(1/\varepsilon)\,\Delta\varepsilon/\Delta T \approx 10^{-5}/°K$. Remedies for these difficulties are available, however. Firstly, the temperature sensitivity of the dielectric constant changes in proportion to $1/T^3$, so significant improvements are obtained by decreasing the temperature by a modest amount. Secondly, the dominant component of the vibrational sensitivity of a dielectric resonator is proportional to the maximum strength of the electromagnetic field on the interior surface of the enclosure, so the vibrational sensitivity can be mitigated by straightforward electromagnetic design analysis in conjunction with diligent mechanical design. The configuration implied by these considerations would not necessarily be suitable for the full range of X band signal source applications, but the availability of compact closed-cycle cryocoolers will make it possible to configure a rack-mountable device that will not depart excessively from existing sources in either cost or overall size. Obviously, most airborne applications would remain out of reach, but

0-7803-0905-7/93 $3.00 © 1993 IEEE

other applications such as shipboard radar master oscillators and phase noise measurement system reference oscillators should be adequately served by such devices.

A fundamental constraint on the design of high Q sapphire dielectric resonators is imposed by the low value of the dielectric constant of sapphire, $\varepsilon \approx 10$. With such a low dielectric constant, the strength of the electromagnetic field external to the resonator itself will be high, and it will generally be necessary to adopt specialized design strategies in order to avoid excessive loss contributions from the shielding currents on the inside surfaces of the enclosure. In order to address this issue, two generic varieties of sapphire dielectric resonator have been extensively studied. The first type is a low order TE mode resonance of a cylindrical sapphire puck centered in a cylindrical conducting enclosure. These resonators suffer Q degradation from enclosure wall loss due to weak confinement of the electromagnetic fields. Some improvement in Q values can be achieved by replacing the conducting end caps of the enclosure by HTS films.

The second generic type of sapphire dielectric resonator is the whispering gallery mode configuration. It consists of a sapphire ring which confines the electromagnetic energy to the dielectric region by a physical mechanism not unlike total internal reflection in optical systems. The modes can be qualitatively described as traveling waves around a bent dielectric waveguide which closes upon itself. The strong electromagnetic energy confinement is due to the existence of only evanescent fields outside of the dielectric structure. The relative field confinement increases rapidly as a function of mode number, where mode number is defined as the number of modal maxima encountered in one complete circuit of the ring. For mode numbers on the order of 7 or more, the relative field energy confinement is strong enough to greatly minimize the Q degradation from enclosure wall loss. Configurations of this type allow Q values approaching the limitations imposed by the internal sapphire loss mechanisms. However, a significant difficulty with these whispering gallery resonators is that in the microwave regime, the required size is such that the mode density in the frequency domain is extremely high. These spurious modes are other whispering gallery modes as well as "hybrid" modes, which are simply the usual empty cavity modes perturbed by the presence of the sapphire ring.

In order to incorporate these whispering gallery mode resonators into robust oscillator configurations, these numerous spurious modes must be controlled. The obvious solution is to judiciously choose the geometry of the resonator to "open a window" in the forest of spurious modes wherein the desired whispering gallery mode may reside. This task requires an extremely accurate means of calculating all modes of the sapphire-enclosure system, due to the high density of modes occurring in the typical system and the relatively small windows possible in this situation. It is also obvious that completion of this task requires incorporating the inherent anisotropy of the sapphire, which is on the order of 20%, to adequately predict the frequencies of the calculated modes. The ability to accurately compute the frequencies of all of the modes contained in the sapphire-enclosure system allows direct determination of an optimal resonator geometry, without having to resort to trial-and-error construction.

The technique described in this paper for accurate calculation of the whispering gallery and hybrid modes is a generalization of the radial mode matching method described in the literature. [6] The extension shown here is the necessary incorporation of the sapphire anisotropy. This generalized radial mode matching technique is rigorous in the sense that the solution is obtained as a series which approaches the exact solution, with the accuracy determined by the number of terms kept in the expansion. In principle, the frequencies and field distributions can be calculated to arbitrary accuracy. Cruder methods, such as finite element techniques, are normally incapable of achieving the tolerances required of the mode frequency determinations, as well as the extremely weak but important evanescent field strengths out at the enclosure walls which determine the Q degradation from wall losses.

In the next section, the anisotropic mode matching method will be described for the whispering gallery resonator geometry. Included in this section will be an application of our results to accurately determine the temperature dependent dielectric constants of sapphire by using measured data of whispering gallery mode frequencies as a function of resonator temperature. The subsequent section contains measured results of an optimized whispering gallery resonator at 9.6 GHz, including Q values and mode spectra. Two-oscillator phase noise measurements of simple oscillators incorporating these resonators will be described and presented. The final section summarizes the current results and describes plans for future work.

Modeling

A. Statement of the Problem

The physical resonator model consists of a ring of dielectric material of rectangular cross-section symmetrically placed in a conducting cylindrical enclosure as illustrated by Figure 1. The dielectric is not isotropic, and is assumed to have different values in the "z" and "transverse" dimen-

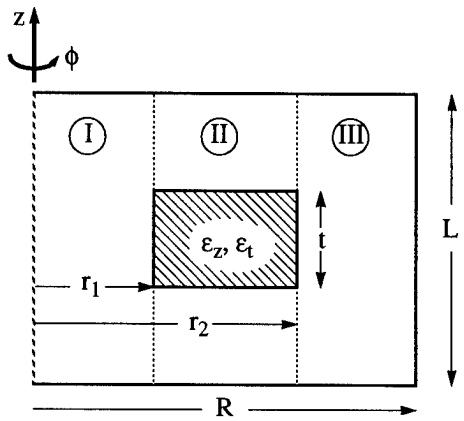


Figure 1. Dimensions and parameters of the physical resonator model. (Half cross-section)

sions. Determination of the resonant modes of this system requires solution of Maxwell's equations subject to the imposed boundary conditions. For a source-free region and assuming all fields have harmonic time dependence of the form $e^{i\omega t}$, Maxwell's equations take the form

$$\vec{\nabla}\times\vec{H} = i\omega\varepsilon_0\,\underset{\sim}{\varepsilon}\cdot\vec{E} \qquad (1)$$

$$\vec{\nabla}\times\vec{E} = -i\omega\mu_0\vec{H} \qquad (2)$$

$$\vec{\nabla}\cdot\vec{H} = 0 \qquad (3)$$

$$\vec{\nabla}\cdot(\underset{\sim}{\varepsilon}\cdot\vec{E}) = 0 \qquad (4)$$

where the permittivity tensor $\underset{\sim}{\varepsilon}$ has the form

$$\underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_t & 0 & 0 \\ 0 & \varepsilon_t & 0 \\ 0 & 0 & \varepsilon_z \end{bmatrix} \qquad (5)$$

for a dielectric with the symmetry properties of sapphire and the c-axis aligned along the coordinate z-axis.

Solution of Eq.(1-4) is clearly complicated by the fact that the permittivity tensor is not homogeneous in addition to its anisotropy, i.e. the relative dielectric constant varies with position inside the enclosure, being unity in free space and greater than one inside the sapphire. This positional $\rho - z$ dependence of the dielectric constants results in Maxwell's equations not being separable over the entire enclosure region. A standard technique for reintroducing separability of the partial differential equations is the use of radial mode matching expansions. This procedure involves separating the resonator structure into three distinct regions, labeled by I, II, III in Figure 1, and solving for a normal mode expansion in each region. Since the dielectric constant is only a function of z in each individual region, the partial differential equations (1-4) regain their separability. The added complexity of solving the problem in this fashion derives from the fact that the individual normal mode expansions of the three regions must be matched along the two boundary surfaces shown by the dotted lines in Figure 1.

Application of radial mode matching to the analogous problem where the dielectric is fully isotropic is well documented in the literature. Here, the anisotropic calculation will be shown in only a schematic fashion in order to show the differences that occur when the anisotropy is included.

B. Radial Mode Matching - TM components

Introduction of vector potentials often simplifies formulation of electromagnetic boundary value problems. In particular, these potentials easily allow the general normal mode series solutions to be decomposed into two orthogonal subsets, the TE and TM solutions. Of course, the full solution satisfying the physical boundary conditions normally contains contributions from both sets.

To first investigate the dynamic equations for the TM fields, the magnetic vector potential $\vec{A}$ is defined by

$$\vec{H}^e = \frac{1}{\mu_0}\vec{\nabla}\times\vec{A} \qquad (6)$$

which clearly satisfies Maxwell's Eq. (3) and the super e denotes TM fields. Substitution of this expression for $\vec{H}$ into Eq. (2) demonstrates that an associated scalar potential ϕ can be defined which allows the $\vec{E}$ field to be given by

$$\vec{E}^e = -\vec{\nabla}\phi - i\omega\vec{A}. \qquad (7)$$

Substituting these expressions for the $\vec{E}$ and $\vec{H}$ fields into Eq. (1) allows derivation of the dynamical equation for the vector potential

$$\nabla^2\vec{A} + k_0^2\varepsilon_t\vec{A} = \frac{1}{\varepsilon_t}(\vec{\nabla}\varepsilon_t)(\vec{\nabla}\cdot\vec{A}) - k_0^2\Delta\varepsilon A_z\hat{n}_z$$

$$-\Delta\varepsilon\frac{\partial}{\partial z}\left(\frac{1}{\varepsilon_t}\vec{\nabla}\cdot\vec{A}\right)\hat{n}_z \qquad (8)$$

where $\Delta\varepsilon = \varepsilon_z - \varepsilon_t$, n_z is a unit vector in the z-direction, $k_0^2 \equiv \omega^2\varepsilon_0\mu_0$ and the gauge choice is

$$\vec{\nabla}\cdot\vec{A} + i\omega\varepsilon_0\mu_0\varepsilon_t\phi = 0 \qquad (9)$$

to eliminate the scalar field ϕ. Note that Eq. (8) is valid for all three regions I, II, and III, however the dielectric constants will of course be different in each respective region.

Restriction to the TM subset is now straightforward. Choosing $\vec{A}$ parallel to the z-axis, where $\vec{A} = A\,\hat{n}_z$ gives the scalar equation

$$\nabla^2 A + k_0^2\varepsilon_z A = \frac{\varepsilon_z}{\varepsilon_t^2}\left(\frac{\partial\varepsilon_t}{\partial z}\right)\left(\frac{\partial A}{\partial z}\right) - \frac{\Delta\varepsilon}{\varepsilon_t}\left(\frac{\partial^2 A}{\partial z^2}\right) \qquad (10)$$

which is clearly separable. For a solution ansatz of the form

$$A(\rho, \phi, z) = R^e(\rho)\, Z^e(z)\, \Phi^e(\phi) \quad (11)$$

the resulting three independent ordinary differential equations are

$$\frac{d^2\Phi^e}{d\phi^2} + m^2\Phi^e = 0 \qquad (12)$$

$$\frac{d}{dz}\left(\varepsilon_t^{-1}\frac{dZ^e}{dz}\right) + \left(k_0^2 + \frac{\lambda^e}{\varepsilon_z}\right)Z^e = 0 \quad (13)$$

$$\frac{1}{\rho}\frac{d}{d\rho}\left(\rho\frac{dR^e}{d\rho}\right) - \left(\lambda^e + \frac{m^2}{\rho^2}\right)R^e = 0 \quad (14)$$

where λ^e and m are separation constants to be determined by the relevant boundary conditions imposed on the physical fields by the resonator geometry.

The vanishing of tangential $\vec{E}$ at the metal enclosure surfaces defined by $z = \pm L/2$, combined with Eqs. (7,9,11) gives the condition

$$\frac{dZ^e}{dz} = 0 \quad \text{at} \quad z = \pm\frac{L}{2} \qquad (15)$$

which is true for the equations in all three regions I, II, III. By similar algebraic manipulations, the vanishing of tangential $\vec{E}$ at $\rho = R$ leads to the condition

$$R^e = 0 \quad \text{at} \quad \rho = R \qquad (16)$$

which applies to solutions in region III. Furthermore, singlevaluedness of the physical fields requires the separation constant m to be an integer, which also corresponds to the whispering gallery mode number discussed previously.

Note that application of the boundary condition Eq. (15) on Z^e will determine the constants of separation λ^e, or eigenvalues, for each of the three separate regions. In fact, a whole series of solutions exists in each region

$$\lambda^e \rightarrow \lambda_p^{eN} \quad \text{where} \quad p = 1, 2, 3 \ldots \infty \qquad (17)$$

and the ansatz in each region takes the form

$$A_N(\rho, \phi, z) = \sum_{p=1}^{\infty} C_p^N\, R_N^e(\rho;\lambda_p^{eN})\; Z_N^e(z;\lambda_p^{eN})\; \Phi_N(\phi) \qquad (18)$$

where N specifies the region, and the C_p^N are the coefficients defining the linear combination of eigenfunctions which will be necessary to satisfy the boundary conditions joining the three regions. Explicit details concerning the determination of the eigenvalues λ_p and eigenfunctions for the simple geometry of Figure 1 are contained in the literature. [6]

C. Radial Mode Matching - TE components

Determination of the TE fields is similar to the procedure followed in the previous subsection for the TM fields. However, in the TE case an electric vector potential $\vec{F}$ is defined by

$$\vec{E}^h = -\frac{1}{\varepsilon_0}\vec{\nabla}\times\vec{F} \qquad (19)$$

which satisfies Maxwell's Eq. (4) if the fields are eventually restricted to TE modes. Substitution of Eq. (19) into Eq.(1) demonstrates that an associated scalar potential ϕ can be defined which allows the $\vec{H}$ field to be given by

$$\vec{H}^h = -\vec{\nabla}\phi - i\omega\varepsilon_t\vec{F} \quad . \qquad (20)$$

Substituting these equations for $\vec{E}$ and $\vec{H}$ into Eq. (2) leads to the dynamical equation for the electric vector potential

$$\nabla^2\vec{F} + k_0^2\varepsilon_t\vec{F} = 0 \qquad (21)$$

with the gauge choice

$$\vec{\nabla}\cdot\vec{F} + i\omega\mu_0\varepsilon_0\phi = 0 \qquad (22)$$

to eliminate the scalar field. Restriction to the TE subset of fields is implemented by choosing $\vec{F}$ parallel to the z-axis, $\vec{F} \equiv F\,\hat{n}_z$. For a solution ansatz of the form

$$F(\rho, \phi, z) = R^h(\rho)\, Z^h(z)\, \Phi^h(\phi) \quad (23)$$

the field equation separates into the three independent ordinary differential equations

$$\frac{d^2\Phi^h}{d\phi^2} + m^2\Phi^h = 0 \qquad (24)$$

$$\frac{d^2Z^h}{dz^2} + (k_0^2\varepsilon_t + \lambda^h)\, Z^h = 0 \qquad (25)$$

$$\frac{1}{\rho}\frac{d}{d\rho}\left(\rho\frac{dR^h}{d\rho}\right) - \left(\lambda^h + \frac{m^2}{\rho^2}\right)R^h = 0 \quad (26)$$

where λ^h and m are separation constants to be determined by the resonator boundary conditions.

The vanishing of the tangential $\vec{E}$ field on the resonator enclosure conductive surface when combined with Eqs.(19,22,23) gives the conditions

$$Z^h = 0 \quad \text{at} \quad z = \pm\frac{L}{2} \qquad (27)$$

for all three regions, and

$$\frac{dR^h}{d\rho} = 0 \quad \text{at} \quad \rho = R\,. \qquad (28)$$

As before, m is required to be an integer, and the application of boundary condition Eq. (27) generates a series of eigenvalues given by

$$\lambda^h \rightarrow \lambda_p^{hN} \quad \text{where} \quad p = 1, 2, \ldots\infty \quad (29)$$

and the ansatz in each region takes the form

$$F_N(\rho, \phi, z) = \sum_{p=1}^{\infty} D_p^N\, R_N^h(\rho;\lambda_p^{hN})\; Z_N^h(z;\lambda_p^{hN})\; \Phi_N(\phi) \qquad (30)$$

where N specifies the region and D_p^N are the coefficients defining the linear combination of eigenfunctions which will be required to satisfy the boundary conditions joining the three regions. Again, explicit details concerning the determination of the eigenvalues and eigenfunctions for the simple geometry of Figure 1 are contained in the literature.

D. Physical Mode Solutions

Equations (18) and (30) give the series of eigenmode solutions of Maxwell's equations valid in the three individual regions of Figure 1. It remains to join these solutions at the region interfaces. This will specify the, as yet, undetermined coefficients C_p^N and D_p^N in the expansions. The boundary conditions that must be satisfied are the physical field continuity equations at $\rho = r_1$ and $\rho = r_2$. Specifically, the continuity of tangential $\vec{E}$ and $\vec{H}$ fields at the interfaces yields the conditions

$$E_z^{eI}\big|_{\rho=r_1} = E_z^{eII}\big|_{\rho=r_1} \qquad (31)$$

$$E_\phi^{eI}\big|_{\rho=r_1} + E_\phi^{hI}\big|_{\rho=r_1} = E_\phi^{eII}\big|_{\rho=r_1} + E_\phi^{hII}\big|_{\rho=r_1} \qquad (32)$$

$$H_z^{hI}\big|_{\rho=r_1} = H_z^{hII}\big|_{\rho=r_1} \qquad (33)$$

$$H_\phi^{eI}\big|_{\rho=r_1} + H_\phi^{hI}\big|_{\rho=r_1} = H_\phi^{eII}\big|_{\rho=r_1} + H_\phi^{hII}\big|_{\rho=r_1} \qquad (34)$$

and

$$E_z^{eIII}\big|_{\rho=r_2} = E_z^{eII}\big|_{\rho=r_2} \qquad (35)$$

$$E_\phi^{eIII}\big|_{\rho=r_2} + E_\phi^{hIII}\big|_{\rho=r_2} = E_\phi^{eII}\big|_{\rho=r_2} + E_\phi^{hII}\big|_{\rho=r_2} \qquad (36)$$

$$H_z^{hIII}\big|_{\rho=r_2} = H_z^{hII}\big|_{\rho=r_2} \qquad (37)$$

$$H_\phi^{eIII}\big|_{\rho=r_2} + H_\phi^{hIII}\big|_{\rho=r_2} = H_\phi^{eII}\big|_{\rho=r_2} + H_\phi^{hII}\big|_{\rho=r_2}. \qquad (38)$$

These equations for the field constraints yield linear relations between the coefficients C_p^N and D_p^N in the expansions of Eqs. (18) and (30) respectively. Determination of these relations requires calculation of the physical $\vec{E}$ and $\vec{H}$ fields from the vector potential expansions, and the subsequent use of the orthogonality conditions of the eigen-

functions $Z_N^{e,h}$ to obtain scalar relations between the coefficients. Implementation of this procedure yields equations of the sort

$$C_p^I = \left[M_{I,II}^e\right]_{pp'} C_{p'}^{II} \qquad (39)$$

$$D_p^I = \left[M_{I,II}^h\right]_{pp'} D_{p'}^{II} \qquad (40)$$

from Eqs. (31) and (33), and equations of the form

$$\left[R_{II,I}^e\right]_{pp'} C_{p'}^I + \left[R_{II,I}^h\right]_{pp'} D_{p'}^I = \left[X_{I,II}^e\right]_{pp'} C_{p'}^{II} + \left[X_{I,II}^h\right]_{pp'} D_{p'}^{II} \qquad (41)$$

from Eq. (32), for example. The matrices M, R, X, etc are functions of the constants determined by the resonator geometry and properties. The constraints imposed on the coefficients can be rewritten in a (deceptively) simple form

$$\left[\Gamma\right]_{pp'} \begin{bmatrix} C^{II} \\ D^{II} \end{bmatrix}_{p'} = 0 \qquad (42)$$

where equations of the sort of Eqs. (39,40) have been used to eliminate the occurrence of the coefficients from regions I and III. In order that these constraints imposed upon the expansion coefficients are satisfied, it is required that

$$\det(\Gamma) = 0 \ . \qquad (43)$$

The physical solutions are thus found by varying the eigenmode frequency ω, and searching for those discrete frequencies where Eq. (43) is satisfied. For each of these frequencies, Eq. (42) can be solved for the expansion coefficients of the vector potentials, and the physical fields determined. These expressions for the fields can then be used to calculate such mode properties as dielectric and wall losses, and relevant perturbation phenomena.

E. Sapphire Dielectric Constants

Accurate determination of eigenmode frequencies, i.e. to less than a tenth of a percent, requires an accurate knowledge of the sapphire dielectric constants as a function of temperature. The analysis described in the previous subsections has been used in conjunction with measurements of eigenmode frequencies as a function of temperature to accurately determine these constants for the sapphire used in our work. To determine $\varepsilon_t(T)$, the frequency of an $m=0$ pure TE mode of a simple Courtney-style sapphire puck in a copper enclosure was measured between $T=40K$ and $T=300K$. The room temperature resonator dimensions were: $t=1.0623\ cm$, $r_1=0.0\ cm$, $r_2=0.6915\ cm$, $L=1.0626\ cm$, and $R=2.5400\ cm$. Given the accurately measured sapphire and enclosure dimensions, and correcting for the thermal contraction of the cooled sapphire at each measurement temperature, [7] the parameter $\varepsilon_t(T)$ was adjusted in the modeling program to match the calculated eigenmode frequency with the measured frequency. The results of this procedure are displayed in Figure 2.

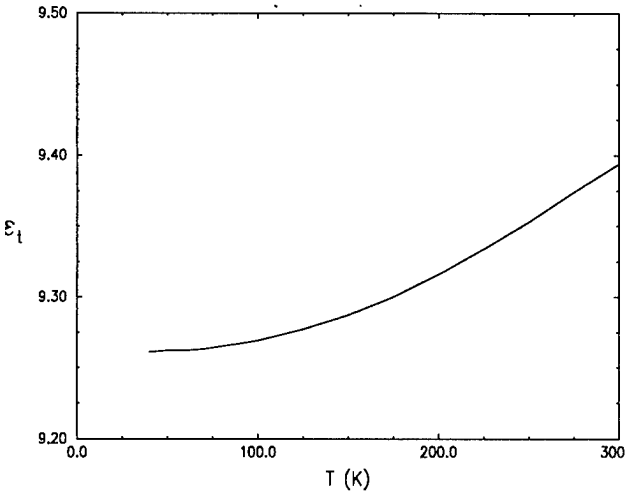


Figure 2. ε_t vs. temperature, as determined from the measured mode frequencies using the modeling program.

To determine $\varepsilon_z(T)$, the frequency of an $m=7$ whispering gallery mode with electric field strongly polarized in the z-direction was measured between $T=40K$ and $T=300K$. The room temperature dimensions of the sapphire ring and copper enclosure were: $t=0.7584\ cm$, $r_1=0.9204\ cm$, $r_2=1.2709\ cm$, $L=2.3000\ cm$, and $R=2.5400\ cm$. As before, for the specified geometry, and correcting for the thermal contraction of the cooled sapphire at each temperature, the parameter $\varepsilon_z(T)$ was adjusted in the modeling program to match the calculated eigenmode frequency with the measured frequency. The results of this procedure are displayed in Figure 3.

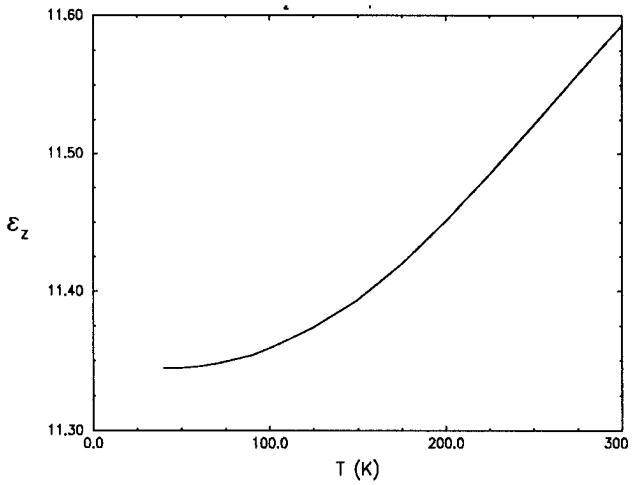


Figure 3. ε_z vs. temperature, as determined from the measured mode frequencies using the modeling program

Polynomial fits to the curves in Figures 2 and 3 yield the following accurate expressions for $\varepsilon_t(T)$ and $\varepsilon_z(T)$:

$$\varepsilon_t(T) = 9.2614 - 3.5467\times10^{-5}T + 3.4466\times10^{-7}T^2 + 9.6381\times10^{-9}T^3 - 1.8255\times10^{-11}T^4 \quad (44)$$

$$\varepsilon_z(T) = 11.3492 - 1.9942\times10^{-4}T + 1.5977\times10^{-6}T^2 + 1.7237\times10^{-8}T^3 - 3.7723\times10^{-11}T^4 \quad (45)$$

Measurements

Following the development of suitable analytical design tools, a number of other issues must be resolved before a complete design can be adopted :

- Confirmation of computational accuracy
  - mode frequencies
  - loss coefficients
- Investigation of extraneous loss contributions
- Thermal design
- Mechanical/vibrational design
- Oscillator system design
- Cryogenics

At the time of this report, we have made substantial progress on the first two steps but most of the other activities remain in preliminary stages. Our ultimate goal is to operate the resonator with a closed-cycle cryocooler which could be installed in a rack-mounted instrument. However, for development purposes, a closed cycle cryocooler system would be extremely inconvenient because of the lengthy cooldown time. To facilitate development measurements, we have constructed a pair of continuous-flow cryostats which use liquid helium as the coolant. These are built with rigid transfer tubes attached directly beneath them, so the entire assembly can be lowered over a liquid helium storage dewar. This results in simple construction, ample working volume, efficient operation and the ability to cool from room temperature to below 10 K in about 20 minutes.

A. Mode Frequency Confirmation

Clearly, the simplest test of the accuracy of the design calculations is simply to check the mode frequencies of sapphire resonators placed in suitable enclosures. A sapphire ring and enclosure with dimensions: $t = 1.033\ cm$, $r_1 = 0.986\ cm$, $r_2 = 1.691\ cm$, $L = 2.900\ cm$ and $R = 3.55\ cm$ have been constructed. These dimensions were selected by means of an iterative optimization procedure based on the computational package described above. The optimization objective was to produce a resonator with a $WGH_{7,0,0}$ mode at 9.60 GHz with an unloaded Q of more than 20 million at 77 K and the furthest possible separation between this mode and its nearest neighbors. A key aspect of this optimization is the adjustment of the chamber size. If the chamber is small, the mode density near 9.6 GHz will be low and it will be easy to make small adjustments to the various dimensions and produce a large 'window', but the conductive losses on the chamber walls will be high. A larger chamber will have small wall losses with a very dense spectrum of spurious modes. According to the calculations, the selected dimensions will produce a 'window' with a width of 500 MHz. The sapphire was

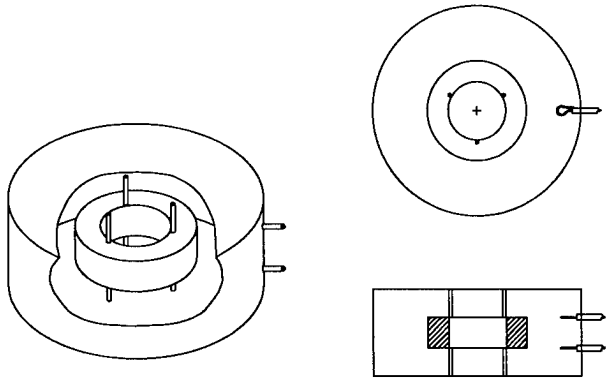


Figure 4.Three views of the sapphire whispering gallery resonator and its enclosure showing the support designed for Q and mode-frequency measurements.

purchased as HEMLUX grade material from Crystal Systems. [8] It was finished to the final size and shape by Jae Glass and Ceramic Grinding. [9] It was left in an unpolished condition with an estimated surface finish of ~30 microinch.

For the purposes of these tests, the sapphire was held in place with an arrangement of 1 mm diameter sapphire rods. These are placed so that three rods project vertically through the lower cover, and the ring is placed so that it rests on them. Then the upper cover is installed and three more rods are inserted through the upper cover and external springs are used to supply a substantial clamping force to hold the ring firmly in place. By observing the frequency shift that occurs when the upper rods are inserted and removed, it was verified that the presence of these rods had negligible influence on the various mode frequencies.

Figure 5 shows two transmission spectra obtained with this configuration. For both measurements, the center frequency was set to the $WGH_{7,0,0}$ whispering gallery mode. In the upper trace, the sweep span was 5 GHz and for the lower trace, it was 0.6 GHz. For the upper trace, only relatively low Q modes are visible, because the resolution exceeded the bandwidth of the whispering gallery modes by several orders of magnitude. In fact, the $WGH_{7,0,0}$ mode is visible in the lower trace only because the network analyzer was programed to dwell at exactly the previously measured frequency of that mode. The prediction error for the frequency of the $WGH_{7,0,0}$ mode was no more than one part per thousand. As can be seen from Figure (5b), the location of the nearest neighbor modes proved to be in completely satisfactory agreement with the calculations. It is evident from Figure (5a) that the presence of a 500 MHz wide gap in the mode spectrum is an anomaly, and must be intentionally designed into the resonator mode spectrum. This vividly demonstrates the virtue of design by computation as opposed to trial-and-error construction.

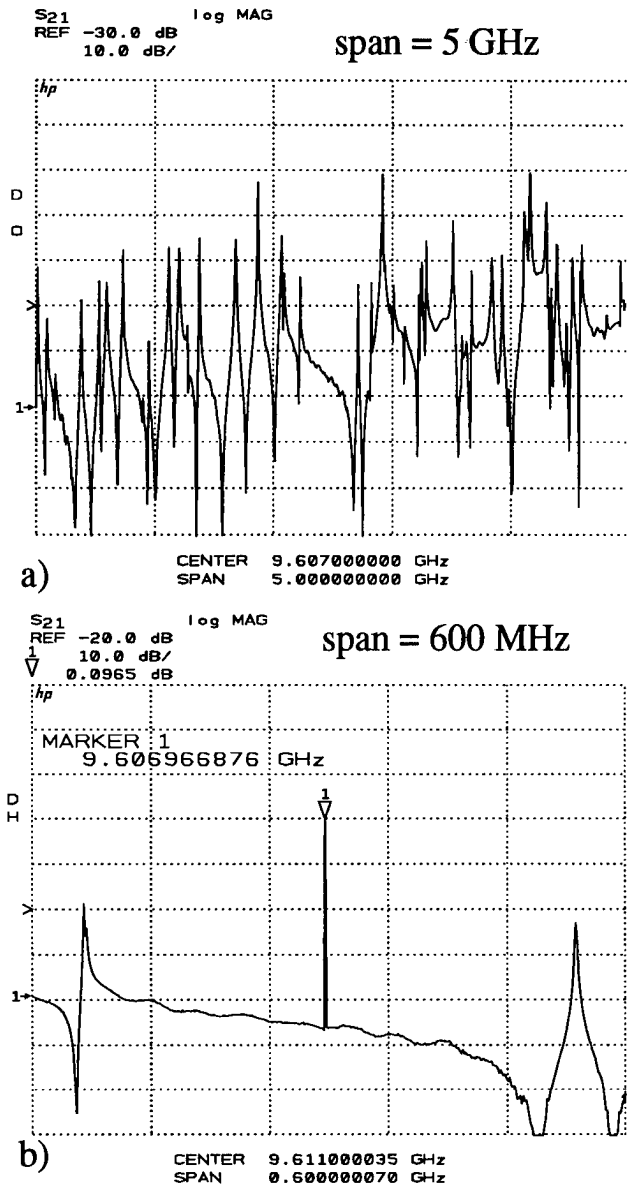


Figure 5. Transmission spectra measured in the vicinity of the $WGH_{7,0,0}$ mode.

B. Residual Losses and Loss Coefficient Confirmation

A more stringent test of the mode calculations is obtained by verifying the loss coefficients associated with the resistive losses of the enclosure. In order to perform this check, a brass test enclosure was fabricated. Q values of the TE_{011} modes of the empty brass and copper chambers were measured in order to estimate the surface resistance of the two materials at low temperature. Then Q measurements were performed on a sapphire ring placed first in one and then the other enclosure. Based on this procedure, the experimentally determined loss coefficient for the $WGH_{7,0,0}$ mode was $1.16\times10^{6}\Omega$. The predicted value for this coefficient was $1.78\times10^{6}\Omega$. The discrepancy between these two values is consistent with experimental uncertainties in the various steps of the measurement procedure.

With a reasonably accurate estimate of the loss contribution from the walls, it is next possible to address the question of whether or not there are extraneous contributors to the overall loss. For the temperature range of interest (40 K to 80 K) the intrinsic loss tangent of sapphire is known,[5] but there is very little published information on the specific requirements that must be met in order to obtain loss values that are consistent with the intrinsic values. At sufficiently low temperatures, there is likely to be

some residual loss mechanism that will exceed the intrinsic losses. For our purposes, we are primarily interested in temperatures between 30 and 80 K. For the resonators that we have studied, there has generally been a moderate amount of extraneous loss evident at temperatures below 70 K, but the precise origin remains unclear. Figure 6 displays the Q values for two modes of one of the ring resonators along with the intrinsic losses estimated from the data of Braginsky, et al. In spite of the fact that sapphire's intrinsic dielectric loss increases with frequency, the mode with the higher frequency achieves significantly higher Q values. These results clearly show that the $WGH_{7,0,0}$ mode is strongly influenced by some extrinsic loss factor. This is due, in part, to the fact that the mode with the higher index will provide stronger confinement. However, the experimentally verified wall loss contribution at 9.6 GHz is not quite strong enough to fully account for the observed Q. Surface cleanliness and surface damage may be significant factors, but further work will be necessary to resolve this question.

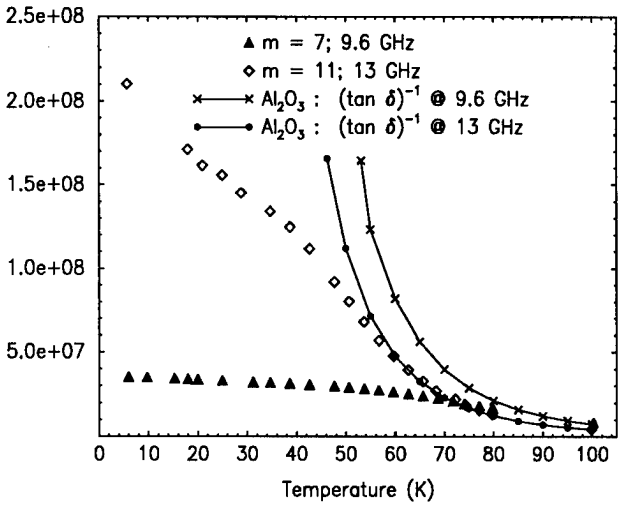


Figure 6. Q values as a function of temperature for two of the whispering gallery modes compared to the reciprocal of the loss tangent of sapphire at the corresponding frequencies.

C. Preliminary Phase Noise Measurements

We have not yet attempted to finalize either the thermal or mechanical designs for the enclosure and support arrangements, but, in order to obtain baseline information relevant to this design process, we have begun performing a series of preliminary phase noise measurements on oscillators incorporating cryogenic sapphire resonators. We anticipate that the whispering gallery configuration will ultimately offer excellent vibrational immunity due to the fact that the electromagnetic field levels at the inside surfaces of the

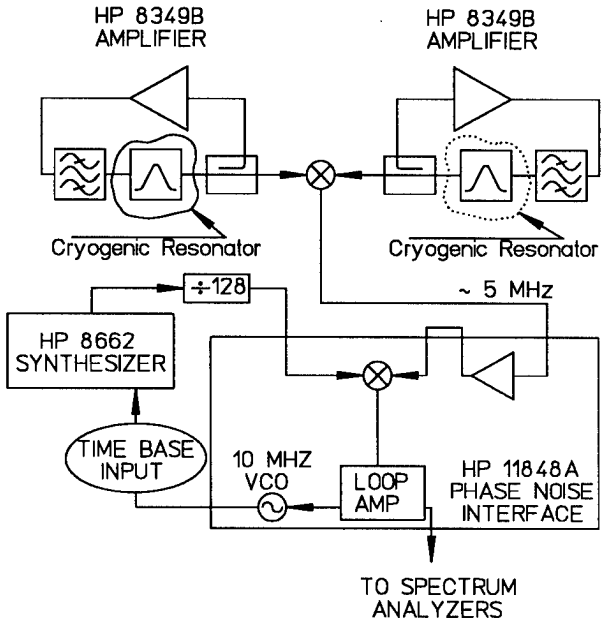


Figure 7. Measurement configuration for evaluating the phase noise of a pair of oscillators.

enclosure are extremely low. However, the existing mechanical support using 1 mm sapphire posts (as described above) is known to provide minimal lateral stiffness, so it is reasonable to expect that the existing design will display significant vibrational sensitivity. The expedient approach to developing an improved design is to adopt phase noise measurement protocols in conjunction with the existing design and then iteratively evaluate trial designs by means of phase noise measurements performed on each trial configuration.

