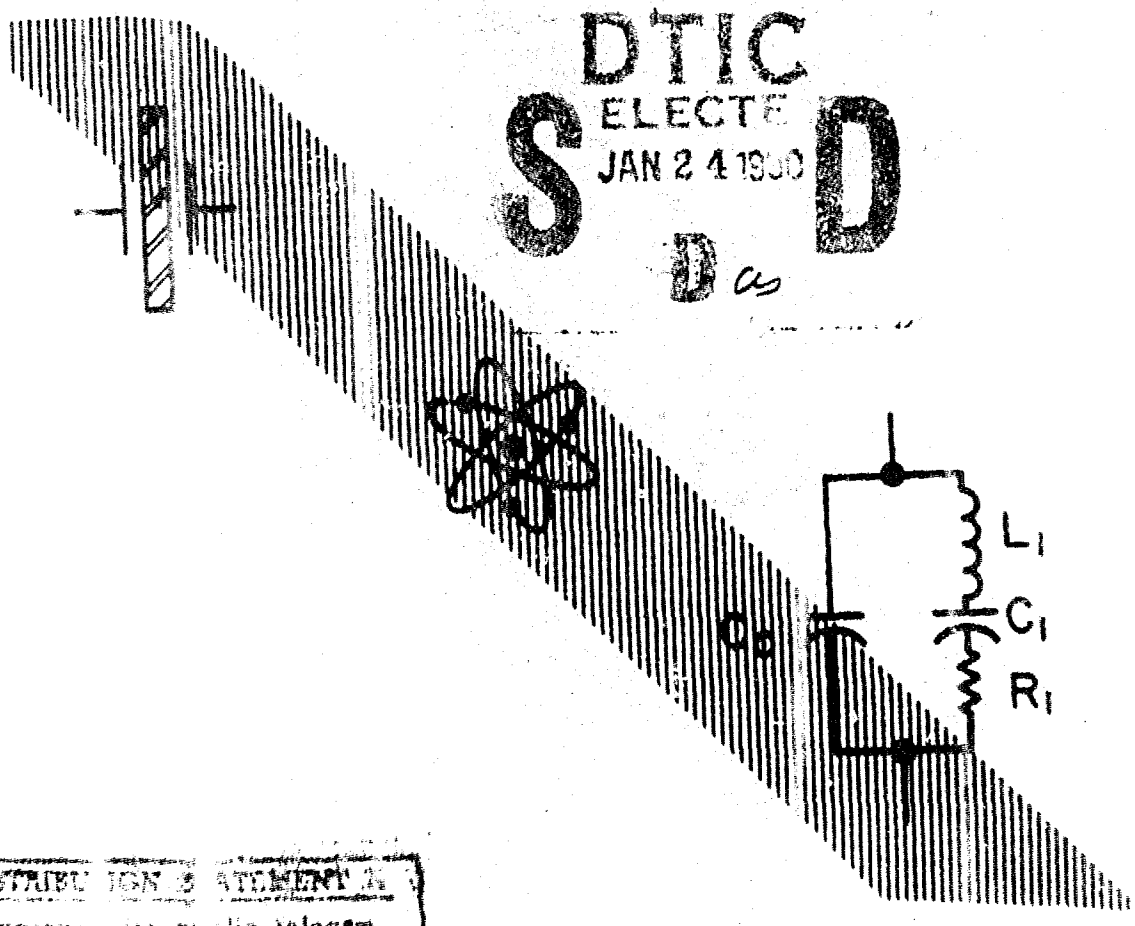
PROCEEDINGS
OF THE
39TH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL
1985

AD-A217 404

DTIC
ELECTE
JAN 2 4 1990
S D
D cs



DISTRIBUTION STATEMENT A
Approved for public release;
Distribution unlimited

29 - 31 MAY 1985

85CH2186-5

58-60781

PROCEEDINGS

OF THE

39th ANNUAL
Control
FREQUENCY SYMPOSIUM

1985

Co-Sponsored by the

U.S. ARMY ELECTRONICS RESEARCH
AND DEVELOPMENT COMMAND
ELECTRONICS TECHNOLOGY AND
DEVICES LABORATORY

and

THE INSTITUTE OF ELECTRICAL AND
ELECTRONICS ENGINEERS, INC.
SONICS AND ULTRASONICS GROUP

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By per ltr	
Distribution/	
IEEE out of stock	
Availability Codes	
Dist	Avail and / or Special
A-1	

IEEE Catalog No. 85CH2186-5

Library of Congress No. 58-60781

29 - 31 May 1985

Marriott Hotel

Philadelphia, Pennsylvania

90 01 23 227

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of US 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, 21 Congress Street, Salem, MA 01970. Instructors are permitted to photocopy isolated articles for noncommercial classroom use without fee. For other copying, reprint or republication permission, write to Director, Publishing Services, IEEE, 345 East 47th Street, New York, NY 10017. All rights reserved. Copyright © 1985 by the Institute of Electrical and Electronics Engineers, Inc.

THIRTY-NINTH ANNUAL FREQUENCY CONTROL SYMPOSIUM

Co-Sponsored by

U.S. ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND
ELECTRONICS TECHNOLOGY AND DEVICES LABORATORY

and

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.
SONICS AND ULTRASONICS GROUP

SYMPOSIUM CHAIRMEN

General Chairman.....Dr. John R. Vig, U.S. Army, ERADCOM
Technical Program Chairman.....Dr. Samuel R. Stein, Ball Corporation
Finance Chairman.....Dr. Thomas E. Parker, Raytheon Research Division

TECHNICAL PROGRAM COMMITTEE

MR. D. ALLAN
National Bureau of Standards

DR. A BALLATO
U.S. Army ERADCOM

MR. M. BLOCH
Frequency Electronics, Inc.

DR. J. BROWN
Fisher Controls International, Inc.

DR. L. CUTLER
Hewlett-Packard Company

MR. L. DWORSKY
Motorola, Inc.

DR. E. EERNISSE
Quartex, Inc.

MR. M. FRERKING
Rockwell International

DR. H. HELLWIG
Frequency an ' Time Systems, Inc.

DR. W. HORTON
Piezo Technology, Inc.

MR. C. JENSIK
Piezo Crystal Company

DR. R. KAGIWADA
TRW-DSSG

MR. D. KEMPER
Quarztek, Inc.

MR. D KOEHLER
Sandia National Laboratory

MR. J. KUSTERS
Hewlett-Packard Company

MR. T. LUKASZEK
U.S. Army ERADCOM

DR. T. MEEKER
Bell Laboratories

DR. T. PARKER
Raytheon Research Division

MR. V. ROSATI
U.S. Army ERADCOM

MR. L. RUEGER
Applied Physics Laboratory

MR. S. SCHODOWSKI
U.S.Amy ERADCOM

DR. R. SYDNOR
Jet Propulsion Laboratory

DR. S. STEIN
Ball Corporation

DR. J. VIG
U.S. Army ERADCOM

MR. J. WHITE
Naval Research Laboratory

DR. G. WINKLER
U.S. Naval Observatory

DR. N. YANNONI
U.S. Air Force RADC

TECHNICAL SESSION CHAIRMEN

PLENARY SESSION
S. Stein, Ball Corporation

QUARTZ RESONATOR STUDIES I
R. Besson, Ecole Nationale Sueprieure de Macanique et des Microtechniques, France

CRYSTAL OSCILLATORS
M. Frerking, Rockwell International

CRYSTAL MEASUREMENTS AND APPLICATIONS
B. Smythe, Piezo Technology

OSCILLATORS AND SIGNAL PROCESSING
B. Parzen, San Diego, CA

QUARTZ RESONATORS STUDIES II
R. Peach, GEC Research Laboratories, United Kingdom

RESONATOR PROCESSING I
J. Staudte, Xtal Equipment Corporation

NOISE IN RESONATORS, FILTERS AND OSCILLATORS—A PANEL DISCUSSION
J. Barnes, Austron, Inc.

Panelists:

A. Benjaminson, ST Research, Inc.
M. Driscoll, Westinghouse, Inc.
J. Gagnepain, Laboratoire de Physique et Metrologie des Oscillateurs, France
B. Parzen, Consultant
R. Smythe, Piezo Technology, Inc.
C. Stone, Brightline, Inc.
F. Walls, National Bureau of Standards

RESONATOR DESIGN
A. Ballato, U.S. Army ERADCOM

FILTERS
B. Kinsman, Motorola, Inc.

SAW DEVICES
K.F. Kau, TRW

FREQUENCY STABILITY
J. Gagnepain, Laboratoire de Physique et Metrologie des Oscillateurs, France

RESONATOR PROCESSING II
T. Lukaszek, U.S. Army ERADCOM

CESIUM BEAM FREQUENCY STANDARDS
N. Ramsey, Harvard University

RESONATOR PROCESSING III
R. Ward, Quartztronics, Inc.

TIME DISTRIBUTION AND TRANSFER
C. Wardrip, Bendix Field Engineering

CRYSTAL RESONATOR NONLINEARITY
R. Filler, U.S. Army ERADCOM

PRECISION FREQUENCY SOURCES
J. Vanier, National Research Council, Canada

PROPERTIES OF QUARTZ AND OTHER MATERIALS
J. Vig, U.S. Army ERADCOM

HYDROGEN MASERS AND RUBIDIUM OSCILLATORS
M. Ohtsu, Tokyo Institute of Technology, Japan

TABLE OF CONTENTS

1985 Award Winners.....1

Awards Program.....2

ATOMIC FREQUENCY STANDARDS

A Beam Reversal Experiment on NBS-6 Primary Cs Standard Including Rabi Pulling Evaluation.....3
A. DeMarchi, Instituto Elettrotecnico Nazionale G. Ferraris, G.D. Rovera, ITAL CABLE Grant
R. Drullinger and D.A. Howe, National Bureau of Standards

Experimental Studies on Majorana Transitions in a Cs Atomic Beam Frequency Standard.....8
A. Bauch and T. Heindorff, Physikalisch-Technische Bundesanstalt, Federal Republic of Germany

A Recirculating Oven for Atomic Beam Frequency Standards.....13
R.E. Drullinger, D.J. Glaze, and D.B. Sullivan, Time and Frequency Division,
National Bureau of Standards

Optically Pumped Small Cesium Beam Standards: A Status Report.....18
A. Derbyshire, R.E. Drullinger, M. Feldman, D.J. Glaze, D. Hilliard, D.A. Howe,
L.L. Lewis and J.H. Shirley, National Bureau of Standards
I. Pascaru and D. Stanciulescu, Frequency Electronics, Inc.