For the purposes of these measurements, we place each resonator and a general purpose laboratory microwave amplifier (HP 8349B) in a simple two-port feedback oscillator circuit. All components of the oscillators remain at room temperature with the exception of the resonators themselves. In order to enforce oscillation on the intended modes, two-pole cavity filters with bandwidths of approximately 50 MHz are placed in the circuits. Eventually, these can be replaced with simpler filters with somewhat larger bandwidths. Phase noise evaluations are performed by use of an HP 3048A phase noise measurement system. For the purposes of this measurement, the 3048A system consists of an 11848A phase noise interface, an HP 8662 microwave synthesizer and a computer workstation with software suitable for performing calibrated phase noise measurements by means of a phase-locked loop configuration. In order to obtain optimal performance from the microwave synthesizer, it is set to an appropriate frequency near 640 MHz and its output is passed through a frequency divider set for a divide ratio of 128. Two free-

running cryogenic sapphire oscillators are operated with a frequency separation of approximately 5 MHz. The output signals from the two oscillators are applied to a mixer and the resultant 5 MHz signal is delivered to the 3048A system. In order to measure the phase noise of the 5 MHz signal, the 3048 system establishes a phase-locked loop which applies the loop error signal to a 10 MHz VCO (the 10 MHz 'B' oscillator was used for these measurements) which serves as an external time base for the 8662 synthesizer. Based on separate measurements, we believe that the far-from-carrier (f >= 2kHz) measurement floor for this particular system is approximately -165dBc/Hz.In the measurements performed so far, the thermal and vibrational sensitivity are readily apparent. In order to diminish the thermal regulation problems, most measurements have been performed at 35 K or below. The present temperature regulation is based on a commercial silicon diode thermometer and temperature controller, and the temperature stability achieved is approximately ±0.001 K. The silicon diode thermometers are not optimized for stable temperature regulation in this range, so they will be replaced with thin film resistance thermometers.

Figure 8 displays the results of a phase noise measurement for two free-running oscillators utilizing the $WGH_{7,0,0}$ mode near 9.6 GHz. The loaded Q was approximately $2.\times10^7$ and the resonator temperatures were approximately 28 K. For offset frequencies between 10 Hz and 1 kHz, the phase noise is dominated by vibrational sensitivity which is attributable to the excessive lateral compliance of the present mechanical support arrangement. In order to substantiate the role of mode confinement in limiting vibrational sensitivity, phase noise measurements were also performed using the $WGH_{11,0,0}$ modes at 13 GHz in the same resonators used for the measurement shown in Figure 8. The results, displayed in Figure 9, show a significant reduction in the vibrational spurs under virtually identical measurement conditions. For the 13 GHz measurement, the loaded Q values were approximately $3.\times10^7$ and the temperature was 35 K. Minor improvements in the measurement configuration resulted in a slightly lower far-from-carrier floor for the data in Figure 9. The $1 / f^3$ component of the phase noise is consistent with the measured value of the loaded Q and the measured amplifier 1 / f phase noise, so there is no evidence that the frequency flicker of the resonators themselves has yet been revealed. This plot has not been modified to reference a single oscillator, so, according to the usual assumptions, one can assume that the performance of an individual oscillator is at least 3dB lower than the plotted values.

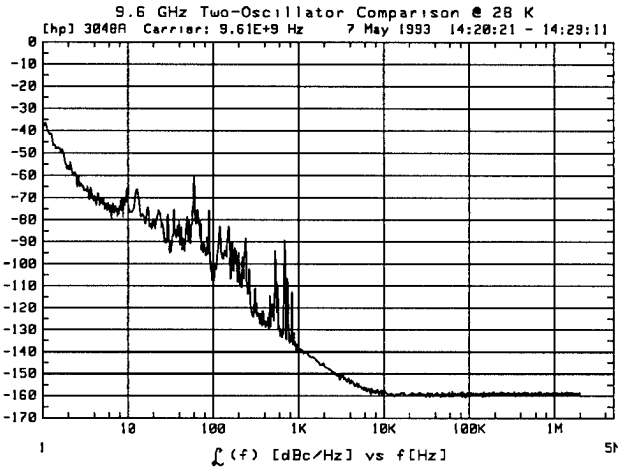


Figure 8. Phase noise of the beat signal from two cryogenic sapphire oscillators near 9.6 GHz.

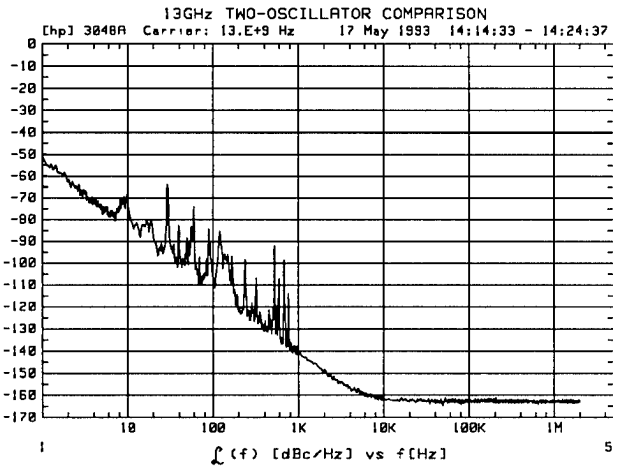


Figure 9. Phase noise of the beat signal from two cryogenic sapphire oscillators near 13 GHz.

Conclusion

It has been verified that accurate calculation of resonance modes in a whispering gallery geometry can be used to place a specified mode precisely within a clean window in the relatively dense mode spectrum. Oscillators built incorporating these whispering gallery resonators have demonstrated very encouraging phase noise performance.

Further work will include a careful mechanical design to improve microphonic performance, and studies of potential sources of residual Q degradation beyond the intrinsic sapphire loss levels.

Acknowledgment

It is a pleasure to acknowledge the continued support and encouragement of R. Temple, T. Faulkner, L. Cutler, and J. Hollenhorst of Hewlett-Packard throughout this work. We also thank J. Kral for help on the cryostat design and construction.

References

[1] M. M. Driscoll, "Low-Noise Microwave Signal Generation Using Bulk- and Surface-Acoustic-Wave Resonators," IEEE Trans. Ultrason. Ferroelec. Freq. Contr., 35, 426 (1988).

[2] M. M. Driscoll, J. T. Haynes, S. S. Horwitz, R. A. Jelen, R. W. Weinert, J. R. Gavaler, J. Talvacchio, G. R. Wagner, K. A. Zaki, and X.-P. Liang, "Cooled Ultra-high Q, Sapphire Dielectric Resonators for Low Noise, Microwave Signal Generation," Proc 45th Annual Symp. on Freq. Contr., 700 (1991).

[3] A. J. Giles, A. G. Mann, S. K. Jones, D. G. Blair, and M. J. Buckingham, "A Very High Stability Sapphire Loaded Superconducting Cavity Oscillator," Physica B, 165, 145 (1990).

[4] G. J. Dick and J. Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Dielectric Ring Resonator for Ultra-Low Noise," IEEE Trans. Ultrason. Ferroelec. Freq. Contr., 37, 339 (1990).

[5] V. B. Braginsky, V. S. Ilchenko, and Kh. S. Bagdassarov, "Experimental Observation of Fundamental Microwave Absorption in High-Quality Dielectric Crystals," Phys. Lett. A, 120, 300 (1987).

[6] D. Kajfez and P. Guillon, "Dielectric Resonators," Artech House, (1986).

[7] Amer. Instit. of Physics Handbook, McGraw-Hill, NY (1972).

[8] Crystal Systems Inc., Salem, Mass.

[9] Jae Glass and Ceramic Grinding, Sunnyvale, Ca.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

CLOSED LOOP TESTS OF THE NASA SAPPHIRE PHASE STABILIZER

David G. Santiago and G. John Dick
California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Pasadena, California 91009, U.S.A.

Abstract

We have developed a cryogenic Sapphire Phase Stabilizer (SPS) to meet microwave oscillator phase noise requirements. The SPS employs a high Q, X-band sapphire dielectric "whispering gallery" mode resonator as a discriminator to stabilize a quartz crystal oscillator. At an untuned frequency of 7.9449 GHz with a loaded Q of 6 million we previously reported an "open loop" discriminator noise floor (referred to 100 MHz) of approximately $S_{\phi}(f) = -110\ dB/f^3(/Hz)$ for offset frequencies from f= 1 Hz to f= 1 kHz.

Precise tuning of the sapphire resonant frequency now allows implementation of simplified control loops together with suppressed-carrier phase sensing circuitry. These improvements make possible an ultra-low noise demonstration of closed-loop SPS performance. From 1 Hz to 1 kHz a comparison of the SPS with a quartz crystal reference oscillator of the highest quality showed only the noise of the reference oscillator. To our knowledge these results represent the lowest phase noise in a closed loop or active sapphire oscillator to date at temperatures achievable with liquid nitrogen (77K or higher).

Introduction

The SPS is based on a novel technology consisting of a cooled sapphire "whispering gallery" mode X-band resonator operating in the temperature range from 70 Kelvin to 300 Kelvin - values achievable by means of radiative and thermoelectric cooling.

Sapphire oscillator technology is presently under development in a number of laboratories, with experimental results for designs at 300K[1,2,3], 77K[1,3,5], 35K[6], and liquid helium temperature[7,8,9] being reported. Oscillator configurations include discriminators, active oscillators, and stabilized local oscillators (STALO) with dc or ac (Pound) frequency sensing. Suppressed carrier techniques have been applied to active (bridge) and passive (STALO) configurations.

The SPS employs a sapphire resonator operating at 80 Kelvin in a suppressed-carrier STALO configuration. We previously reported an open loop measurement of the phase noise of a quartz oscillator of the highest quality, with a noise floor reduced by 10 to 45 dB for offset frequencies from 1 Hz to 1000 Hz[1]. By accurately tuning the sapphire resonator, and with simplified circuitry thus allowed, we are now able to make a similar closed loop comparison.

While the previous tests used an untuned sapphire element, our system is designed to operate critically coupled at 8.1000 GHz with a loaded Q of 15 million at 77 Kelvin. Advantages of this system include a simplified electronic configuration resulting in reduced phase noise. A 3-stage methodology has been developed to provide the accurate frequency tuning the SPS system requires.

Resonator Tuning

The sapphire "whispering gallery" mode resonator frequency can be tuned in three stages. The first stage of tuning is reduction of the physical size of the sapphire resonator wheel by machining down the outer diameter and thickness. The $WGH_{10,0,0}$ mode used in our resonator confines ten azimuthal cycles within the sapphire wheel. Reduction of the wheel's dimensions simply reduces the available path length for the 10 cycles thereby increasing the resonant frequency. The goal is to reduce the sapphire wheel such that its natural resonant frequency for the $WGH_{10,0,0}$ mode at 77K is just above 8.1000 GHz. The remaining two tuning methods can then be used to lower the resonant frequency to 8.1000 GHz. The machining process was done in steps so as not to overshoot our capability to fine tune the resonant frequency. Figure 1 shows the resonant frequency of the resonator as predicted and as measured for each iteration of machining. A simple model was used to choose the reduction in each of the sapphire's dimensions and predict the new resonant frequency.

The second tuning method reduces the resonator's operational frequency by suspending a sapphire disk above the resonator wheel as in Figure 2. The additional dielectric material present in the resonant fields decreases the resonant frequency. Mechanical tuning with the sapphire disk used in our experiment reduced the resonant frequency up to ≈30 MHz (if positioned directly on top of the resonator). The sapphire tuning disk is attached to a micrometer drive so mechanical tuning can be done while the resonator is operating. This technique provides a large tuning range, but is susceptible, and very

This work was carried out at the Jet Propulsion Labooratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

0-7803-0905-7/93 $3.00 © 1993 IEEE

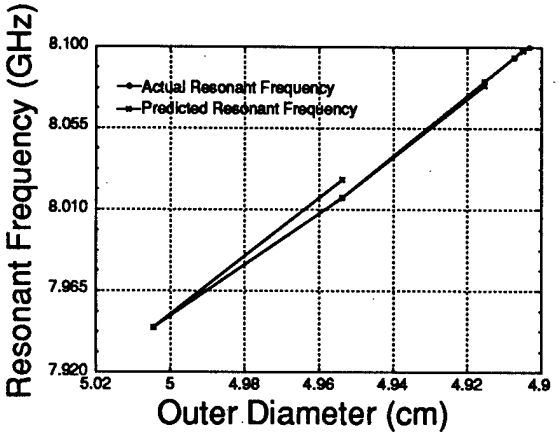


Figure 1. Resonant Frequency as a Function of Resonator Outer Diameter

sensitive, to vibration. The tuning disk must be held extremely rigidly so as to remain parallel to the resonator wheel. The resonator's frequency stability is determined by the tuning disk's mechanical/vibrational stability[10].

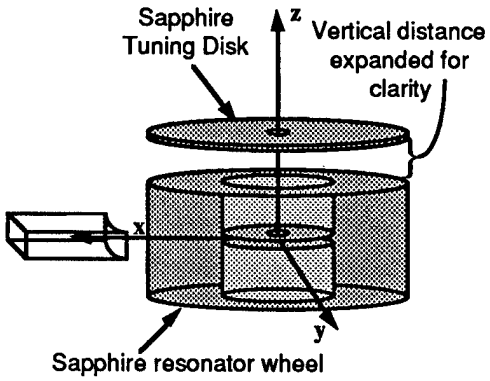


Figure 2. Mechanical Tuning with disk above resonator wheel

The third tuning method provides fine tuning of the resonator's frequency by thermally controlling the resonator. The thermal expansion of the sapphire reduces the resonant frequency. The sapphire wheel as pictured in Figure 2 sits on a copper post wrapped with a heater wire. The copper post sits in a metal cylinder of lesser thermal conductivity. This cylinder is heat sunk to the outer copper resonator containment can which is cooled by liquid nitrogen. In this fashion we can cool the sapphire to liquid nitrogen temperature or warm it to a few Kelvin above 77K. Figure 3 shows an example of thermally tuning resonant frequency. Resonant frequency and resonator temperature are shown as a function of the heater power applied in the resonator.

Figure 3 shows approximately 300 kHz of tuning capability. Therefore the coarse tuning methods must bring the resonant frequency to within several hundred kHz. For the heater assembly currently installed in the resonator we calculate a maximum operational heater

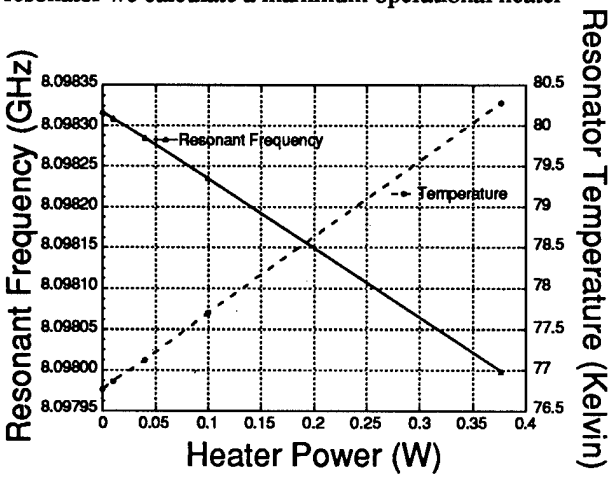


Figure 3. Resonator Response to Heater Power

wire current as seen in Figure 4. At maximum heater power the resonator would operate at ≈86K with resonant frequency reduced by ≈1 MHz.

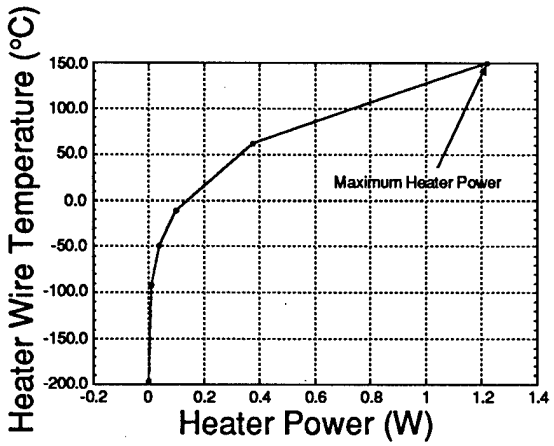


Figure 4. Radiation Limit of Heater Wire Temperature vs. Heater Power

For reduced vibrational sensitivity and therefore optimum phase noise performance, we opted to machine the sapphire resonator wheel to within 500 kHz above 8.100 GHz so that thermal tuning brings the resonator to the operational frequency.

Coupling and Quality Factor

Operation at the design frequency is essential for optimization of the electronic configuration, but more critical to the performance as a low noise stabilized oscillator are the characteristics of the discriminator. The discriminating power is determined by the high Q cavity operating in a critically coupled condition. Critical coupling with a loaded Q of 15 million requires both a sapphire intrinsic Q and resonator coupling Q of 30

million. Our design uses waveguide coupling ports operating below natural cutoff due to Teflon dielectric loading of the waveguide. The length of the Teflon insert determines the coupling to the sapphire resonator wheel. Loading the entire waveguide port produces an overcoupled condition and therefore a low coupling Q. Loading a portion of the waveguide port can provide us the critical coupling we desire, although other more subtle aspects of the rf transmission system also effect the coupling Q.

The sapphire resonator's intrinsic Q is expected to be approximately 30 million at 77K, but the actual intrinsic Q is affected by impurities, resonator geometry and alignment, and temperature. We have operated the sapphire resonator very near critical coupling at loaded Q's around 12 million. The coupling and quality factors are determined from simple measurements. The time constant required for a 3 to 1 decay in signal voltage amplitude (τ_{31}) is measured by pulsing the X-band drive signal and monitoring the decaying reflected signal from the resonator. The maximum voltage amplitudes of the reflected resonator input and output (V_{in},V_{out}) are also measured. Given the resonant frequency (f) the following calculations are performed to determine the measured (loaded), intrinsic and coupling time constants (τ_m,τ_i,τ_c), and the respective quality factors (Q_m, Q_i, Q_c).

$$\tau_m = \frac{\tau_{31}}{2 \cdot \ln(3)} \quad (1)$$

$$\tau_i = \tau_m \cdot \left[1 + \frac{1}{\left(2 \cdot \frac{V_{out}}{V_{in}} - 1 \right)} \right] \quad (2)$$

$$\tau_c = \tau_i \cdot \left(2 \cdot \frac{V_{out}}{V_{in}} - 1 \right) \quad (3)$$

$$Q_m = 2 \cdot \pi \cdot f \cdot \tau_m \quad (4)$$

$$Q_i = 2 \cdot \pi \cdot f \cdot \tau_i \quad (5)$$

$$Q_c = 2 \cdot \pi \cdot f \cdot \tau_c \quad (6)$$

These parameters are regularly monitored to evaluate the various systematic changes made to satisfy design and performance goals.

Electronics

The high Q resonator is implemented in the Sapphire Phase Stabilizer system sketched in Figure 5 which includes suppressed carrier feedback circuitry[11]. The SPS circuitry is designed for optimum sensitivity and phase noise performance at critical coupling. When under or overcoupled the signal returned from the resonator is not a small (nominally zero) signal. Therefore implementation of a rf feedback amplifier requires pre-processing of the resonator return signal. Attenuation and filtering successfully artificially produce the small signal feedback, but some of the suppressed carrier sensitivity is sacrificed.

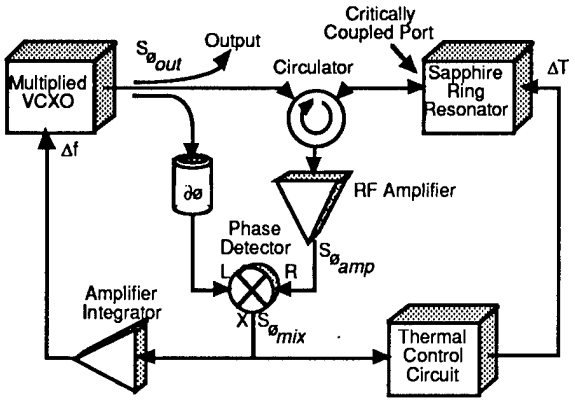


Figure 5. Block Diagram of Sapphire Phase Stabilizer (SPS)

Phase Noise Measurements

Tests of the SPS were made at 80 Kelvin using the circuit shown in Fig. 5 without the rf feedback amplifier and with the circulator replaced by a 3 dB hybrid coupler. The hybrid allows lower noise operation at the expense of signal strength. The under coupled resonator with a loaded Q of approximately 8 million dictated initial testing without suppressed carrier implementation. The current tuning step gave an operational frequency of 8.0995 GHz which forced use of a frequency synthesizer during phase noise measurements. Figure 6 shows the SPS phase noise (referred to 100 MHz) measured at 100 MHz and at X-band. The phase noise measured at 100 MHz was limited by the HP3325A synthesizer used to generate the measurement system's local oscillator frequency. Measurement of the phase noise at X-band, however, reduced the system sensitivity to the synthesizer noise, thereby lowering the measurement system noise floor. Unfortunately the X-band measurement is also limited by the measurement system noise floor, so the actual phase noise performance of the SPS without the rf feedback amplifier is not known at this time.

Also displayed in Figure 6 is the phase noise of the Vectron VCXO used in the SPS. This phase noise was measured by using the resonator as an open loop discriminator. The open loop discriminator noise floor was not measured at that time, but is expected to be similar to previous measurements [1].

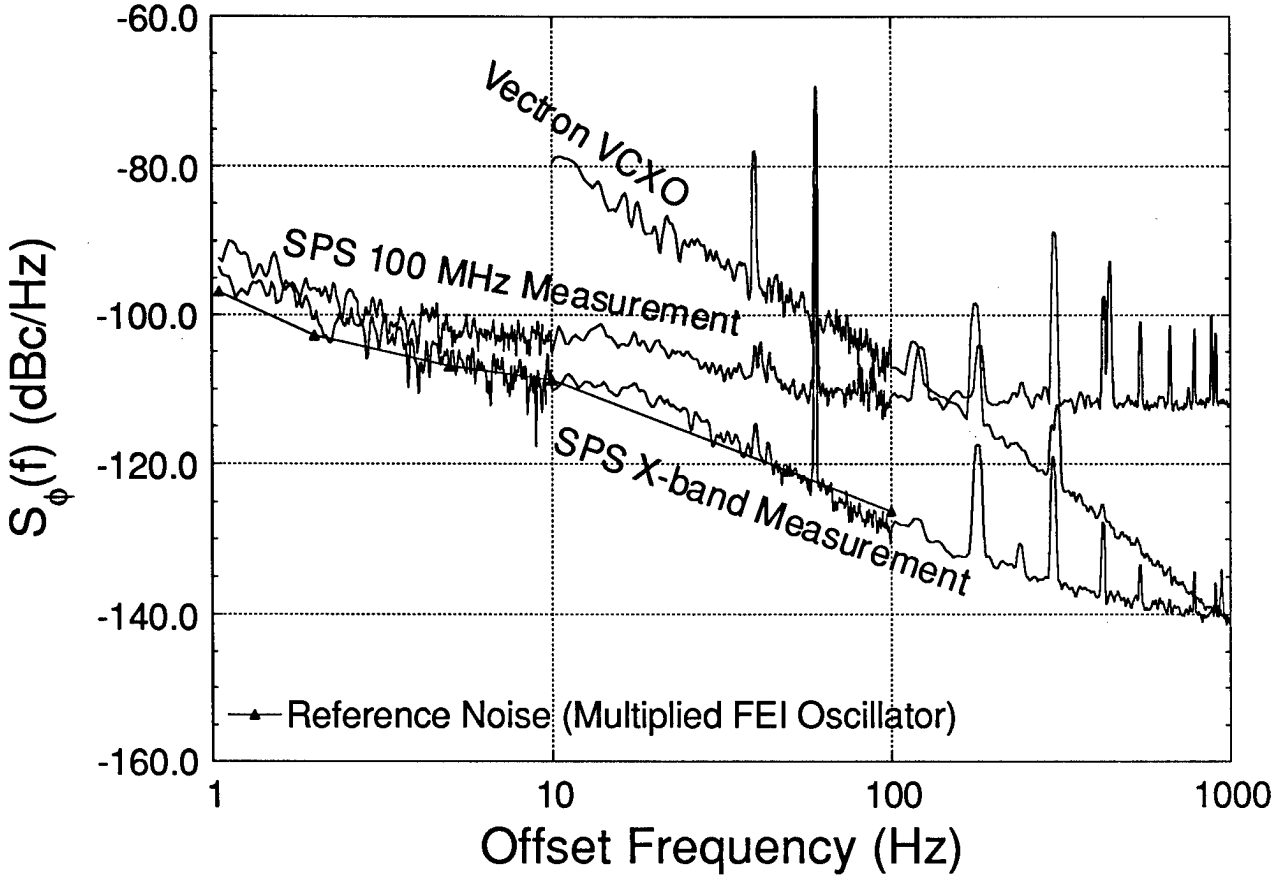


Figure 6. Closed Loop SPS Phase Noise referred to 100 MHz

Conclusion

A closed-loop demonstration of Sapphire Phase Stabilizer performance has been made possible by accurate tuning of the sapphire resonator. In this demonstration, the SPS output frequency was compared with an ultra-low noise quartz crystal reference oscillator, with the results showing only the phase noise of the reference oscillator. This represents the lowest phase noise measurement for a sapphire oscillator operating at temperatures achievable by cooling with liquid nitrogen.

Actual phase noise for the SPS was not measured in this test, only determined to be less than that of the reference oscillator. A second sapphire resonator is presently being tuned to allow this performance to be accurately determined.

Acknowledgments

The authors would like to express their gratitude to R. T. Wang for cryogenic and rf expertise, A. Kirk for assistance with phase noise tests, and T. K. Tucker for assistance with system assembly.

References

[1] D. G. Santiago and G. J. Dick, "Microwave Frequency Discriminator with a Cooled Sapphire Resonator for Ultra-Low Phase Noise," *Proc. 1992 IEEE Frequency Control Symposium*, 176-182 (1992).

[2] D. P. Tsarapkin, "An Uncooled Microwave Oscillator with 1-Million Effective Q-Factor," *These Proceedings.*

[3] M. E. Tobar, A. J. Giles, S. Edwards and J. Searls, "High Q TE Stabilized Sapphire Microwave Resonators for Low Noise Applications," *These Proceedings.*

[4] M. M. Driscoll and R. W. Weinert, "Low Noise, Microwave Signal Generation Using Cryogenic, Sapphire Dielectric Resonators: An Update," *Proc. 1992 IEEE Frequency Control Symposium*, 157-162 (1992).

[5] A. G. Mann, et al, "Ultra-Stable Cryogenic Sapphire Dielectric Microwave Resonators," *Proc. 1992 IEEE Frequency Control Symposium*, 167-171 (1992).

[6] C. A. Flory and R. C. Taber, "Microwave Oscillators Incorporating Cryogenic Dielectric Resonators," *These Proceedings.*

[7] M. E. Tobar and D. G. Blair, "Phase Noise Analysis of the Sapphire Loaded Superconducting Niobium Cavity Oscillator," *These Proceedings.*

[8] V. I. Panov and P. R. Stankov, "Frequency Stabilization of oscillators with high-Q leucosapphire dielectric resonators," *Radiotekhnika i Electronika* 31, 213, (1986), (In Russian).

[9] R. T. Wang and G. J. Dick, "Improved Performance of the Superconducting Cavity Maser at Short Measuring Time," in *Proc. 44th Ann. Symposium on Frequency Control*, 89-93 (1990).

[10] M. E. Tobar and D. G. Blair, "Phase Noise Analysis of the Sapphire Loaded Superconducting Niobium Cavity Oscillator," to be published in *IEEE Trans. Microwave Theory and Techniques*, vol. 42, No 2, 1994.

[11] G. J. Dick and J. Saunders, "Method and Apparatus for Reducing Microwave Oscillator Output Noise," U.S. Patent #5,036,299, July 30, 1991.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

AN UNCOOLED MICROWAVE OSCILLATOR WITH 1-MILLION EFFECTIVE Q-FACTOR

D.P.Tsarapkin

Radio Engineering faculty
Moscow Power Engineering Institute
Moscow 111250, RUSSIA

Abstract: Low phase noise output signals are very important for the most of microwave communication systems and decidedly critical for CW Doppler radars.

The best centimeter waves sources based on quartz crystals and uncooled sapphire disc dielectric resonators excited on a "whispering gallery" mode (DDR) have near carrier spectral purity performance which is accordingly 20-25 and 10-15 dB worse than need.

The use of cryogenic cooling to improve DDR Q-factor is troublesome mostly. So it looks profitable to get another method to grow Q-factor better. The effective Q-factor increasing in bridge type circuits near their balance phenomenon gives a promising way.

Our calculations reveal the possibility of 10-15 dB flicker phase noise decreasing in a bridge type oscillator.

The report deals with a super low phase noise S-band bipolar transistor oscillator based on a shielded sapphire DDR (SDDR) used as a filtering circulator bridge. With shielding sizes optimized SDDR Q_o-factor was $\geq 3 \cdot 10^5$ at a room temperature. An effective (loaded) oscillator Q-factor more than $3 \cdot Q_o$, i.e. about 1 million was realized. The oscillator pushing as low as 250 Hz/V and $\mathcal{L}$(1 kHz) near -130 dB/Hz were reached.

It seems the data reported are nowadays tops for uncooled microwave oscillators.

Introduction

Future microwave communication systems and radars will have a need for extremely low phase noise.This need is a permanent challenge to specialists.

The best centimeter waves sources based on a quartz crystal oscillator have spectral purity performance characterized by $\mathcal{L}$(1 kHz) = -(107..112) dB/Hz which is 20-25 dB worse than need.

As well known [1] there are two main ways to improve spectral purity: to decrease an active device phase noise and to increase oscillator circuitry Q-factor. The first way for the particular case of a near carrier noise leads to choice a bipolar transistor (BT) or a Gunn diode as an active device following regime optimization. Sapphire disc dielectric resonators (DDR) excited on a "whispering gallery" mode enable us to obtain directly at microwave frequencies resonator Q value as high as up to $2 \cdot 10^6/f(\text{GHz})$ at room temperature and even much more in the liquid nitrogen to liquid helium temperature range [2-3]. A properly designed C-band BT oscillator stabilized with uncooled sapphire DDR has the ratio of the single-sideband power of the phase noise in a 1-Hz bandwidth 1 kHz away from the carrier (i.e. $\mathcal{L}$(1 kHz)) $\approx$ -120 dB/Hz. It is much better than for a quartz based source but still worse a level wanted.

If cryogenic cooling is undesirable one needs to get another method to grow Q-factor better. The effective Q-factor increasing in bridge type circuits near their balance phenomenon gives a promising way [4].

Bridge Oscillator Analysis

Let us bear in mind a feedback oscillator (fig.1) which consists of two main parts: a forward-amplifying stage and a frequency-selective feedback network.

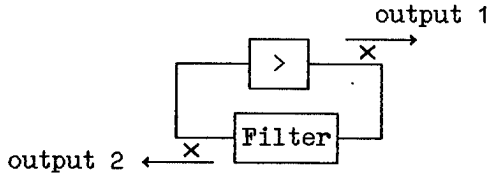


Figure 1 Block diagram of a feedback oscillator.

0-7803-0905-7/93 $3.00 © 1993 IEEE

As shown by many authors an oscillator phase noise level depends on an oscillator effective (loaded) Q-factor as

$$\mathcal{L}(F) \propto Q^{-2}. \qquad (1)$$

The reason of Q-factor appearance in this formula lies in two basic facts:
- an oscillator frequency deviation is directly proportional to steepness of a filter phase-frequency characteristic;
- for a single tank near its resonance the last parameter is equal

$$\varphi' = d\varphi/df \approx -\xi = -2FQ/f_o, \qquad (2)$$

where f_o - a resonant frequency;
$F = (f - f_o)$ - an offset frequency;
f - a current frequency.
So in general case expression (1) has to be rewritten as

$$\mathcal{L}(F) \propto (\varphi')^{-2} \qquad (3)$$

and the problem of phase noise lowering reduces to selection of the circuits with a maximum steep phase characteristic slope. A circuit theory prompts us that the searched circuit might be a bridge one.

Main Properties of Bridge Circuits

According to definition a bridge circuit, or simply a "bridge" is a linear circuit which has at least one zero of a transfer function at a finite frequency. The simplest resonant bridge has one pole and one zero. Its transfer function looks like

$$K_b(p) = C(p + \delta)/(p + a), \qquad (4)$$

where $p = j\xi_b = j2(f - f_b)Q_s/f_b = d/d\tilde{t}$ - a differential operator along normalized time $\tilde{t}$; C - constant;
f_b - a bridge central frequency;
$Q_s = Q_o$ - a Q-factor scale meaning;
$-\delta$, $-a$ - zero/pole co-ordinates. For the sake of simplicity the roots $-\delta$ and $-a$ are considered as real.

A bridge phase-frequency characteristic (PFC) summarizes the pole and zero PFCs. According (4) a full bridge phase shift is

$$\varphi_b = \varphi_p + \varphi_z + \varphi_c =$$

$$-\mathrm{arctg}(\xi_b/a) + \mathrm{arctg}(\xi_b/\delta) + \arg(C). \qquad (5)$$

Hence it follows (ignoring φ_c')

$$\varphi_b' \cong \varphi_p' + \varphi_z' = -[1 + (\xi_b/a)^2]^{-1}$$

$$+ \mathrm{sign}(\delta)\cdot[1 + (\xi_b/\delta)^2]^{-1} \qquad (6)$$

Oscillator stability criteria demand $\varphi_b' < 0$. For $|\delta| < a$ it means a necessity to place the zero into a right-hand semi-plane. Changing then $\delta \to -\delta$ we get at the central frequency

$$\varphi_b'(f_b) = \varphi_{bo}' = -(1/a + 1/\delta). \qquad (7)$$

From (2), (7) an effective bridge Q-factor Q_{ef} normalized with respect to Q_o is equal

$$q_b = Q_{ef}/Q_o = -\varphi_{bo}' = 1/a + 1/\delta, \qquad (8)$$

i.e. with a bridge going to a balance ($\delta \to 0$) an effective bridge Q tends to infinity.

From (4) bridge resonant insertion losses characterized by

$$L_{bo} = K_b^{-1}\,|_{f=f_o} = a/(C\delta). \qquad (9)$$

So the greater an effective bridge Q the greater insertion losses.

Bridge Selection

The first attempt to build a microwave bridge oscillator [4] has failed because of an unsuccessful bridge design.

One can find two main principles of a microwave bridge realization:
- summing up two parallel channels with different transfer functions (fig.2a);
- a use of resonator reflected wave frequency behavior features (fig.2b).

It can be shown that the least losses for a chosen effective Q_b meaning are provided with the second way using a reflecting resonator as a circulator load.

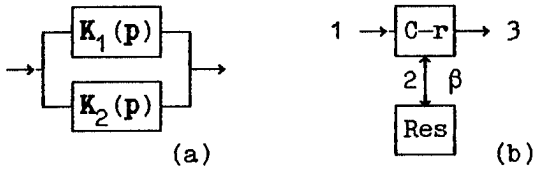


Figure 2 Two bridge options.

That is why the case at fig.2b only is delivered further.

For a circulator-type bridge (CB) with a resonator coupling constant $\beta > 1$

$$\mathbf{K}_b(\mathbf{p}) = -[\mathbf{p} - (\beta - 1)]/[\mathbf{p} + (1 + \beta)]. \quad (10)$$

A pole-zero diagram for this particular case is shown at fig.3.

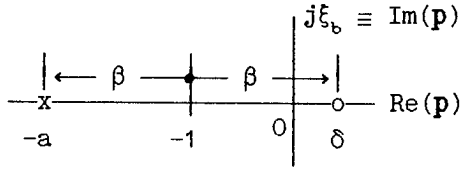


Figure 3 A pole-zero diagram of a circulator-type bridge.

At the central frequency

$$L_{bo} = (\beta + 1)/(\beta - 1), \quad (11)$$

$$q_b = 2\beta/(\beta^2 - 1). \quad (12)$$

With the offset increasing, $|\mathbf{K}_b|$ grows too tending to unity:

$$\lim_{|\xi| \to \infty} |\mathbf{K}_b(j\xi)| = 1. \quad (13)$$

Changing at (11) right-hand side β with q_b according to (12), it is possible to express q_b right through insertion losses:

$$q_b = (L_{bo}^2 - 1)/(2L_{bo}), \quad (14)$$

and on the contrary, L_{bo} via q_b:

$$L_{bo} = q_b + (q_b^2 + 1)^{1/2}. \quad (15)$$

One can check that a feedback filter quality parameter taken in a form $Q_{ef} \cdot K_o$ product is twice to CB benefit as compared with a transmitting resonator (TR) with symmetric optimized coupling.

More detail comparison a "classic" and "bridge"-type loop oscillator is made at fig.4.With L_o going to infinity, TR (curve 1) increases its Q_{ef} from $Q_o/2$ (-6 dB as against Q_o) for optimum partial coupling 0.5 at a point L_{opt}= 6 dB up to Q_o (0 dB); in contrast with it the CB Q_{ef} (curve 2)

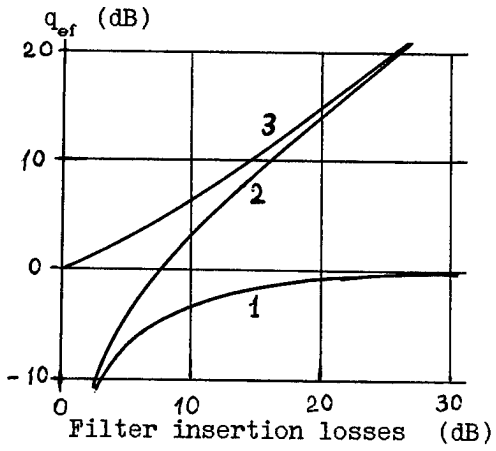


Figure 4 Normalized available Q_{ef} for a single transmitting resonator and a circulator-type bridge.

exceeds Q_o for L_o > 7.66 dB. revealing monotonous increase of a difference between CB and TR Q_{ef} (curve 3)

Amplifier

The greater the bridge Q_{ef} the greater insertion losses. At first sight it does not matter because always possibly to compensate the losses increasing amplifier gain. However as was found recently [5] a microwave BT stage multiplicative phase noise level is about directly proportional to stage power gain. So going closer to bridge balance to increase oscillator Q_{ef} improves $\mathcal{L}$ almost at the same rate as goes worse the stage gain growth.

Fortunately this contradiction may be solved with a multistage amplifier. In a case of n identical stages each of which characterized by gain G_{u1} and a power spectral density of phase fluctuations

$$S_\varphi = C_\varphi \cdot G_{u1}^2 \quad (16)$$

(C_φ - a characteristic constant) total noise equals

$$S_{\varphi n} = C_\varphi \cdot n \cdot G_u^{2/n}, \quad (17)$$

where $G_u = nG_{u1}$ - total gain. For given G_u minimum noise takes place if

$$n = n_{opt} = 2 \cdot \ln(G_u) \approx 0.23 \cdot G_u \text{ (dB)}, \quad (18)$$

i.e. a two-stage amplifier is the optimum decision for G_u = 8.7 dB, n_{opt} = 3 for 13 dB and so on. In fact both cases give very close results if G_u < 20 dB.

Corrector

Outside bridge balance the feedback loop gain is about L_{bo} times more. It creates a severe problem of a possible oscillator instability.

The loop total phase shift is $\pm 2\pi$ rad changed at the nearest "parasitic" frequencies f_p. The CB itself due to its higher Q in comparison with the amplifier adds $\approx (\pm\pi)$ rad having at the parasites

$$|K_m(f_p)| \equiv K_{mp} \approx 1. \qquad (19)$$

Because of signal delay in an amplifier its selectivity used to be insufficient to prevent parasitic oscillations. This problem can be solved by inserting into the loop of a special corrector. This role can effectively play any dielectric resonator (DR) with high enough Q-factor. The DR Q calculations are based at an equivalent amplifier stage circuit.

The corrector insertion losses are very harmful because of additional power gain needed to compensate them (recall - more gain means more noise). Nevertheless our calculations reveal a possibility of multiplicative phase noise decreasing at 10-15 dB in a bridge-type oscillator using two-stage amplifier.

Noise Filtering

A complete oscillator block-diagram at fig.5 incorporates three possibilities to get output power:
- from the transmission port (output 1);
- after CB from the amplifier input (output 2);
- from the filtered port (output 3).

The bridge-type oscillator circuitry features modify seriously usual phase noise frequency dependencies. Using Leeson method one can derive the next expressions writing at different points the bridge oscillator phase noise normalized with respect to $S_\varphi^{amp}/2$:

$$\text{- output 1: } S_{\varphi 1} = \frac{\delta^2 \cdot [\xi^2 + (2+\delta)^2]}{4\xi^2(1+\delta)^2}, \qquad (20)$$

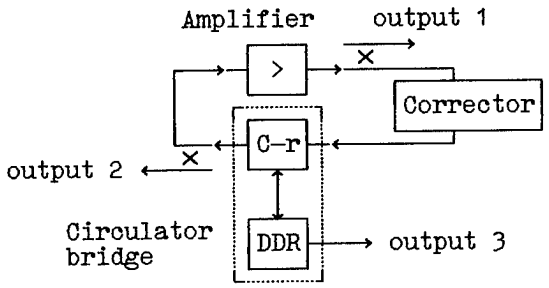


Figure 5 Complete block-diagram of the DDR-based CB-type loop oscillator.

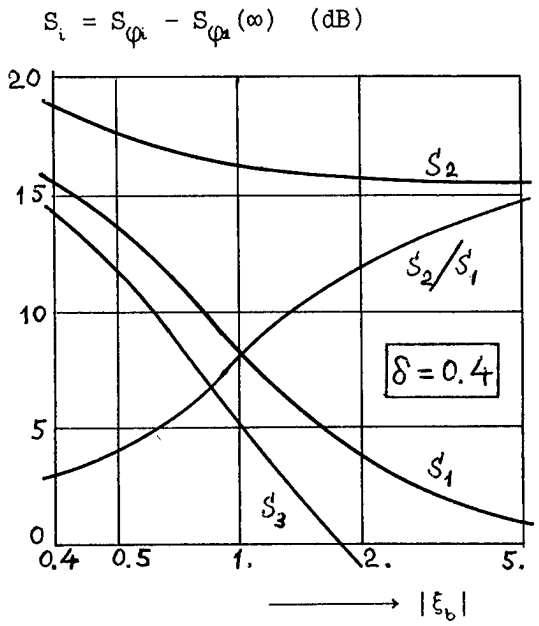


Figure 6 Filtering properties of bridge oscillator circuitry.

$$\text{- output 2: } S_{\varphi 2} = \frac{(2+\delta)^2 \cdot (\xi^2 + \delta^2)}{4\xi^2(1+\delta)^2}, \qquad (21)$$

$$\text{- output 3: } S_{\varphi 3} = \frac{(2+\delta)^2 \cdot [\xi^2 + (2+\delta)^2]}{4\xi^2(1+\delta)^2(1+\xi^2)}. \qquad (22)$$

Here ξ_b is signed as ξ for brevity.

The phase noise curves at fig.6 are additionally moved at the value $S_{\varphi 1}(\infty)$ to emphasize filtering. It is easy to see just the same noise levels at all points

with $|\xi_b| \to 0$. But for finite offsets the noise level is much lower at the point 1, i.e. at an amplifier output in comparison with its input. This unusual phase noise behavior is a straight result of gaining bridge transfer with removing from bridge balance. The least noise takes place after transmission through DDR (point 3).

Experimental Aspects

To check the theory an S-band BT oscillator using the 100 × 20.8 mm size shielded sapphire DDR [6] was built and tested.

For the particular disc format

$$F_d = D_d/2h = 2.4$$

(D_d - a disc diameter, h - a disc height) an optimum shielding diameter $D_{sh} \approx 1.6D_d$ at HE_{711} mode. With optimized copper-made shielding DDR Q_o exceeds $3 \cdot 10^5$ at room temperature. This value is lower than the expected value $\approx 5 \cdot 10^5$ mainly due to some small damages at the cylindric surface of a disc.

A two-stage microstrip amplifier uses BT KT647A-2 at the first stage and more powerful BT KT643A-2 at the second one. The overall small signal gain is $\approx$ 27 dB. It means the CB losses to be 15-16 dB, i.e. the oscillator Q_{ef} is 9-10 dB more than a DDR Q_o and the particular Q_{ef} value realized is about 1 million.

The output power is taken by means of a 3-dB hybrid. To save one additional ferrite isolator the corrector was placed before the hybrid which led to $\approx$ 3-dB loss of output.

The experimental data reveal good stabilization. The oscillator pushing as low as 250 Hz/V and a phase noise single side band spectral density near -130 dB/Hz at 1 kHz offset were reached. To measure phase noise we needed to amplify the output up to $\approx$ 100 mW to improve a noise measuring set-up noise floor.

It seems the data reported are nowadays tops for uncooled microwave oscillators.

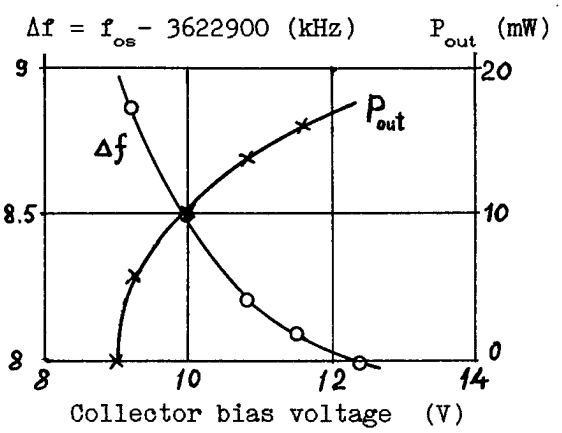


Figure 7 Some experimental data.

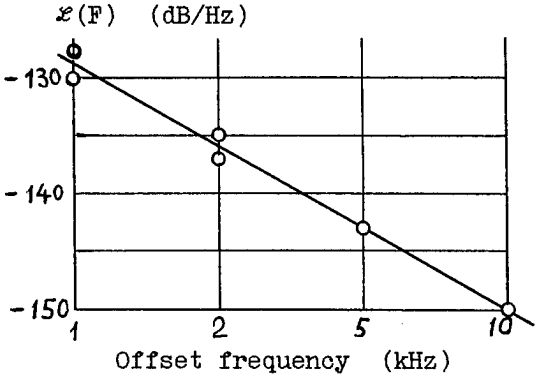


Figure 8 The measured bridge-type S-band oscillator phase noise.

References

[1] D.B.Leeson, "Short Term Stable Microwave Sources," Microwave Journal, vol. 13, pp. 59-69, June 1970.

[2] E.N.Ivanov, I.N.Mukhtarov and D.P. Tsarapkin, "High-Q Disc Dielectric Resonators," Radiotekhnika i Electronika, vol.28, pp.1658-1659, Aug. 1983, (In Russian).

[3] V.B.Braginsky, V.S.Ilchenko and Kh.S. Bagdassarov, "Experimental observation of fundamental microwave absorption in high quality dielectric crystals," Phys.Lett.A, vol.120, pp.300-305, Mar. 1987.

[4] W.R.Sooy, F.L.Vernon and J.Munushian "A Microwave Meacham Bridge Oscillator," IRE Nat. Convent. Rec., vol.7, Pt.3, pp.68-78, 1959.

[5] A.A.Borodulin and D.P.Tsarapkin, "Regime Impact at KT647A-2 Transistor Amplifier Phase Noise," Electronic Technique, s. "Microwave Electronics," N.8(432), pp.23-25, 1990 (In Russian).

[6] V.F.Vzyatyshev, V.S.Dobromyslov, V.I.Kalinichev et al., "Screened Dielectric Resonators with Azimuth Oscillations," Izvectija vis. uch.zav. - Radiofizika, vol.30, pp.79-88, N.1, 1987 (In Russian).

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

A THALLIUM-BASED SUPERCONDUCTING CAVITY FOR MICROWAVE OSCILLATOR APPLICATIONS

S. C. Han*, B. L. Zhou*, C. P. McCarroll*, T. G. Hammersley**, T. Apgar**,
P. N. Arendt***, and D. W. Cooke***

* Xsirius Superconductivity, Inc., 1110 N. Glebe Road, Suite 620, Arlington, VA 22201
** Communication Techniques, Inc., 9 Whippany Road, Whippany, NJ 07981
*** Los Alamos National Laboratory, Los Alamos, NM 87545

Abstract

A thallium-based high temperature superconducting cavity is developed to stabilize an X-band microwave oscillator. The 10 GHz cavity has unloaded Q values of 5.7×10^4 at 77 K and 1.5×10^5 at 20 K. A 10 GHz free running oscillator stabilized by the cavity has single sideband phase noise of -82, -114, and -132 dBc/Hz at 1, 10, and 100 kHz frequency offset from the carrier, respectively.

Introduction

Microwave oscillators with low phase noise are required for a number of radar system applications. For Doppler radars designed to detect small cross section targets in a clutter environment, single sideband phase noise near the carrier has a critical impact on the system performance.

In this paper, we report recent experimental results on the use of a thallium-based high temperature superconducting (HTS) cavity for low noise oscillator applications. Use of high Q superconducting cavity resonators enables realization of oscillators with considerably lower phase noise than their counterparts using conventional dielectric resonator oscillators (DROs). We will discuss cryogenic integration of the superconducting cavity oscillator following a brief summary of the cavity fabrication and performance testing.

Cavity Resonator

Microwave resonators have been fabricated by a number of groups using HTS thin films deposited on planar dielectric substrates.[1,2] A Ku-band oscillator using a planar HTS resonator has also been reported.[3] In contrast to the previous work, the present work utilizes a three dimensional superconducting cavity resonator for oscillator fabrication. In general, a three dimensional metal cavity has a higher Q factor than a planar resonator structure because of the greater geometric factor. A superconducting cavity can provide extremely high Q which would lead to a significant reduction of the single sideband (SSB) phase noise near the carrier.[4]

Practical realization of the HTS cavity is achieved by depositing high quality superconducting thin films on the interior surfaces of a three dimensional metallic cavity. Since the deposition method (magnetron sputtering) used is line of sight, cylindrical cavities are excluded from use because of the difficulty in obtaining uniform coating. This limitation leads to investigation of alternative cavity geometries which are amenable to the deposition process.

The cavity geometry employed in the present work has a clamshell configuration as discussed in [4]. Each cavity consists of two halves with concave surfaces jointed opposing to each other. The diameter to depth ratio of the cavity is approximately 5:1. The cavity geometry is designed such that the quasi-TE_{011} mode resonates at 10 GHz. The electric field of this mode circulates in the horizontal plane; consequently, there is no additional loss across the joint of the two half cavities. This ensures a high Q factor to be achieved which is important for low phase noise oscillators.

Two different clamshell cavities are employed in the present work. One cavity is made of high purity copper, and the other of Consil 995 alloy. The copper cavity is coated with gold to prevent oxidation. The Consil cavity is covered with a thallium-based HTS

0-7803-0905-7/93 $3.00 © 1993 IEEE

thin film using a post-annealing magnetron sputtering process.[4] The resonant cavity is enclosed by gold or the HTS thin film coating. The cavity Q is mainly limited by the surface resistance of the conductive coating. The unloaded Q values of both cavities were tested from room temperature (300 K) down to 20 K. The results are shown in Figures 1 and 2.

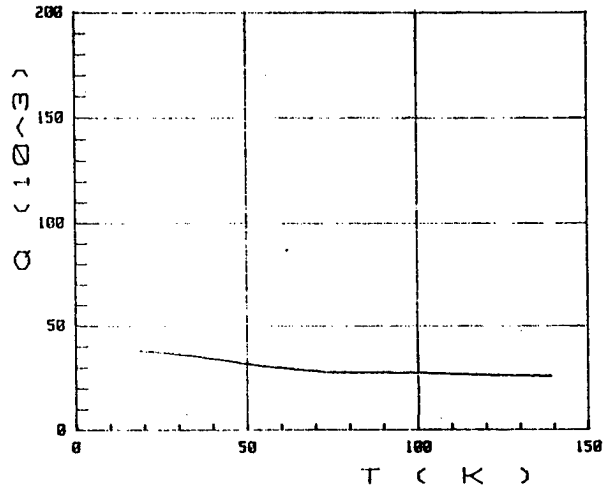


Fig. 1 Unloaded Q as a function of temperature for the gold coated copper cavity

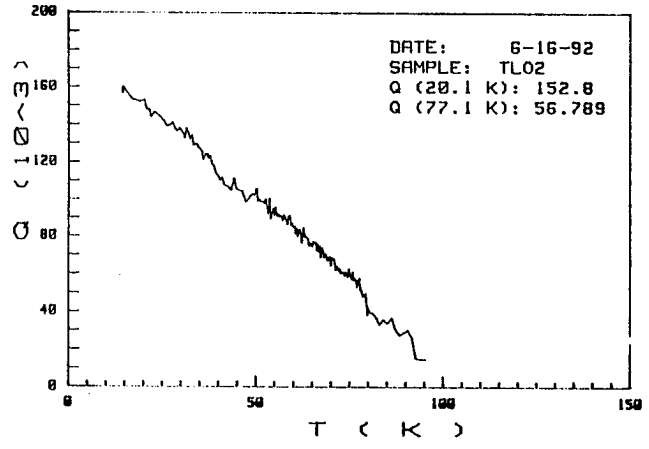


Fig. 2 Unloaded Q as a function of temperature for the HTS cavity

The Q value of the 10 GHz HTS cavity is 5.7×10^4 at 77 K and 1.5×10^5 at 20 K. In comparison, the Q value of the gold coated copper cavity is approximately 2.7×10^4 at 77 K and 3.7×10^4 at 20 K. While the gold/copper cavity has a higher Q than the HTS cavity at room temperature, its Q value does not increase as rapidly as the HTS cavity does. Due to the superconducting transition, the HTS cavity Q starts to rise sharply once the temperature drops below 100 K. This sharp increase gives rise to the higher Q value than the gold/copper cavity.

Cryogenic Integration of Oscillator

The basic configuration of the oscillator consists of a microwave amplifier and a cavity resonator as the feedback and frequency selection device. Details of the circuit design and analysis are described in a separate paper.[6] Briefly, the oscillator is a free running source using a transmission topology with a GaAs MESFET transistor amplifier. A microstrip directional coupler was utilized to provide power output. The oscillator was first tuned and tested at room temperature. Subsequently, it was cooled down to low temperatures and tested again at 77 K.

A cryogenic system was specially designed for cooling the microwave oscillator. It is based on an APD closed cycle refrigerator using cyclic adiabatic expansion of helium gas as the cooling mechanism. The environment around the cold head was redesigned to accommodate the cavity oscillator. First, the cryostat expander was reversed upside down so that the cavity could be placed on top of the cold head. This makes it easy for mounting the cavity during testing. Second, a double metal shroud was built around the cold head to provide improved thermal isolation between the cold head and the ambient. Electric connections to the cavity oscillator, both dc and rf, were realized using vacuum feedthrough. The cryogenic system configuration is illustrated in Figure 3.

During cooling, a vacuum environment was maintained between the cavity oscillator and the metal shroud. This minimized thermal leakage and as a result the system was able to cool the cavity oscillator down to 20 K. During warming, the vacuum chamber was backed filled with helium gas to significantly increase the rate for warming up. An additional heating element was attached to the cold head for rapid warming up or maintaining a fixed cryogenic temperature.

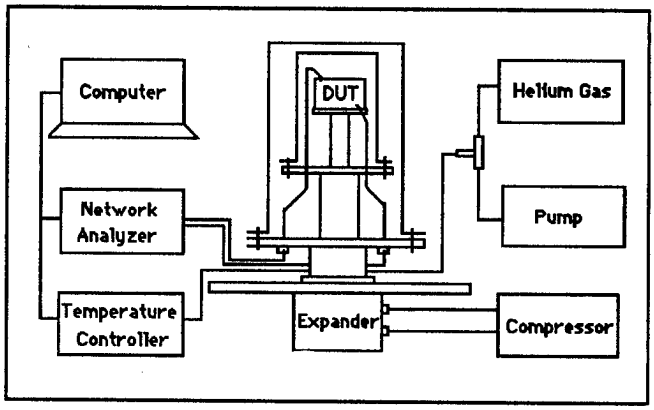


Fig. 3 Illustration of the cryogenic system configuration

Two different oscillators were built, one using the gold coated copper cavity and the other using the HTS cavity. Both oscillators were tested at room temperature (300 K) and cryogenic temperature (77 K). At 300 K, the gold/copper cavity resonator has a single sideband of -84, -117, and -133 dBc/Hz at 1, 10 and 100 kHz offset from the 10 GHz carrier. The phase noise of the HTS cavity oscillator at room temperature is -82, -114, and -132 dBz/Hz as shown in Figure 4.

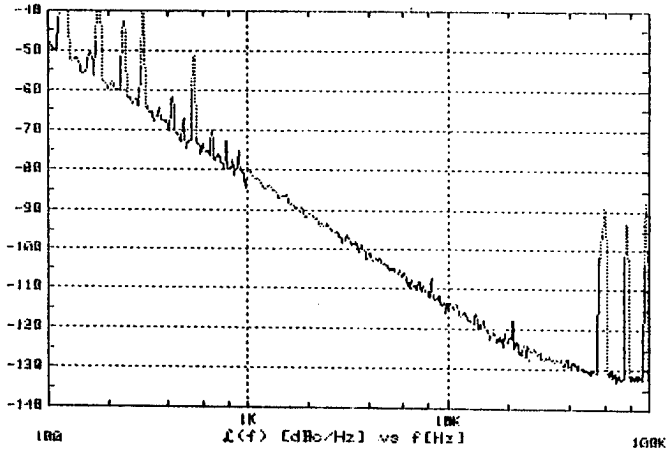


Fig. 4 Single sideband phase noise of the HTS cavity oscillator at room temperature

The phase noise of the HTS cavity oscillator at room temperature is slightly higher than that of the gold/copper oscillator, as indicated in Table 1. This reflects the fact that the HTS coating has higher surface resistance than the gold coating. The HTS coating is still conductive, although not superconductive, at room temperature. In addition, the Consil alloy base cavity is highly conductive due to the high percentage (99.5%) of silver. These factors are responsible for the low phase noise of the HTS cavity oscillator at room temperature. This feature could be of considerable interest for practical applications in case of an unexpected cryostat failure.

Table 1 Single sideband phase noise of oscillators using gold/copper and HTS cavities at room temperature (dBc/Hz)

Cavity Type	Offset Frequency		
	1 kHz	10 kHz	100 kHz
Au/Cu	-84	-117	-133
HTS	-82	-114	-132

When the HTS cavity oscillator was cooled down to low temperatures, the resonant mode jumped from the 10 GHz mode to another mode at 9.3 GHz. The mode jump was mainly caused by the variation in phase velocity during cool down which deviated from the optimum condition for obtaining the desired 10 GHz TE_{011} mode. The oscillator which was previously tuned to oscillate at room temperature would require additional tuning in order for low temperature operations. Effort is currently under way to achieve this objective.

Conclusions

A three dimensional high temperature superconducting cavity resonator is cryogenically integrated for microwave oscillator applications. Unloaded Q values of 5.7×10^4 and 1.5×10^5 have been achieved at 77 K and 20 K, respectively. A 10 GHz free running oscillator stabilized by the cavity has single sideband phase noise of -82, -114, and -132 dBc/Hz at 1, 10, and 100 kHz frequency offset from the carrier, respectively.

References

[1] C. Wilker, Z. Y. Shen, P. Pang, D. W. Face, W. L. Holstein, A. L. Mattews, and

D. B. Laubacher, "5 GHz High-Temperature-Superconductor Resonators with High Q and Low Power Dependence up to 90 K", IEEE Trans. Microwave Theory Tech., vol. 39, pp. 1462-1467, 1991.

[2] M. S. Schmidt, R. J. Forse, R. B. Hammond, M. M. Eddy, and W. L. Olson, "Measured Performance at 77 K of Superconducting Microstrip Resonators and Filters", IEEE Trans. Microwave Theory Tech., vol. 39, pp. 1475-1479, 1991.

[3] J. W. Smuk, M. G. Stubbs, and J. S. Wight, "Hybrid Semiconductor/High Temperature Superconductive Ku-band Oscillator and Amplifier MICs", IEEE MTT-S Int. Microwave Symposium Digest, pp. 485-488, 1992.

[4] R. G. Rogers, "Low Phase Noise Microwave Oscillator Design", Artech House, Boston & London, p.12, 1991.

[5] P. N. Arendt, D. W. Cooke, N. E. Elliott, E. R. Gray, K. M. Hubbard, H. Piel, and G. R. Swain, "A Tl-based High-temperature Superconducting X-band Cavity", Journal of Superconductivity, vol. 5, No. 4, pp. 431-435, 1992.

[6] T. G. Hammersley, T. Apgar, B. L. Zhou, S. C. Han, C. P. McCarroll, "X Band Superconducting Cavity Oscillator", submitted to IEEE Microwave and Guided Wave Letters, Jan. 1993.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

THE FLIGHT PERFORMANCE OF THE GALILEO ORBITER USO

DAVID MORABITO, TIMOTHY KRISHER, and SAMI ASMAR
Jet Propulsion Laboratory,
California Institute of Technology
Pasadena, Calif. 91109

Abstract

Radiometric tracking Doppler measurements of the signal transmitted by the Galileo spacecraft using an Ultra-Stable Oscillator (USO) as a frequency reference have been acquired by antennas of the NASA Deep Space Network (DSN) between Dec. 1989 and Nov. 1991. These measurements serve two purposes; 1) the scientific investigation of the gravitational redshift phenomenon as the spacecraft moves in and out of the gravitational fields of massive bodies in the solar system, as predicted by Einstein's theory of general relativity, and 2) engineering evaluation of the USO frequency and frequency stability for calibration purposes, and to evaluate the health and performance of the USO. These calibrations serve as a baseline for Radio Science experiments, such as the Redshift Observation and occultations of Jupiter and its satellites.

The USO is a dual oven-controlled oscillator with an AT-cut quartz crystal (SiO_2) resonator. The output frequency of 19.124980 MHz is multiplied by 120 to produce the transmitted signal frequency (~2.29 GHz).

There were 82 data acquisition passes conducted between launch (Oct. 1989) and Nov. 1991. Each pass consisted of about two hours of Doppler data sampled at one per sec, received by the DSN antennas using Hydrogen-masers as frequency and timing references. The Doppler were converted into estimates of spacecraft transmitted frequencies and frequency residuals after removal of the spacecraft trajectory and other effects.

The USO-referenced frequencies showed a significant positive logarithmic increase shortly after initial turn-on, which is probably due to the liberation of contamination on the crystal vibrator surfaces acquired during the long period of inoperation prior to launch. A least-squares fit to an aging model (combination logarithmic curve and linear drift) was applied to the estimated frequencies of the 76 passes acquired during the first in-flight USO on-off cycle (Dec. 1989 to Aug. 1991). The residuals showed a large systematic variation shortly after initial turn-on. Since the aging behavior during this period is complex, the first twelve passes were removed, and the model was fit to the remaining 64 passes. The resulting post-fit residuals showed no significant systematic variation. It is believed that the variations of the 17 mHz rms residual scatter are due to the random walk of the USO or possible unknown mismodeling. It took a little over a year of continuous operation before the USO reached its linear aging realm (-5.6 x 10^{-12}/day). The USO was powered-off and back on in Aug. 1991 with no adverse effect on its performance.

The average Allan deviations at 1-s and 10-s correlate well with spacecraft range, and fall as τ^{-1}. This is consistent with white PM noise dominating at these time scales due to the low SNR's expected from using the Low Gain Antennas. The Allan deviations at 100-s and 1000-s agree with pre-flight USO stability results at these time scales. The Galileo USO appears to be healthy and functioning normally.

I. Introduction

This is an article on the efforts to characterize the instrument used for the Galileo Radio Science investigation. It discusses the performance of the Ultra Stable Oscillator (USO), which, when the spacecraft and ground elements of the instrument are configured in the one-way mode, is the limiting error source observed in the received Doppler

0-7803-0905-7/93 $3.00 © 1993 IEEE

data at time intervals where there is sufficient signal power.

Galileo was originally scheduled to be launched in May 1986, but was delayed due to the Challenger accident. Galileo was launched on Oct. 18, 1989 and the USO was turned-on on Dec. 5, 1989. Galileo acquired gravity assists from Venus (Feb. 1990) and Earth (Dec. 1990, Dec. 1992) and will go into orbit around Jupiter in Dec. 1995. Between Dec. 1989 and Nov. 1991, 94 passes have been scheduled for two-hour data acquisition periods of the Galileo spacecraft where the USO was the signal source. However, not all scheduled passes resulted in valid data; twelve passes were lost for various reasons (six were aborted after a spacecraft safing anomaly early in 1990, four were lost due to a crashed sequence in March-April 1991, and two were lost due to unrecoverable data problems). Data for the remaining 82 passes were acquired by the DSN and delivered to the Galileo Radio Science Team (RST) for analysis in the form of Archival Tracking Data File (ATDF) tapes. Estimates of the spacecraft transmitted frequency and frequency stability were made for the 82 passes conducted during these first two years of cruise. Each pass consists of about two hours of doppler data sampled at 1/s, and Hydrogen-masers served as the frequency and timing references. The received doppler data were converted into estimates of spacecraft transmitted frequencies and frequency residuals after accounting for spacecraft trajectory and other effects.

II. Purpose

The USO-referenced data were acquired for two purposes; 1) the scientific investigation of the redshift phenomenon: the frequency shift as the spacecraft moves in and out of the gravitational fields of massive bodies in the solar system as predicted by Einstein's theory of general relativity, and 2) engineering evaluation of the USO frequency and frequency stability for calibration purposes and to evaluate the health and performance of the USO. These calibration data serve as a baseline for Radio Science experiments such as the Red Shift Observation, Solar Corona Experiment, and occultations of Jupiter and its satellites. This article will focus only on the engineering aspects of the USO data. The scientific results are discussed in [1]. For discussions of the expected scientific results of the Galileo Radio Science investigations, the reader is referred to [2,3].

The goal of the USO analysis is, thus, to establish the USO-referenced spacecraft transmitted frequency and the frequency stability associated with it and to build a database containing the statistics and all parameters relevant to the measurements.

III. Spacecraft Configuration

The Galileo spacecraft configuration for these Radio Science tests is the normal cruise configuration, with the exception of a command which enables the USO to be the radio downlink reference; this is the mode required for the redshift experiment and future occultation investigations. No configuration changes such as changing the telemetry modulation index were expected during a pass. The majority of the tests were performed on a "quiet" spacecraft where no maneuvers or other motion was permitted to occur.

The USO, a science payload instrument integrated with the spacecraft's telecommunications subsystem, was manufactured by Frequency Electronics Inc., New York between 1975 and 1976. The USO which resides in the Radio Frequency Subsystem (RFS) of the Galileo orbiter is serial number #4, from the same lot as the USOs flown on Voyagers 1 and 2. The Voyager 1 USO failed in Nov. 1992 (after fifteen years of continuous service), engendering concerns about the survival lifetime of the Galileo USO. The USO is a dual oven controlled device with an AT-cut quartz crystal (SiO_2) resonator. The design output frequency is 19.125000 MHz. When the USO is the downlink signal source, it drives the S-band exciter. The output frequency is multiplied in the transponder by a factor of 120 to produce the 2.29 GHz transmitted signal. The S-band Traveling Wave Tube Amplifier (TWTA) amplifies the signal it receives from the S-band exciter to either one of two power levels; the high-power mode (27 watts) or the low power mode (9 watts). The TWTA has routinely been configured to the high-power mode. The S-band TWTA provides the amplified RF output to the HGA/LGA transmit switch of the S-band antenna switch, which connects the outputs of the S-band TWTAs to either the HGA or the LGA. If the HGA is ever to be successfully deployed, an

X-band (8.4 GHz) transmitted signal will also become available and its frequency will be 11/3 of the S-band signal frequency.

All of the data acquired were at 2.3 GHz and Right Circularly Polarized. The signal was transmitted from the spacecraft via either LGA-1 or LGA-2 (Low Gain Antennas). LGA-1 is located on the spacecraft spin-axis in front of the HGA tip sunshade. LGA-2 is located on a boom 3.58 m away of the spin axis, and is pointed in the direction opposite that of HGA and LGA-1. Both LGAs work at 2.3 GHz only. LGA-2 was utilized for the period right after launch and the period right after the Earth I encounter. The time periods of the spacecraft LGA configuration for the data set analyzed in this article are given below;

10/18/89 to 03/15/90	LGA-2
03/15/90 to 12/08/90	LGA-1
12/08/90 to 01/31/91	LGA-2
01/31/91 to 11/30/91	LGA-1

The Galileo spacecraft is a spin-stabilized spacecraft which rotates at about three rpm.

IV. Ground Data System Configuration

The DSN configuration was also that of normal cruise tracking. The radio metric data were sampled at a rate of one per second using a loop filter bandwidth of 10 Hz. For all of the passes, hydrogen-masers were the frequency and timing references at the ground stations. The overall stability of the ground system (frequency reference, receiver, cables, etc.) is expected to be about 1×10^{-15} at 1000 sec, which is far more stable than the USO.

The following data products are received for each pass; an ATDF tape containing the closed-loop Doppler data, a pass folder from the DSN containing copies of frequency predictions, operator logs and related material, and a spacecraft trajectory vector file from the Galileo Navigation Team.

V. Analysis Software and Techniques

The analysis was performed on the Radio Science Support System RODAN (Radio Occultation Data Analysis Network) computer which includes a Prime 4050 computer, a Floating Point Systems (FPS) array processor, two magnetic tape drives and other peripherals. The system is accessible by a set IBM PS/2 terminals as well as two Sun workstations.

Although the analysis tools used for this work were inherited from the Voyager Project, there were enough differences in configuration and procedures in the Galileo Project that non-trivial modifications were made to the software and techniques. The analysis software was upgraded to estimate more accurately the spacecraft transmitted frequency and the frequency residuals used for the stability analysis, including installation of code to model the effects introduced by a spinning spacecraft.

The data were processed by the STBLTY software program set. The functions of each of the component programs (see Fig. 1) are described below.

The program OCEP reads the doppler counts and doppler extractor reference frequency from the input Archival Tracking Data File (ATDF) tape, converts these to sky frequencies, and writes the time-tagged frequency data and related information to the F50 file.

The program GETTRAJ reads the spacecraft state vectors from the Navigation Team provided Celestial Reference Set (CRS) trajectory file, performs vector manipulations and light-time solutions, and writes the F45 output file containing the time-tagged spacecraft position and velocity vectors relative to the sun and the observing ground station. Details of the orbit determination solutions determined by the Galileo Navigation Team used to generate the trajectory are given elsewhere [4].

The program RESID reads the output files of the two previous programs (F50 sky frequencies and F45 trajectory vectors). Model sky frequencies are estimated from the trajectory vectors, and are corrected for troposphere, spacecraft spin, gravitational redshift and, if applicable, the spinning off-axis LGA-2 antenna induced doppler signature. The spacecraft transmitted frequency is estimated at the time tag of the first data point, and the frequency residuals are computed by differencing the observed sky frequencies from the estimated sky frequencies for each data point. The troposphere model is a simple zenith path delay translated to the line-of-sight elevation angle. A spacecraft spin model correction is applied to the estimated downlink frequency and

residuals. The magnitude of this correction depends upon whether the spacecraft was in all-spin mode (0.0481 Hz) or dual-spin mode (0.0525 Hz). The sign of this correction depends upon whether the signal source was LGA-1 or LGA-2. This program writes the residual frequencies and related information to the F52 output file.

Relativistic effects, including the gravitational redshift, are modeled in the analysis, and removed from the data. The results of the scientific analysis of these effects, the first test of the solar redshift with an interplanetary spacecraft, indicates that the total frequency variation as predicted by general relativity has been verified to an accuracy of 0.5% and the solar gravitational redshift to 1% (Ref. 1). Therefore, it can be stated with a reasonable degree of confidence that these effects have been removed to the stated accuracies.

The program STBLTY reads the F52 file, computes and applies a bias correction to the spacecraft transmitted frequency estimate relative to the center of weight of the residuals over the pass, and computes phase, Allan deviation, frequency and phase power spectral densities (PSDs), other relevant statistics and writes selected information for the current pass onto a data base summary file. This program also produces plots of relevant data quantities.

Once all of the USO passes have been processed through the above programs, the data base file will contain a set of records for each processed pass. This file is then processed through the programs FITUSO and USOSMRY.

The program FITUSO reads the frequencies from the data base, and fits and removes an aging model. The resulting post-fit residuals are written onto a previously blank field of the data base. The user can specify how many passes to skip, how many passes to accept and where the break between logarithmic and linear aging behavior occurs.

The program USOSMRY reads the USO data base and displays graphically any requested quantities from a menu of available data types.