Fluorescent Light Shift in Optically Pumped Cesium Standards.....22
J. Shirley, National Bureau of Standards

Cesium Beam Frequency Standard for the PLSS Program.....24
D. Silvermetz, I. Pascaru and M. Meirs, Frequency Electronics, Inc.

A Rubidium Clock Model.....29
R.P. Frueholz and J.C. Camparo, The Aerospace Corporation

A Highly Stabilized Semiconductor Laser and its Application to Optically Pumped Rb Atomic Clock.....43
M. Ohtsu, M. Hashimoto and H. Ozawa, Tokyo Institute of Technology, Japan

An Ultra-Miniature Rubidium Frequency Standard.....54
K. Chiba and T. Hashi, Fujitsu Limited, Japan

Evaluation of the Rubidium Atomic Frequency Standard Developed in India.....59
G.M. Saxena, A. Chatterjee (Mrs.), D.S. Sachdeva and B.S. Mathur, National Physical Laboratory, India

Experimental Results on a Frequency Standard Based on a Rubidium 87 Maser.....64
M. Têtu, R. Brousseau, N. Cyr, A. Michaud, P. Tremblay, B. Villeneuve, Université Laval, Canada

Measurement and Interpretation of Hydrogen Maser Quality Parameters.....72
E.M. Mattison and R.F.C. Vessot, Smithsonian Astrophysical Observatory
W. Shen, Shaanxi Observatory, The People's Republic of China

Properties of Low-Expansion Materials For Hydrogen Maser Cavities.....75
E.M. Mattison and R.F.C. Vessot, Smithsonian Astrophysical Observatory
S.F. Jacobs, University of Arizona

Design for a Subcompact Q-Enhanced Active Maser.....80
R.R. Hayes and H.T.M. Wang, Hughes Research Laboratories

Doppler-Free Two-Photon Laser Spectroscopy of HgII.....85
J.C. Bergquist, D.J. Wineland, W.M. Itano, H. Hemmati, H.U. Daniel, and G. Leuchs,
National Bureau of Standards

Performance of Laser-Induced Resonance Raman Clock.....88
P.R. Hemmer,Rome Air Development Center, G.P. Ontai, Captain, U.S. Army,
A. Rosenberg and S. Ezekiel, Massachusetts Institute of Technology

Errors in Servo Systems Using Sinusoidal Frequency (Phase) Modulation.....91
F.L. Walls, National Bureau of Standards

FREQUENCY STABILITY

1/f Frequency Fluctuations in Acoustic and Other Stable Oscillators.....97
T.E. Parker, Raytheon Research Division

TABLE OF CONTENTS (Continued)

Time Scale Stabilities Based on Time and Frequency Kalman Filters.....107
J.A. Barnes, Austron, Inc., and D.W. Allan, National Bureau of Standards

The Fractal Dimension of Phase and Frequency Noises:
Another Approach to Oscillator Characterization.....113
J.J. Gagnepain, J. Groslambert, R. Brendel, Laboratoire de Physique et Métrologie des Oscillateurs, France

Frequency Stability Characterization From the Filtered Signal of a Precision Oscillator.....119
P. Tremblay and M. Têtu, Laboratoire de Recherche sur les Oscillateurs, Université Laval, Canada

The Analytic Signal Representation of Oscillators with Application to Frequency Stability Analysis.....127
L.G. Bernier and F.E. Gardiol, Laboratoire d'Electromagnétisme et d'Acoustique, Ecole Polytechnique Fédérale, Switzerland

Application of Spectrum Estimation in Phase Noise Measurement.....132
D. Wulin, X. Sanbao, Northwest Telecommunications Engineering Institute
People's Republic of China

NOISE IN RESONATORS, FILTERS AND OSCILLATORS PANEL (A DISCUSSION SESSION)

A Comparison of Frequency Noise of Quartz Resonators.....135
J.J. Gagnepain, Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S. associe â l'Université de Franche-Comté-Besançon, France

Phase Noise in Crystal Filters.....138
R.C. Smythe, Piezo Technology, Inc.

Designing Crystal Oscillators for Improved Phase-Noise Performance.....140
A. Benjaminson, S.T. Research Corp.

TIME DISTRIBUTION AND TRANSFER

Development of a GPS Time-Transfer Receiver and Time Compaison Results.....145
M. Imae, M. Uratsuka, C. Miki, T. Morikawa, K. Akatsuka and K. Yoshimura, Radio Research Laboratory, Ministry of Posts and Telecommunications, Japan

Commercial GPS Receiver for Time and Frequency Equipment Applications.....150
R.L. Lewis, G.F. Knoernschild and N.B. Hemesath, Collins Government Avionics Division, Rockwell International Corp.

The State-of-the-Art Medium Terminal (SAMT) Time and Frequency Distribution System.....153
A. Vulcan and M.B. Bloch, Frequency Electronics, Inc.

OSCILLATORS

A 9.2 GHz Superconducting Cavity Stabilized Oscillator.....159
B. Komiyama, Radio Research Laboratory, Japan

Reducing TCXO Error After Aging Adjustment.....166
R.L. Clark, McCoy Electronics Co.

Results of Continued Development of the Differential Crystal Oscillator.....176
T. Adachi, M. Hirose, Y. Tsuzuki, Yokohama National University, Japan

Recent Developments in Synchronization and Tracking with Synchronous Oscillators.....183
T. Flamouropoulos, M.H. White, Fairchild Laboratory of Lehigh University
V. Uzunoglu, Fairchild Communications and Electronics Co.

TABLE OF CONTENTS (Continued)

A Novel Synthesizer for Miniature SSB Radio Equipment.....189
C.K. Richardson, Plessey Electronic Systems Research, United Kingdom

Hybrid Miniature Oven Quartz Crystal Oscillator.....193
J. Ho, Frequency Electronics, Inc.

Low Noise, VHF Crystal-Controlled Oscillator Utilizing Dual, SC-Cut Resonators.....197
M.M. Driscoll, Westinghouse Defense and Electronics Center

A Satellite Oscillator for Very Precise Orbitography: The Doris Program.....202
A. Debaisieux, J.P. Aubry, E. Gerard, CEPE, and M. Brunet, CNES, France

An Oscillator for the GPS Program.....212
G. Marotel, G. Caret and J.P. Aubry, CEPE, France

PROPERTIES OF QUARTZ AND BERLINITE, RADIATION EFFECTS, AND ETCHING

Characteristics of Natural, Swept Natural, and Cultured X- and Z-Growth Quartz Material in High Temperature, High Stress Applications.....223
J.A. Kusters and G.S. Kaitz, Hewlett-Packard Co.

The Growth of High Quality Quartz in Commercial Autoclaves.....230
A.F. Armington, Rome Air Development Center, and J. Balascio, Motorola Components Division

Crystal Growth, Physical Characterization and B.A.W. Devices Applications of Berlinite.....234
J. Détaint, J. Schwartzel, CNET/PAB/BAG/MCT Bagneux, E. Philippot, J.C. Jumas, Université de Montpellier, A. Zarka and B. Capelle, Université de Paris VI, and J. Doukhan, Université de Lille, France

The Influence of Temperature and Electric Field on the Etch-Channel Density in Swept Cultured Quartz.....247
J.G. Gualtieri, U.S. Army ERADCOM

Defect Centers in Irradiated and Swept Quartz.....255
A. Kahan and H.G. Lipson, Rome Air Development Center

Radiation Effects in Quartz: Low Doses and Defect Production Mechanisms.....259
L.E. Halliburton, C.Y. Chen and S.D. Tapp, Oklahoma State University

Radiation Effects in the Acoustic Loss Spectra of AT-Cut Quartz Crystals.....266
J.J. Martin, H.B. Hwang, and J. Bahadur, Oklahoma State University

Quartz for the National Defense Stockpile.....270
R.A. Laudise, AT&T Bell Laboratories

Etching Study of AT-Cut Cultured Quartz Using Etchants Containing Fluoride Salts, Hydrofluoric Acid, and Ammonium Bifluoride.....271
A.J. Bernot, Sperry Corp.

Chemical Polishing in Etching Solutions That Contain Surfactants.....276
R.J. Brandmayr and J.R. Vig, U.S. Army ERADCOM

Effect of Crystal Orientation on the Surface Texture of Chemically Etched Quartz Plates, The Case of Cuts Close to the AT-Cut.....282
C.R. Tellier, Laboratoire de Chronométrie, Electronique et Piezoélectricité, ENSMM, France

Chemically Milled VHF and UHF AT-Cut Resonators.....292
J.R. Hunt and R.C. Smythe, Piezo Technology, Inc.

TABLE OF CONTENTS (Continued)

Etch Processing of Bulk and Surface Wave Devices.....301
J. Dowsett, F.G. Dwyer and F. Stern, STC Components Ltd., and R.A. Heinecke and A.H. Truelove, STC Technology Ltd., United Kingdom

QUARTZ AND OTHER PIEZOELECTRIC RESONATORS—THEORY, DESIGN & PROCESSING

The Amplitude-Frequency Effect In SC-Cut Resonators.....311
R.L. Filler, U.S. Army ERADCOM

Experimental Evaluation of the Effective Non-Linear Elastic Constant for Trapped Energy and Contoured Resonators.....317
R.C. Smythe and P.E. Morley, Piezo Technology, Inc.

The Evaluation of the Coefficient of Nonlinear Resonance for SC-Cut Quartz Resonators.....325
H.F. Tiersten and D.S. Stevens, Rensselaer Polytechnic Institute

Laser Processed Miniature $LiTaO_3$ Resonators and Monolithics Filters.....333
R. Lefevre, L. Jenselme and D. Servajean, CNET, France

A Progress Report on Manufacturing Methods and Technology for Production of High-Stability, Vibration-Resistant Quartz Crystal Units.....338
J.C. Korman, Frequency Electronics, Inc.

Cut and Grind.....342
A.W. Warner, Jr., J. Tsaclas and J. Korman, Frequency Electronics, Inc.

Low Profile Glass Packaged Crystal Unit.....345
M. Sato, J. Eguchi, S. Ishigami, Tokyo Denpa Co., Ltd., and K. Yamamoto
Science University of Tokyo, Japan

Strip Type Resonator of Lithium Tetraborate.....351
Y. Fujiwara, M. Ono, M. Sakai and N. Wakatsuki, Fujitsu Limited, Japan

Recent Developments on Membrane Bulk-Acoustic-Wave Resonators.....356
J.S. Wang, A. Kong, K.F. Lau and K.H. Yen, TRW Electronic Systems Group

An Air-Gap Type Piezoelectric Composite Thin Film Resonator.....361
H. Satoh, Y. Ebata, H. Suzuki and C. Narahara, Toshiba Research & Development Center, Japan

Processing of a Five Resonator VHF Crystal Device.....367
C.W. Shanley, L.N. Dworsky, J.A. Whalin, G.C. Clifford and M.N. Scansaroli, Motorola, Inc.