VI. Analysis Results

A. General Single Pass Results

Eighty-two of ninety-four scheduled USO-referenced data acquisition passes were processed through the STBLTY program set. Each pass was typically two hours in duration. All data processed were closed-loop data acquired by the Deep Space Stations sampled at 1/s. Table 1 displays the year, day number, UTC start time, UTC end time, DSS station ID, signal level (AGC), and estimated spacecraft transmitted frequency for each pass. Fig. 2 displays a typical plot of the residuals after the removal of the spacecraft trajectory, the gravitational redshift, a simple troposphere model, and the effect of the spacecraft spin when LGA-1 is the signal source. Fig. 3 displays the frequency residuals after every 60 points have been averaged, allowing long period trends to be easily examined. Fig. 4 displays the reconstructed phase for the residuals of Fig. 2. Fig. 5 is the log of the Allan deviation for the residuals of Fig. 2. Figs. 6 and 7 display the logs of the frequency spectral density and phase spectral density, respectively, for the residuals of Fig. 2. The spikes at about 0.05 Hz are related to the 3 rpm spacecraft spin.

B. LGA-2 Induced Spin Doppler Example

When LGA-2 is the signal source, there is a significant sinusoidal signature present in the received doppler due to LGA-2 being mounted on a boom located 3.58 meter from the spacecraft spin axis. Fig. 8 displays a typical plot of frequency residuals for a pass where LGA-2 was used, after the trajectory and other effects were removed. Fig. 9 clearly illustrates the sinusoidal signature for a selected 200 sec period.

In order to remove this signature from the data, a three-parameter sinusoidal model was iteratively fit to the Doppler residuals of Fig. 8. This model includes an amplitude, a frequency and a phase offset. After the model was successfully fit and removed, the resulting "post-fit" residuals (see Fig. 10) yielded Allan deviations consistent with those of LGA-1 passes, suggesting that the three-parameter sinusoidal model is sufficient for removal of the spin-induced Doppler signature.

For three passes where LGA-2 was the signal source 91-01-14 (91-014), 90-01-09 (90-009) and 90-12-10 (90-345), dynamic activity occurred on-board the spacecraft which introduced phase shifts into the data. The result is that the fit of the sinusoidal model failed to remove all of the induced off-axis LGA-2 Doppler signature, resulting in residuals as exemplified in Fig. 11 for pass 91-014. In this specific case, tape recorder motion was known to occur, where the envelope changes in Fig. 11 correlate with the tape recorder start and stop times. Fig. 12 illustrates the resulting degradation to the Allan deviation (compare with Fig. 5). Pass 90-345 occurred after the Earth I flyby where several activities occurred on-board the spacecraft which introduced dynamic motion. Commands which introduced dynamic motion on the spacecraft also occurred during pass 90-009. In these cases, the dynamic activity was assumed to be symmetric about the center of weight of the spacecraft and thus appeared not to have biased the estimated spacecraft transmitted frequencies which are determined at the first data point, and then translated to the center of weight of the residuals over the full data span. The Allan deviations for these passes were however degraded.

C. Solar Interference Example

Fig. 13 displays the residuals for pass 91-03-05 (91-064) where several solar disturbances occurred during that day including the period of the data acquisition. The degradation in the observed Doppler noise measurements from the closed-loop system are consistent with those expected from solar interference (R. Woo, private communication). This was the only pass in this data set where this behavior was observed. The estimated spacecraft transmitted frequency appears reasonable, however, the Allan deviations were degraded as expected.

D. Stability Analysis Results

The Allan deviation is the recommended measure of oscillator stability for time domain signal processing. The Allan deviations are displayed in Figs. 14-17 for 1, 10, 100 and 1000 sec time intervals, respectively. Outlier passes which contain degraded data have been removed. Outlier passes include a) the three passes where dynamic events were known to occur on the spacecraft 900109 (90-009), 901210 (90-345), and 910114 (91-014), b) the suspected solar interference pass 910305 (91-064), c) three passes where the cause of the degradation is not known, but is possibly due to ground equipment problems 900330 (90-089), 911019 (91-292) and 911130 (91-334), and d) the first two passes after initial turn-on (89-341, 89-350) where the significant increasing frequency dynamics adversely dominated the Allan deviation measurements at 1000 sec.

The measured means and errors of the Allan deviations are given in Table 2 for each time interval (outlier passes were removed as well as three passes where there were insufficient data to estimate Allan deviations at 1000 sec). The pre-flight Allan deviation measurements of the Galileo USO for these time intervals are also presented in Table 2. The pre-flight measurements were performed at the JPL Hydrogen Maser Test Facility on May 1, 1980 [5]. Fig. 18 is a plot of the averages of the measured flight Allan deviations with the pre-flight measurements superimposed.

For 1 and 10 sec, the measured flight Allan deviations significantly exceed the pre-flight Allan deviations. This was expected since it is known that wideband thermal system noise (white phase noise) dominates at these time intervals due to the low signal-to-noise ratios resulting from using Galileo's low gain antennas. The flight Allan deviations for individual passes agree with estimates derived from the measured signal levels, system noise temperatures (~20 K) and receiver bandwidth (10 Hz). The Allan deviations at 1 sec (Fig. 14) and 10 sec (Fig. 15) also correlate with spacecraft range (compare with Fig. 19). If the HGA ever becomes available, and if there is an opportunity to turn off the telemetry modulation to increase the signal strength, there will better visibility of the true USO performance at the 1-sec and 10-sec time intervals.

The average of the flight Allan deviations at 100 and 1000 sec agree with the corresponding pre-flight values in Table 2 and Fig. 18. This implies that the flight data are dominated by the true behavior of the USO at these time intervals. The apparent large scatter of the 100-sec (Fig. 16) and 1000-sec (Fig. 17) Allan deviations is

in agreement with theoretical values using equations derived by Lesage and Audoin [6].

The hump at 70-sec in the pre-flight values in Fig. 18 has been attributed to the thermal oscillation of the inner oven current driven by noise (G. Wood, private communication). The consistent behavior of the flight data between 100 s and 1000 s with the pre-flight measurements at these time intervals implies that this effect is visible in the flight data. In the absence of this thermal cycling, one would then expect the Allan deviation behavior to be flat over these time scales, and thus be consistent with flicker frequency noise. The expected noise due to unmodeled media fluctuations lies well below this level at these time intervals.

The Allan deviation behavior for the flight data measurements can be broken down into several regions. The Allan deviation in the 1 to 10 sec region falls roughly as τ^{-1}, characteristic of white phase noise dominating at these time intervals. The USO behavior here is masked by this noise at these time scales. The behavior of the region from 100 to 1000 sec can be interpreted as being consistent with the known inner-oven thermal cycling effect being superimposed on a flicker frequency noise floor (using the pre-flight measurement information). Between 1000 sec to about 40000 sec, one can assume a continuation of the flicker noise floor upon inspection of the pre-flight measurements and error bars.

E. Signal Levels

The Galileo passes have relatively low signal levels which run from about -140 dBm shortly after launch and around Earth I flyby to as low as -170 dBm. The receiver threshold is typically about -176 dBm. As previously discussed, the Allan deviations derived from the measured signal levels using thermal noise theory are consistent with the measured Allan deviations at 1 and 10 sec. The signal level behavior also correlates well with spacecraft range (Fig. 19).

After correcting the received signal levels for receiver station antenna gain, space loss, spacecraft LGA antenna gain, and telemetry state carrier suppression, the resulting spacecraft transmitted power levels at the Radio Frequency Subsystem/Antenna interface of the spacecraft were computed. These mean values were in good agreement with the expected power levels, and were within the known calibration uncertainties at the ground stations.

F. Analysis of Spacecraft Transmitted Frequency Measurements

The frequency transmitted by the spacecraft was estimated for all 82 passes. Fig. 20 displays the USO-referenced spacecraft transmitted frequencies as estimated by STBLTY. Each point on the plot is the USO frequency estimated at the first time tag for that pass and then corrected to the center of weight of all of the residuals over that pass. The assigned uncertainties of the estimated frequencies run about 3 mHz and are dominated by the uncertainty inferred by not performing an ionosphere correction. The plot is annotated with the times the USO was powered on and off. The time axis is in days since Jan. 1, 1989. The USO was initially powered-on in flight on December 5, 1989 (DOY 339). There was one instance of cycling the USO off (91-217) and then back on (91-228) in this data set.

A preliminary pass of a few minute duration was conducted shortly after initial turn-on on Dec. 5, 1989 (and after the inter-oven current was allowed to stabilize) in order to verify operation of the USO. Here the frequency was observed to be increasing at a very high rate, exhibiting dynamic behavior due to early impurity migration and/or stress relief resulting in the very poor stability expected shortly after turn-on. The first valid USO pass was conducted on Dec. 7, 1989, several hours after initial turn-on.

Changes in the USO frequency with time are referred to as resonator aging. In general, the principal causes of aging are stress relief in the mounting structure of the crystal unit, mass transfer to or from the resonators surface due to adsorption or desorption of contamination, changes in the oscillator circuitry, and possibly changes in the quartz [7]. Aging effects seen in this data set likely include surface liberation of impurities, impurity migration across and within the crystal, and stress relief. The significant positive logarithmic increase in frequency shortly after initial turn-on (see Fig. 20) is probably due to the liberation of

contamination on the crystal resonator surfaces which were acquired during the long period of inoperation prior to launch (R. Sydnor, private communication). The USO was powered off for much of the time 1986-1989 that the spacecraft was dormant on the ground prior to launch. The linear region is expected to be reached after the USO has been turned on for a sufficiently long enough period of time. The curvature observed for the six passes conducted after the USO was cycled off and back on in Aug, 1991 could be attributed to stress relief and migration of impurities.

During the first USO-on cycle (Dec. 1989 to Aug. 1991), 76 passes of USO data were acquired. An aging model was removed from the estimated spacecraft transmitted frequencies for this period, so that the resulting residuals could be analyzed, and remaining error sources could be identified. The model removed was that of a combination logarithmic curve and a linear aging drift;

$$\hat{f}_{S/C} = C_1 \log_e[C_2(t-t_0)+1] + C_3+C_4(t-t_0) \quad for\ t < t_b$$

$$\hat{f}_{S/C} = C_5+C_6(t-t_b) \quad for\ t \geq t_b$$

where t_0 is the time tag of the first data point, and t_b is the time tag of the first data point of where the linear aging realm begins. C_1-C_6 are constant coefficients, which were estimated by least-squares analysis.

The initial attempt to fit this model over the data acquired during the first USO-on cycle, resulted in the residuals displayed in Fig. 21. A large systematic variation is evident for passes occurring shortly after the USO was powered on, suggesting that the model is insufficient for this period. The rms scatter of the residuals is about 48 mHz with a χ^2_n of 272. Since the behavior of the USO is known to be complex during this period and not easily modeled (initial stabilization phenomena masks the aging behavior), the first twelve passes were deleted, and the model was fit to the remaining 64 passes. The corresponding post-fit residual plot of Fig. 22 displays significantly less scatter and no significant systematic variation. The rms residual scatter of this fit was 17 mHz with a χ^2_n of 35. It is believed that the 17 mHz variations are due to the random walk of the USO or mismodeling. In an attempt to identify any mismodeling, the post-fit residuals were examined against troposphere correction, spacecraft-earth-sun angle, elevation angle, spacecraft range, signal level (AGC), gravitational redshift correction, station ID, USO oven current (from telemetry), and spacecraft temperatures (from telemetry). No apparent correlations or trends were detected.

The 17 mHz rms scatter of the estimated flight frequencies about the aging model is consistent with the level of random walk inferred from extrapolating the pre-flight Allan deviation measurements as $\tau^{1/2}$ to a weekly time interval. However, a portion of the 17 mHz scatter could possibly be attributed to some yet to be identified mismodeling.

The logarithmic time constant C_2^{-1} was estimated to be about 71 days. The turnover of the curve occurred 259 days after initial turn-on. The slope of the linear aging region was estimated to be -1.50×10^{-7} Hz/sec which is in agreement with the measured slope of the Voyager 2 USO (-1.28×10^{-7} Hz/sec) [8]. This translates to an aging rate over a one day period of -5.6×10^{-12}, which shows that the USO is drifting well below its specification of 5×10^{-11}/day [9].

In an attempt to verify when the USO had reached its linear aging realm, different subsets of the last several passes prior to first turn-off in Fig. 20 were subjected to linear fits of frequency versus time. Initial fits of the last 31 passes prior to first turn-off (90-262 to 91-154), and the last 21 passes (90-360 to 91-154) displayed a signature in the residuals with significant curvature. A linear fit of the last 19 passes (91-006 to 91-154) showed no significant residual signature and yielded results consistent with the combination logarithmic/linear combination fit discussed previously. This implies that the Galileo USO required a little over a year of operation time prior to reaching its linear aging realm.

It is preferred that once the USO is switched on, it is left on so that it is allowed to reach the linear aging realm where it should remain for the duration of the Galileo activities. The linear aging realm allows USO-referenced (one-way) Radio Science experiments such as the gravitational redshift experiment

and planetary occultations to be accurately calibrated. Since August 1991, the USO was switched off and back on in support of cooling turns as part of the effort to free the High Gain Antenna (HGA) stuck ribs (in conjunction with corresponding warming turns). The USO and its heater were powered-off on Aug. 5, 1991 (91-217). There were six passes conducted after the USO was powered back on Aug. 16, 1991 (91-228). The estimated spacecraft transmitted frequencies for these passes display a smooth continuity in time (see Fig. 23). A simple four-parameter aging model was fit to these frequencies yielding a post-fit rms scatter of 3 mHz. Although the number of points and time period were insufficient to infer an accurate linear aging rate, this result suggests that the USO is behaving in a reasonable manner during this period. The USO was again turned off after this time period. If the USO had been allowed to remain on, it was expected that not as long of a period would be required for it to reach its linear aging realm as was the case during the first data set (89-341 to 91-154). The Allan deviations appear consistent with those of the first on-off data set, although two of the passes have somewhat higher values.

G. Retrace

If the USO and its heater are turned off, and then turned back on, the crystal will oscillate at a different frequency which is difficult to predict. This phenomena known as "retrace" is defined as the nonrepeatability of the frequency vs. temperature characteristic at a fixed temperature upon on-off cycling of the oscillator [7]. An example of "retrace" is the 12 Hz jump between passes on 91-154 and 91-247 (see Fig. 20). Between these passes, the USO and its oven were turned off for an eleven day period in Aug. 1991. Several mechanisms which can cause retrace include strain changes, changes in the quartz, oscillator circuitry changes, contamination redistribution in the crystal enclosure and apparent hysteresis [7].

H. Assessment of Environmental Effects

The frequency of the USO can change due to variations in environmental parameters which include temperature, pressure, acceleration, magnetic field, and radiation. The crystal frequency is also dependent on the power level. See [7], [10] and [11] for discussions on the effects of the various phenomena on the behavior of crystal oscillator devices.

The Galileo USO was designed to minimize the effects of magnetic fields. A worst case estimate of the magnetic field of the spacecraft in the environment of the USO found that the resulting fluctuations in frequency were expected to be negligible for the 10 to 1000 sec time intervals [12].

Since the USO is oven controlled so as to maintain a constant temperature, and the crystal temperature is designed to operate at an optimum point on the f vs T curve, noise due to temperature fluctuations is expected to be insignificant.

The majority of the USO passes were conducted during quiescent periods on the spacecraft when there was no scheduled dynamic activity. The few known exceptions of dynamic activity involved no net thrusts on the spacecraft. Acceleration effects on the spacecraft USO are therefore considered to be negligible.

It is assumed that in the deep space environment, changes due to atmospheric pressure and humidity are virtually nonexistent. Pre-launch testing showed that the Galileo USO exhibited spikes of less than 1 mHz during changes in pressure [13].

Due to the high level of shielding, no charged particles are expected to hit the crystal during cruise. High energy particles are expected to be stopped by the lead shielding producing photons which could possibly hit the crystal (G. Wood, private communication). The level of radiation reaching the crystal during the cruise phase is expected to cause negligible shifts in frequency. The USO frequency could however shift about 1 Hz during passage through the Jovian radiation belts [14].

VII. Conclusion

The Galileo USO appears to be healthy and functioning in a reasonable manner based on the analysis presented in this article. The evaluation of the Radio Science instrument will continue for the duration of the mission.

Acknowledgements

P. Eshe, R. Herrera and T. Horton

of the Radio Science Support Team provided valuable support. The comments and discussions provided by R. Sydnor and G. Wood of JPL, and R. Filler of LABCOM, New Jersey were greatly appreciated. The authors would also like to thank the Galileo Flight Team and Galileo Navigation Team for their support. The NASA/Deep Space Network generated the data used in this analysis. This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, sponsered by the National Aeronautics and Space Administration.

References

1. Krisher, T. P., Morabito, D. D., and Anderson, J. D., "The Galileo Solar Redshift Experiment", Physical Review Letters, Vol. 70, pp. 2213-2216, April 12, 1993.

2. Howard, H. T., Eshleman, V. R., Hinson, D. P., Kliore, A. J., Lindal, G. F., Woo, R., Bird, M. K., Volland, H., Edenhofer, P., Pätzold, M, and Porsche, H., "Galileo Radio Science Investigations", Space Science Reviews 60, 565-590 (1992).

3. Anderson, J. D., Armstrong, J. W., Campbell, J., K., Estabrook, F. B., Krisher, T. P., and Lau, E. L., "Gravitation and Celestial Mechanics Investigations with Galileo", Space Science Reviews 60, pp. 591-610 (1992).

4. Pollmeier, V. M., and Kallemeyn, P. H., "Galileo Orbit Determination from Launch through the First Earth Flyby", Proc. of the 47th Ann. Meeting of the Inst. of Nav., Williamsburg, Pa., June 10-12, 1991, pp. 9-16.

5. Kirk, A. "Frequency Stability Measurements of Galileo Project High Stability Crystal Oscillators" JPL 331-TRAK-800527, May 1, 1980.

6. Lesage, P. and Audoin, C., "Characterization of Frequency Stability: Uncertainty Due to the Finite Number of Measurements", IEEE Trans. on Inst. and Meas., Vol. IM-22, June 1973.

7. Vig, J. R., "Introduction to Quartz Frequency Standards" Research and Development Technical Report SLCET-TR-91-1 (Rev. 1) October 1992, Army Research Laboratory, Electronics and Power Sources Directorate, Fort Monmouth, N.J. 07703-5601, USA.

8. Asmar, S. W., and Eshe, P. M., "Evaluation of the USO Performance - Final Report", JPL IOM Voyager-RSST-90-121, January 17, 1990.

9. A. Gussner, "Summary of Galileo USO Testing", JPL IOM 3364-80-080, Aug. 20, 1980

10. Howe, D. P., "Frequency Domain Stability Measurement," National Bureau of Standards, Technical Note 679, U.S. Dept. of Commerce, PB-252-171, March 1976.

11. Parzen, B., Design of Crystal and Other Harmonic Oscillators, John Wiley & Sons, 1983.

12. Postal, R. "A Concern of USO Stability as a Function of Magnetic Field", JPL IOM 3362-87,019, June, 3, 1987.

13. Gussner, A., "Summary of Galileo Ultra Stable Oscillator (USO) Testing", JPL IOM 3364-80-080, Aug. 20, 1980.

14. Wood, G. E., "Radiation Testing of Ultra Stable Oscillator S/N 004", JPL IOM 3396-76-095, August 20, 1976.

Table 1. Galileo USO Pass Summary

YEAR	DOY	START HR:MN:SC	END HR:MN:SC	DSS ID	AGC (dBm)	S/C TRANSMITTED FREQ - 2294997000 (Hz)
89	341	21:34:22	23:27:59	14	-149.5	690.321
89	350	00:02:40	01:57:59	14	-142.4	696.287
89	360	22:35:42	00:00:59	14	-145.5	699.954
90	2	17:09:45	18:59:59	14	-149.0	701.425
90	9	17:37:22	19:26:08	14	-149.6	702.669
90	15	17:34:22	19:28:11	14	-153.1	703.597
90	19	16:10:46	17:57:50	14	-156.1	704.134
90	28	16:00:09	18:00:00	14	-154.3	705.128
90	32	15:33:51	17:29:59	14	-155.8	705.508
90	37	10:50:57	12:12:36	63	-155.0	705.895
90	44	03:00:00	05:19:59	43	-151.6	706.383
90	46	00:19:29	01:04:57	43	-154.3	706.523
90	49	19:44:31	21:22:29	43	-153.9	706.855
90	56	02:10:12	03:26:48	43	-157.4	707.295
90	58	23:08:28	01:01:01	43	-158.5	707.457
90	61	00:17:51	01:55:53	43	-162.2	707.586
90	68	02:12:12	03:42:02	43	-159.5	707.987
90	76	02:19:06	03:28:03	43	-165.0	708.376
90	78	00:41:32	02:27:47	43	-166.5	708.452
90	83	19:13:51	20:56:11	43	-167.5	708.708
90	89	00:09:38	02:03:16	43	-166.8	708.888
90	97	22:18:28	23:55:22	43	-165.3	709.205
90	104	22:14:52	23:56:39	43	-167.6	709.422
90	110	00:47:05	02:26:45	43	-168.0	709.570
90	113	20:46:51	22:01:00	43	-169.0	709.671
90	121	23:21:02	00:57:56	43	-166.5	709.875
90	128	09:26:40	11:04:42	63	-168.3	710.014
90	136	23:22:03	01:16:30	43	-167.0	710.111
90	139	19:55:00	21:48:51	43	-166.9	710.166
90	149	19:17:45	21:02:08	43	-165.0	710.337
90	155	19:18:09	19:45:01	43	-165.7	710.395
90	162	18:14:46	19:43:47	14	-166.6	710.457
90	172	15:18:15	17:04:29	14	-164.8	710.566
90	176	16:19:35	18:05:46	14	-165.7	710.587
90	183	18:20:15	20:07:25	43	-165.6	710.634
90	193	14:48:37	16:34:51	14	-162.8	710.704
90	197	18:19:45	20:02:33	43	-165.0	710.720
90	206	21:17:32	23:00:05	43	-165.3	710.737
90	213	21:12:49	22:03:33	43	-165.6	710.769
90	221	20:43:23	22:29:37	43	-163.7	710.778
90	228	18:14:37	19:59:44	43	-162.8	710.775
90	233	15:17:38	17:03:53	14	-168.1	710.785
90	243	19:44:31	21:24:40	43	-165.5	710.764
90	250	01:41:02	02:45:10	63	-162.7	710.757
90	253	19:13:26	20:55:34	43	-161.1	710.735
90	262	19:26:55	21:11:16	43	-160.5	710.731
90	268	00:30:00	03:40:00	63	-162.5	710.706
90	274	18:56:25	20:43:34	43	-160.1	710.658
90	281	17:12:34	18:54:36	43	-156.9	710.612
90	303	01:44:22	03:23:19	63	-153.0	710.495
90	312	01:16:55	02:57:56	63	-148.5	710.432
90	321	03:18:28	04:58:11	63	-145.1	710.326
90	330	01:13:20	02:57:56	63	-145.0	710.271
90	345	02:14:52	03:42:02	42	-138.4	710.169

Table 1. Galileo USO Pass Summary (continued)

YEAR	DOY	START HR:MN:SC	END HR:MN:SC	DSS ID	AGC (dBm)	S/C TRANSMITTED FREQ - 2294997000 (Hz)
90	350	08:18:28	10:12:02	61	-147.7	710.115
90	360	17:09:14	18:57:41	43	-150.3	709.991
91	4	07:32:13	07:49:48	63	-149.4	709.922
91	6	18:01:47	19:58:45	43	-149.1	709.885
91	14	16:55:11	18:46:30	43	-152.5	709.792
91	16	17:04:22	18:59:59	43	-151.3	709.769
91	19	16:05:14	17:59:59	43	-152.5	709.743
91	21	18:04:52	19:58:57	43	-153.1	709.709
91	26	16:07:05	17:50:51	43	-154.8	709.657
91	29	06:07:42	07:58:54	63	-156.1	709.638
91	33	15:58:12	17:43:56	43	-158.1	709.577
91	36	06:07:11	07:58:50	63	-162.6	709.548
91	39	06:05:08	08:01:13	63	-160.9	709.507
91	43	15:05:08	16:59:59	43	-162.2	709.459
91	49	21:21:14	23:13:13	43	-160.6	709.370
91	55	15:35:23	17:30:30	43	-160.7	709.300
91	64	14:05:14	16:02:30	43	-158.6	709.173
91	73	16:23:17	18:13:59	43	-160.2	709.068
91	81	15:08:37	17:02:30	43	-155.4	708.929
91	109	19:02:38	20:45:14	43	-155.4	708.563
91	141	00:02:40	01:59:52	63	-157.6	708.161
91	154	23:04:06	00:59:41	61	-164.6	707.997
91	247	16:39:14	18:23:50	63	-167.5	696.304
91	259	21:28:12	23:11:46	14	-164.8	696.556
91	275	00:59:23	02:43:34	43	-165.9	696.678
91	292	15:15:11	17:00:01	63	-168.2	696.728
91	318	03:54:25	05:29:56	43	-168.1	696.772
91	334	20:41:17	22:10:45	14	-165.9	696.769

Table 2. Summary of Galileo USO Allan Deviations

Time Interval (sec)	Measured Flight x 10^{12}	Measured Pre-launch x 10^{12}
1	29.4 ± 1.1	0.56
10	3.93 ± 0.17	0.56
100	0.90 ± 0.03	1.1
1000	0.71 ± 0.03	0.68

Note: Uncertainties are errors in the mean.

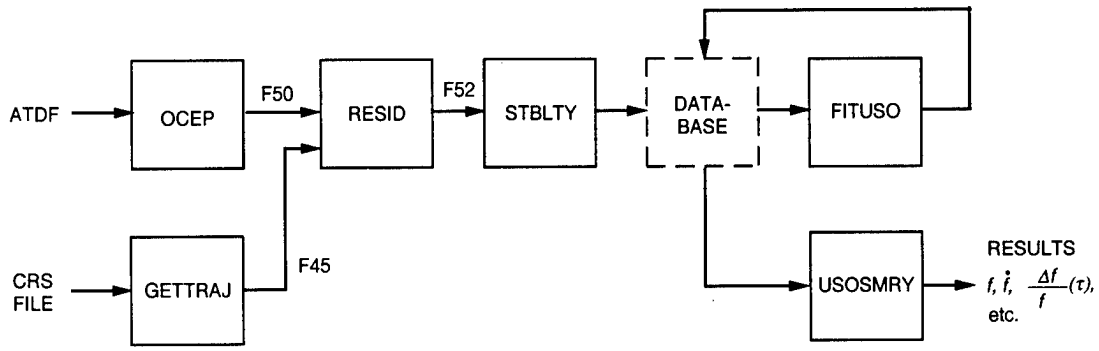


Fig. 1. STBLTY program-set block diagram.

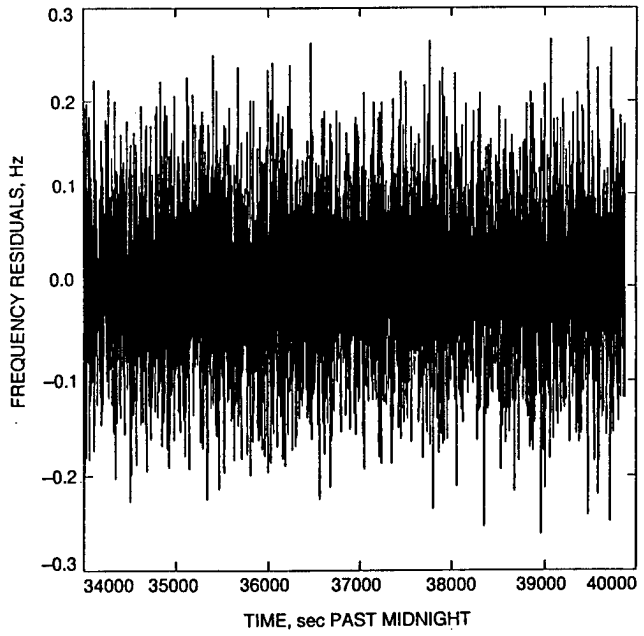


Fig. 2. Frequency residuals of sampled 1/sec Doppler for the USO pass of May 8, 1990.

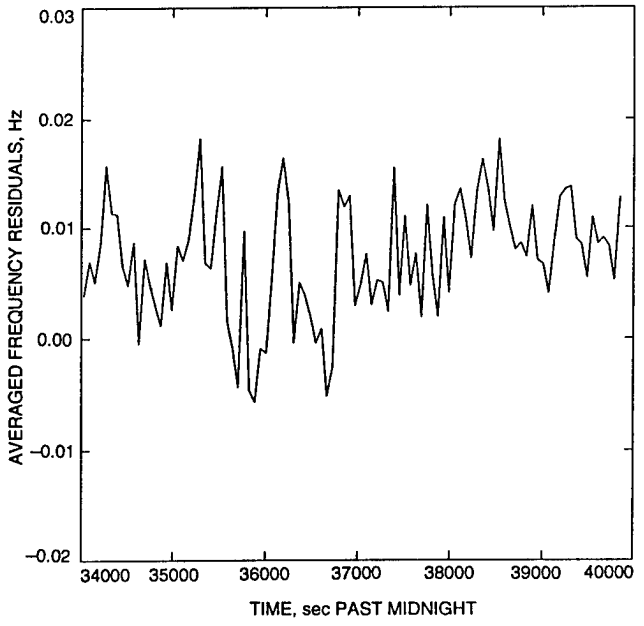


Fig. 3. Frequency residuals averaged every 60 sec for the USO pass of May 8, 1990.

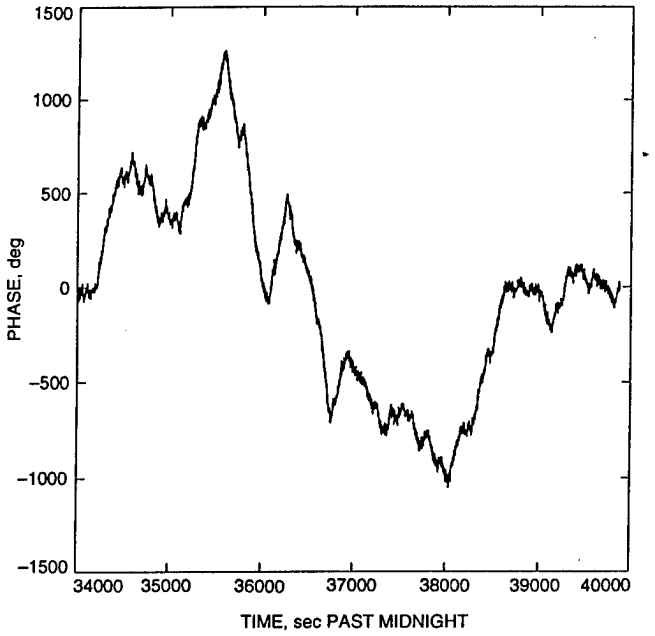


Fig. 4. Phase reconstructed from frequency residuals for the USO pass of May 8, 1990.

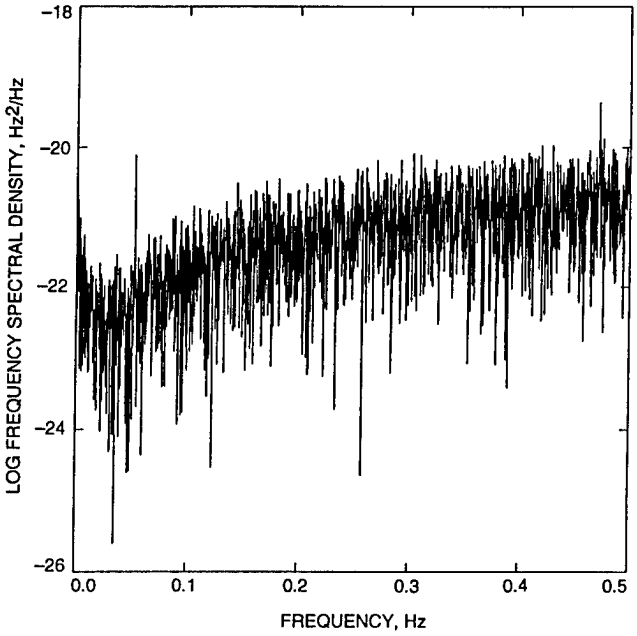


Fig. 6. Log of frequency spectral density of frequency residuals for the USO pass of May 8, 1990.

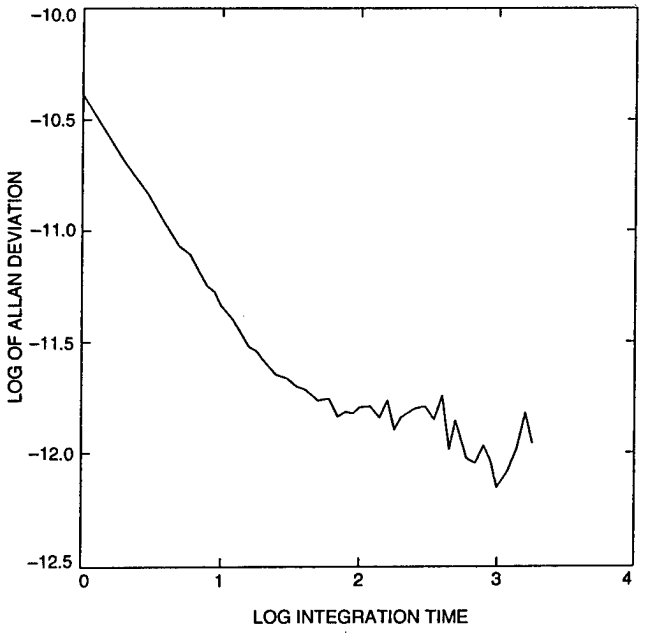


Fig. 5. Log of Allan deviation of frequency residuals for the USO pass of May 8, 1990.

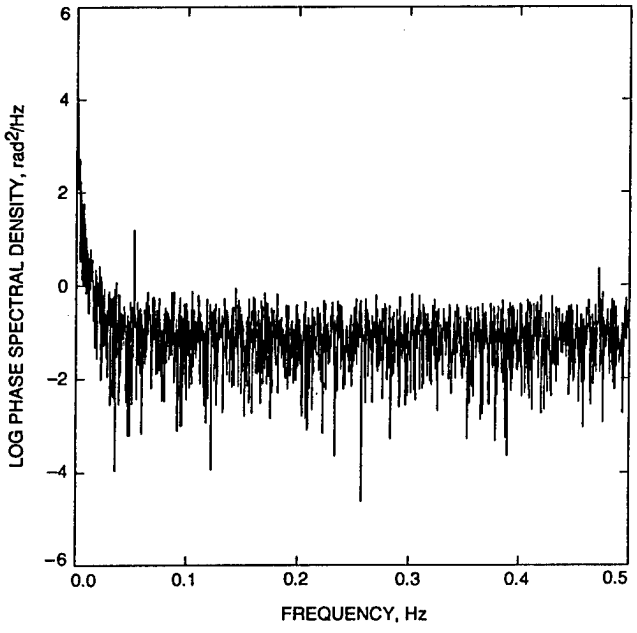


Fig. 7. Log of phase spectral density of frequency residuals for the USO pass of May 8, 1990.

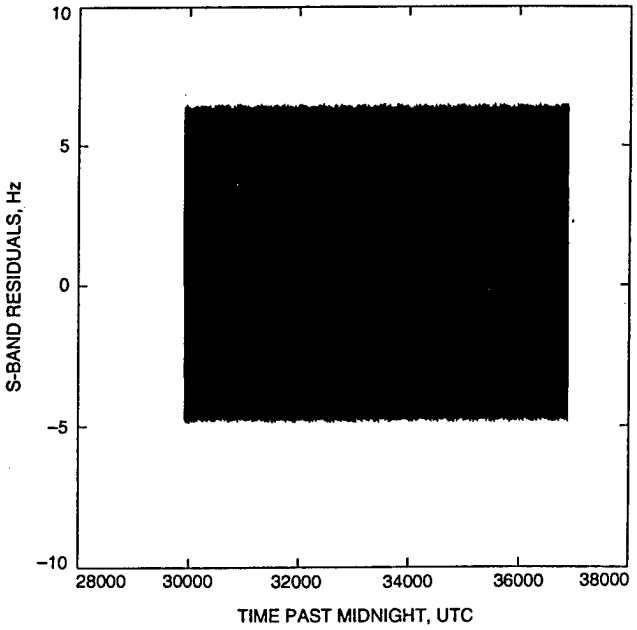


Fig. 8. Frequency residuals of sampled 1/sec Doppler for USO pass of December 16, 1990, where LGA-2 was the spacecraft antenna.

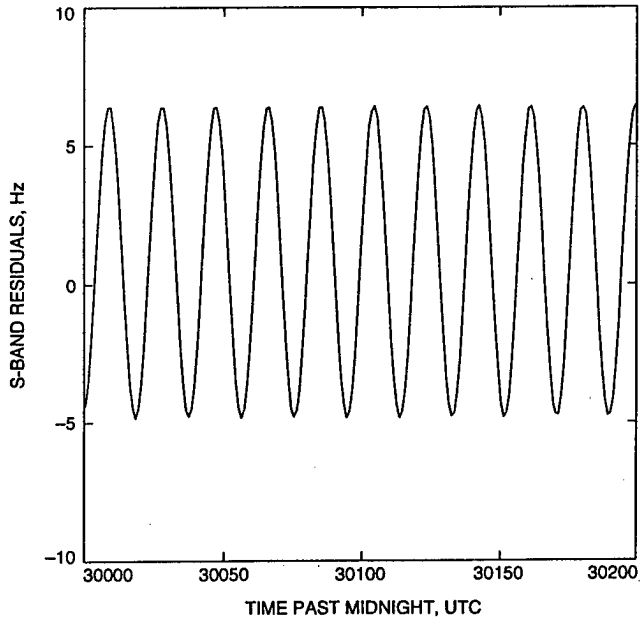


Fig. 9. Selected 200-sec period of frequency residuals of sampled 1/sec Doppler for USO pass of December 16, 1990, where LGA-2 was the spacecraft antenna.

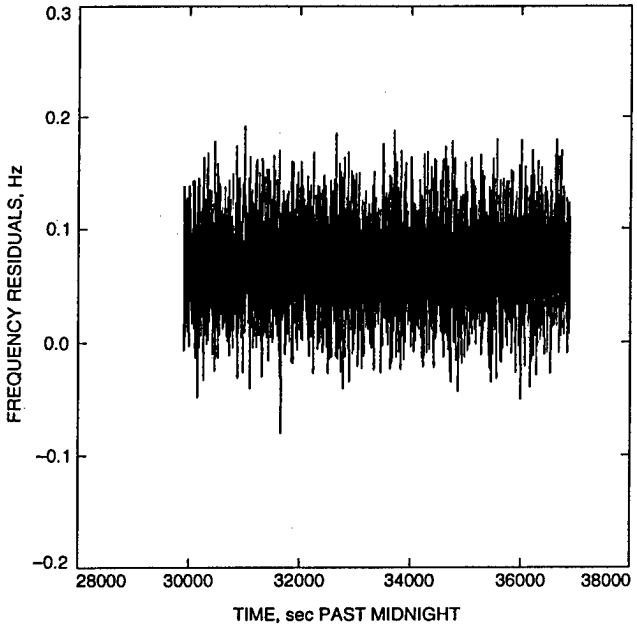


Fig. 10. Frequency residuals after removing the sinusoid fit from residuals displayed in Fig. 8 for USO pass of December 16, 1990, where LGA-2 was the spacecraft antenna.

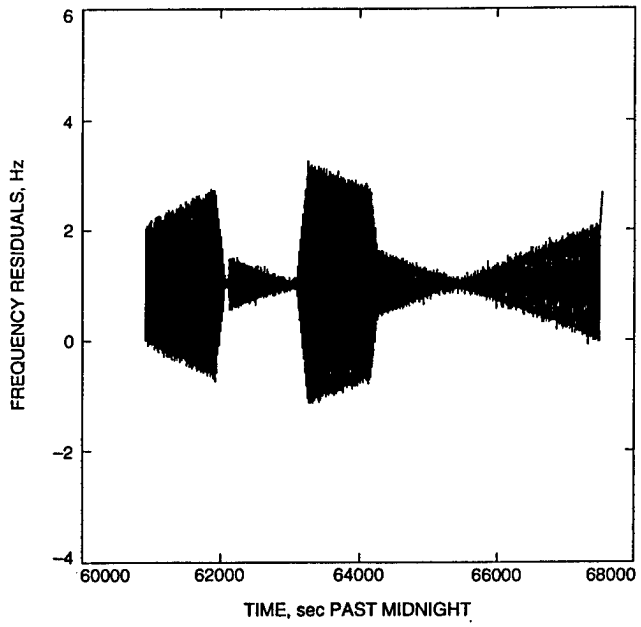


Fig. 11. Frequency residuals of sampled 1/sec Doppler for the USO pass of January 14, 1991, where LGA-2 was the signal source and dynamic motion occurred on board the spacecraft.

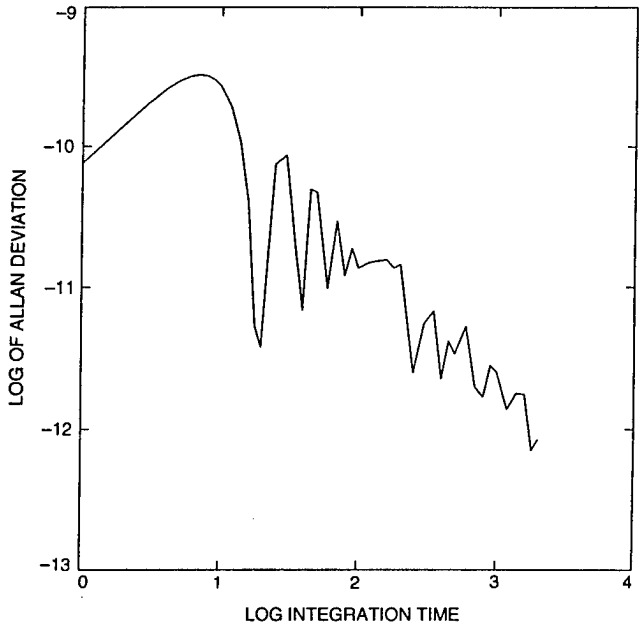


Fig. 12. Log of Allan deviation of frequency residuals of Fig. 11.

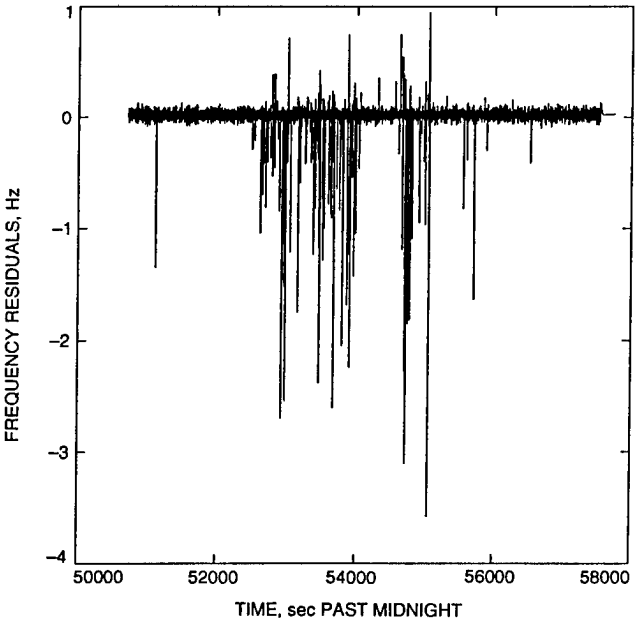


Fig. 13. Frequency residuals of the USO pass of March 5, 1991, where solar activity was known to have occurred.

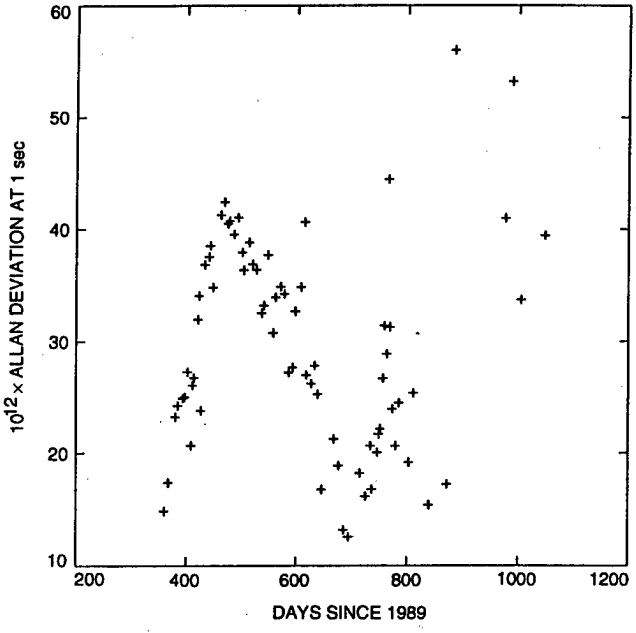


Fig. 14. Allan deviations at 1 sec for 73 USO passes (outliers have been removed).

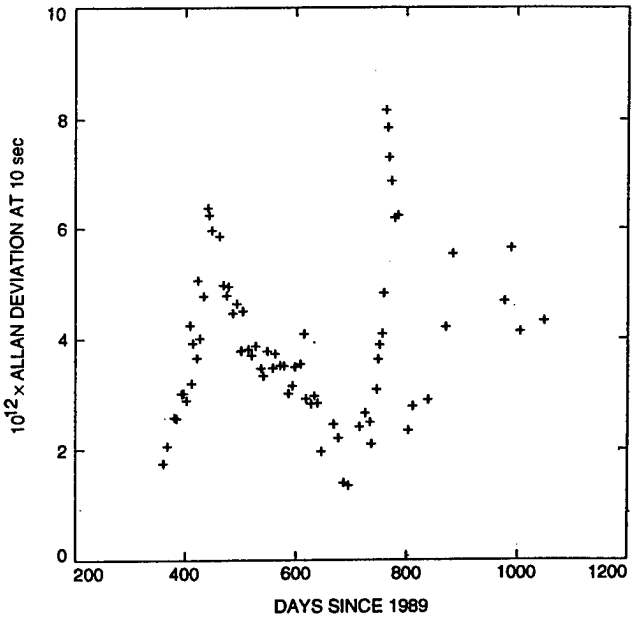


Fig. 15. Allan deviations at 10 sec for 73 USO passes (outliers have been removed).

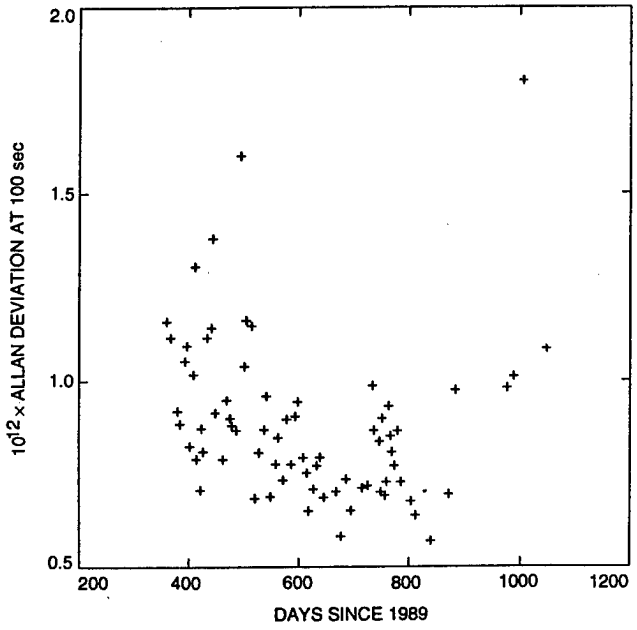


Fig. 16. Allan deviations at 100 sec for 73 USO passes (outliers have been removed).

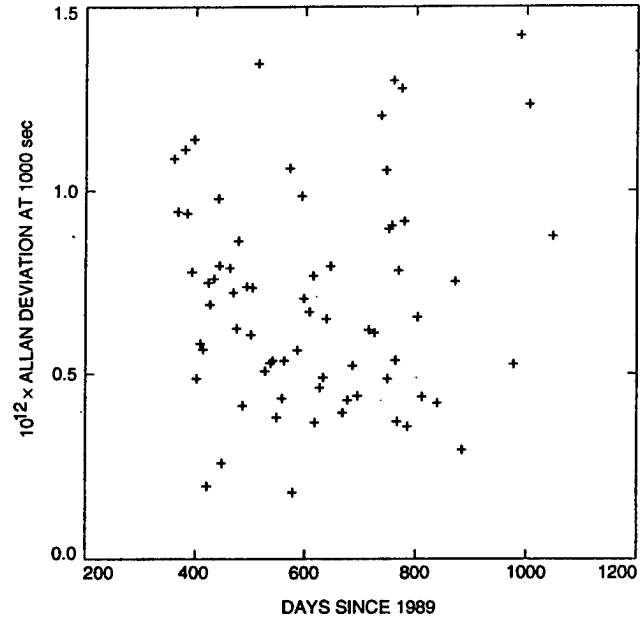


Fig. 17. Allan deviations at 1000 sec for 70 USO passes (outliers have been removed).

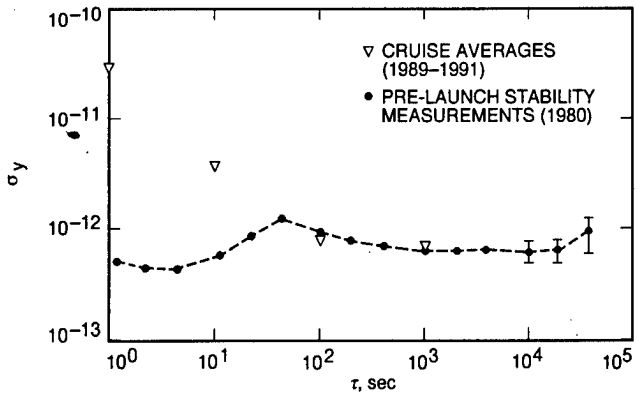


Fig. 18. In-flight USO-pass Allan-deviation measurement averages for 1, 10, 100 and 1000 sec superimposed with preflight Allan deviation measurements.

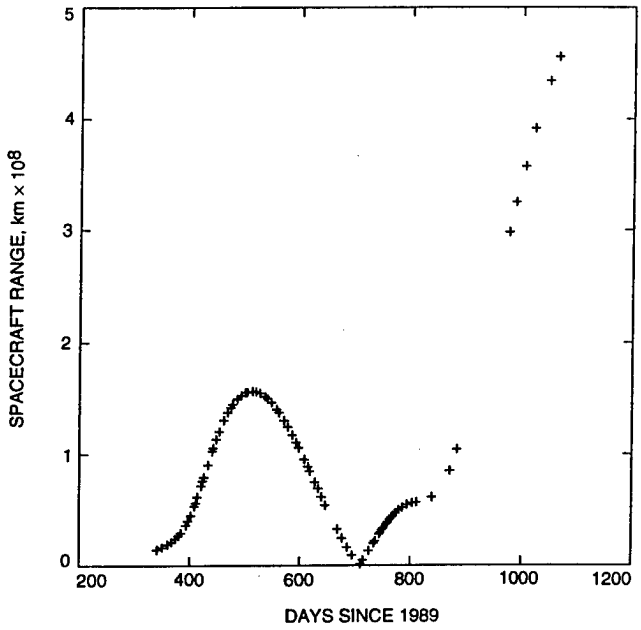


Fig. 19. Spacecraft range for each USO pass. Launch occurred at day 291; the dip at 707 days after 1989.0 was the Earth 1 flyby.

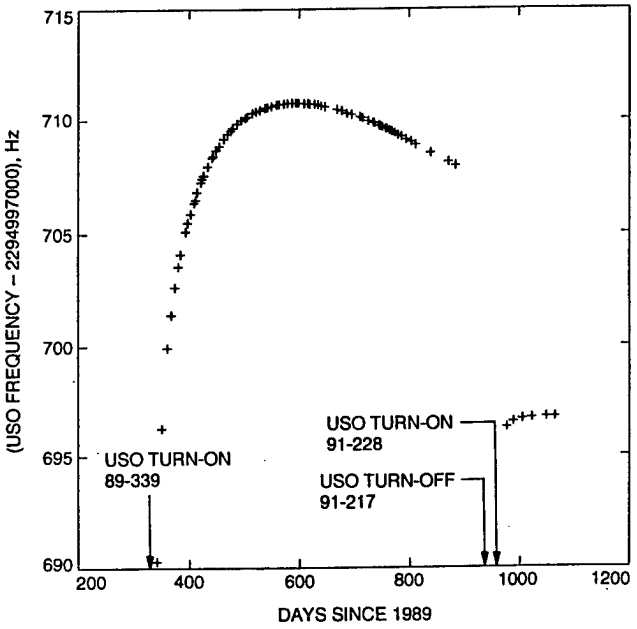


Fig. 20. Estimated spacecraft transmitted frequencies for all 82 USO passes as determined by STBLTY.

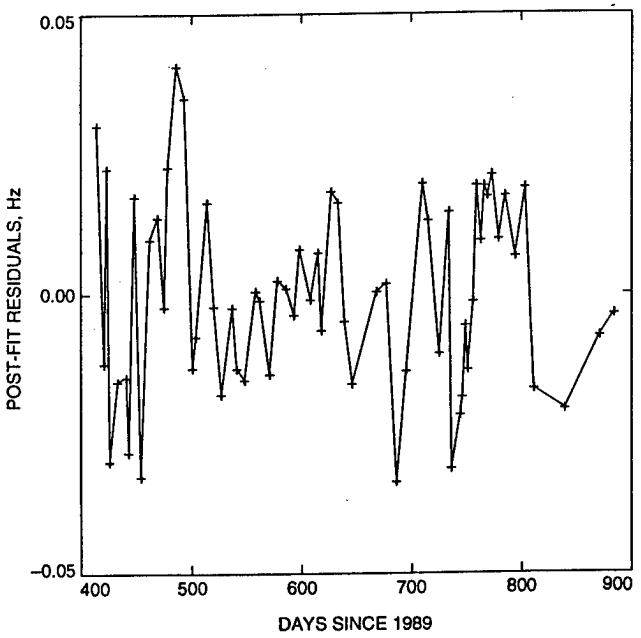


Fig. 22. Post-fit residuals of estimated spacecraft transmitted frequencies of the first USO on–off cycle for 64 passes (first 12 removed) after fitting and removing an aging model.

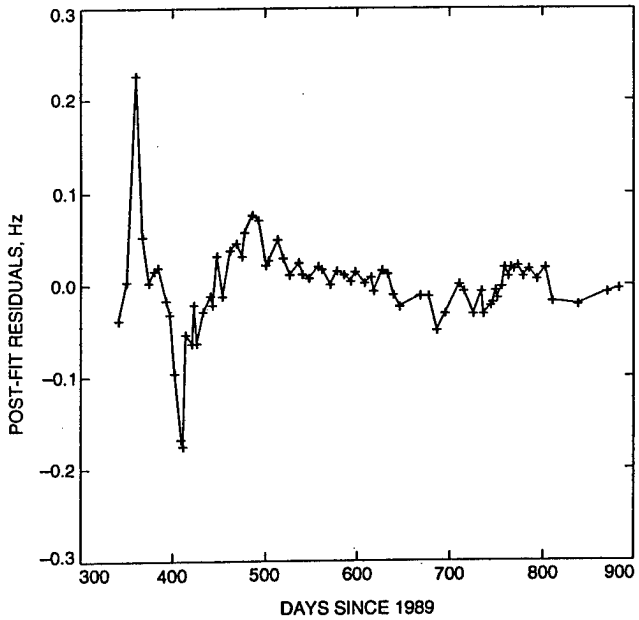


Fig. 21. Post-fit residuals of estimated spacecraft transmitted frequencies of the first USO on–off cycle for all 76 passes after fitting and removing an aging model.

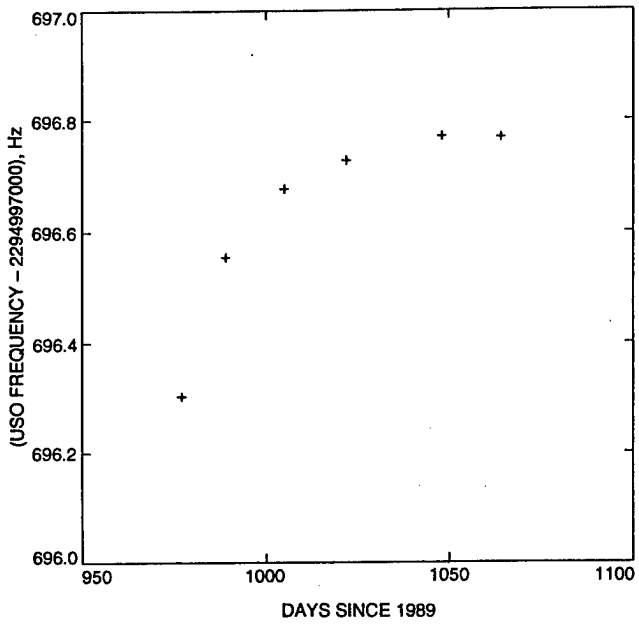


Fig. 23. Estimated spacecraft transmitted frequencies of 6 USO passes conducted after the first USO on–off cycle.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

FACTORIZATIONAL SYNTHESIS OF SAW BANDPASS FILTERS

Alexander S. Rukhlenko

Department of Semiconductors Physics, Belorussian State University
4, Fr. Skorina Avenue, 220800, Minsk, Republic of Belarus, CIS

Abstract: Design of surface acoustic wave (SAW) filters with linear phase (LP), prescribed nonlinear phase (NLP), minimum phase (MP), and optimum minimum phase (OMP) characteristics by factorizing (splitting) the filter frequency response is considered. The universal factorization algorithm based on the Z-transform roots searching, lexicographical sorting, and sharing between the transducers is proposed in the paper. A feature of the algorithm is that irrespective of the overall phase specifications both weighted IDT tend to have minimum phase characteristics being less sensitive to the tap-weighting inaccuracies.

The design examples of LP, NLP, MP, and OMP SAW filters are presented which confirm the flexibility and efficiency of the design procedure.

Introduction

In the first order approximation the frequency response (FR) of a SAW filter can be written as the following product [1]

$$F(\omega)=\xi(\omega)\,F_1(\omega)\,F_2^*(\omega) \tag{1}$$

where the functions $F_i(\omega)$, $i=1,2$, are attributed to the IDT array factors, and the asterisk denotes the complex-conjugation. The skewing factor $\xi(\omega)$ is introduced to account for the IDT element factors and/or the multistrip coupler FR, etc. Without the loss of generality we may suppose hereafter $\xi(\omega)=1$ as this factor can be implicitly accounted for by a proper predistortion of the function $F(\omega)$.

To simplify the design the frequency response $F_1(\omega)$ of one of the transducers is usually supposed to be given a priori while the other's $F_2(\omega)$ is optimized to satisfy the filter transfer function specifications [2, 3].

An alternative design simplification is to use two identical weighted IDT with the "halved" frequency response

$$F_1(\omega)=F_2^*(\omega)=\sqrt{F(\omega)}\ . \tag{2}$$

But in general case both functions $F_1(\omega)$ and $F_2(\omega)$ are to be found separately from the design, any a priori restrictions to them resulting in a filter length increase.

For linear phase SAW filters composed of two weighted IDT in conjunction with the multistrip coupler the factorizational synthesis technique [4-6] based on the Z-transform roots searching and sharing was elaborated. The present paper generalizes the factorization technique to the NLP design including minimum phase SAW filters.

Basic Z-transform Properties of a SAW Filter Response

Supposed the functions $F_i(\varphi)$, $i=1,2$ to be the complex trigonometric (exponential) polynomials of the order N_i, the overall transfer function $F(\varphi)$ is also the trigonometric polynomial of the order $N=N_1+N_2-1$, i.e.

$$F(\varphi)=F_1(\varphi)F_2^*(\varphi)=\sum_{k=0}^{N-1}A_k e^{jk\varphi} \tag{3}$$

where $\varphi=\pi\omega/\omega_*$ is an angle variable, ω_* being a transducer synchronous frequency. The complex trigonometric polynomial (3) can be converted to the algebraic polynomial $F(z)$ using Z-transform [7]

$$F(z)=F_1(z)F_2(1/z)=\sum_{k=0}^{N-1}A_k z^k,\ z=re^{j\varphi}\ . \tag{4}$$

0-7803-0905-7/93 $3.00 © 1993 IEEE

The frequency response $F(\varphi)$ is a Z-transform $F(z)$ evaluated on the unit circle $z=e^{j\varphi}$.

According to the fundamental theorem of algebra the scaled polynomial (4) can be expressed in terms of their roots z_i as follows

$$F(z)=\prod_{i=1}^{N-1}(z-z_i)=\prod_i M_i(z)\prod_j D_j(z)\prod_k Q_k(z) \qquad (5)$$

where for the real-valued coefficients A_k the roots (zeros) $z_i=r_i e^{j\varphi_i}$ appear in the complex-conjugated pairs (couples)

$$D_j(z)=D(z,z_j)=(z-z_j)(z-z_j^*),\ \varphi_j\neq\pi n \qquad (6)$$

or degenerate into the single zeros (monozeros) on the real axis of the Z-plane

$$M_i(z)=M(z,z_i)=z-z_i,\ \varphi_i=\pi n. \qquad (7)$$

Moreover, to ensure a phase linearity each zero couple at the point z_i off the unit circle ($r_i\neq 1$, $\varphi_i\neq\pi n$) must have its reciprocal at the point $1/z_i$ with the same scaled magnitude response but with the phase reversal composing a linear phase quadruplet

$$Q_k(z)=Q(z,z_k,1/z_k)=D(z,z_k)D(z,1/z_k)= \\ =(z-z_k)(z-z_k^*)(z-1/z_k)(z-1/z_k^*) \qquad (8)$$

when combined together (Fig.1).

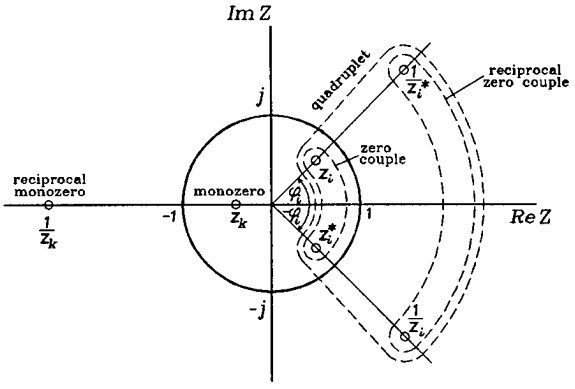


Fig. 1. Z-transform monozeros, couples of zeros, and linear phase quadruplet

Thus, the zeros distribution on the Z-plane depends highly on the SAW filter magnitude shape and phase characteristic required. While in the general case of the NLP filter zero couples $D_i(z)$ may occur both inside as well as outside the unit circle providing the desired magnitude and phase approximation, all zeros off the unit circle must occur in quadruplets to ensure a phase linearity. Minimum (or maximum) phase SAW filters have zeros inside (or outside) the unit circle. Zero couples on the unit circle correspond to the real roots of the overall transfer function $F(\varphi)$ at the points φ_i and always have linear phase response.

The factorized form (5) allows to observe visually how each zero contributes to the overall transfer function. One could effectively control a SAW frequency response by the withdrawal, substitution, and addition of zeros.

An illustrative example of the elemental root frequency response $D_i(\varphi)$ is shown in Fig. 2 for $r_i=0.5$, 0.75, and 1.0 where $\varphi_i=\pi/4$. One can see that moving a zero z_i off the unit circle results in the deterioration of the outofband attenuation.

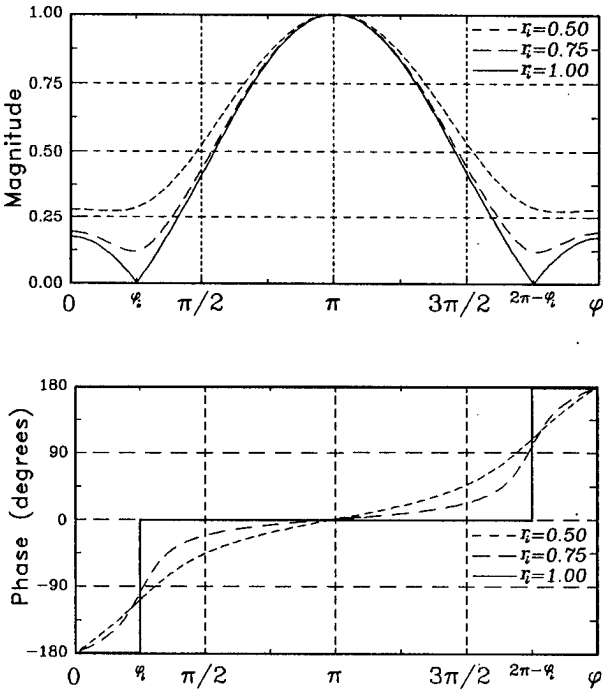


Fig. 2. Frequency response $D_i(\varphi)$ of the elemental zero couple

Factorizational Synthesis Algorithm

The ultimate purpose of the algorithm is to determine the separate transducer responses $F_1(\varphi)$, $F_2(\varphi)$, and the transducers tap-weights from the overall transfer function $F(\varphi)$.

The design procedure starts from the optimization of the function $F(\varphi)$ to meet the desired magnitude and phase specifications. The known FIR synthesis techniques, for example the Remez exchange algorithm [7] or linear programming [8] may be directly applied to the LP design without any adaptations. But to meet the supplementary phase or group delay requirements one should optimize separately the real and image part using one of the tolerance transformation scheme [8-11] or resort to nonlinear programming [12].

The functions $F_1(\varphi)$ and $F_2(\varphi)$ so far unknown can be found from (5) by sharing the roots z_i in the systematic manner between IDT. There are many possible ways of factorizing but one should take into consideration the following:

1) a stopband attenuation depends on the zeros distribution density. The closer are the roots allocated on the unit circle, the higher is a stopband attenuation;

2) moving the root off the unit circle improves the passband shape factor but deteriorates the stopband attenuation.

Hence, while factorizing it is preferable to maintain in the stopband nearly the even zeros density on (or near) the unit circle and to minimize the magnitude difference of the transducers elemental root factors in the passband.

The following factorization algorithm may be applied.

1. Find Z-transform roots $z_i=e^{j\varphi_i}$ using the root solving program for high-order polynomials [13].

2. Sort the roots lexicographically in their angles φ_i and radii r_i.

3. Assign every second root in the filter stopband and all passband roots inside the unit circle to the input IDT, the others being attributed to the output IDT.

4. Restore the coefficients of the complex trigonometric polynomials $F_1(\varphi)$ and $F_2^*(\varphi)$ by using the multiple recurrent convolution of the elemental root factors $M_i(\varphi)$ and $D_j(\varphi)$ or by performing the discrete Fourier transform.

5. Invert in time the coefficients of the output IDT to account for the complex-conjugation of the function $F_2(\varphi)$ in the expression (1).

The algorithm is general and may be applied both LP and NLP designs.

A feature of the factorization technique above is that after the step 5 both IDT have their roots inside or on the unit circle satisfying the minimum phase condition [7]. Such transducers are less sensitive to the tap-weighting inaccuracies [14] if compared to the conventional design where two identical weighted IDT are used.

Another feature is that quadruplets are splitted into inside and outside zero couples with the same scaled magnitude response to be attributed to the input and output IDT. Thus, the minimal difference in the passband magnitude shape for both transducers is attained.

Linear Phase Design

The factorizational synthesis is illustrated by an example of a SAW bandpass LP filter design with the -3 dB fractional bandwidth of 50% and the -3/-40 dB shape factor of 1.5, the filter center frequency being 25 MHz. The design results are presented in Fig. 3 and 4 where for comparison the conventional design results using two identical IDT are also shown. The design specifications are the same for both designs, the stopband attenuation being -60 dB and the passband peak-to-peak ripple being 0.1 dB.

The Z-transform roots distribution is shown in Fig. 3 where bars and circles denote zeros attributed to the input and output IDT respectively. We can see that there are 4 monozeros and 3 LP quadruplets off the unit circle, other 20 LP zero couples allocated on the unit circle in the case of factorizational synthesis. In the case of the conventional design all zeros (4 monozeros, 3 quadruplets, and 12 couples of zeros) are double (Fig.3b).

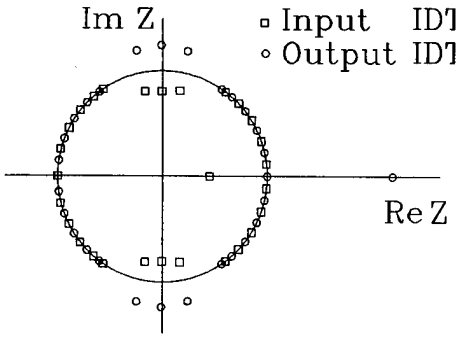
The filter frequency responses versus a normalized frequency $\nu=\varphi/2\pi=\omega/2\omega_*$ are shown in Fig. 4 where the curves 1 and 2 correspond to the responses $F_1(\omega)$ and $F_2(\omega)$ of the input and output IDT, the curve 3 being the filter transfer function $F(\omega)=F_1(\omega)F_2^*(\omega)$.

At the same approximation accuracy the filter length is $N=57$ ($N_1=N_2=29$) in the factorizational synthesis and $N=81$ ($N_1=N_2=41$) in the conventional design. In both cases the IDT synchronous frequency $\omega_*=2\omega_0$ that corre-

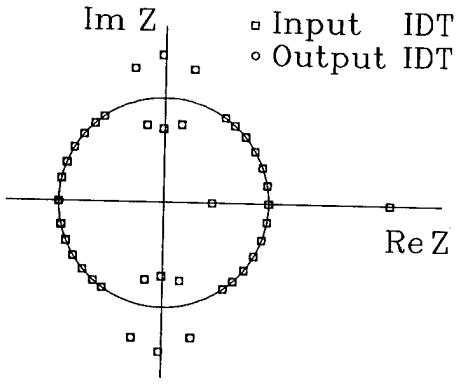
sponds to the IDT structures with splitted electrodes, ω_0 being the filter central frequency.

The gain in the filter length due to the factorization is of 30% for this example if compared to the conventional design. The shorter filter length is due to a more rational arrangement of zeros and extremuma in a filter stopband (Fig. 4). Namely, extremuma of one transducer are allocated in the proximity of zeros of the other transducer and vice versa while all extremuma and zeros are superimposed each other in the conventional design.

In the case of factorizational synthesis both transducers have identical minimum phase group delay time characteristics (Fig.4a) to cancel each other in the overall filter frequency response due to the complex-conjugation. Contrary to that, the transducers have linear phase characteristics in the conventional design.

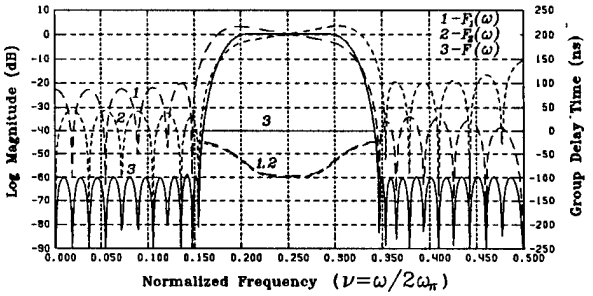


(a) factorizational synthesis

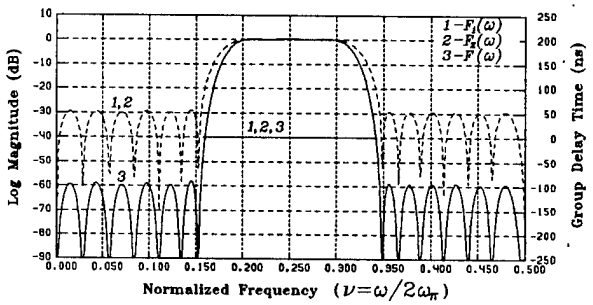


(b) conventional design

Fig. 3. Z-transform roots distribution of a SAW bandpass LP filter



(a) factorizational synthesis



(b) conventional design

Fig. 4. Frequency response of a SAW bandpass LP filter

Nonlinear Phase Design

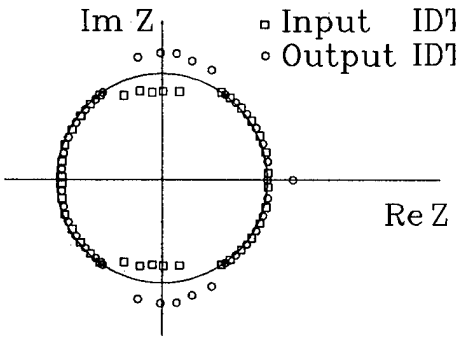
The nonlinear phase design is illustrated by the factorizational synthesis of a dispersive (chirp) SAW bandpass filter with the same magnitude specifications as in the previous example.

The design results are presented in Fig. 5 where the advanced tolerance transformation scheme [11] was used for the NLP filter transfer function synthesis.

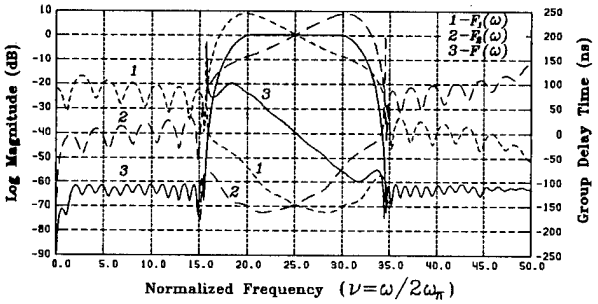
Due to the supplementary group delay time requirements the filter length was increased to $N=77$, the transducer lengths being $N_1=N_2=39$. Contrary to the previous examples, the transducers have different group delay time characteristics to ensure required approximation, the group delay time passband approximation error being less than 5 ns (Fig. 5b).