Highly Stable, Ovenized Bulk Shear Mode Resonators.....372
B.R. McAvoy and S.V. Krisnaswamy, Westinghouse R&D Center
H.L. Salvo, Jr., and R.A. Moore, Westinghouse Defense and Electronics Center

A Novel Miniature ZT-Cut Resonator.....375
J.W. Hermann, Swiss Center for Electronics and Microtechnology, Inc., Switzerland

Characteristics of a Quartz Crystal Tuning Fork with Shortened Arm Length for High Frequencies.....381
M. Okazaki, and H. Tohma, Nihon Dempa Kogyo Co., Ltd.,
Y. Tomikawa, Yamagata University, Japan

Analysis of Trapped Energy Resonators with Tabs.....386
H. Sekimoto, Tokyo Metropolitan University
H. Nakata and M. Miura, Nihon Dempa Kogyo Co., Ltd., Japan

A Variational Method for the Design of Trapped Energy Resonators.....392
R.C. Peach, GEC Research Laboratories Limited, United Kingdom

TABLE OF CONTENTS (Continued)

Further Results on the Contour Dependence of the Frequency-Temperature Characteristic of SC-Cut Resonators.....400
J.A. Kosinski, U.S. Army ERADCOM

Mechanical Couplings Involving Discontinuities of the Frequency-Temperature Curves of Contoured Quartz Resonators.....405
R. Bourquin, B. Dulmet, G. Genestier, Ecole Nationale Superieure de Mecanique et des Microtechniques, France

Frequency-Temperature Behavior of Flexural and Thickness-Shear Vibrations of Rectangular Rotated Y-Cut Quartz Plates.....415
Y.K. Yong, Rutgers University, P.C.Y. Lee, Princeton University

A note on "Ballato's Angle Increment".....427
J.H. Sherman, Jr., Lynchburg, Va.

Vibrations of Piezoelectric Discs Under Initial Stresses.....431
M.C. Dökmeci, Istanbul Technical University, Turkey

An Analysis of Doubly-Rotated Contoured Quartz Crystal Resonators.....436
D.S. Stevens and H.F. Tiersten, Rensselaer Polytechnic Institute

Generalized Equation for the Force-Frequency Characteristics of Circular Quartz Plates with Three-Point Support and its Application to Supporting of an SC-Cut Plate.....448
N. Oura, N. Kuramochi, Y. Miyazaki, M. Yamashina, and S. Suzuki
Tokyo Institute of Technology, Japan

Vibrations of Doubly-Rotated Piezoelectric Crystal Strip with a Pair of Electrode-Plated, Traction-Free Edges....453
P.C.Y. Lee and J.P. Hou, Princeton University

Simple Thickness Plate Modes Driven by Lateral Felds.....462
A. Ballato, E.R. Hatch, M. Mizan, T. Lukaszek, and R. Tilton, U.S. Army ERADCOM

Lateral Field Resonators.....473
A.W. Warner, Jr., and B. Goldfrank, Frequency Electronics, Inc.

Applications of Total Process Control Techniques in the Production of High Precision Quartz Resonators.....475
J.A. Kusters and C.A. Adams, Hewlett-Packard Co.

FILTERS AND SAW DEVICES

VHF Monolithic Crystal Filters Fabricated by Chemical Milling.....481
R.C. Smythe, M.D. Howard and J.R. Hunt, Piezo Technology, Inc.

The Motorola Multi-Pole Monolithic Filter Project.....486
L.N. Dworsky and C.S. Shanley, Motorola, Inc.

Monolithic Crystal Filters Having Improved Intermodulation & Power Handling Capability.....491
M.D. Howard, R.C. Smythe and P.E. Morley, Piezo Technology, Inc.

Monolithic Filters Using Ion Etched Fundamental Mode Resonators Between 60 and 100 MHz.....504
J. Brauge, M. Fragneau and J.P. Aubry, CEPE, France

Scattering Parameters of Electrically Loaded SAW Group Type Unidirectional Transducers.....514
A.R. Reddy, Indian Telephone Industries, R&D
S.K. Lahiri, Department of Electronics and ECE, India

A New All Quartz Package for SAW Devices.....519
T.E. Parker, J. Callerame and G.K. Montress, Raytheon Research Division

TABLE OF CONTENTS (Continued)

CRYSTAL MEASUREMENTS AND TRANSDUCERS

A System for Precision Parameter Measurements on Quartz Crystal Resonators and Bipoles.....527
R.C. Peach and S.E. Morris, GEC Research Limited, United Kingdom

Investigation of Quartz Crystal Thickness Shear and Twist Modes Using a New Noninterferometric Laser Speckle Measurement Method.....535
S. Hertl, E. Benes, L. Wimmer, and M. Schmid, Technische Universität Wien, Austria

A Study of Flexural, Anharmonic and Thickness-Shear Modes of Vibrations in Quartz Resonators Using Scanning Electron Microscope.....544
H. Bahadur, Oklahoma State University, R. Parshad, National Physical Laboratory, India

Enhanced Composite Resonator Analysis and its Application to the Quartz Crystal Microbalance.....556
E. Benes, Institut für Allgemeine Physik, Technische Universität Wien, Austria
K.C. Harms, AVL Gesellschaft für Verbrennungskraftmaschinen und MeBtechnik m.b.H., Austria
G. Thorn, Leybold-Heraeus, G.m.b.H., Federal Republic of Germany

Transient Analysis of Piezoelectric Transducer Response.....571
A.H. Banah, King Abdulaziz University, Saudi Arabia

A Filled Thermal System Utilizing A Gas Density Sensing Quartz Crystal Tuning Fork.....575
R.W. Ward, E.P. EerNisse, Quartztronics, Inc.

AUTHOR INDEX.....581

SPECIFICATIONS AND STANDARDS GERMANE TO FREQUENCY CONTROL.....583

PROCEEDINGS AVAILABILITY INFORMATION.....585

1985 AWARD WINNERS

THE CADY AWARD

The Cady Award was presented to John A. Kusters, Hewlett-Packard Co., for his contributions to the development of SC-cut and other doubly rotated quartz resonators. The award was presented by Errol EerNisse, Quartex, Inc.

THE RABI AWARD

The Rabi Award was presented to Professor Norman Ramsey, Harvard University, for his contributions to the development of atomic frequency standards. The award was presented by Robert Vessot, Smithsonian Astrophysical Observatory.

THE SAWYER AWARD

The Sawyer Award was presented to Thrygve Meeker, AT&T Bell Laboratories for his contributions to the theory and design of piezoelectric quartz devices. The award was presented by Warren Smith, AT&T Bell Laboratories and Baldwin Sawyer of Sawyer Research Products, Inc.

Samuel R. Stein, Technical Program Chairman; Thrygve Meeker, Sawyer Award winner; Prof. Norman Ramsey, Rabi Award winner; John A. Kusters, Cady Award winner; and John R. Vig, General Chairman, at the award dinner.

NOMINATING PROCEDURES FOR THE CADY & RABI AWARDS

INTRODUCTION

In early 1983, the program committee voted to create two awards that recognize outstanding contributions in all fields covered by the Annual Frequency Control Symposium. One, the Cady Award, named after Walter Guyton Cady, is to recognize outstanding contributions related to piezoelectric frequency control devices. The other, the Rabi Award, named after Professor I. I. Rabi, is to recognize outstanding contributions related to fields such as atomic and molecular frequency standards, time transfer, and frequency and time metrology. Each award consists of $500.00, and a limited edition original print and certificate in a leather binder. The awards are presented to the recipients at the Symposium Award Dinner.

ELIGIBILITY CRITERIA

Either award is open to any worker in any of the fields of endeavor traditionally associated with the Annual Frequency Control Symposium. The nominee should be responsible for significant contributions of a technical nature to the field selected. No posthumous awards will be made. The time span over which the contributions have occurred is not limited.

The significance of the contributions may be measured, in part by:

- The degree of initiative, ingenuity and creativity displayed.
- The quality of the work and degree of success attained.
- The overall importance of the work and impact of the frequency control and associated communities.

NOMINATIONS

Anyone may nominate another for either award. Each nomination should include the following:

1. Name of nominee
2. Current address of nominee.
3. Name of award for which nominated.
4. Description of accomplishments, including:
    a. Initiative, ingenuity, and creativity;
    b. Quality and degree of success;
    c. Importance of the work and impact on the frequency control and associated communities;
    d. Proposed citation, one or two sentences (see examples on the first pages of any Proceedings since 1983, under "Awards".)

It is strongly suggested that the nomination not exceed two typewritten pages. Nominations for the award should be submitted, by the date announced for the submission of summaries, to the Chairman of the Technical Program Committee:

Dr. Leonard S. Cutler
Hewlett-Packard Laboratories
1651 Page Mill Road
Palo Alto, CA 94304

SELECTION OF RECIPIENTS

The selection of the recipient for each award will be made by the Technical Program Committee during the spring meeting. The decision of the committee is final. If, in the opinion of the committee, no suitable nominee exists, no award will be given.

A BEAM REVERSAL EXPERIMENT ON NBS-6 PRIMARY Cs STANDARD INCLUDING RABI PULLING EVALUATION

A. DeMarchi*, G. D. Rovera**, R. Drullinger°, and D. A. Howe°

Abstract

An improvement in the evaluation of the Cs beam primary frequency standard NBS-6 is being attempted through a reevaluation of Rabi pulling using a recently published theory. Time of flight distribution measurements and frequency measurements at various C-field values have been performed in both beam directions. This allows us to model Rabi pulling and hence more clearly study other systematic effects.

Results presented here show the relative magnitude of the two effects (Rabi pulling and cavity phase shift). Zero crossings of Rabi pulling are identified and demonstrated to be near the C-field settings which give minimum power dependence. The results are preliminary in the sense that full evaluation of Rabi pulling should include analysis of the effect in the two beam directions as a function of source and detector position, in order to better guarantee configuration repeatability. It is expected that total evaluation accuracy of NBS-6 might improve to a few times 10^{-14} once the indicated experiments are properly performed and analyzed.

Introduction

The end-to-end cavity phase shift is among the major uncertainty sources in primary Cs beam frequency standards. It comes from the residual asymmetry existing in the most carefully built Ramsey cavity and its value is given by the formula [1]:

$$\frac{\delta\nu}{\nu} = \frac{\ell}{2\pi L} \cdot \frac{\langle \delta_\phi(\tau) \rangle}{\langle \tau \rangle} \qquad (1)$$

where L and ℓ are the lengths of Ramsey and Rabi cavities, as shown in fig. 1, τ is the time of flight (TOF) through a single Rabi cavity for a given velocity group and $\delta_\phi(\tau)$ is the average phase difference of the microwave interrogating field experienced by that group between first and second cavity. The size of this effect can be several 10^{-13} in well designed long cavities [2].