The uncompensated zero couples off the unit circle causes the transducer responses $F_1(\varphi)$ and $F_2(\varphi)$ to be

highly asymmetrical in the filter passband if compared to the previous examples.



(a) Z-transform roots distribution

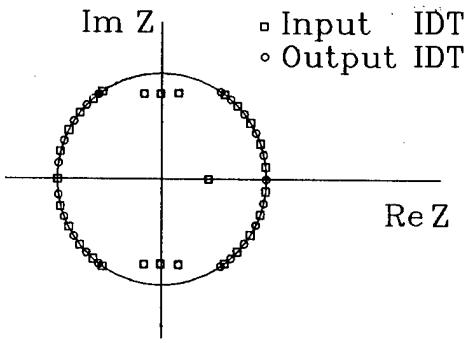


(b) frequency response

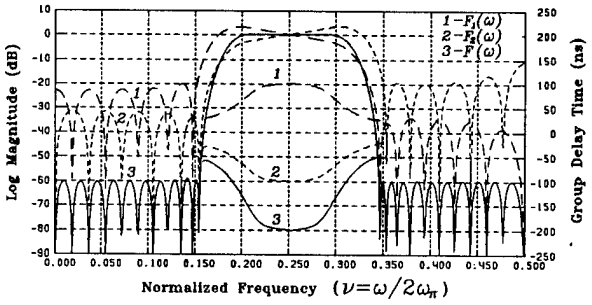
Fig. 5. Design results of a chirp SAW bandpass filter

Minimum Phase SAW Filter Design

Due to the intrinsic minimum phase feature peculiar to the factorization algorithm proposed it may easily be adopted to the minimum phase SAW filter design starting from the LP or NLP prototypes. Before factorizing one should only transform all the outside roots into their reciprocals inside the unit circle. As there will be no outside zeros now, a modified factorization algorithm should be applied where at the step 3 every second root irrespective of its location is assigned to the input IDT, the others being attributed to the output IDT. Virtually, starting from the LP prototype the same result could be obtained by omitting the last step of the factorization algorithm above.



(a) Z-transform roots distribution



(b) frequency response

Fig. 6. Design results of a SAW bandpass MP filter

A MP response and its prototype always have equal filter length and exactly the same magnitude response that supplies an excellent vehicle to compare their properties if necessary. A MP design allows to decrease the filter group delay time that could prove to be useful in some applications.

The Z-transform roots distribution and the frequency response of the MP filter designed from the LP prototype (Fig. 4a) are shown in Fig. 6.

The filter has 3 double passband zero couples and one double monozero inside the unit circle, the others being allocated on it. As in the case of LP design the filter length is $N=57$, the transducer lengths being $N_1=N_2=29$. The transducers have the opposite group delay characteristics to be summed up in the overall transfer function due to the complex-conjugation that results in minimizing the filter group delay at the central frequency.

Optimum Minimum Phase SAW Filter Design

A minimum phase design above is not optimum. There is a way to reduce further the MP filter length by applying the more sophisticated procedures [15-17] of the optimum MP design.

The major design steps are: 1) the synthesis of the squared optimum magnitude LP (OMLP) prototype $|F(\varphi)|^2$ of the order $2N$-1, all its double zeros being off and on (or near) the unit circle; 2) the synthesis of the optimum minimum phase response $F(\varphi)$ by retaining zeros inside or on the unit circle if available; 3) splitting the function $F(\varphi)$ into the separate IDT responses $F_1(\varphi)$ and $F_2(\varphi)$ using the modified factorization algorithm above.

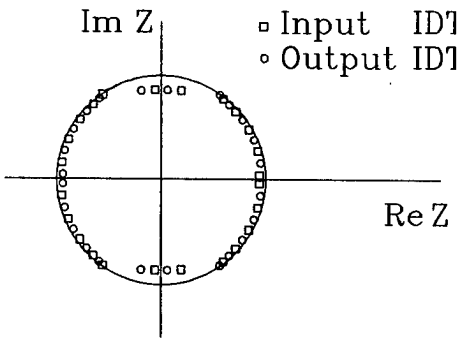
The squared OMLP prototype $|F(\varphi)|^2$ may be synthesized by the Hermann-Shuessler technique [15-17] using the Remez exchange algorithm [7] for example. The modifications of the McClellan's computer program [7] to overcome divergence problems for very high-order trigonometric polynomials are discussed elsewhere [18].

The design results of the OMP SAW filter with the same magnitude specifications as for the LP filter above are presented in Fig. 7. The modified Hermann-Schuessler procedure [16] and the McClellan's computer program [7] were used to design squared LP optimum magnitude response prototype.

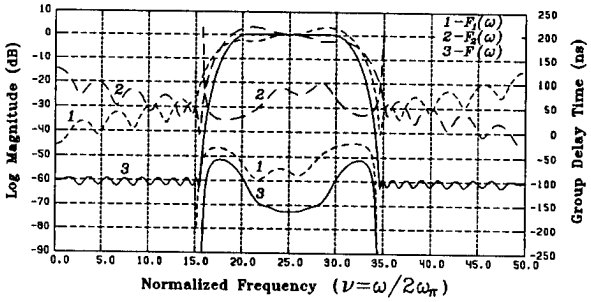
The Z-transform roots distribution after one of each double zeros has been retained is shown in Fig. 7a and the frequency response is presented in Fig. 7b.

At the same design specifications the OMP filter length N=49 is shorter if compared to the previous examples, the transducer lengths being $N_1=N_2=25$. The filter has 4 passband couples of zeros inside the unit circle, other 20 couples of stopband zeros being near the unit circle also inside it.

It is worthy to note that apart from the delay at the central frequency the passband group delay time characteristics for optimum and non-optimum design are very close.



(a) Z-transform roots distribution



(b) frequency response

Fig. 7. Design results of a SAW bandpass optimum MP filter

Conclusion

A general design procedure for LP, NLP, MP, and OMP SAW bandpass filters based on the Z-transform roots searching, lexicographical sorting, and sharing between IDT has been proposed in the paper. The factorization algorithm results in the minimum phase transducers that are less sensitive to the tap-weighting inaccuracies. The problems of the factorizational synthesis of the SAW filters with various phase requirements has been discussed.

Given the design specifications, the factorizational design leads uniquely to the minimum-length SAW filters that are usually of 20-30% shorter if compared to the conventional design where two identical weighted IDT are used.

The filter length may be further reduced in the optimum MP filters. It may be recommended to use such filters in the applications where the dispersionless requirements are not too severe. Moreover, a reasonable compromise between LP and MP designs may be attained in the quasi-minimum phase filters containing a small portion of zeros outside the unit circle.

References

[1] D.P. Morgan, Surface-Wave Devices for Signal Processing. Amsterdam: Elsevier, 1985, ch. 4, 5.

[2] V.M. Dashenkov, A.S. Rukhlenko, "Optimal Design of Surface Acoustic Wave Filters," Izvestija Vuzov, Ser. Radioelektronika, Vol. 27, no. 7, pp. 76-78, July 1984 (in Russian).

[3] A.S. Rukhlenko, "Optimal and Suboptimal Design of SAW Bandpass Filters Using the Remez Exchange Algorithm," to be published in the IEEE Trans. Ultrason., Ferroelec., and Freq. Contr.

[4] M. Morimoto, Y. Kobayashi, M. Hibino, "An Optimal SAW Filter Design Using FIR Design Technique," in Proc. 1980 IEEE Ultrason. Symp., pp. 298-301.

[5] C.C.W. Ruppel, E. Ehrmann-Falkenau, H.R. Stocker, W.R. Mader, "A Design for SAW-Filters with Multistrip Couplers," in Proc. 1984 IEEE Ultrason. Symp., pp. 13-17.

[6] A.S. Rukhlenko, "SAW Filter Factorizational Synthesis," Thezisy Dokladov XIII Vsesojuznoj Conferencii po Akustoelektronike i Kvantovoj Akustike, Chernovtsy, 1986, Pt. 2, p. 157 (in Russian).

[7] L.R. Rabiner, B. Gold, Theory and Application of Digital Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1975, ch. 3.

[8] C.C.W. Ruppel, E. Ehrmann-Falkenau, H.R. Stocker, R. Velth, "Optimum Design of SAW-Filters by Linear Programming," in Proc. 1983 IEEE Ultrason. Symp., pp. 23-26.

[9] P.M. Jordan, B. Lewis, "A Tolerance-Related Optimized Synthesis Scheme for the Design of SAW Bandpass Filters with Arbitrary Amplitude and Phase Characteristics," in Proc. 1978 IEEE Ultrason. Symp., pp. 715-719.

[10] K. Steiglitz, "Design of FIR Digital Phase Networks," IEEE Trans. Acoust., Speech, and Signal Processing, Vol. 29, no. 2, pp. 171-176, Apr. 1981.

[11] A.S. Rukhlenko, "Design of SAW Bandpass Filters with Arbitrary Magnitude and Phase Characteristics," to be published in Proc 1993 IEEE Ultrason. Symp.

[12] T. Kodama, "Optimization Techniques for SAW Filter Design," in Proc. 1979 IEEE Ultrason. Symp., pp. 522-526.

[13] T. Yoshikawa, M. Mitani, "On the Algorithm for Finding the Zeros of Very-High-Degree FIR Filters Using the Finite Bit Method," IECE Trans. Japan, Vol. J 63-A, no. 7, pp. 444-450, July 1980.

[14] M. Feldmann, J. Henaff, "Design of SAW Filter with Minimum Phase Response," in Proc. 1978 IEEE Ultrason. Symp., pp. 720-723.

[15] O. Hermann, W. Schuessler, "Design of Nonrecursive Digital Filters with Minimum Phase," Electron. Lett., Vol. 6, no. 11, pp.329-330, May 1970.

[16] H. Ueda, T. Aoyama, "A Practical Procedure for Designing FIR Filters with Optimum Magnitude and Minimum Phase," in Proc. 1979 IEEE Int. Symp. Circ. and Syst., pp. 24-25.

[17] P.P.N. Yang, M.S. Song,and M.J. Narasimha, "On the Design of Optimal Narrowband Linear and Minimum Phase FIR Filters," In Proc 1987 IEEE Symp. Acoust., Speech, and Signal Processing., p.884-886.

[18] E. Goldberg, R. Kurshan, and D. Malah, "Design of Finite Impulse Response Digital Filters with Nonlinear Phase Response," IEEE Trans. Acoust., Speech, and Signal Processing., Vol. 29, no. 5, pp. 1003-1010, Oct. 1981.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

X RAY TOPOGRAPHY STUDY OF GALLIUM PHOSPHATE CRYSTALS AND RESONATORS

B.Capelle*, A.Zarka*, J.Schwartzel**, J.Detaint**, Y.Zheng*,
A.Ibanez***, E.Philippot***

*L.M.C.P. Universités P VI et P VII, CNRS, 4 Place Jussieu 75252 Paris France
**CNET Laboratoire de Bagneux, 196 Av.H.Ravera 92225 Bagneux France
***L.P.C.M.S. Université de Montpellier, CNRS-URA D0407, case 03, 34095 Montpellier France

Abstract.

Using the synchrotron radiation delivered by the DCI storage ring at the LURE (Orsay, France), we have studied by the X-ray topography technique the crystalline perfection of $GaPO_4$ samples grown in different conditions. The growth of the gallium phosphate on a berlinite seed is analysed by X-ray section topography. It was observed different vibration modes of an approximately rhombohedral face cut and a near AT cut thickness shear resonators.

Introduction.

The gallium phosphate ($GaPO_4$) is, as berlinite ($AlPO_4$), a material which is a close piezoelectric analog of quartz and possesses similar cuts. It was shown that this material has very attractive properties for the applications to piezoelectric devices, due to its large coupling coefficients and to the existence of compensated cuts [1,2]. Moreover, for the AT cut, the first results obtained seem to indicate a value of coupling coefficient higher than those of quartz and berlinite. This induces the possibility to obtain larger band-pass of filters. On the other hand, there is no α-β transition for this material and, then, it can be used in a large range of temperature (as much 900°C), for sensor devices for instance.

We present here an X-ray topography study of crystals obtained with different growth approachs, and some vibration modes of gallium phosphate resonators. The plates of GaP04 were studied by X-ray transmission topography using the white beam delivered by the DCI storage ring at LURE. We have used conventional X-ray topography to characterize this material and stroboscopic X-ray topography [3] to analyse the vibration modes.

All the gallium phosphate crystals presented in this paper have been made by the group of Etienne Philippot at Montpellier.

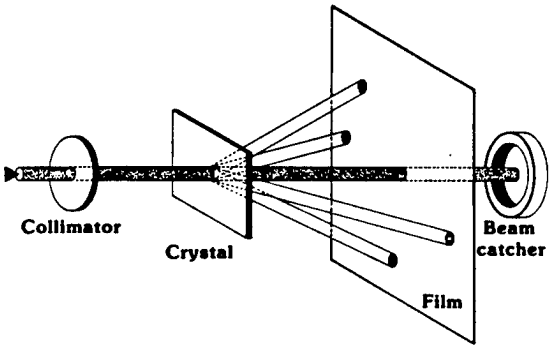


Figure 1- The transmission Laue set-up.

The experimental set-up.

The figure 1 presents the principle of the experimentl set-up which is a Laue set-up. The sample is placed in the white X-ray incident beam delivered by the synchrotron source. On the film, we obtain a Laue pattern on which the crystal is simultaneously imaged with different diffraction vectors. Two types of topographs can be obtained. The first one is a traverse topograph which is obtained when a large incident X-ray beam is used, the second one is a section topograph when the incident beam is limited by a fine slit of about fifteen microns wide. With this second type of topograph we can see the section of the crystal. Let us see a pattern made in the vertical plane and shown on the figure 2.

0-7803-0905-7/93 $3.00 © 1993 IEEE

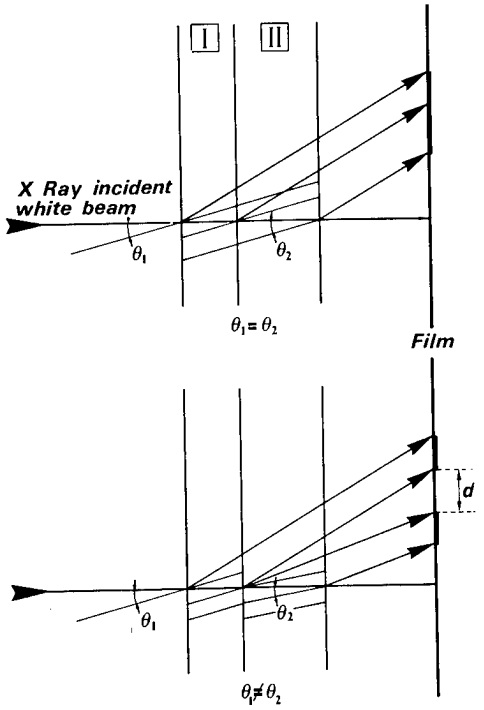


Figure 2- Section topograph of a crystal made with two parts without distortions in a and with distortions in b.

We assume that the crystal is made with two parts which are in our case an epitaxial growth in I and a substrate in II, this crystal is a slice perpendicular to the plan of this pattern .

In a this is the case where the two parts are the same structure and where there are no distortions. The diffracting planes are the same for the two parts. In this case, the angles $\theta 1$ and $\theta 2$ between the incident beam and the diffracting planes of the two parts respectively are equal, and the diffracted beams by the two parts are parallel. Then, we obtained on the film only one image for the two parts.

In b this is the case where the diffracting planes are not identical for the two parts because of different structures or parameters, distortions or a bad epitaxial growth. In this case the two angles $\theta 1$ and $\theta 2$ are different and the diffracted beams are not parallel because they make an angle equal to $2\theta_1$, or $2\theta_2$ with the incident beam. Then, we obtained, on the film, two separated images for the two parts of the crystal. If we know the distance between the crystal and the film, we can measure d on the film and calculate the disorientation $\Delta\theta$ between the two parts.

We will see an example of this effect farther.

$GaPO_4$ crystals growth.

The $GaPO_4$ crystals growth has been carried out through the three methods already used for berlinite [4,5]:

-the slow heating method (SHT);

-the horizontal composite gradient method (HTG);

-the vertical reverse temperature gradient method (VTG).

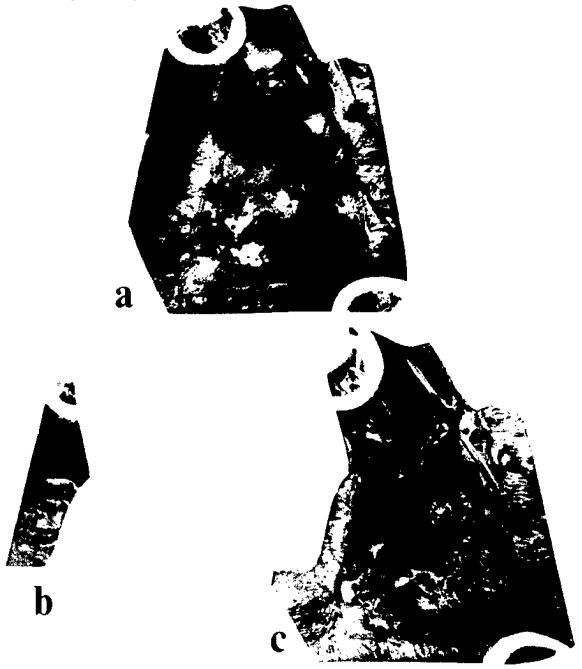


Figure 3- Topographs of a gallium phosphate crystal obtained by successive growths from a small spontaneous nucleation by SHT method in phosphoric acid.

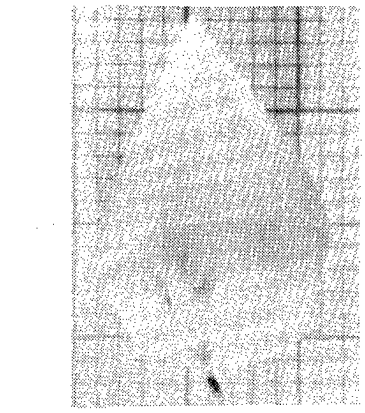
The figure 3, in a, shows a topograph of a gallium phosphate crystal obtained by successive growths from a small spontaneous nucleation by slow heating method in phosphoric acid. Many defects appear, dislocations around the crystal and several twins mixed in the central part. Due to the different values of the structure factor for the two electrical twin domains for fixed reflections it is possible to see on two different topographs each twin domain.

This is the case for the two topographs on the figures 3c and d, where one isolated electrical twin can be seen. The other twins in the central part of this crystal are optical twins.

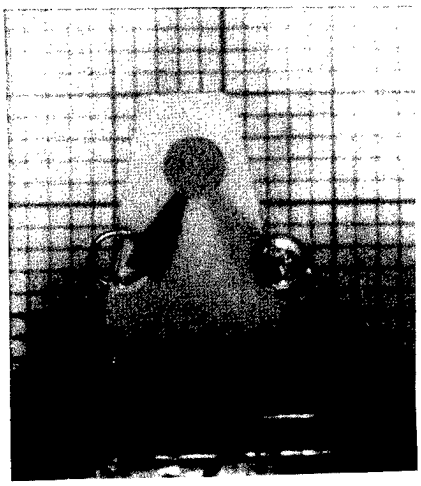
An another example of this type of crystal is shown on the figure 4. Note that, like in the precedent case, the constrast in the central part of this crystal are very complicated to analyse because the superposition of many defects.



Figure 4- Topograph of a gallium phosphate crystal obtained by successive growths from a small spontaneous nucleation by SHT method in phosphoric acid.



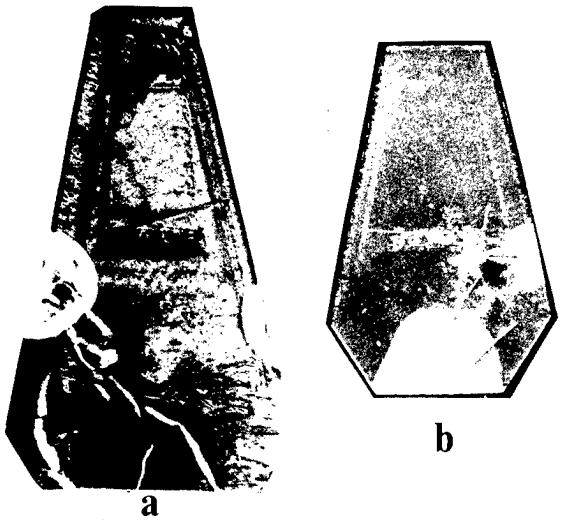
a



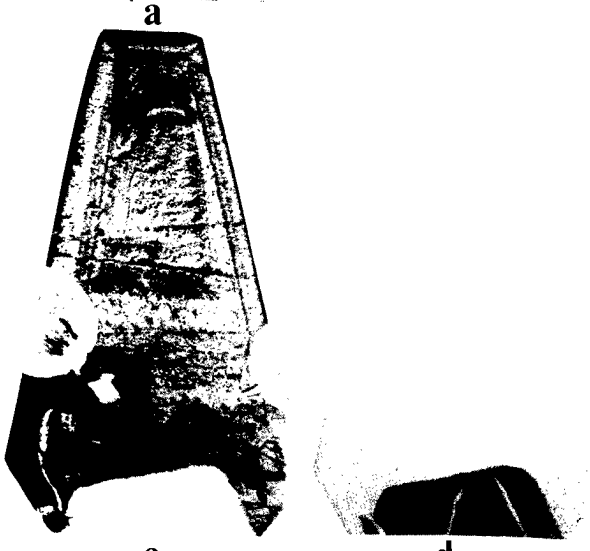
b

Figure 5- Gallium phosphate crystal obtain from a seed selected among plates cut in the crystals obtained from the spontaneous nucleation in a and in b a slice of this crystal.

The first step to improve the crystalline quality of this material is to select seeds among plates cut in the crystals obtained from the spontaneous nucleation. That way the crystal on the figure 5a has grown, on the bottom we can see the fixation of the seed. The SHT method in phosphoric acid has been used. This method where the temperature is slowly increased from about 150 to 180°C with a speed equal to 0.5 to 2°C/day, is now well known. One's major drawback is the impossibility to obtain large crystals in one operation because of the limited quantity of gallium phosphate available in the solution. In b is presented a slice parallel to a face of the rhombohedron of this crystal that we have study.



a b



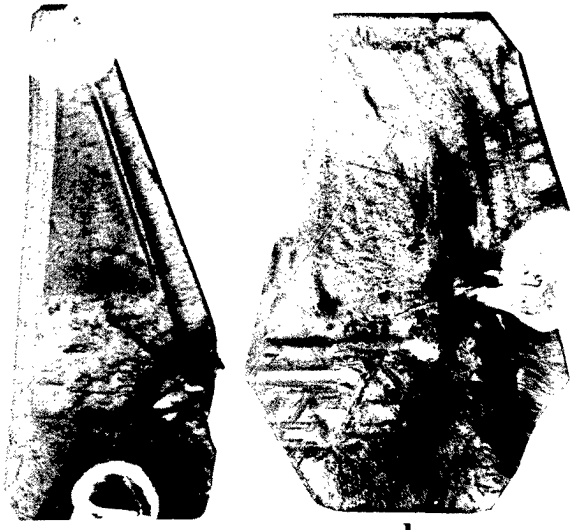
c d

Figure 6- Topographs in a, c, d and optical observation in b, of the slice of the figure 5b.

The optical observation of this slice shows (Fig.6b) an electrical twin, probably due to the fixation of the seed, and a strong black and white contrast due to the fixation of the first growth from a small spontaneous nucleation. On the topograph (Fig.6a) strong contrasts around the fixation of the first growth and, in the same area of the fixation of the seed that the twin, some cracks appear in white contrast .

In c and d two topographs obtained with two different diffraction vectors show the two electrical twin domains. The quality of this crystal is better, in particular in the top part, than one of the crystals of the first generation.

Anothers slices of the same crystal are shown on the figure 7. We see again the electical twin and a strong contrast due to the strains induced by the fixation of the first growth which exist in all the crystal . These defects make unusable the half of the crystal. But the top part of the sample in a seems better. The quality of the sample in b is worse because this slice is cut in the middle of the crystal and therefore is near the seed. The quality becomes better when the slice is cut far from the seed.



a b

Figure 7- Topographs of two others slices of the crystal of the figure 5a.

This way does not seem to be the good way to obtain crystals with a good quality and large sizes. Also another approach is necessary.

This new approach consists of the use of a berlinite seed for the growth of the gallium phosphate. Indeed, large berlinite seeds (four to eight centimeters long) are now available.

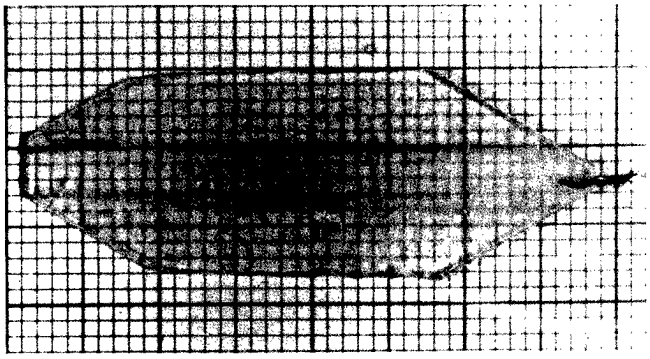
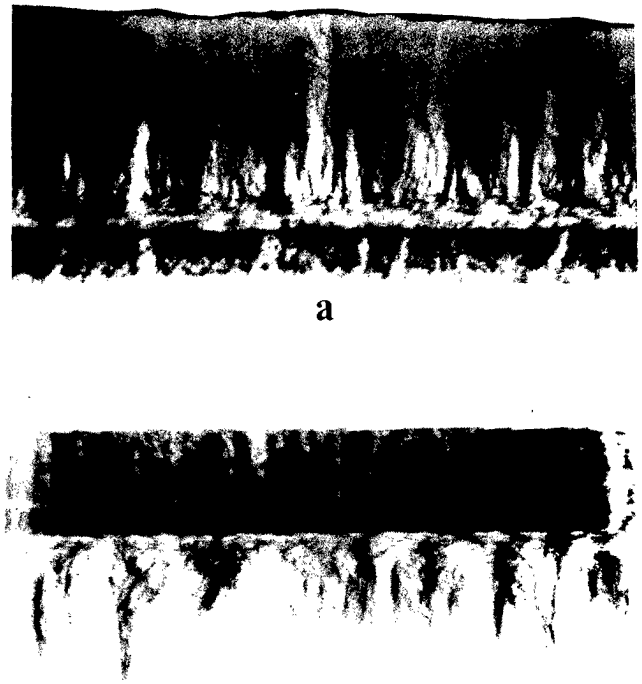


Figure 8- Flat gallium phosphate crystal obtained from Z berlinite seed in sulfuric acid and VTG method.

We become by a growth with a Z berlinite seed in sulfuric acid and VTG method. V_X being the only significant growth rate, flat crystals are obtained with a very thin epitaxy of gallium phosphate in the Z direction. An example is given in figure 8. We continue the growth from this crystal in phosphoric acid and composite gradient method to have a significant V_Z growth rate. This method, with nutrient in the colder side of the glass autoclaves, allows to obtain larger crystals than the SHT method. The fluid circulation, from cold to hot zone due to



a

b

Figure 9- Topographs of the crystal of the figure 8 showing the berlinite seed.

the horizontal temperature gradient, is done at constant temperature or coupled to a slow heating rate of 0.25 to 1.0°C/day. In both cases, quite good results are obtained for the crystal growth experiments.

Then we can cut gallium phosphate X seeds for a growth in sulfuric acid and the VTG method or Z seeds for a growth in phosphoric acid and the SHT method or the composite gradient method.

Topographs of the X zone and the top part of the berlinite seed and of the berlinite seed and a part of the $\bar{X}$ zone of the crystal shown in the figure 8 are in the figure 9 a and b respectively. We can see, at the interface between the berlinite seed and the gallium phosphate X zone, strains which decrease when the distance to the seed increases. To analyse the epitaxial growth of the gallium phosphate on the berlinite seed the X-ray section topography is a very good method.

The results are shown in the figure 10. In e the same traverse topograph that in the figure 9a and in a, b, c and d different section topographs obtained with a fine horizontal slit.

In c and d two section topographs made at the seed-level. We have here the case where the crystal is made with 3 parts, a berlinite substrate, the seed, and two epitaxial gallium phosphate growth on both sides. These two section topographs are obtained with two different diffraction vectors. Due to the strains, separated images for the different parts are obtained. In d we can see, at the bottom, the image of the berlinite seed and above the two images of the epitaxial gallium phosphate. The aspect of these images is due to the different values of the strain at each point of the interface. In c we have the same thing, but the seed image is like a ribbon. The two other section topographs show that the quality increase far from the seed. In b we have many defects and in a the quality becomes better.

If we put a fine vertical slit we obtain one another set of section topographs (Fig.11).

In d a traverse topograph with the X zone, the seed and the $\bar{X}$ zone indicates the position of the vertical slit. The section topographs, in a, b and c, are obtained simultaneously with different diffraction vectors; h' is the projection of the diffraction vector on the film. In a, the image of the seed appears very well on the left, the images of the two gallium phosphate layers are not separated because the strong strains give over large images. On the top the gallium phosphate of the X zone becomes better progressively far from the seed. At the interface between the seed and the X zone no

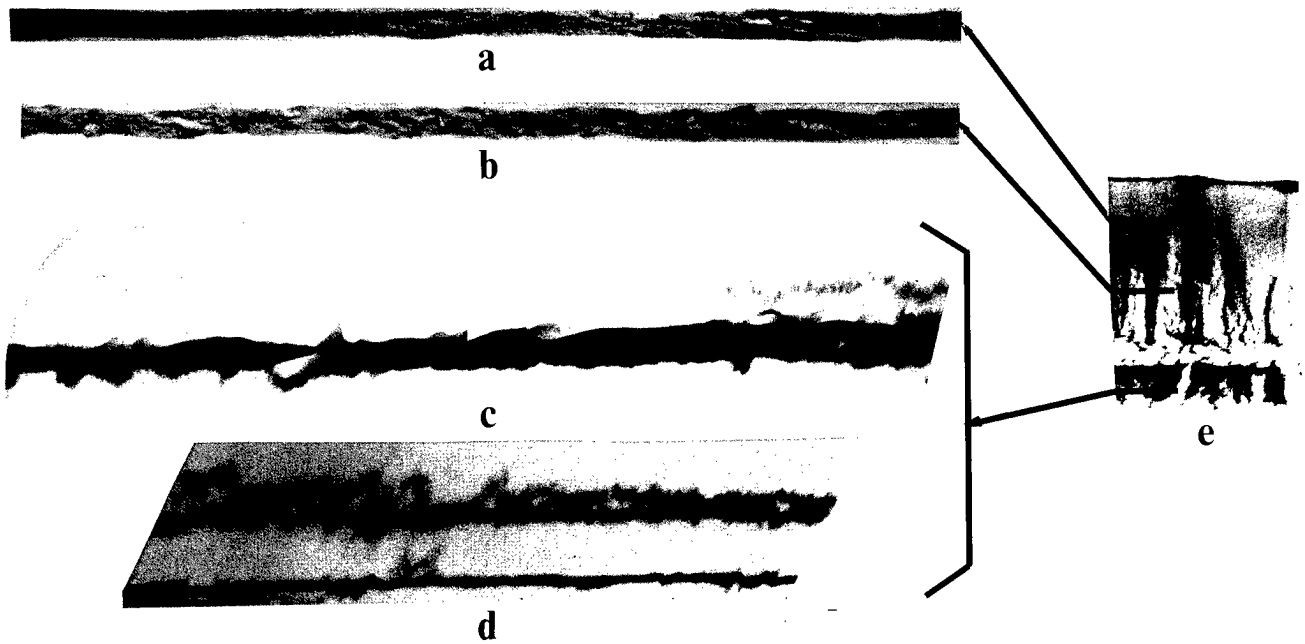


Figure 10- Section topographs of an epitaxial growth of gallium phosphate from berlinite seed obtained with an horizontal slit.

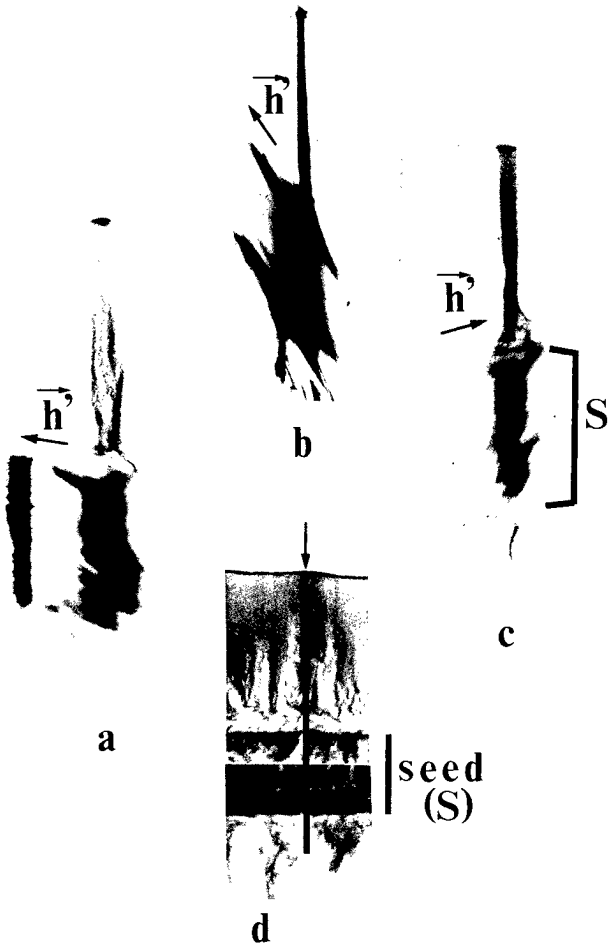


Figure 11- Section topographs of an epitaxial growth of gallium phosphate from berlinite seed obtained with an vertical slit.

image appears showing that there is an amorphous zone, alone the epitaxial gallium phosphate seems to be unbroken. This amorphous zone is also visible on the traverse topograph. In b and c two anothers section topographs. In c, the image of the seed, on the left, is very pale and the unbroken gallium phosphate between the seed and the X zone can be seen again.

This result is very interesting and shows how to obtain large good seeds of gallium phosphate.

First examples of growth from this type of seed are shown in the figure 12. The better crystal of these two crystals, in b, has been obtained in phosphoric acid and SHT method with successive growths. The conditions are not yet optimized but the quality is good. We have no twins and we can see individual dislocations. We have used the same method to obtain the crystal in a, but it presents more dislocations.

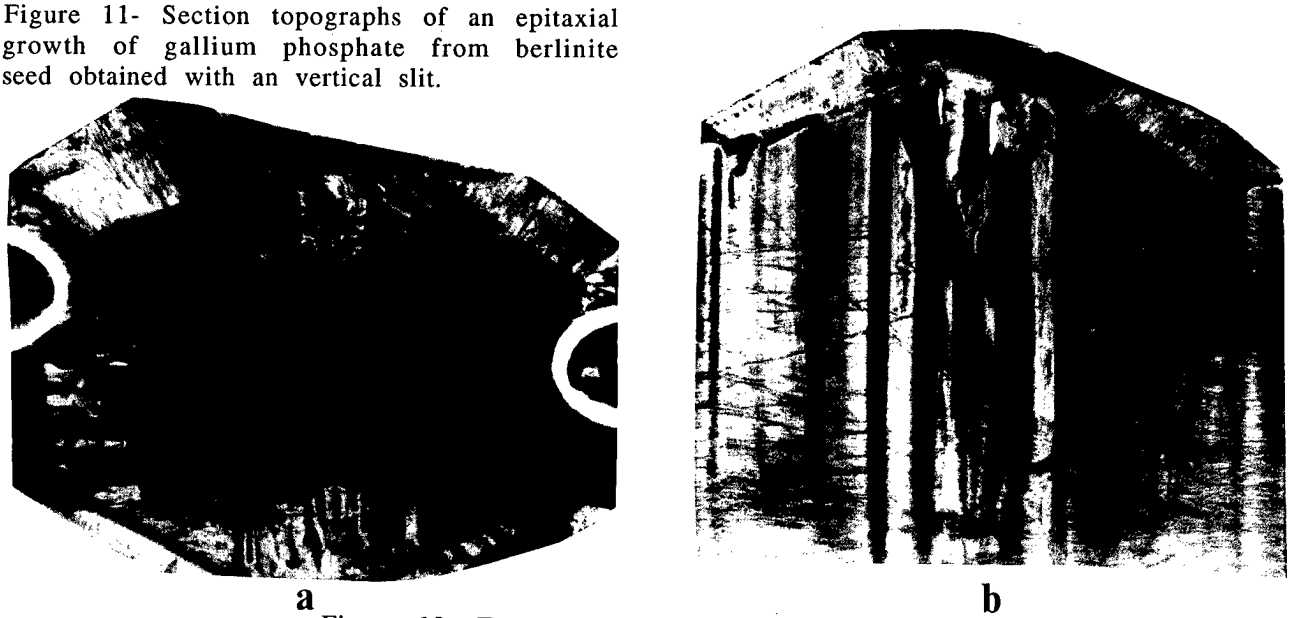


Figure 12- Topographs of two crystals growth from gallium phosphate seeds obtained from a growth on a berlinite seed.

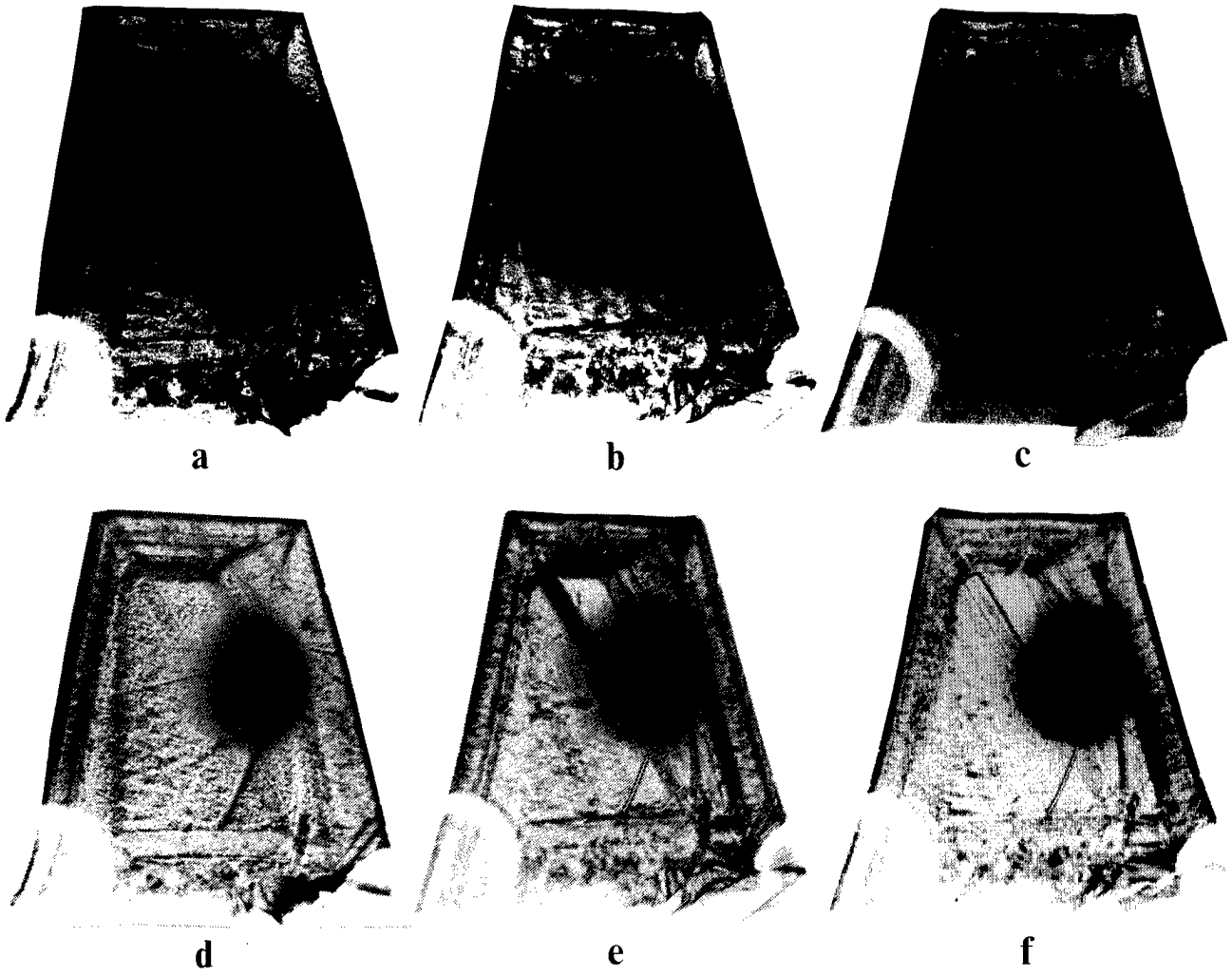


Figure 13- In a, b and c three topographs of the fundamental mode (about 6 MHz) of a gallium phosphate resonator recorded simultaneously in the same Laue pattern with three different diffraction vectors. In d, e, f three topographs of the third overtone.

Resonators

We have made some observations of vibration modes of an approximately rhombohedral face cut and a near AT thickness shear resonators.

The crystal shown in the figure 13, a slice approximately parallel to a face of the rhombohedron, has been obtained from a seed cut in a crystal obtained from a spontaneous nucleation by the SHT method in phosphoric acid. The three topographs of the fundamental mode (about 6 MHz) in a, b and c have been recorded simultaneously in the same Laue pattern with three different diffraction vectors. We can see in b the apparition of a flexion mode coupled with the principal one. This is the third overtone which is on the three topographs d, e, f recorded also in the same Laue pattern.

The modes interact with the defects, mainly growth sector boundaries, which shift the modes with regard to the electrodes. The electrodes are visible in c.

We have also observed a crystal growth from a gallium phosphate seed obtained from a berlinite seed. The figure 14 shows the response curve of a resonator made with this crystal. It is a Y - 15° cut, the frequency of the fondamental mode is about 6 MHz and the Q factor is equal to 11.000. The principal mode is weakly attenuated and some anharmonic modes including one asymmetrical can be seen.

The topographs of these modes are shown in the figure 15. In a the fondamental mode not very confined but more than in the quartz, appears with coupled modes on the top of the topograph. In b we can observe the asymmetrical anharmonic mode and in c the fifth overtone mode. All these modes interact with the dislocations like you can see it in particular in the middle of the topograph in c, moreover the boundaries of the vibrating parts are diffuse.



Figure 15- Topographs of modes corresponding to the figure 14. In a the fondamental mode appears with coupled modes on the top of the topograph. In b the asymmetrical anharmonic mode and in c the fifth overtone mode.

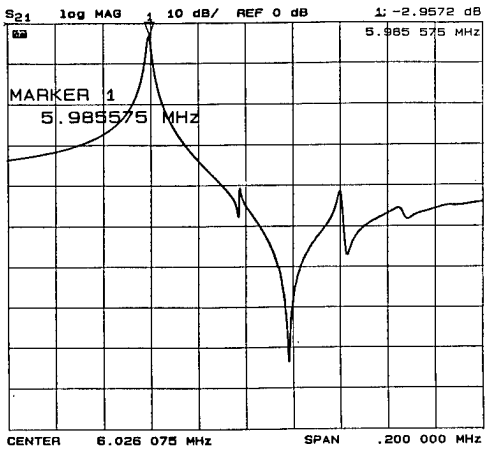


Figure 14- The response curve of a Y - 15° cut gallium phosphate resonator. The frequency of the fondamental mode is about 6 MHz and the Q factor is equal to 11.000.

CONCLUSION

In this investigation, it has been demonstrated the feasability of gallium phosphate epitaxy on large berlinite seeds in sulphuric acid medium. From these new seeds, the crystal growth of large gallium phosphate crystal in phosphoric acid medium can be undertaken.

From these first characterizations, the good quality of the gallium phosphate epitaxy on berlinite seeds appears. This method is a good solution to obtain gallium phosphate crystals with a good quality and large sizes.The first results of the resonators are very promising.

[1] P.Krempl, J.Stadler, W.Wallnöfer, W.Ellmeyer and R.Sclic; Proc. 5th European Frequency and Time Forum, Besançon (1991) p.143.

[2] J.Detaint, J.Schwartzel, C.Joly, B.Capelle, A.Zarka, Y.Zheng, Y.Toudic, E.Philippot; Proc. 6th European Frequency and Time Forum, Noordwijk (1992) p.383-388.

[3] A.Zarka, B.Capelle, J.Detaint , J.Schwartzel; Proc. 41st Annual Frequency Control Symposium, Philadelphia (1987)

[4] J.C.Jumas, A.Goiffon, B.Capelle, A.Zarka, J.C.Doukhan, J.Schwartzel, J.Detaint, E.Philippot; J. of Cryst. Growth,80 (1987) 133.

[5] E.Philippot, A.Goiffon, M.Maurin, J.Détaint, J.Schwartzel, Y.Toudic, B.Capelle, A.Zarka; J. of Cryst. Growth, 80 (1990) 713-726.

1993 IEEE INTERNATIONAL FREQUENCY CONTROL SYMPOSIUM

APPLICATION OF LOW COST FREQUENCY STANDARDS FOR COMMERCIAL AND MILITARY GPS

L. D. Vittorini, Rockwell International,
Collins Avionics & Communications Division, Cedar Rapids, Iowa 52498 USA

Brent Robinson, Rakon Ltd, Auckland, New Zealand

INTRODUCTION

Explosive growth in Global Positioning System (GPS) usage is imminent as the 24SV constellation approaches completion this year. The growth potential in the commercial sector will be fueled, in large part, by the significant size, power and cost reductions made possible through recent ASIC and VLSI technology advancements. Low-cost, single card implementations for SPS[1] use, such as the NAVCORE V manufactured by Rockwell International, are now readily available for OEM type installations.

An equal, if not more significant, contributor to the success of these miniature implementations has been the advances in TCXO[2] technology. Manufacturing improvements achieved by some vendors for AT-cut crystals has advanced such that their g-sensitivity and AVAR[3], two GPS critical performance parameters, can approach that of SC[4] crystals. This has spawned significant improvement in the GPS performance level achievable in a less expensive, smaller form-factor.

The significance of this is evident when viewed from the limited demonstrations of GPS as a force multiplier during the 1991 Gulf War. The near-term effect has been to raise the "GPS consciousness" level of the military in general. The result is an avalanche of tactical system designers seeking to extend capability via embedded GPS. Some of these applications are of the disposable variety including guided munitions such as the Tomahawk cruise missile and SLAM[5], as well UAV's[6].

This leads to the conclusion that it is an opportune time to revisit how the GPS frequency reference affects the various functions necessary to do GPS. This should to be followed by a trade study to determine where compromises may be made allowing use of a lower cost standard. The result will foster a smaller, lower-cost implementation for use in the aforementioned applications.

The pertinent parameters of a frequency reference for GPS will be discussed and include:

- Temp stability
- Temp gradient
- Allan variance - AVAR
- SSB Phs Noise
- Aging
- Warmup Time
- Hysteresis
- Frequency gradient
- Acceleration sensitivity
- Frequency Perturbations
- Voltage sensitivity

The relevance of these parameters will be discussed as they relate to the pertinent GPS system parameters of:

- Time-To-First-Fix performance
- Carrier tracking loop performance
- Code tracking loop performance

[1]SPS - Standard Positioning Service
[2] TCXO - Temperature Compensated Crystal Oscillator
[3]AVAR - Allan variance
[4]SC - Stress Compensated
[5]SLAM - Standoff Land Attack Missile
[6]UAV - Unmanned Aerial Vehicle

0-7803-0905-7/93 $3.00 © 1993 IEEE

TIME-TO-FIRST-FIX PERFORMANCE

The TTFF/TTSF (Time-to-[First/Subsequent]-Fix) refers to the time required by the receiver to acquire the C/A[7] code on four SV[8] and either begin tracking the C/A code or do a handover to the P/Y[9] code and track for PPS[10] equipment. The initial conditions the receiver starts from when attempting this function determines how the available hardware can best be utilized to minimize TTFF. Generally speaking, it is desirable to minimize the initial uncertainties which translate directly into frequency error, minimizing the TTFF. High performance systems generally require minimization of the time necessary to acquire and track. This is especially true for some stand-off weapon use where the mission length is short. From a hardware standpoint, more correlators allow a quicker search, but DSP methods such as FFT's may be employed to perform quicker frequency cell (bin) type searches.

The TTFF/TTTSF performance of a GPS receiver is governed by five components of frequency standard performance:

- *Temperature Stability*
- *Warmup*
- *Aging*
- *Hysteresis*
- *Temp Gradient*

Portable units are generally more sensitive to temperature stability and hysteresis for TTFF/TTSF since extending battery life may require the set to be operated in a standby mode and reacquire (TTSF) on a periodic basis, like the PLGR[11]. The issue here is the number of frequency windows that must be searched, and how fast, to find the correct code position. The worst case initial conditions will heavily influence this requirement. Figure 1 depicts the cell approach typically used in Direct Sequence Spread Spectrum acquisition schemes. The particulars vary somewhat between manufacturers.

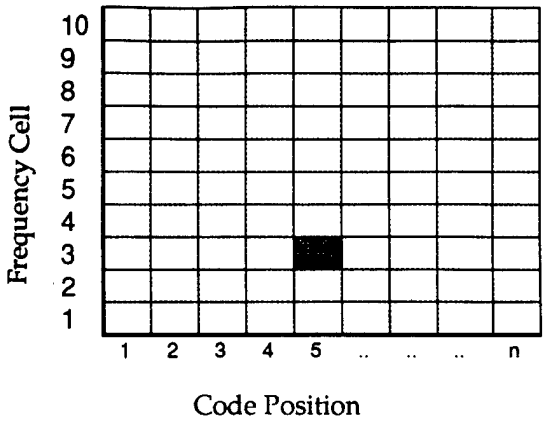


Figure 1. Acquisition Search Cells

Temperature Stability The overall temperature stability becomes a frequency uncertainty that must be searched to find the required frequency/code phase position to track the GPS signal. The operating temperature range of the receiver and the subsequent frequency standard induced uncertainty is a prime consideration in the acquisition process. A $\pm$ 2 ppm TCXO induces a total of some 6 KHz of uncertainty alone. This could translate into a greatly increased TTFF in the worst case. The temperature performance of most TCXO's produced today is 2 ppm over a -30° to +85°C range. The yield for a given spec, and hence cost, is proportional to the performance. Yield, in the present context, refers to the probability that a given percentage of manufactured units will have a frequency deviation, over the operating temperature range, less than that required by the specification. It is possible to make significant improvements in temperature performance for the standard two-port compensation technique by paying attention to minute details in the manufacturing process. See figures 14 and 15 for examples of production yield realized for a production batch intended for use in the NAVCORE V. Figure 14 illustrates yield for static temperature and Figure 15 for a dynamic slew at 2°C/min. Note the similarity between the two. The correlation between static and dynamic performance observed in Rakon TCXO's is NOT representative of "typical" performance from most manufacturers. These results are indicative of some significant manufacturing process improvements Rakon has achieved, which results in a high level of performance uniformity.

[7]C/A - Coarse/Acquisition
[8]SV - Satellite Vehicle
[9]P/Y - Precise/Encrypted
[10]PPS - Precise Positioning Service
[11]PLGR - Precise Lightweight GPS Receiver

Warmup Time: The time necessary for the stabilization of frequency before the acquisition process can begin is a critical specification which continues to gain in importance. The ovenized units typically employed in conventional receivers typically spec their warmup times on the order of several minutes. This is the time required for the oven to stabilize the frequency within a spec, typically 10 ppb (.01 ppm). However, it is not necessary for the frequency to stabilize completely before acquisition can begin. Empirical testing of ovens used in military applications shows that even at -55 °C, a good ovenized oscillator will be within several tens of ppb within a few seconds after turn-on. Translating this into a workable figure, 30 ppb is equivalent to less than 48 Hz @ L1 with a frequency window typically 10-15 times that. *A priori* knowledge of this fact allows the system designer to compensate for the frequency/code positions that might be missed as the standard warms up. Low-cost TCXO's will reach similar stability within a few seconds. Figure 2 shows a typical warm-up characteristic for a Rakon TCXO. Other TCXO's have been tested with similar results.

Aging: For tactical weapons, the storage time will normally comprise the largest error component in the form of frequency standard aging. For example, a 0.25 ppm/yr aging component over 5 years will introduce an uncertainty of 1.25 ppm, or nearly 2KHz at L1 (1575MHz). For a 1ms PDI[12], this is at least three frequency cells or "windows", depending on the overlap selected. Since the samples are generally formed from a digitally sampled signal, the equivalent process is that of an IDF[13] function, so the frequency domain response is a $\text{sinc}^2(x)$ function. For an overlap of less than 3 dB, one frequency window will be no greater than 700 Hz.

A typical aging specification for a TCXO is 1 ppm/year. This translates into approximately two frequency windows/yr for 1 msec search PDI (F_{pdi}= 1000Hz), a window being at least 700 Hz wide per the above discussion. High quality ovens are typically on the order of 10X better, mostly due to crystal pre-aging and more rigid mounts. However, there are several factors in favor of a non-ovenized oscillator. First of all, the aging characteristic is accelerated by heating. However, for the typical storage scenario, this has negligible importance. The SC crystals used in ovens, while having a much higher level of processing, are also typically aged a great deal more than a typical AT crystal. Similar performance can be obtained from well manufactured AT crystals by utilizing a longer pre-aging period. Stress relaxation in the mount and crystal lattice imperfections and relaxation all contribute. Figure 3 exhibits how pre-aging can dramatically improve the long term frequency drift. The characteristic shown was plotted for crystals aged @ 125°C before installation into an oscillator and then aged @ 85°C for a period of 50 days. Measurements were taken every few hours. The data was extrapolated to 1000 days. Consider also that most units will be at considerably lower storage temperatures and thus age at proportionately slower rates. Future study will determine the impact of storage temperature on aging rate. The other point of note here is that since properly aged & manufactured units will have monotonically increasing frequency with time, it should be possible to establish a model, given *a priori* knowledge of the pre-aging the crystals were subjected to(REF 1).

Hysteresis: Hysteresis, or retrace, is the phenomena whereby the frequency does not return to the same point the standard was the last time it was at that same temperature. It is of concern mostly for portable applications which use a standby mode and affects both the frequency uncertainty and the code position error growth since the set last tracked. This effect is magnified by running the oscillator under a temperature gradient. Generally speaking, for AT crystals, the larger the gradient, the larger the hysteresis, up to some limit. The limit depends almost entirely on the quality of the crystal processing and pre-aging. There are several reasons gradient effects magnify hysteresis, but often the main contributor is thermal lag between the thermistor compensation network and the crystal itself. This can be better or worse, depending on the thermal mass of the oscillator as a whole. By reducing component sizes, the resultant thermal coupling will be better matched, resulting in better overall performance under temp gradients. However, when evaluating TCXO vendors for GPS applications, one should be aware that test data taken under static conditions can be considerably better than that under even a mild temperature gradient. *Proper evaluation for a GPS application should include something other than static data supplied by the manufacturer.* Figure 4 shows good hysteresis

[12]PDI - Predetection Interval
[13]IDF - Integrate and Dump Filter

performance of under 0.25 ppm, typical of a Rakon TCXO.

Temp Gradient: For TTFF/TTSF, gradients manifest themselves in the form of the frequency drifting out of the search window before you've searched all the code. This is nontrivial to spec, but it is possible to compute the maximum allowable gradient, given a complete definition of the acquisition process. Figure 5 illustrates a spreadsheet application designed for just such a purpose and the corresponding output plot for a simple two correlator search process. This spreadsheet allows bounding of the requirement in terms of a tangible specification for the TCXO. TTFF gradient effects are significant mostly in conditions where the ambient temperature is changing rapidly while acquisition is taking place. Worst case initial conditions and the acquisition scheme for TTFF/TTSF will drive this requirement

Perhaps the most important point relevant to the acquisition discussion is that the faster your search capability, the less importance the aforementioned effects assume. To be sure, future generations of GPS equipment will have considerably more tolerance of frequency standard related uncertainties, rendering these performance issues with respect to TTFF somewhat irrelevant.

CARRIER TRACKING LOOP PERFORMANCE

Temperature: GPS tracking loop performance is largely unaffected by slowly varying frequency vs. temperature since the loop bandwidth is typically set to 20 Hz or less.

Frequency Perturbations: Frequency perturbations, a phenomenon seen more commonly in low-cost TCXO's, can cause the tracking loop to lose lock. The effect perturbations have on carrier tracking will depend on the loop type and order. Two prominent loop types are discussed. Loop order also plays a significant role here since the typical 3rd order loop can estimate frequency, Doppler and acceleration errors, with a stand-off to Jerk (g's/sec).

These phenomena typically manifest themselves over a specific temperature range. Figure 6 illustrates a typical example of this phenomena. It only becomes evident when a dynamic temperature slew is utilized to screen for it. Static testing will likely fail to detect it in most instances. One way to combat perturbations, however, is to decrease the PDI, which effectively widens the detection bandwidth (T=1/f). This allows proportionately larger disturbances at the expense of loop noise which translates into Delta Pseudorange error.

FLL/AFC (Frequency Lock Loop): For a receiver using a FLL with a 20msec PDI and differential detection to demodulate the 50Hz GPS data stream, the discrimination range is $\pm$ 12.5 Hz. Any 1/4 cycle phase change is detected as a bit transition so only a fraction of this may be budgeted for perturbation. A frequency step of greater than 12.5 Hz @ L1 (1575 MHz), or .016 ppm, occurring within a 20msec period will cause a bit error. The pure dynamic range is at least $\pm$1/2 cycle, so the loop will likely not lose lock even if a bit error has occurred However, the loop may jump onto an adjacent stable lock point causing a frequency error reflected in the delta range as a velocity error. For example, at L1, a 25 Hz sidelobe track results in an error:

$$\Delta R = f\lambda = 4.76 m/s \qquad (1)$$

PLL (Phase Lock Loop): Many GPS manufacturers use a Costas loop for carrier tracking. For the same tracking requirements, PLL's are not as robust as a FLL for frequency disturbances. With impressed data on the signal, as in the GPS case, even less so (1/8 cycle). It is probably necessary to limit jumps to < .01 ppm per PDI interval.

The physics behind frequency perturbation is not completely understood, but surface defects and contour are thought to be the principal contributors. The only real way to protect against perturbations is for the manufacturer to screen for them. Factory tests can be performed to ensure a low probability of occurrence. Higher cost SC crystals typically have these screened out due to a higher level of crystal purity and processing. Low cost AT crystals will likely have a significant probability of occurrence, unless the manufacturer is cognizant of the problem and screens them out. Most TCXO's built for cellular and radio applications are insensitive to these phenomena since they are not functioning as sophisticated clocks, unlike GPS.

Voltage Sensitivity: Power supply ripple sensitivity manifests itself in the form of FM sidebands which can cause implementation loss or worse, loss of the carrier altogether. This is an AM-to-PM type of loss

and can be seen from a degradation in the single sideband phase noise performance of the oscillator. Computing the equivalent system loss only requires knowledge of the multiplication factor N of the synthesizer loop. Using a Bessel function computation of the sideband ratio power of $J_1(x)/J_0(x)$, the equivalent FM-to-PM phase error can be computed:

$$\theta_{pk} = \frac{\partial F}{\partial v}\frac{\Delta V_{pk}}{f_m} f_o \qquad (2)$$

From this, the sideband power level can be specified from the FM relationship:

$$SSB(dBc) = 10\log\left|\frac{\theta_{pk}^2}{4}\right| \qquad (3)$$

Acceleration Sensitivity: Acceleration sensitivity has two dominant effects for consideration in GPS:

1. **Vibration**: Similar to voltage sensitivity, acceleration sensitivity has a phase modulation effect transmitted from vibration. Therefore it is crucial to understand the environment the receiver will encounter as well as to understand the resonances and transmissibility of the PCB on which it is mounted.

2. **Tracking Loop Acceleration Error**: Specific environments, such as those which cause large jerk effects, must be carefully assessed as to their effect on tracking loop performance. For high dynamic, unaided host-vehicle (HV) applications such as fighters and the like, sustained acceleration causes frequency standard velocity errors in the tracking loop that must be tracked without loss-of-lock (LOL). In these cases FLL's, prove superior to the standard Costas loop due to their ability to maintain lock over a wider range and reacquire quickly once lock has been lost. The particulars of this are beyond the intended scope of this discussion, but the point is: *evaluate these effects before setting a requirement.*

In general, ground mobile equipment should conform to MIL-810E as a minimum. It is possible to compute the tracking loop effects of a particular vib profile by constructing a function to encompass the prevalent factors. These have been detailed by Pratt[REF 2] in a recent ION article. Figure 7, illustrates an example of a spreadsheet application developed from the aforementioned article to output an evaluation of the resultant phase error for a given input vibration profile and corresponding acceleration sensitivity Γ_a in ppb/g. Antenna effects are also included as well as effects to consider either a rigid or isolated mount. This paper is recommended as a reference to gain insight into vibration effects.

The crystal mount itself is an important player in determining suitability for non-benign environments. *In general, cantilever and spring mounts are not acceptable for vigorous environmental applications.* Figures 8 and 9 illustrate random vibration profiles for a TCXO with both a conventional spring mount and a ruggedized mount under a 3g rms random vibration from 30-1500 Hz. The spring mount unit had its noise floor degrade over 30 dB in the case shown. At least 20 dB may be expected. The low end, below 20 Hz, seems to exhibit acoustic effects, probably from the crystal banging against its casing. While these affects will be somewhat filtered by the tracking loop bandwidth, it is clear that significant sideband power over the carrier loop bandwidth can cause havoc in the form of significant implementation loss or perhaps loss of the carrier altogether in extreme conditions. The plot for the ruggedized mount shows minimal degradation, especially for lower frequencies where the SSB phase noise power is most significant.

Frequency Gradients: Another issue to consider is the maximum frequency gradient (vs. temp). This rate must be contained and budgeted to ensure the dynamic range of the loop is not dominated by frequency rate. Limiting the allowable drift within a 20 msec period to a few cycles can be extrapolated into a ppm per minute average which can be specified to a manufacturer. The deciding factor will depend somewhat on the expected gradient the set will undergo. A typically accepted temperature dynamic limit is 2°C/min.

Delta Pseudorange: This parameter is measured by the carrier tracking loop. It is computed as the change in phase of the received signal over a specific time interval. The range rate computation accuracy is a function of several variables, one of which is AVAR, or, alternately, SSB phase noise. The interval over which this measurement is made and the budgeting of the allowable error will both play a part in the tracking uncertainty of this variable.

Figure 10 illustrates a simulation of the sensitivity of phase error versus phase noise density for a family of C/N_0[14] values. A typical specification for delta range accuracy is 0.1 cycles in military grade equipment.

CODE TRACKING LOOP PERFORMANCE

The code tracking loop functions in GPS to track either the C/A or P code during carrier track and functions as the sole tracking mechanism in STATE 3 when the carrier tracking loop loses lock under high jamming conditions. This occurs at something less than 30 dB-Hz for the unaided FLL loop used in Rockwell-Collins GPS receivers, and is a function of the induced Jerk error in the 3rd order tracking loop which will eventually drive it out of lock. The main component of interest in an aided code tracking loop, with respect to oscillator specs, is the short term stability or AVAR. High jamming levels with lower C/N_0 and longer integration periods (Extended Range STATE 3) ranging in the 1-60sec range requires better AVAR performance. Typical .01-.03 ppb (1s avg) type performance achieved with SC cuts with an oven is probably not achievable anytime soon, at least for a high yield TCXO. Although all the mechanisms affecting AVAR are not entirely understood, until recently, conventional wisdom (and testing) suggested that AT cuts, with their lower Q compared to SC cuts, would be 10-100X worse than a typical SC cut, for the same overtone. Recent evaluation of TCXO's have shown that improved mounting & processing techniques can potentially yield low-cost TCXO's with AVAR (τ=1s) consistently in the 0.1 ppb range, or less; adequate for many MIL-GPS requirements. Refer to Figure 11 for a typical example from recent testing. More recent analysis and testing have indicated that normal code tracking loop performance is, perhaps, not as sensitive as first thought to AVAR. Empirical testing seems to support this. Figure 12 illustrates some analysis results which show sensitivity of tracking loop noise in meters as functions of C/N_0, AVAR and effective loop bandwidth for an SPS (C/A) case. For a 16 dB-Hz tracking limit and a 0.1 Hz effective loop bandwidth, a 0.2 ppb spec for a 1sec average yields a 1σ position uncertainty of 20 meters. This limit is readily achievable for today's TCXO's. This figure becomes important for GPS/INS applications and operation in an obscured or jammed environment.

Acceleration Sensitivity: At one time there was also a 10-100X disparity between the acceleration sensitivities measured for AT crystals versus doubly-rotated SC cuts. It seems clear that here again, the advantage seen by SC-cuts has shrunk considerably due to improved manufacturing and crystal processing techniques. While the mechanisms that affect acceleration sensitivity are also not completely understood, blank size, electrode type and mounting technique have fairly well understood roles in determining performance. Using improved techniques in these areas, as well as an overall improvement in the sorting & processing steps, acceleration sensitivities as good as those seen typically for SC-cuts oscillators have been observed in Rakon's low-cost TCXO's. Data shown in Figure 13 illustrating this for a recently evaluated Rakon TCXO. The worst case axis is at 1 ppb/g. This curve is fairly typical of performance achieved in Rakon's manufacturing process and represents a significant improvement in "high yield" performance.

CONSIDERATIONS FOR MIL GRADE RCVRS

GPS receivers used in a military environment generally have much more stringent performance requirements than those designed for civilian use (besides operating with the P/Y code). The most obvious and significant of these is operating temperature range. Typically, this is -54/+95°C for a standard Mil product versus a -20/+70° range, such as that required by the handheld GPS PLGR. The two port direct thermistor compensation technique used in many standard TCXO's on the market today, capable of delivering high yield $\pm$1 ppm performance for a 100°C range, would probably suffer a minimum 2 ppm degradation for a suitably rotated angle for a MIL-Temp range (150°C) . This is mostly due to the larger nominal excursion the wider temp range crystals will exhibit. Screening could reduce this, but at the expense of yield (cost).

System TTFF requirements alone have driven some of the more critical specs which were listed in the TTFF section. Typical receivers slated for the military environment have utilized high precision ovens. However, continued improvements in signal

[14] C/N_0 - Carrier-to-Noise Density ratio in dB-Hz

processing hardware and crystal manufacturing and processing will soon make it possible to use much smaller and less expensive oscillators for even the more high precision applications. Other technologies, currently in development, offer promise to eliminate ovens altogether for most radio applications requiring .01 ppm level performance. There are several areas where the necessary improvements are imminent. This should allow GPS manufacturers to take full advantage in the form of much cheaper and smaller military implementations, greatly expanding the applications to which GPS may be economically applied.

OTHER CONSIDERATIONS

RECEIVER AUTONOMOUS INTEGRITY MONITORING (RAIM):

One of the functional modes this operational state requires is for the receiver to "ride" the clock during periods of <4SV NAV, where the receivers builds up position error in the solution as a function of clock drift. A typical 1 ppm TCXO with a worst case gradient slope of .1ppm/°C introduces a velocity error of 300 m/s for a 10° temperature swing over a 10 minute period (1°C/min gradient). This integrates into a 180Km position error, making TCXO's generally impractical for use in these instances. However, it is likely a more benign environment is available and a corresponding 10X improvement in drift to 18 Km, still relatively large. One RAIM spec under evaluation for military use requires 100m clock phase error growth over a 3 minute period. This implies an average error growth rate of 0.7 ppb/min, a figure unlikely to be achieved with a TCXO, even under the most benign conditions.

NAVIGATION

As discussed earlier, normal tracking is not particularly sensitive to temperature drift but the navigation solution, typically estimated and smoothed using a multi-state Kalman filter would generally only have two states for the clock: phase (position) and frequency (velocity), the latter a random-walk model. This implies the acceleration is a white noise process. However, with an inherent clock stability ~100X worse than an oven, the temperature related frequency drift of a TCXO, seen as a clock acceleration component, becomes significant. This drift is particularly troublesome at turn-on and isn't described by a white noise process, thus precluding proper estimation without an additional state to estimate it. However, it is possible to gain back some of the performance by retuning the white noise clock frequency scale factor. The change in velocity estimate over time is principally caused by rate change in the settling of the frequency. Some of this can be isolated by package design, depending on the integration scheme. However, while at one time, signal processing throughput was very dear, this is not the issue it once was. Adding another state to the Kalman filter would seem to be the preferred solution to this problem.

CONCLUSION

In conclusion, the evolution in GPS hardware capability, cost, size and power has spurred renewed interest in the study of the sensitivity of frequency standard parameters and their effect on performance criteria. Improvements in crystal manufacturing and hardware signal processing power will continue to drive the size and cost of the GPS frequency standard down.