Various techniques have been proposed to avoid or minimize cavity phase shift, ranging from the use of superconducting cavities [3] to the dual frequency technique [4]. However the most immediate approach to the problem is still to measure the effect. Several methods have been devised for this purpose. All are based on varying in a controlled way one of the accessible variables affecting its value in (1), namely $\langle\tau\rangle$ or (the sign of) $\langle\delta_\phi(\tau)\rangle$.

*Istituto Elettrotecnico Nazionale G. Ferraris, Torino, Italy.

**ITALCABLE Grant.

°National Bureau of Standards, Time and Frequency Division, Boulder, Colorado

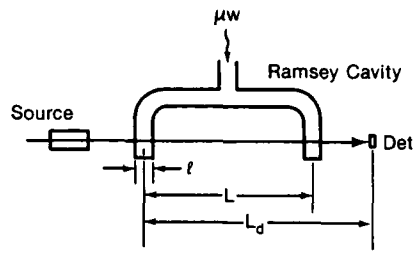


Fig. 1. Schematic and geometrical definitions in an atomic beam with a Ramsey cavity. In NBS-6 ℓ=1cm; L=3.75m; and L_d=4.54m.

In the power shift method the effective $\langle\tau\rangle$ is changed by the microwave power velocity selection, based on the fact that the transition probability depends on τ. Averaging over τ is necessary due to the large TOF window. This method is best applied if coupled with beam reversal [5]. In the pulse method [6,7] the same effect is obtained more directly by introducing a very narrow TOF window with a strobed microwave power, based on the fact that atoms which are excited in both cavities have twice the transition probability of those which are excited in one only.

In the beam reversal method the order of the cavities is changed, which gives in principle a switch in sign for δ_ϕ and a frequency value symmetrically shifted from the unperturbed frequency.

Unfortunately all these experiments are not clean in the sense that other effects of the same order of magnitude may well come into the picture with an intrinsic connection and affect the results. In particular this is true of distributed cavity phase shift [8,9,10] if the beam does not skim the cavity ends, and of Rabi pulling [11]. Both these effects can affect the two methods based on variations of $\langle\tau\rangle$, through the velocity dependence of trajectories (and hence of δ_ϕ) and of the shape of Rabi wings; the power shift method in addition is difficult to interpret because of the dependence of Rabi pulling on microwave power [11], unless used near a zero crossing of the latter. Rabi pulling perturbs also the beam reversal method, due to the fact that TOF distributions are not the same in the two beam directions and may therefore cause different pulling even at the same C-field.

In this paper we report on a beam reversal experiment on the NBS-6 primary cesium standard in which the frequency was measured in both beam directions at various settings of the C-field in order to separate the effect of Rabi pulling from that of cavity phase shift. The latter in fact is expected not to depend on C-field.

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

TOF distribution measurements were performed in both beam directions and corrected for the transfer function of the detector system in order to compare calculated Rabi pulling curves to experimental results. The purpose of this experiment was to obtain high accuracy in the evaluation of phase shift through best fit of theoretical curves and experimental points to yield high accuracy in positioning the null Rabi pulling line, and to gain information about the zero crossings of Rabi pulling. The best operating points, where the power dependence of all residual effects compensate to yield insensitivity to microwave power, are expected to be found near the C-field settings for which the zero crossings occur. Only $\Delta m = 0$ transitions were considered in the analysis, however it appears that intermediate $\Delta m = \pm 1$ transitions may be just as important for Rabi pulling despite their small size, since they are nearer to the central clock transition.

TOF Distributions

Evaluation of TOF distributions in states m = 0, ± 1 was carried out for both beam directions with a time of flight method using L_d as a length basis. Microwave pulses 1.3 ms long were used at a 90 ms repetition rate and the detected signal was then processed in a multichannel signal averager for various microwave power levels for each of the analyzed m sublevels. In fig. 2 the envelopes of these responses are reported for the three central sublevels in the west-to-east beam direction (WE). First and second peaks come from the second and first cavity.

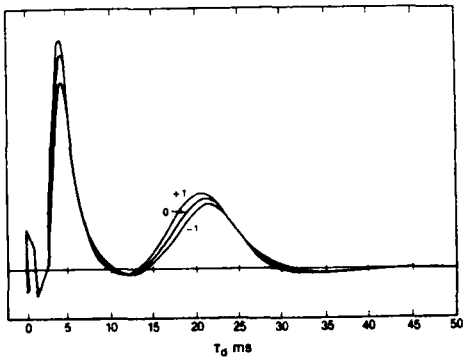


Fig. 2. Envelopes of detected pulse responses at various microwave power levels for the three central $\Delta m=0$ transitions ($m=0,\pm1$). The small feature on the left is leakage into the detector of the microwave pulsing signal.

The waveform distortion effect of the detector system's transfer function is evident in the non ideal ratio between the heights and delays from the microwave pulse of the two peaks. Ideally this ratio should be $\xi = (1-L/L_d)^{-1} = 5.75$ as can be seen from Fig. 1, but the limited bandwidth of the system tends to cut the sharp first peak and alter the relative positioning of the two by introducing a delay which is more important for the sharper peak. Unfortunately the transfer function cannot be easily measured because it incorporates parts of the detection system which are not accessible for controlled excitation from the outside of the tube. A computer program was assembled to perform the deconvolution from the detector system's transfer function. An input waveform $x(t_d)$ of the form

$$x(\tau_d) = \xi\rho(\xi\tau_d) + \rho(\tau_d) \qquad (2)$$

where $\rho(\tau_d)$ is the unknown TOF distribution, was assumed to be the beam response to the microwave pulse, and the calculated output waveform from the assumed detector system was compared to the experimental waveform. Optimization of a smoothed least square fit for this comparison for different zero-pole configurations of the detector system yielded consistently the same detector transfer function for all measured waveforms in both directions, and therefore a high degree of confidence in the capability of the computer program to find real TOF distributions. In fig. 3 the deconvolved distributions for the m = 0 level

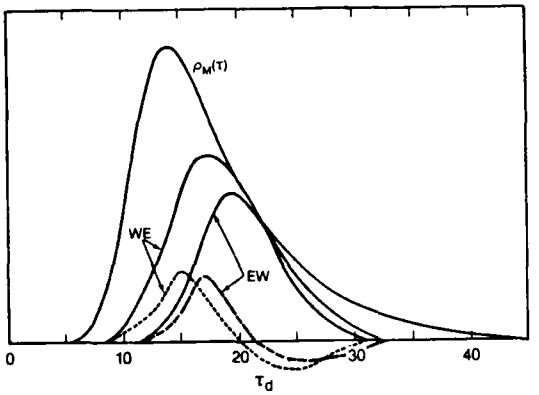


Fig. 3. Real TOF distributions $\rho_o(\tau_d)$ for the two beam directions (solid lines) obtained by deconvolution from the detector system's transfer function are compared to the Maxwellian distribution $\rho_M(\tau_d)$. The broken lines are distributions $\Delta_1(\tau_d)$ as obtained using (3).

in both directions are reported. Comparison with the Maxwellian distribution in the beam, calculated for the oven temperature of 79°C used in the experiments, enables one to recover the relative heights of the two distributions (consistent with measured beam intensities) and to calculate the beam optics selection functions $G_o(\tau_d)$ for the two beam directions, which are shown in Fig. 4. In fig. 3

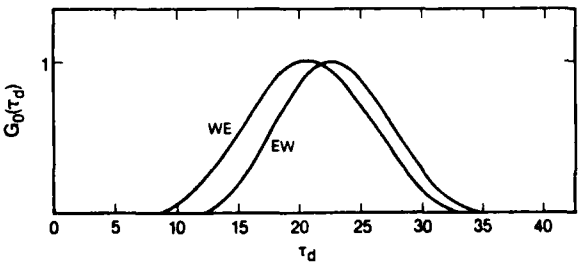


Fig. 4. Beam optics appear different in the two beam directions. TOF windows for the m=0 sublevel as derived from Fig. 3 are shown here.

the differential distributions $\Delta_1(\tau_d) = \rho_1(\tau_d) - \rho_{-1}(\tau_d)$ are also shown for the two directions. Because of the more acceptable signal-to-noise ratio in the acquisition of ρ_0 with respect to the difference Δ, the latter was calculated from the approximated formula

$$\Delta_1(\tau_d) = \varepsilon_1 \tau_d \frac{dG_c}{d\tau_d} \rho_M(\tau_d) \qquad (3)$$

in which ε_1 was evaluated from the asymmetry in the ± 1 transitions saturated intensity [11] using the corrected distributions for evaluation of $\langle\tau\rangle$, and from the relative displacements of calculated $G_{\pm1}(\tau_d)$ with respect to $G_0(\tau_d)$. It turned out that all data were consistent if ε_1 was assumed to be 3% in both directions, which corresponds to an average field of 0.4 T in the deflecting magnets. This value is consistent with the published data of 0.9 T at the tip of the convex pole-piece [12]. The Δ distributions calculated in this way and reported in Fig. 3 were therefore used for the calculation of the Rabi-pulling effect by the formulas given in [11].

Frequency Measurements

Averaging times of 6 to 12 hours were used in frequency measurements, which should in principle provide an uncertainty of few parts in 10^{14} from the NBS-6 stability contribution at the low source temperature used in the experiment. The measurements were taken by comparison with SPHM-4 passive hydrogen maser, which is stable in the 10^{-14} region for that length of time [13], and then referred on a two day basis to the frequency of AT1, which is constructed from an average of commercial standards and is stable in the low 10^{-14} from two days on. The 1σ measurement uncertainty was estimated to be ± 5 x 10^{-14}, at least for the points with small microwave power dependence.

Two series of measurements were taken in the two beam directions at various C-field and microwave power settings; the results for optimum microwave power are reported as a function of the Zeeman frequency of the m = 1 component in fig. 5. Unfortunately the time available for these measurements did not allow us to complete the set at low C fields for the EW direction, it is however evident that the general behavior of the experimental points agrees with the shape of the theoretically calculated curves.

The power dependence of a point near a zero crossing of Rabi pulling and near a maximum are shown in fig. 6 for the WE beam direction.