BIBLIOGRAPHY

1. Filler, Raymond L., *Aging Specification, Measurement and Analysis*, IEEE Freq Control Symposium Proceedings, June 1991

2. Pratt, Anthony R., *g-Effects on Oscillator Performance in GPS Receivers*, Nav Journal of the ION, Vol. 36, No. 1, Spring 1989

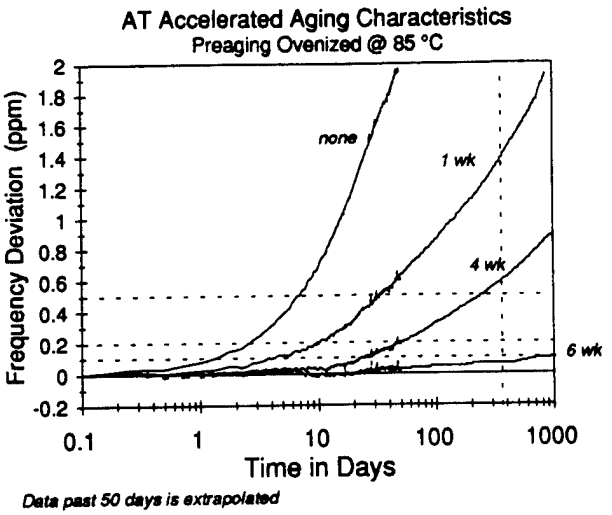


Figure 3

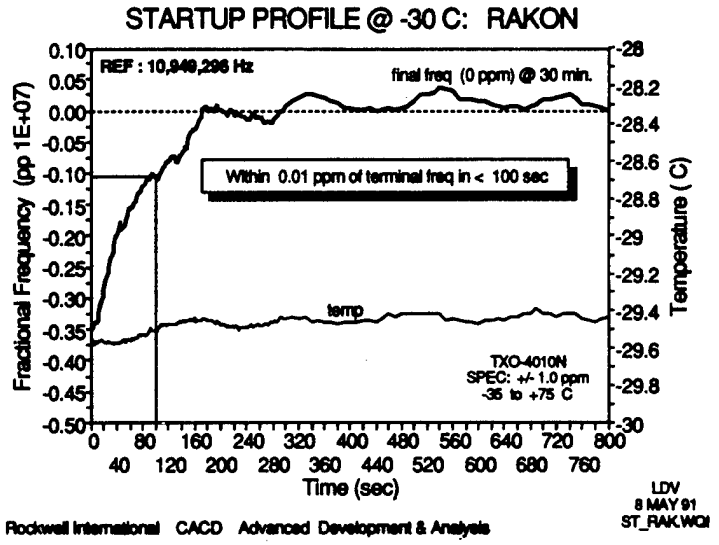


Figure 2

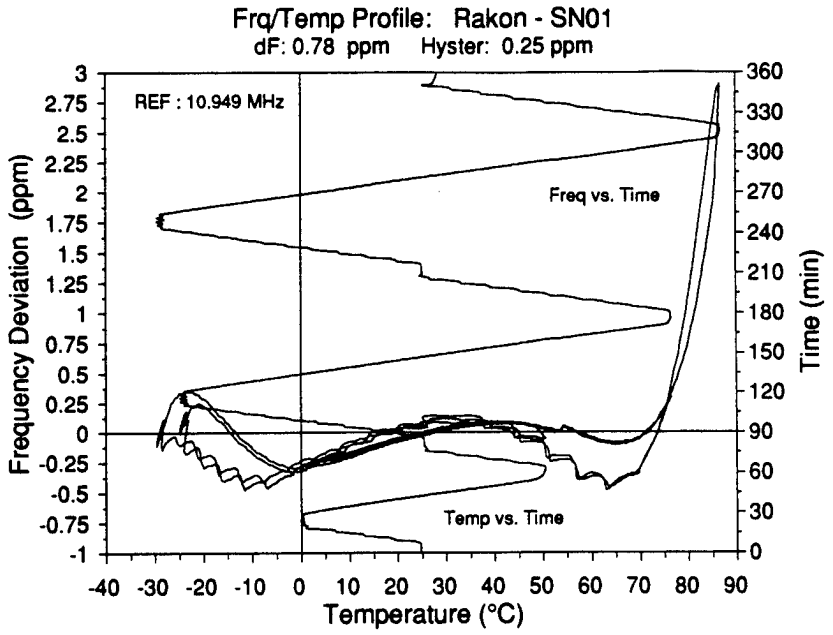


Figure 4

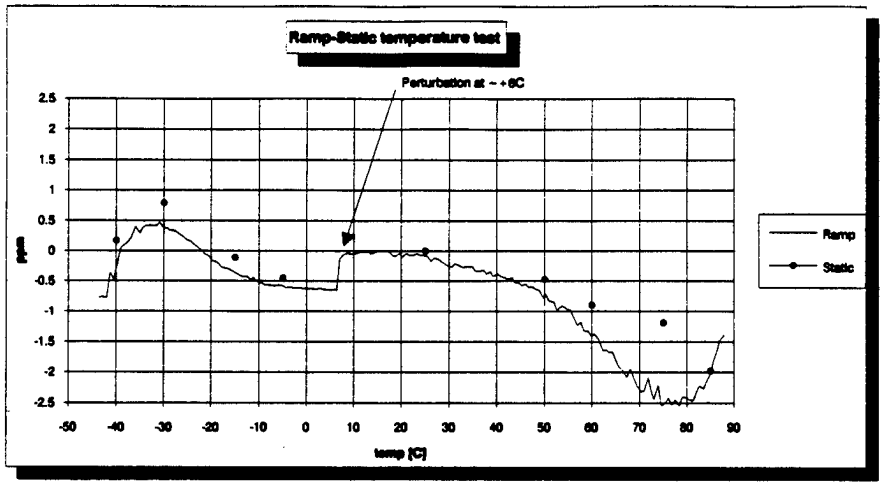


Figure 6

LD Vittorini
High AJ GPS Systems

Fstd_v10.Wb1

GPS Frequency Gradient Allowance

(sec)	(°C/min)	(ppb/day)	(days)	(ppm)	(Hz)	(Hz)	
TSSF	VT	AT_age	Storage	Stability	Age_fs	Tmp_fs	TmpVar (°C)
1000	1.0	2.74	30	1.0	129.5	1575	17

User: Stationary
Max Dop (Hz): 4347.5

System Uncertainties

(mph)	(m)	(m/sec)	(sec)	(m)		154
HV Vel	Position	Velocity	Time	Almanac	Nσ	c/Fo
30	13389	5	10	10	3.00	0.190

Vel (m/sec)
13.4

(m)	(m/sec)	(sec)		(m/sec)	(Hz)	(Hz)	
σ_ρ	σ_vv	σ_vτ	σ_av	σ_vτ	ω(σ_vτ)	Σ	-294
2.28	5.00	1.41	0.00	17.0	89.4	1794	
					mod:	194	

		(Hz)		(m/sec)
PDI	FW	Ovrlap	dF_end	Doppler
1	1000	320	60	60.9
2	500	160	30	30.5
5	200	64	12	12.2
10	100	32	6	6.1
20	50	16	3	3.0

Δω	sinc²(fT)
0.320	-1.52
0.262	-1.00
0.3184	-1.50
0.3655	-2.00
0.4424	-3.00

dF_edge 0.26
dF_power -0.99
Note: 0.5 dB window edge criteria

PDI	Tdwell	F_err	f_a
1	20	66	0.042
(msec)	(msec)	(Hz)	(ppm)
T_srch	Net S/R	∇_ω	f_unc
23	44	0.11	1794
(sec)	(cps)	(ppm/min)	(Hz)

Search Calculator	
Sriom	Nca
50	1023
N_ch	Ns
1	20
M_2	Pfa
5	0.010
M_1	Ncor
1	2
Raw S/R	Net S/R
50	44
Tscrh	
23	

Direct P			1.7E-04
time (s)	# P chips	T (μs)	P chips
30	1317	129	1705

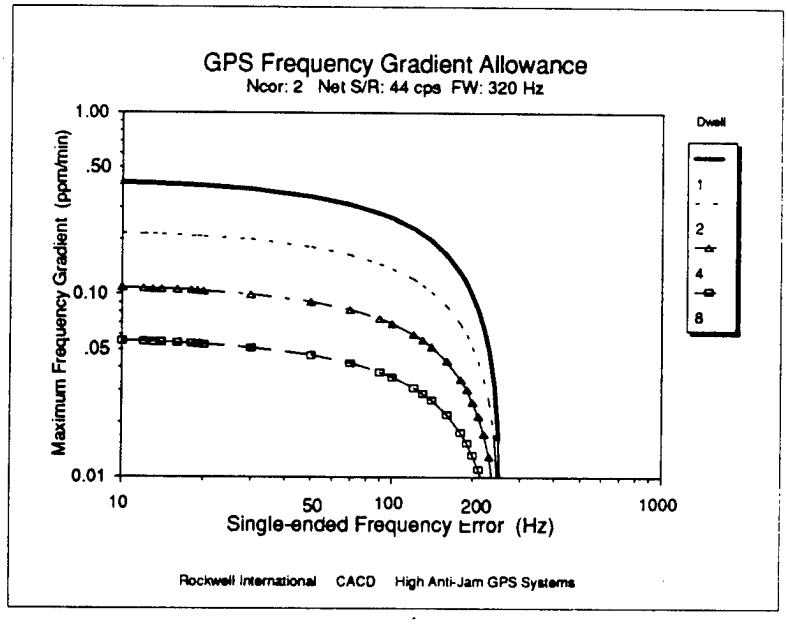


Figure 5

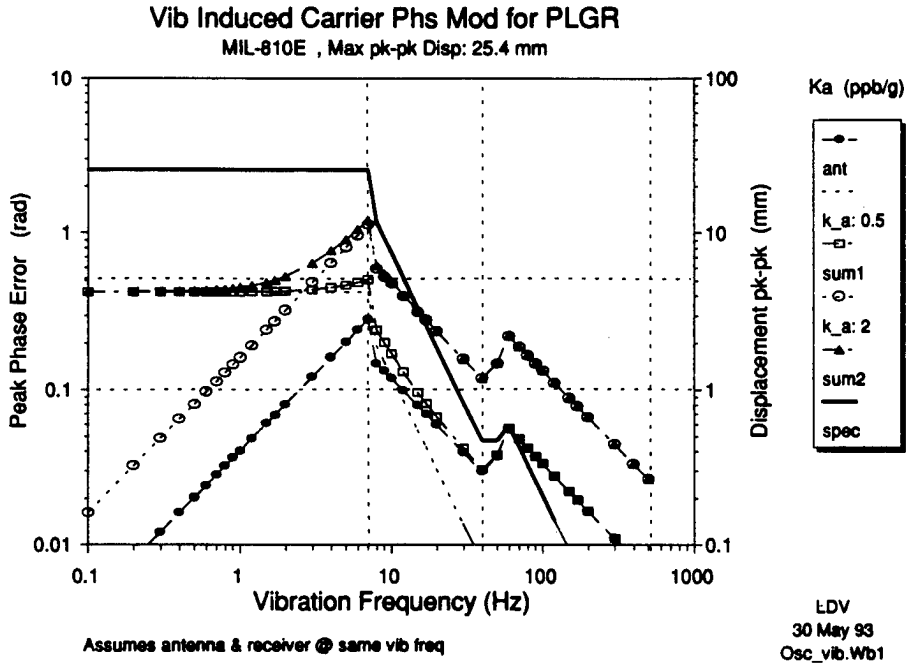


Figure 7

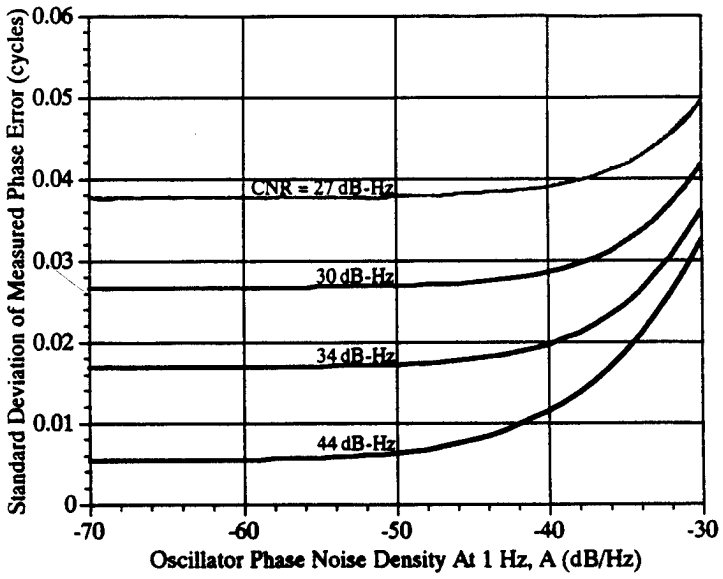


Figure 10

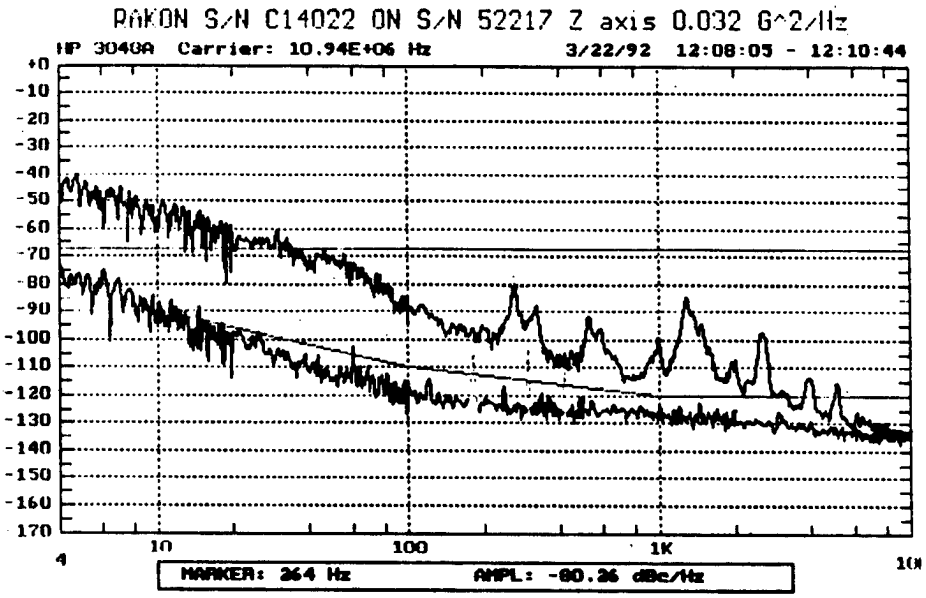


Figure 8

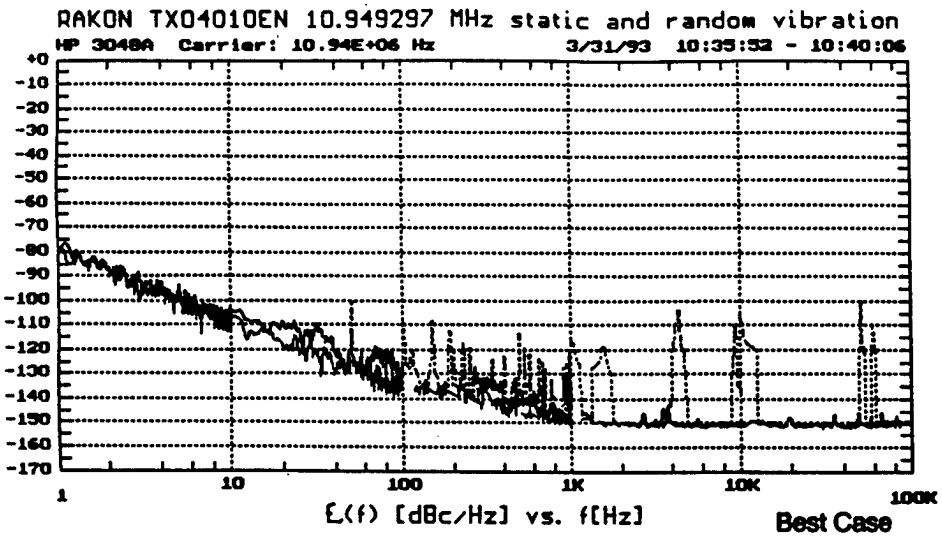


Figure 9

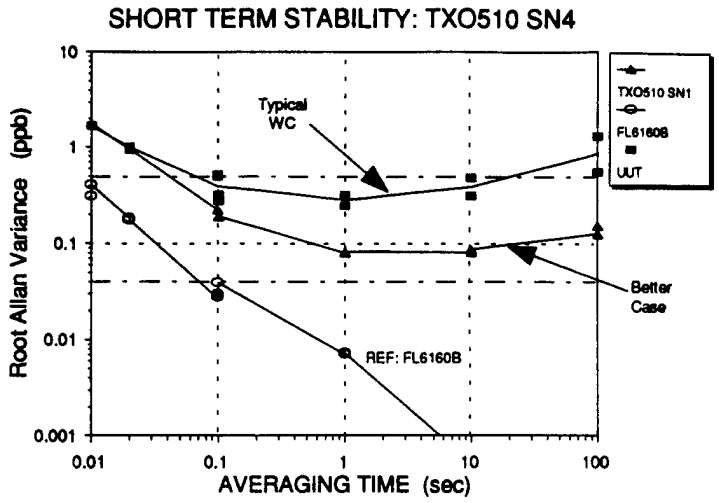


Figure 11

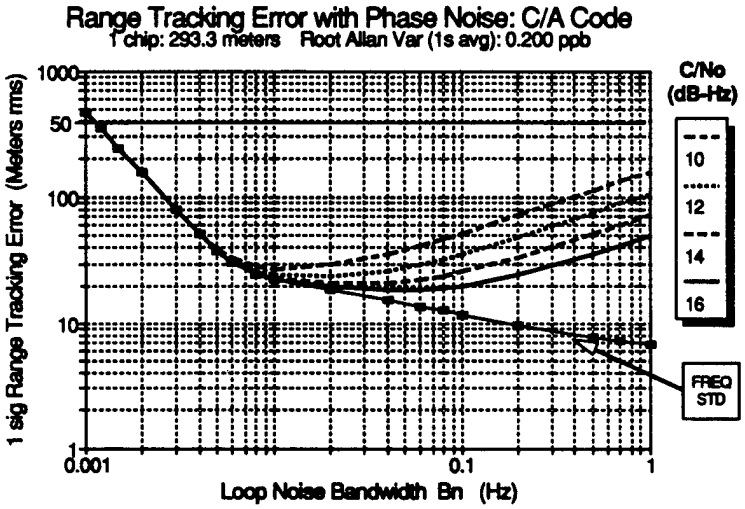


Figure 12

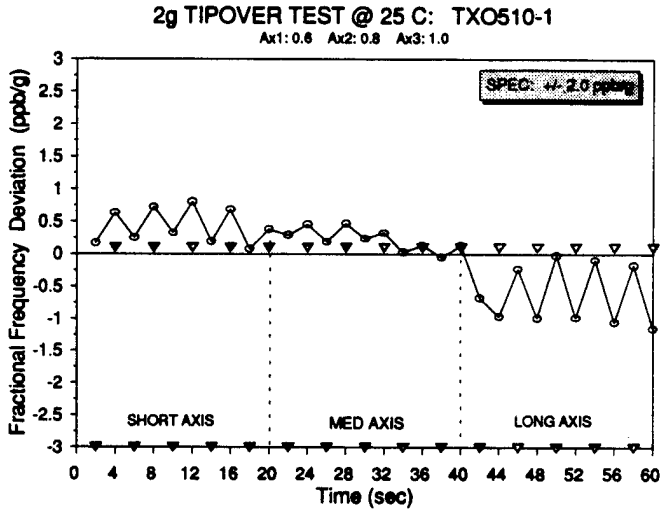


Figure 13

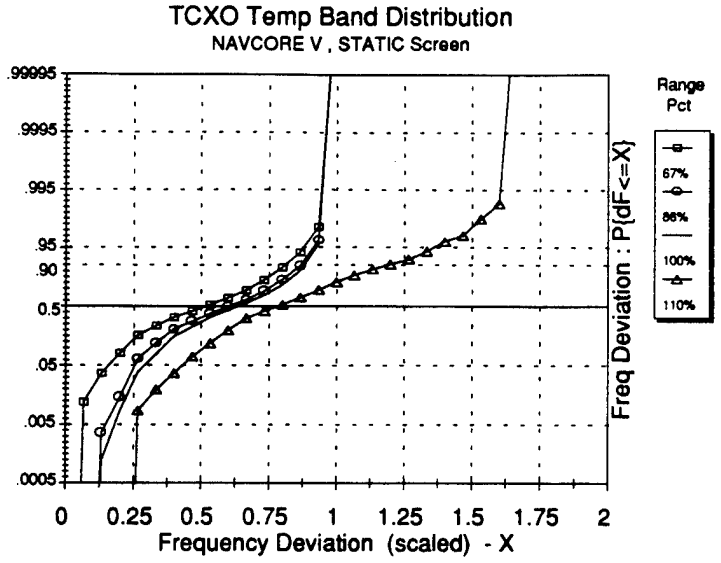


Figure 14

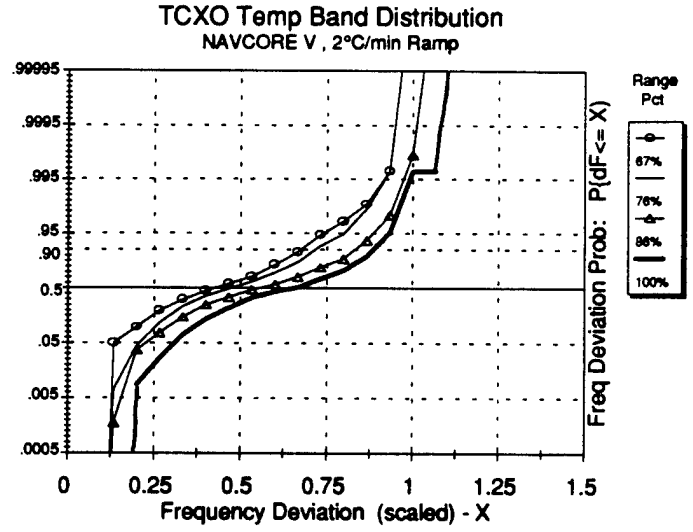


Figure 15

AUTHOR INDEX

Abdrafikov, S. 427
Abramson, I.V. 535
Achille, B. 679
Aliev, V.S. 582
Allan, D.W. 55
Andres, D. 178
Apgar, T. 784
Arendt, P.N. 784
Ascarrunz, F.G. 303
Ashby, N. 2
Asmar, S. 788
Austin, J. 283
Avramov, I. 582,645,728

Bacon, D.F. 283
Bagaev, V. 687
Baghai-Wadji, A.R. 514
Bailey, D.S. 744
Balendra, A. 45
Ballandras, S. 650
Ballato, A. 359,461,541
Barnes, J.A. 270
Bassevich, A.B. 65,97
Bausk, E.V. 656
Bava, E. 109
Bessedina, A.N. 97
Besson, R.J. 609
Bezdelkin, V.V. 339,348,351
Bhaskar, N.D. 87
Bigler, E. 650
Blair, D.G. 757
Bloch, M. 164
Bogdanov, P.P. 65
Bradaczek, H. 416
Brown, A. 55
Buisson, J.A. 23

Caldwell, S.P. 744
Calvetti, G.T. 261
Cambon, O. 381
Camparo, J.C. 114
Campbell, C.F. 472
Capelle, B. 626,632,813
Casaus, L. 603
Cernosek, R.W. 603
Clairon, A. 75
Clark, R.L. 706
Clayton, L.D. 560
Coles, C. 139
Collins, T. 679
Comer, D.T. 706
Cooke, D.W. 784

Danaher, J. 45
Danel, J.S. 587,597
deClercq, E. 75
DelCasale, A. 105
Deleuze, M. 381
DeMarchi, A. 105,109,216
Denisenko, S. 645
Detaint, J. 626,632,813
Dick, G.J. 132,144,148,774
Dikidzhi, A.N. 535
Drezin, D.I. 339
Driscoll, M.M. 186,202,744
Drullinger, R.E. 71,81
Dufour, M. 587,597
Dupuy, M. 597

Edwards, S. 749
EerNisse, E.P. 560
Endres, T.J. 261
Ermolieff, A. 597
Estrick, V.H. 209
Evans, J. 40

Fan, S-J. 353
Ferre, E.S. 303
Ferris, J.E. 371
Filler, R.L. 331
Fisk, P.T.H. 139
Flory, C.A. 763
Frueholz, R.P. 114
Frye, G.C. 603

Gebert, C.T. 603
Gevorkyan, A.G. 65,97
Giles, A.J. 749
Glaze, D.J. 71
Goiffon, A. 381
Gotalskaya, A.N. 339
Gouzhva, Y.G. 65
Grouzinenko, V.B. 548
Guo, W. 321

Halperin, A. 420
Hammersley, T.G. 784
Han, S.C. 784
Hanson, W.P. 186
Hart, D.W. 371
Hashimoto, K. 639
Hendrickson, M. 574
Higgins, T. 733
Hildebrandt, G. 416
Ho, K. 220
Hode, J.M. 632
Hua, Q.D. 40

Ibanez, A. 381,813
Ionov, B. 687

Joly, C. 626

Kabanovich, I.V. 377
Karaulnik, A.E. 548
Kawashima, H. 527,670
Kirkpatrick, J.B. 261
Kline, G.R. 718
Kolosovsky, E.A. 660
Koppang, P.A. 120
Kosinski, J. 359,620
Kosykh, A. 687
Koyama, M. 614
Kozlov, A.S. 660,665
Krisher, T. 788
Kroupa, V.F. 242
Kushner, L.J. 255
Kusters, J.A. 33,155
Kutalev, A.I. 535

Laffey, S. 574
Lakin, K.M. 502,718
Lam, W.K. 283
Lantz, E. 326
Lawn, M.A. 139
Leblois, T.G. 397
Lee, P.C.Y. 461,523
Lee, R.T. 220
Lepek, A. 312
Lepetaev, A. 687
Little, C.P. 155
Lowe, J.P. 71,81
Lu, Y. 359
Luiten, A.N. 757

MacMillan, C.C. 120
Maleki, L. 132,144,148
Mann, A.G. 757
Maradudin, A.A. 514
Marthon, S. 597
Martin, J.J. 371
Martin, K.E. 15
Martin, P. 597
Martin, S.J. 603
Marvin, D. 679
Mateescu, I. 620
Mattison, E.M. 129

McCarroll, C.P. 784
McCarron, K.T. 718
McCaskill, T.B. 23
McClelland, T. 164
Meyer, F. 326
Michel, F. 587
Mitchell, M.A. 603
Mizan, M. 733
Morabito, D. 788
Moshkovsky, A.S. 407
Mussino, F. 216

Nakazawa, M. 541,722
Naraghi, F. 326
Nelson, C.W. 298
Nelson, L.M. 298
Norton, J.R. 609
Novick, A. 81

Oomura, Y. 614
Osterdock, T.N. 33
Owings, H.B. 120

Parker, T.E. 178
Parzen, B. 698
Pascaru, I. 164
Pashchin, N.S. 665
Peters, H.E. 120
Philippot, E. 381,813
Pierre, F. 597
Pisarevsky, Y.V. 348,351
Plachotin, S.V. 535
Pogrebnyak, A. 427
Prestage, J.D. 132,144,148

Reid, W.G. 23
Reinhardt, V.S. 230
Robinson, B. 821
Rovera, G.D. 75
Rubiola, E. 105
Rukhlenko, A.S. 806

Salvo, H.L. 744
Santiago, D.G. 774
Schwartzel, J. 626,632,813
Searls, J. 749
Senyushenkov,P.A. 348,351
Shirley, J. 71
Shmaly, Y.S. 193
Siccardi, M. 216
Silvera, I.F. 129
Silvestrova, I.M. 348,351
Stansberry, S.D. 744
Stassevich, V.N. 339
Stewart, J.T. 476
Stone, C. 164
Sturzebecher, D. 733
Su, W. 331
Sunaga, K. 670
Szekely, C. 81

Taber, R.C. 763
Takeuchi, M. 541
Tanno, M. 722
Taslakov, M.A. 740
Tavella, P. 109
Taziev, R.M. 660
Tellier, C.R. 397
Tiersten, H.F. 431,442
Tjoelker, R.L. 132,144,148
Tobar, M.E. 749
Trippy, M.A. 40
Tsarapkin, D.P. 779
Tyulyakov, A.Y. 65

Uebach, W. 416

Vasiliev, A. 687
Vernotte, F. 326
Vessot, R.F.C. 129
Vig, J.R. 574
Vilar, E. 283
Vittorini, L.D. 821
Vlassov, Y.N. 665
Vyun, V.A. 509

Walls, F.L. 81,289,298,303,312
Walsworth, R.L. 129
Walton, R. 55
Wan, K.W. 283
Ward, R.W. 390
Warren, H.E. 23
Watanabe, Y. 614
Weber, R.J. 472
Wessendorf, K.O. 603,711
Williams, M.S. 40

Xuan, Z. 335

Yakovkin, I.B. 656,665
Yamaguchi, H. 541
Yamaguchi, M. 639
Yang, J. 448
Yevdokimenko, Y.I. 193
Yong, Y-K. 476
Yu, J.D. 461

Zarka, A. 626,632,813
Zavialov, S. 687
Zelenka, J. 523
Zhang, W. 448
Zheng, Y. 813
Zhou, B.L. 784
Zhou, W. 251,335
Zhou, Y.S. 431,442

IEEE Frequency Control Symposium
PROCEEDINGS ORDERING INFORMATION

NO.	YEAR	DOCUMENT NO.	OBTAIN FROM*	PRICE MICROFICHE / HARDCOPY
10	1956	AD-298322	NTIS	$9.00 / $59.00
11	1957	AD-298323	NTIS	9.00 / 66.00
12	1958	AD-298324	NTIS	9.00 / 66.00
13	1959	AD-298325	NTIS	9.00 / 73.00
14	1960	AD-246500	NTIS	9.00 / 50.00
15	1961	AD-265455	NTIS	9.00 / 43.00
16	1962	AD-285086	NTIS	9.00 / 50.00
17	1963	AD-423381	NTIS	9.00 / 66.00
18	1964	AD-450341	NTIS	9.00 / 66.00
19	1965	AD-471229	NTIS	9.00 / 66.00
20	1966	AD-800523	NTIS	9.00 / 66.00
21	1967	AD-659792	NTIS	9.00 / 59.00
22	1968	AD-844911	NTIS	9.00 / 66.00
23	1969	AD-746209	NTIS	9.00 / 43.00
24	1970	AD-746210	NTIS	9.00 / 43.00
25	1971	AD-746211	NTIS	9.00 / 43.00
26	1972	AD-771043	NTIS	9.00 / 43.00
27	1973	AD-771042	NTIS	9.00 / 50.00
28	1974	AD-A011113	NTIS	9.00 / 50.00
29	1975	AD-A017466	NTIS	9.00 / 50.00
30	1976	AD-A046089	NTIS	9.00 / 59.00
31	1977	AD-A088221	NTIS	9.00 / 66.00
32	1978	AD-A955718	NTIS	9.00 / 59.00
33	1979	AD-A213544	NTIS	9.00 / 73.00
34	1980	AD-A213670	NTIS	9.00 / 59.00
35	1981	AD-A110870	NTIS	19.00 / 73.00
36	1982	AD-A130811	NTIS	9.00 / 59.00
37	1983	AD-A136673	NTIS	9.00 / 59.00
38	1984	AD-A217381	NTIS	12.50 / 43.00
39	1985	AD-A217404	NTIS	19.00 / 59.00
40	1986	86CH2330-9	IEEE	70.00 / 70.00
41	1987	AD-A216858	NTIS	17.00 / 59.00
42	1988	AD-A217275	NTIS	17.00 / 59.00
43	1989	AD-A235629	NTIS	19.00 / 66.00
44	1990	90CH2818-3	IEEE	90.00 / 90.00
45	1991	91CH2965-2	IEEE	110.00 / 110.00
46	1992	92CH3083-3	IEEE	120.00 / 120.00
47	1993	93CH3244-1	IEEE	TBD/TBD

*NTIS - National Technical Information Service
5285 Port Royal Road, Sills Building
Springfield, VA 22161, U. S. A.
Tel: 703-487-4650

*IEEE - Inst. of Electrical & Electronics Engineers
445 Hoes Lane
Piscataway, NJ 08854, U.S.A.
Tel: 800-678-4333 or 908-981-0060

An index to all the papers in the Proceedings since 1956, and other Symposium information are available as ASCII files on a 5 1/4" (13 cm) MS-DOS format floppy disk, for $5-, from Synergistic Management, Inc., 3100 Route 138, Wall Township, NJ 07719.

SPECIFICATIONS AND STANDARDS RELATING TO FREQUENCY CONTROL

INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS (IEEE)

Order from: IEEE Service Center
445 Hoes Lane
Piscataway, NJ 08854
(201) 981-0060

176-1987 (ANSI/IEEE) Standard on Piezoelectricity (SH11270)

177-1966 Standard Definitions & Methods of Measurements for Piezoelectric Vibrators

180-1986 (ANSI/IEEE) Definitions of Primary Ferroelectric Crystal Terms (SH10553)

319-1971 (Reaff 1978) Piezomagnetic Nomenclature (SH02360)

1139-1988 Standard Definitions of Physical Quantities for Fundamental Frequency & Time Metrology (SH12526)

DEPARTMENT OF DEFENSE

Order from: Military Specifications and Standards
700 Robbins Ave., Bldg. 4D
Philadelphia, PA 19111-5094
(215) 697-2667/2179 - Customer Service
(215) 697-1187 to -1195 Telephone Order Entry System

MIL-C-3098 Crystal Unit, Quartz, Gen Spec for

MIL-C-49468 Crystal Units, Quartz, Precision, Gen Spec for

MIL-C-24523 (SHIPS), Chronometer, Quartz Crystal

MIL-F-15733 Filters & Capacitors, Radio Interference, Gen Spec for

MIL-F-18327 Filters, High Pas, Band Pass Suppression and Dual Processing, Gen Spec for

MIL-F-28861 Filters and Capacitors, Radio Frequency/Electromagnetic Interference Suppression, Gen Spec for

MIL-F-28811 Frequency Standard, Cesium Beam Tube

MIL-H-10056 Holders (Encl), Crystal Gen Spec for

MIL-O-55310 Oscillators, Crystal, Gen Spec for

MIL-O-39021 Oven, Crystal, Gen Spec for

MIL-S-49433 Surface Acoustic Wave Devices, Bandpass Filter

MIL-S-49433(ER) Surface Acoustic Wave Devices, Gen Spec for

MIL-STD-683 Crystal Units, Quartz/Holders, Crystal

MIL-STD-188-115 Interoperability & Performance Standards for Communications, Timing & Synchronization Subsystems

MIL-STD-1395 Filters & Networks, Selection & Use of

MIL-T-28816(EC) Time Frequency Standard, Disciplined, AN/URQ-23, Gen Spec for

MIL-W-46374D Watch wrist: General Purpose

MIL-W-87967 Watch wrist: Digital

GENERAL SERVICES ADMINISTRATION

Order from: Naval Publication & Form Center or General Services Administration Business Service Centers in major U.S. cities

FED-STD-1002 Time & Frequency Reference Information in Telecommunication Systems

ELECTRONIC INDUSTRIES ASSOCIATION

Order from: Electronic Industries Assoc.
2001 Eye Street, NW
Washington, DC 20006
(202) 457-4900

(a) Holders and Sockets

EIA-192-A, Holder Outlines and Pin Connections for Quartz Crystal Units (Standard Dimensions for Holder Types).

EIA-367, Dimensional & Electrical Characteristics Defining Receiver Type Sockets (including crystal sockets).

EIA-417, Crystal Outlines (Standard dimensions and pin connections for current quartz crystal units-1974).

(b) Production Tests

EIA-186-E, (All Sections), Standard Test Methods for Electronic Component Parts.

EIA-512, Standard Methods for Measurement of Equivalent Electrical Parameters of Quartz Crystal Units, 1 kHz to 1 GHz, 1985.

EIA/IS-17-A, Assessment of Outgoing Nonconforming Levels in Parts Per Million (PPM).

EIA-IS-18, Lot Acceptance Procedure for Verifying Compliance with the Specified Quality Level (SQL) in PPM.

(c) Application Information

EIA Components Bulletin No. CB6-A, Guide for the Use of Quartz Crystal Units for Frequency Control, Oct 1987.

(d) EIA-477, Cultured Quartz (Apr. 81)

EIA-477-1, Quartz Crystal Test Methods (May 1985).

INTERNATIONAL ELECTROTECHNICAL COMMISSION (IEC)

Order from: American Nat'l. Standard Inst.
(ANSI), 1430 Broadway
New York, NY 10018
(212) 354-3300

IEC PUBLICATIONS

122: Quartz crystal units for frequency control and selection.

122-1 (1976) Part 1: Standard values and test conditions. Amendment No. 1 (1983).

122-2 (1983) Part 2: Guide to the use of quartz crystal
units for frequency control and selection.

122-3 (1977) Part 3: Standard outlines and pin connections. Amendment No. 1 (1984)

122-3A (1979) First supplement.

122-3B (1980) Second supplement.

122-3C (1981) Third supplement.

283 (1986) Methods for the measurement of frequency & equivalent resistance of unwanted resonances of filter crystal units.

302 (1969) Standard definitions & methods of measurment for piezoelectric vibrators operating over the frequency range up to 30 MHz.

314 (1970) Temperature control devices for quartz crystal units. Amendment No. 1 (1979)

314A (1971) First supplement.

368: Piezoelectric Filters.

368-1 (1982) Part 1: General information, standard values and test conditions.

368A (1973) First supplement. Amendment No.1 (1977). Amendment No.2 (1982).

368B (1975) Second supplement.

368-2 Part 2: Guide to the use of piezoelectric filters.

368-2-1 (1988) Section One-Quartz crystal filters

368-3 (1979) Part 3: Standard outlines.

368-3A (1981) First supplement.

444: Measurement of quartz crystal unit parameters by zero phase technique in a π-network.

444-1 (1986) Part 1: Basic method for the measurement of resonance frequency and resonance resistance of quartz crystal units by zero phase technique in a π-network.

444-2 (1980) Part 2: Phase offset method for measurement of motional capacitance of quartz crystal units.

444-3 (1986) Part 3: Basic method for the measurement of two-terminal parameters of quartz crystal units up to 200 MHz by phase technique in a π-network with compensation of the parallel capacitance Co.

444-4 (1988) Part 4: Method for the measurement of the load resonance frequency fL, load

resonance resistance RL and the calculation of other derived values of quartz crystal units, up to 30 MHz.

483 (1976) Guide to dynamic measurements of piezoelectric ceramics with high electromechanical coupling.

642 (1979) Piezoelectric ceramic resonators and resonator units for frequency control and selection. Chapter I: Standard values and conditions. Chapter II: Measuring and test conditions.

679: Quartz crystal controlled oscillators.

679-1 (1980) Part 1: General information, test conditions & methods.

679-2 (1981) Part 2: Guide to the use of quartz crystal controlled oscillators. Amendment No. 1 (1985).

689 (1980) Measurements and test methods for 32 kHz quartz crystal units for wrist watches and standard values.

758 (1983) Synthetic quartz crystal. Chapter I: Specification for synthetic quartz crystal. Chapter II: Guide to the use of synthetic quartz crystal. Amendment No. 1 (1984).

862: Surface acoustic wave (SAW) filters.

862-1-1 (1985) Part 1: General Information, test conditions and methods.

862-3 (1986) Part 3: Standard outlines.

CONSULTATIVE COMMITTEE ON INTERNATIONAL RADIO (CCIR)

Order from: International Telecommunications Union
General Secretariat - Sales Section
Place des Nations
CH-1211 Geneva
SWITZERLAND

Ask for CCIR 17th Plenary Assembly, Volume VII, "Standard Frequencies and Time Signals (Study Group 7)", which contains all of the following documents:

RECOMMENDATION 457-1	Use of the Modified Julian Date by the Standard-Frequency and Time-Signal Services
RECOMMENDATION 458-1	International Comparisons of Atomic Time Scales
RECOMMENDATION 460-4	Standard Frequency and Time Signal Emissions [Note: defines the UTC system]
RECOMMENDATION 485-1	Use of Time Scales in the Field of Standard-Frequency and Time Services
RECOMMENDATION 486-1	Reference of Precisely Controlled Frequency Generators and Emissions to the International Atomic Time Scale
RECOMMENDATION 535-1	Use of the Term UTC
RECOMMENDATION 536	Time Scale Notations
RECOMMENDATION 538	Frequency and Phase Stability Measures
REPORT 580-2	Characterization of Frequency and Phase Noise
RECOMMENDATION 583	Time Codes
RECOMMENDATION 685	International Synchronization of UTC Time Scales
RECOMMENDATION 686	Glossary