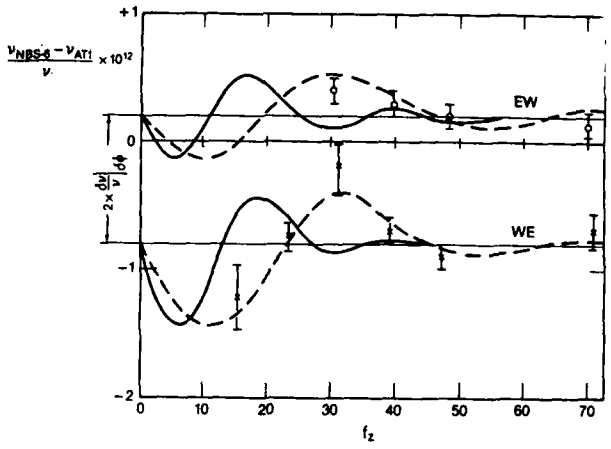


Fig. 5. Frequency measurements, corrected for C-field shift, are reported here for the two directions (0=EW, x=WE) as a function of the Zeeman frequency of the m=±1 Δm=0 transition. The solid curves are Rabi-pulling curves calculated from the TOF distributions of Fig. 3. It appears that a stretching of these curves by a factor of approximately 2 is necessary to match experimental points. It is suggested that the spurious Δm=±1 transitions may have an important role in this disagreement.

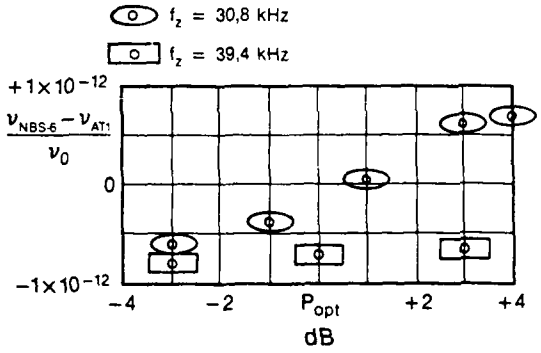


Fig. 6. Microwave power dependence of frequency for two C-field settings in the WE direction. The standard shows a much lower power dependence near the zero crossings of Rabi pulling.

Effect of the Δm = ± 1 Transitions

In fig. 5 the full curves are the result of calculations based on the measured TOF distributions, corrected for the detector system's transfer function. It turns out that a stretching of these curves (shrinking of the Zeeman axis) by about a factor of 2 would provide a very good match of theoretical curves and experimental points. The broken curves in fig. 5 have been obtained by stretching the solid curves by a factor of 1.85.

This fact seems to suggest that the Δm = ± 1 field dependent transitions, which are in principle small because they are excited by the unwanted oscillating B field component perpendicu-

lar to C-field, but are twice as near to the (0,0) transition than the (1,1) and (-1,-1), may have a strong effect due to the differential slope of their wings.

This hypothesis seems surprising if the size of these transitions is observed at a high C-field at which they can be isolated, as in fig. 7. However an actual comparison of measured lineshapes shows in fig. 8 that it may well be the case that these $\Delta m = \pm 1$ transitions have a greater pulling effect than the neighboring $\Delta m = 0$ transitions. Further analysis and experimental data are clearly necessary to realize an accurate description of this aspect of the problem.

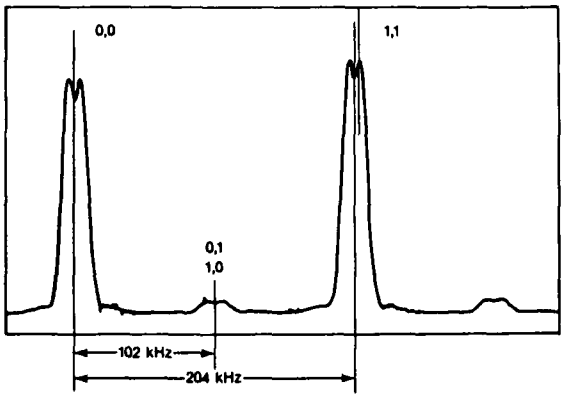


Fig. 7. Tube response at high C-field. Power is about 1 dB above optimum.

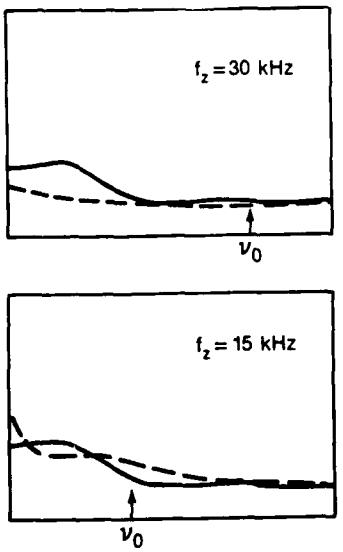


Fig. 8. Comparison of the wings of a $\Delta m = \pm 1$ transition (solid line) and a $\Delta m = 0$ transition (broken line) as they appear superimposed near center frequency for the indicated first side line Zeeman frequency.

Conclusions

Although not everything is understood in the results of the measurements presented in this paper, it is clear from fig. 5 that an accurate positioning of the zero Rabi pulling can be achieved in both directions of the beam, resulting in a value for the end-to-end cavity phase shift, $\delta\phi$, which appears to have an uncertainty of less than $\pm 5 \times 10^{-14}$. We will therefore state as our result

$$\frac{\delta\nu}{\nu}\Big|_{\delta_\phi} = (4.5 \pm 0.5) \times 10^{-13} \qquad (4)$$

This is quite consistent with previously published recent results for the cavity phase shift of NBS-6 shown in figure 9.

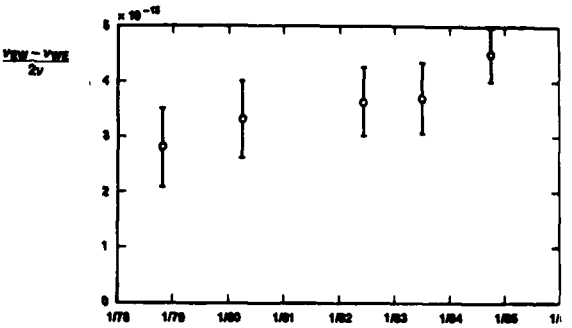


Fig. 9. Recent published results for the cavity phase shift of NBS-6.

The change in the measurements with time has been often interpreted as a drift of the phase difference itself possibly due to unequal cesium deposition. Although this is quite possible we suggest here that care must be taken in drawing conclusions because the Rabi pulling effect has not been adequately accounted for in the past. However, it turns out that all these measurements have been taken at C-field values for which the Rabi pulling is relatively small.

The Rabi pulling is strongly dependent on microwave power at most C-field values and the field values where this dependence vanishes may depend on anything affecting beam optics: from the field strength of selecting magnets (and their demagnetization with time), to source and detector position (with each beam reversal), to changes introduced by major redesign of tube ends as in the modification of NBS-5 into NBS-6. It can be expected that once the effect of all these variables on the zero crossing positions of the Rabi pulling is understood and repeatability can be guaranteed to the necessary degree, the documented accuracy claims for NBS-6 may improve to less than 10^{-13} and possibly to few parts in 10^{-14}, the residual uncertainties being probably due in such a case to electronics and C-field inhomogeneity. The long term stability should also benefit from operation at a zero pulling point because of the small frequency dependence on microwave power.

Acknowledgments

Two of the authors (A. De Marchi and G. D. Rovera) would like to thank the National Bureau of Standards for making this experiment possible by allowing access to NBS-6.

References

[1] R. L. Lacey, Proc. 22nd Annual Symp. on Freq. Contr. 545-558 (1968).
[2] A. De Marchi, G. P. Bava, Metrologia 20, 33-36 (1984).
[3] D. J. Wineland, Metrologia 13, 121-123 (1977).
[4] S. Jarvis, Jr., D. J. Wineland, H. Hellwig, J. Appl. Phys. 48, 5336-5337 (1977).
[5] H. Hellwig, J. A. Barnes, D. J. Glaze, Proc 27th Annual Symp. on Freq. Contr., 309-312 (1973).
[6] H. Hellwig, S. Jarvis, D. Halford, H. E. Bell, Metrologia 9, 107-112 (1973).
[7] D. A. Howe, H. E. Bell, H. Hellwig, A. De Marchi, Proc 28th Annual Symp. on Freq. Contr., 363-372 (1974).
[8] S. Jarvis, Jr., NBS Tech. Note 660, NBS (1975).
[9] D. W. Allan, H. Hellwig, S. Jarvis, Jr., D. A. Howe, R. A. Garvey, Proc. 31st Annual Symp. on Freq. Contr., 555-561 (1977).
[10] A. Bauch, T. Heindorff, IEEE Trans. Instrum. Meas., IM-34, 136-138 (1985).
[11] A. De Marchi, G. D. Rovera, A. Premoli, Metrologia 20, 37-67 (1984).
[12] D. J. Glaze, H. Hellwig, D. W. Allan, S. Jarvis, Jr, Metrologia 13, 19-28 (1977).
[13] F. L. Walls, K. B. Persson, Proc. 38th Annual Symp. on Freq. Contr., 416-419 (1984).

EXPERIMENTAL STUDIES ON MAJORANA TRANSITIONS IN A CS ATOMIC BEAM FREQUENCY STANDARD

A. Bauch and T. Heindorff
Phys.-Techn. Bundesanstalt, Braunschweig, FRG

Abstract

In the PTB's experimental frequency standard CSX, provided with auxiliary solenoids along the beam path, the influence of Majorana transitions on the beam current, the atomic velocity distribution and the clock frequency was measured. The observed effects - the number and the velocity of atoms having undergone a σ transition is substantially altered - can be explained by a theory based on Majorana's formalism. Up to now, neither a theoretical description of the frequency shifts of some 10^{-11} found in experiment exists, nor the cause of the oscillations of these shifts can be given.

1. Introduction

It is well known that the so-called "Majorana transitions" between the Zeeman levels of the hyperfine states F = 4 and F = 3 are induced in Cs atoms moving in an inhomogeneous magnetic field. In cesium beam frequency standards this is probably due to the imperfections of the quantization field, especially in the vicinity of the state-selecting magnets. Majorana transitions are regarded as being harmful to the performance of the standard and are suppressed by trimming the magnetic fields. They are usually taken into account in the error budgets with a tiny contribution. On the other hand, some results have been reported which indicate that the magnetic field distribution along the path of the atoms caused frequency shifts whose magnitude cannot be explained satisfactorily /1,2,3/.

For almost a year, research work on Majorana transitions has been in progress at the PTB. In the course of this time, we have collected a great variety of results, some of them quite unexpected. We are still far from achieving a complete understanding of this field, therefore this paper is intended to be a report on the current state of our work.

2. Apparatus

Up to now, the experiments have been carried out mainly using the CSX apparatus, the PTB's "experimental frequency standard". In Fig. 1 a plot of its vertical section is given. The arrangement is cylindrically symmetric. Four-pole and six-pole magnets are used for state selection. The clock transition is detected by the flop-out method. The longitudinal C field is generated by a cylindrical coil which extends over the magnet systems. It is surrounded by a coaxial Mumetal shielding.

The design results in an excellent homogeneity of the C field. We do not consider there is any probability of Majorana transitions within the interaction region so far. To be able to simulate field variations possibly present in frequency standards we mounted two pairs of small solenoids, one close to the exit of the polarizer magnet and one close to the entrance of the analyzer magnet. Thus two static magnetic fields may be generated at the two places S1 and S2 and the influence of field variations, which are well known from the geometry of the coils and the currents used may be studied.

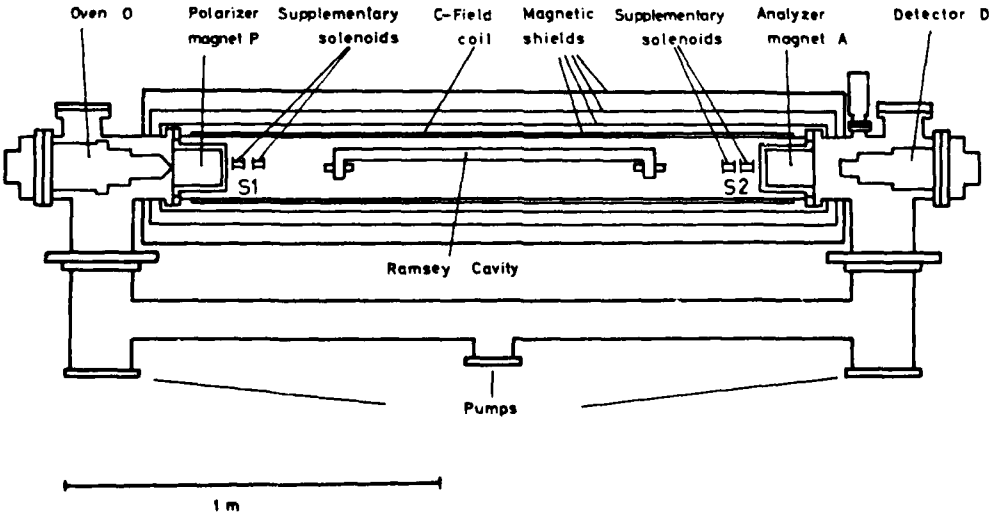


Fig. 1: Schematic diagram of the CSX apparatus (vertical section).

CH2186-0/85/0000-0008$1.00©1985IEEE

A more detailed description of the CSX can be found in /4/. A table of its characteristic data is given below.

Table I. Characteristic CSX operating data.

Diameter of atomic beam	1.2 mm
Total beam current	100 pA
Ramsey peak amplitude	13 pA
Interaction length	L = 791 mm
Mean atomic velocity	416 m/s
Half-width of velocity distribution	60 m/s
Mean C field strength	8.36 A/m
Magnetic field inhomogeneity in the cavity region	0.004 A/m

3. Rotation of the quantization field

The magnetic field seen by the atoms is a superposition of

- the residual earth magnetic field
- the stray field of the magnets
- the C field
- the auxiliary fields.

The first two contributions were measured before the completion of the CSX. With a suitable choice of the C field and the auxiliary fields we can achieve a smooth variation of the longitudinal component of the total magnetic field, designated H_z. The direction of H_z is fixed parallel or antiparallel to the beam direction. This is the CSX's regular mode of operation.

A certain combination of the above fields leads to a zero point of H_z on the axis. The quantization axis changes its direction and is rotated by π with reference to the beam direction. We refer to this field configuration as a "zero point" in the following sections although almost all atoms are subject to a small transversal field. This is just the situation treated by Majorana in 1932 /5/.
As the beam is confined to a small region around the axis of the apparatus, transversal fields are rather small, thus from Majorana's work we conclude that most of the atoms will undergo a spin-flip: The quantum number m_F, defined with reference to the local field axis changes its sign.

How can this be verified? The imaging properties of the state-selector magnets are dependent on the atoms' quantum numbers F, m_F and their velocities v, therefore the mean velocities of the beam atoms contributing to the seven σ transitions ($\Delta F = 1$, $\Delta m_F = 0$) differ, ranging from 360 m/s to 465 m/s in the CSX. In the regular mode of operation the atoms in the state F = 4, $m_F = 3$ are those having the highest mean velocity as expected. It is determined from the linewidth of the (4,+3)-(3,+3) transition. If the atoms undergo a spin-flip before entering the cavity, this linewidth is no longer the largest one. The atoms' quantum numbers are interchanged so that now, the linewidth of the (4,+3)-(3,+3) transition is the smallest of all (on the right in Fig. 2).

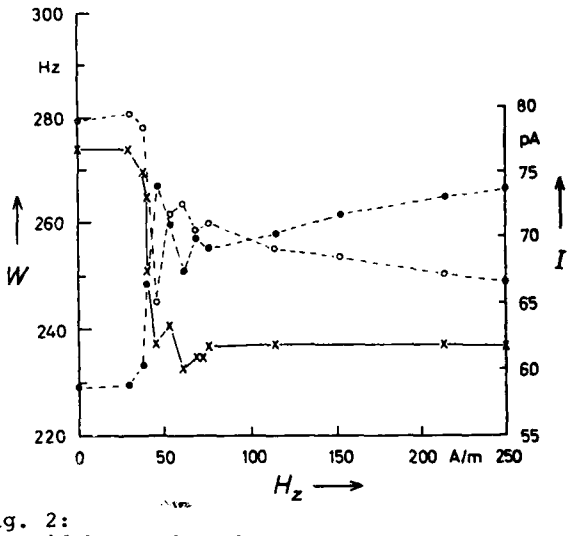


Fig. 2:
Linewidth W and Cs beam current I as a function of the longitudinal component H_z of the auxiliary fields. o Linewidth of (4,3) - (3,3) transition, • linewidth of (4,-3) - (3,-3) transition, X Cs beam current. The connecting lines are shown only for illustration.

Gradually introducing this zero point by changing the currents in the coils, we measured the linewidth as a function of these currents. The result is shown in Fig. 2.

A reduction of the Cs beam current by an amount of typically 25 % was also observed. This can be easily understood if the velocity selectivity of the analyzer magnet is taken into account. If the atoms undergo a spin-flip between P and A there is a mismatch of the atoms' quantum numbers m_F and their velocities compared with the regular mode of operation. Thus a certain number of atoms are no longer focussed on the detector in addition to those being flipped from the state (4,4) to the state (4,-4). As in the CSX this effect makes an appreciable contribution only in a flop-out design and only if the velocity distributions of the atoms in the different m_F states are quite narrow.

Of course, the terminology "spin-flip" is only an abridgement of what actually happens. If this were strictly true, the clock transition would not be affected at all, because there is no difference between $m_F = \pm 0$.

Fig. 3 shows velocity distributions of the atoms contributing to the clock transition. The areas are normalized with respect to the absolute amplitude of the resonance signal. The narrow distribution is the Fourier transform of a Ramsey pattern recorded in the regular mode of operation. The wider one is obtained if a zero point is introduced at S1. The effect of transitions between the Zeeman sublevels due to transversal fields can clearly be seen.

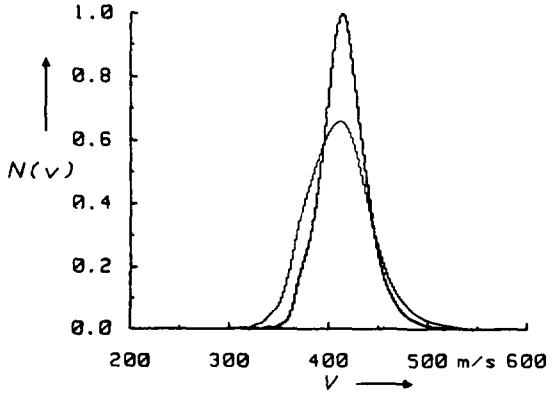


Fig. 3:
Velocity distribution of (4,0) atoms. The areas under the curves are proportional to the Ramsey signal amplitude in normal operation of the CSX (narrow distribution) and in spin-flip mode at S1 (wide distribution).

Is the clock transition frequency affected in this case? Within the statistical uncertainty of $5 \cdot 10^{-14}$ the frequency values measured in the two modes of operation agree with each other.

4. Double rotation of the quantization axis

With a pair of coils mounted at S1 or S2 two zero points of H_z can be generated. A plot of the computed auxiliary field H(z) is given in Fig. 4. The position of the coils is indicated and the two zero points of H(z) can be read out: the first one between the two auxiliary fields, the other one between the second auxiliary field and the C field, which is assumed to be constant. In general the computed values of the fields should not be taken too literally as no stray fields are included in the computation. In the dashed part the fields are only an estimate, as the stray fields of the magnets will predominate in this region.

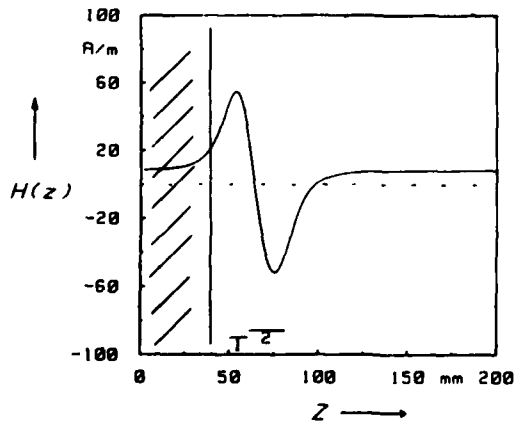


Fig. 4:
Calculated Z dependence of the longitudinal auxiliary field H(z) generated by two solenoids 1,2 mounted at S1 or S2.

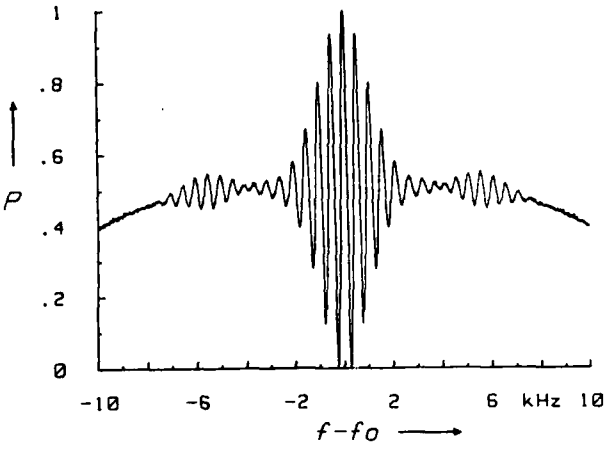


Fig. 5:
Normalized Ramsey pattern of the (4,0)-(3,0) resonance with additional modulations produced with two zero points of the magnetic field along the Cs beam path.

At certain settings of the values of the fields we observe additional modulations on the Rabi pedestal of the transitions. An example is given in Fig. 5. A measurement of the velocity distribution by the pulse method /6/, /7/ and a Fourier analysis revealed structures in the velocity distribution. They change appreciably if the current in coil 2, for example, is changed by small amounts. Fig. 6 shows evaluated velocity distributions. The current in coil 2 is raised in steps of 1 % of a suitable initial value. The dip structures are shifted from left to right and new structures periodically arise.

We concluded that a velocity-selective transition process existed affecting the atoms moving through the field structure. No analytical expression can be given of the transition probability in this special case. Following an idea of Schröder, Baum /8/ we made a numerical analysis integrating the time-dependent Schrödinger equation stepwise. The Hamiltonian is exactly the same as used by Majorana, and the magnetic field is given as a function of the Z coordinate in discrete steps of 0.1 mm. We considered a two-state system and computed the transition probability between the two states in the approximation of small changes of the state populations. A typical result is given in Fig. 7. As a function of the atoms' velocity v the transition probability P, equivalent to $\cos^2 \alpha/2$ in Majorana's notation, is plotted for a certain setting of the fields.

The velocity distributions of the atoms in the state $(4, m_F)$ may be determined by convoluting the different velocity distributions of the atoms with the transition probabilities computed from the above result $P = \cos^2 \alpha/2$ using the Majorana formula (4) in /4/.

If the current in coil 2 is raised stepwise, the oscillatory function of P is shifted, until the shift almost corresponds to one

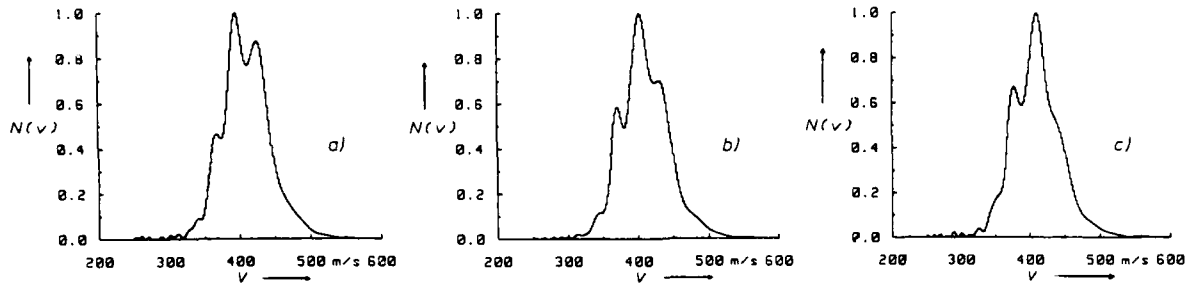


Fig. 6: Selection of velocity distributions of (4,0) atoms. The different structures are the result of slight variation of the current in coil 2 in a magnetic field configuration as shown in Fig. 4 (see text).

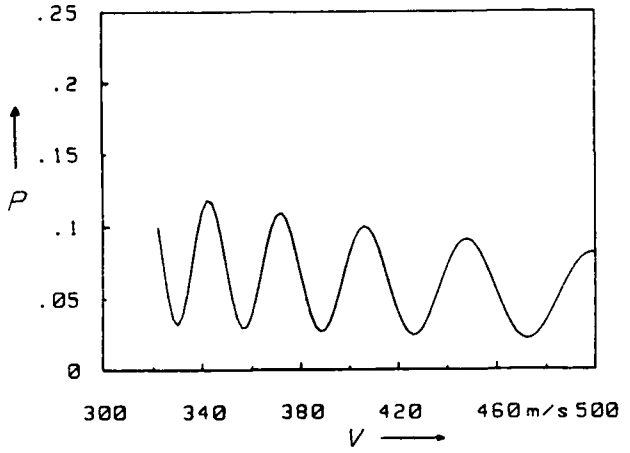


Fig. 7:
Transition probability P in a two-state system as a function of the velocity v of the atoms moving in a magnetic field configuration as shown in Fig. 4.

period. It is just this which happens whenever the number of precessions of the spin around the field axis between the two zero points is increased by one.

If the current is raised appreciably, the number of precessions for a given velocity increases and to the same extent the periodicity of P gets narrower. Thus the structures in the velocity distributions become more and more complicated until they are obliterated. This was verified in the experiment.

Finally, it is of great interest to look at the clock transition frequency. This is affected severely as can be seen in Fig. 8. Relative shifts of up to $\pm$ $5 \cdot 10^{-11}$ can be unambiguously reproduced. In this figure the relative frequency difference between the CSX and a commercial atomic clock is plotted as a function of the field generated by coil 2. The periodicity found here is similar to that derived from Fig. 7, if P is computed as a function of the field H_z where the (fixed) mean velocity of the atoms is used in the computation.

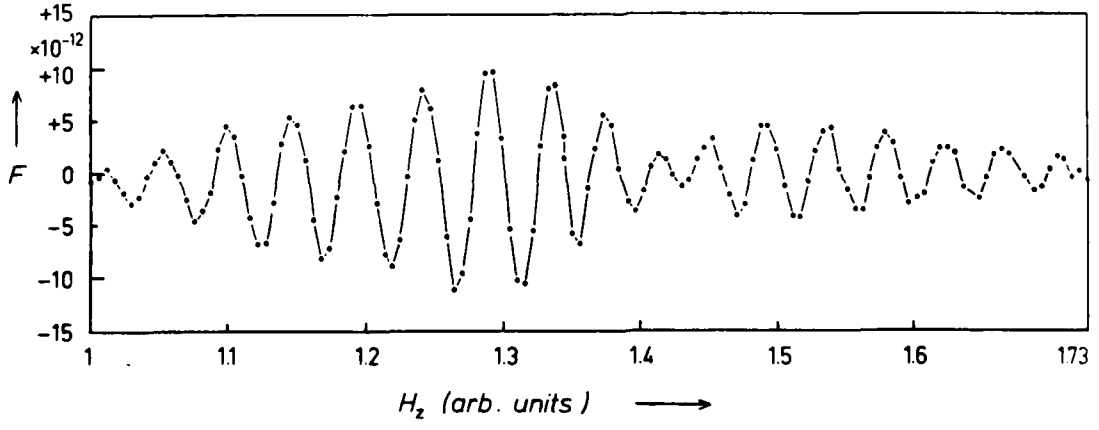


Fig. 8: Dependence of the relative CSX frequency F = F(CSX-R) on the magnetic field of coil 2 in a two-zero-point configuration, shown in Fig. 4. R: reference clock. Connecting lines are given only for clarification.

At present we are engaged in a theoretical treatment of the atomic transition process involving interaction terms other than the usual radio-frequency interaction. We surmise that a coherent population within the Zeeman sublevel is created which is coupled in a still unknown manner to the rf interaction. No further results can be given at present.

5. Conclusion

It has been experimentally shown that certain variations of the magnetic fields along the beam path of an atomic frequency standard may cause severe shifts of the clock transition frequency. These variations may be avoided by a proper trimming of the magnetic fields near the state-selecting magnets. Nevertheless, it should be of the greatest interest to elucidate the mechanism involved so that a complete estimate of the error budget of the standard may be given.

This work was financially supported by the Bundesministerium für Wirtschaft.

The authors are greatly indebted to Messrs. Raith, Baum, Steidl and Sillmen, University of Bielefeld, for the loan of a computer program and for fruitful discussions concerning the calculations.

References

/1/ D.W. Allan, H. Hellwig, S. Jarvis, D.A. Howe, R.H. Garvey, "Some causes and cures of frequency instabilities (drift & noise) in Cesium beam frequency standards", in Proc. 31st. Annual Symp. on Frequency Control, 1977, pp. 555-561

/2/ G. Becker, "Research on Cs beam frequency standards at the PTB: Beam optics, Majorana transitions", IEEE Trans. Instr. Meas., vol. IM-27, pp. 319-325, 1978

/3/ S. Urabe, K. Nakagiri, Y. Ohta, M. Kobayashi, Y. Saburi, "Majorana effect on atomic frequency standards", IEEE Trans. Instr. Meas., vol. IM-29, pp. 304-310, 1980

/4/ T. Heindorff, A. Bauch, R. Schröder, "Performance of the new PTB experimental atomic beam resonance apparatus CSX", PTB-Mitt., vol. 94, pp. 318-326, 1984

/5/ E. Majorana, "Atomi orientati in campo magnetico variabile", Nuovo Cimento, vol. 9, pp. 43-50, 1932

/6/ H. Hellwig, S. Jarvis,Jr., D.J. Glaze, D. Halford, H.E. Bell, "Time domain velocity selection modulation as a tool to evaluate Cesium beam tubes", in Proc. 27th Annual Symp. on Frequency Control, 1973, pp. 357-366

/7/ A. Bauch, T. Heindorff, "Leistungsabhängigkeit, Magnetfeld- und Dopplerkorrektion am Frequenznormal CSX der PTB", PTB-Mitt., vol. 95, pp. 93-97, 1985

/8/ W. Schröder, G. Baum, "A spin flipper for reversal of polarization in a thermal atomic beam", J. Phys. E: Sci. Instr., vol. 16, pp. 52-56, 1983

A RECIRCULATING OVEN FOR ATOMIC BEAM FREQUENCY STANDARDS*

R. E. Drullinger, D. J. Glaze, and D. B. Sullivan

Time and Frequency Division
National Bureau of Standards
Boulder, Colorado 80303

Summary

This paper describes a simple recirculating oven which produces an atomic beam which can be better collimated than that from a conventional oven with equivalent collimation ratio. The oven is spill proof and requires only modest power for operation. Under suitable conditions the total beam flux can be significantly less than for conventional cesium ovens. This translates into more efficient use of the cesium charge and less contamination of the beam tube.

Introduction

While high intensity molecular beam sources have been advanced considerably during the past several decades, the basic design of the lower intensity sources for atomic beam (cesium) frequency standards have remained largely fixed. This reflects the fact that such sources have performed well and development work has naturally focused on other aspects of the systems. However, recent demands for standards of higher reliability, longer life, more rugged design, and lower power consumption have caused us to take another look at the beam source and to ask whether design improvements can materially affect any of the performance factors mentioned above. This question led us to the study of a rather simple recirculating oven which departs significantly in design from conventional ovens. While adequate life tests will require more time, performance tests show the oven to have excellent collimation characteristics which bear on both tube contamination and efficiency of use of the cesium charge.

The present design of the oven involves one significant disadvantage: the forward end of the oven must be kept at a temperature near the melting point of cesium to minimize the off-axis flux. It appears that design modifications can ameliorate this disadvantage, but such modification is only conceptual and remains to be demonstrated.

The concept for recirculation in an atomic beam source is not new [1,2]. What we introduce here is a particularly simple implementation of the recirculation concept which also yields a very narrow beam profile as well as several other advantages.

The on-axis flux in atomic beam ovens depends primarily on the source vapor pressure. It is the control of the off-axis flux which differentiates three classes of ovens (underlined below). Ideally, the collimation of the beam should involve simple geometric shadowing, that is, the collimator should just cut off the source emission in undesirable directions. However, it is difficult to achieve this end without introducing certain undesirable characteristics. For example, a carbon collimator can be used to absorb every cesium atom which strikes it, thus achieving the desirable end, but the carbon soon saturates and the cesium deposited on the walls is either re-evaporated or, if it sticks, causes a change in the size or shape of the collimator. This type of oven, which we call a dark-wall oven, demonstrates a key problem in oven design, that is, dealing with the flux which strikes the walls of the collimator.

Conventional ovens use arrays of long narrow tubes to achieve good collimation. The array of narrow tubes allows for higher beam flux and for a good length-to-diameter (collimation) ratio in a short oven. To prevent these tubes from building up deposits of cesium, they are maintained at an elevated temperature and atoms which strike the wall are then re-evaporated with a $\cos(\theta)$ distribution (θ is the angle with respect to the normal to the surface) [3]. This re-emission from the walls broadens the beam profile well beyond that produced by dark-wall ovens [4], but such ovens have nonetheless proven to be very workable in cesium standards. Because all cesium is eventually re-emitted from the walls, we call this type of oven a bright-wall oven.

The recirculating oven described in this paper captures the flux which strikes the collimator walls and returns it through capillary action for re-use by the source. Of course, these walls are saturated with cesium and emit atoms with a $\cos(\theta)$ distribution also, but at a rate which is commensurate with the vapor pressure and temperature at each point on the wall. This emission rate is not a function of the rate of arrival of atoms at the wall, but can be greater or less than that rate depending upon the temperature of the wall. With certain precautions the beam profile from such an oven can be made significantly narrower than that of a bright-wall oven with the same collimation ratio and on-axis intensity, and the integrated emission can be considerably smaller.

Recirculating Oven Design

The recirculating oven is shown schematically in Figure 1a. The entire structure is fabricated of a porous matrix (tungsten or molybdenum) filled with the source liquid (cesium for our case). The forward-facing surface at the source temperature T_2 emits a flux which is collimated by the tube whose temperature varies linearly between this source surface and the front end which is held at a temperature T_1 just above the melting point of cesium (301.5 K). Flux striking the walls is drawn into the porous material and returned to the source region by capillary action. As noted above, any element of the wall will radiate flux at a rate which is commensurate with the wall temperature and the associated vapor pressure at that temperature. The relative merit of the

*Contribution of the National Bureau of Standards, not subject to copyright.

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

recirculating oven thus depends on the relationship between the vapor-pressure profile along the recirculating-oven walls and the same profile for the bright-wall oven. This issue is addressed in the next section.

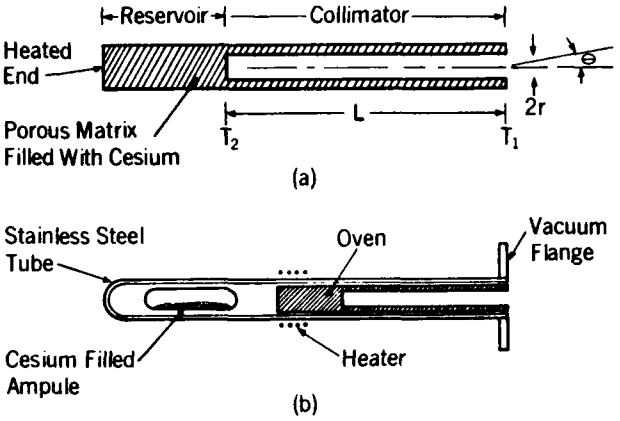


Figure 1. a) Schematic of a recirculating oven. The parameters shown are those used in the calculations. The interior end surface (at the temperature T_2) is the primary source of useful beam flux. b) Simple arrangement for filling and using the recirculating oven. With the heater energized, the cesium ampule is broken by crimping the tube and the liquid metal is absorbed into the porous matrix.

Figure 1b shows the particularly simple experimental arrangement which we have used to mount and charge the porous-tube ovens for performance and life tests. The cesium charge is introduced into the porous matrix by breaking the ampule (crimping the stainless tube) while heating the whole system so that the cesium is molten. The volume of cesium in the ampule is kept just below the volume of the void space in the oven matrix. This assures that all of the cesium (when drawn into the porous matrix) is bound by capillary force and the oven is then "spill-proof". The simple arrangement for mounting the oven has one significant drawback. It does not provide for easy access for measurement of the oven temperatures. T_1 is probably very near the temperature of the vacuum flange to which it is attached, but the temperature along the tube adjacent to the heater places only an upper bound on the temperature T_2. T_2 is most likely lower and we will treat it as an adjustable parameter, although we require that it be close to and bounded by the temperature adjacent to the heater.

The collimator for the particular recirculating oven described in this paper has an internal diameter (2r) of 2 mm. Several ovens of the shape shown in Figure 1a were constructed, but the one described in this paper has different diameter reservoir and collimator sections. The collimator length L is 4.5 cm and the approximate outside diameter is 2.5 mm. The oven was fabricated in 2 pieces. The reservoir is a solid, porous cylinder of diameter 5.3 mm and length 1.3 cm with a short hole in the center of one end which provides for an interference fit with the collimator tube. The collimator is porous, 80%-dense tungsten and the reservoir, also tungsten, is 50% dense.

An important consideration in the final design of an oven of this type is the distribution of cesium within the matrix. Ideally, as the cesium is depleted, the source surface and collimator should remain saturated with the liquid. With careful design it appears that this can be achieved. Where a temperature gradient exists in this capillary matrix, there is a force on the contained liquid in a direction opposite to the gradient, that is, the liquid is forced from the hot toward the cold region. With proper reservoir design and heater placement, the source and collimator can be made to be the last oven portions to "dry out" and the oven output should be quite flat to this point.

Modeling of Performance

The simple concepts used here to model performance are drawn from Ramsey [3] and Giordmaine and Wang [4]. The latter reference develops descriptions of the beam profiles for both dark-wall and bright-wall ovens. The profile for the recirculating oven can be obtained through suitable modification of their equation (16) which describes the emission from each element of the oven wall. The important distinctions between the three classes of ovens (dark-wall, bright-wall and recirculating) involve assumptions about wall emission and these should be restated.

1. Dark-Wall Oven - The assumption for a dark-wall oven is simple: Every atom which strikes a wall is absorbed (no re-emission).

2. Bright-Wall Oven - As stated earlier the atoms which strike the wall in a bright-wall oven are re-emitted with a cos(θ) distribution. For a collimator of uniform cross section this process of absorption and re-emission of atoms leads to a vapor pressure which varies linearly between the pressure at the source and zero at the emitting end of the tube [4]. If position along the tube is measured relative to the forward end, then the rate at which atoms are emitted from a wall-surface element at a distance z from the end of the tube is proportional to z. This assumption is not strictly valid at the front of the tube where an end correction should be made, but it seems to provide a good description of the central portion of the beam profile [4]. Some caution should be exercised in applying the model to the prediction of total oven emission since a large part of the total is emitted at large angles where the front end of the oven is important.

3. Recirculating Oven - The vapor pressure of cesium at each point along the wall of a recirculating oven is taken to be consistent with the temperature at that point on the wall. Nesmeyanov [5] gives an exponential form for vapor pressure as a function of temperature which seems to fit the vapor pressure data quite well and is convenient for performing the calculations on the recirculating ovens. In calculating the flux emitted by a wall element, the mean velocity of the atoms emitted by the wall must be

included. This velocity is proportional to the square root of the temperature of the wall [3]. For the simple uniform tube of figure 1a, the temperature can be assumed to vary linearly between T_1 and T_2.

Following the assumptions outlined above (in 3) the beam intensity as a function of the angle θ can be written as

$$I(\theta) = Ka^2\cos(\theta)[p(T_2)/(T_2)^{1/2}] \times$$

$$[\cos^{-1}(Ls) - Ls(1 - L^2s^2)^{1/2}]$$

$$+ Ka \sin(\theta) \int_0^u [p(T)/(T)^{1/2}] [1 - z^2s^2]^{1/2} dz.$$

where $K = 1.458 \times 10^{23}\ m^{-2}s^{-1}$ is a constant (including Planck's constant and the mass of the cesium atom), a is the tube radius, p(T) is the pressure as a function of temperature [5], $T = T_1 + z(T_2 - T_1)/L$, L is the length of the collimator, $s = \tan(\theta)/2a$ and the upper limit u of the integral depends on the interval of θ that is considered. For $\theta < \tan^{-1}(2a/L)$, $u = L$ and for $\tan^{-1}(2a/L) < \theta < \pi/2$, $u = 2a/\tan(\theta)$. The intensity is in units of atoms/second/steradian and all other quantities are expressed in SI units. The first term of the expression gives the contribution from the source at T_2 which is just the dark-wall-oven result and the second term describes the emission from the walls. The equation is integrated numerically to compare with the experimentally determined beam profiles and a second integration over all solid angles in the forward hemisphere yields the total flux emitted by the oven.

Experiment

The beam profile for the recirculating ovens was determined using a scanning apparatus which consists of a fixed hot wire detector with a 1 mm^2 aperature at a distance of 55.2 cm from the oven, the axis of which can be tilted up to ~ 5° to either side of alignment with the detector. With this apparatus, beam profile can be determined with a relative precision (resolution) of a few percent. The absolute accuracy of the intensity measurements is approximately 10%. With the half-meter source/detector separation and the 1 mm^2 detector area, the system averages the beam flux over an angle of about 0.1° yielding an angular resolution which is more than adequate for the present studies. A dark-wall oven was constructed using a carbon collimator and the peak intensity and profile agreed well with that predicted by the dark-wall theory providing confidence in the measurement methods. While the ovens could be scanned in two directions (tilted about two axes), their cylindrical symmetry always resulted in essentially identical results, so scanning was normally performed along only one of the axes.

Results

Figure 2 shows the peak beam intensity (on-axis flux) of the recirculating oven as a function of source temperature compared with theory. This peak intensity depends only on the source temperature as long as the mean free path in the beam is long compared to any of the oven dimensions. The saturation at higher temperatures might appear to result from cesium-cesium collisions within the source since the decrease in mean free path with increasing pressure should eventually limit the beam flux. However, as will be described below, the experiments do not support this interpretation.

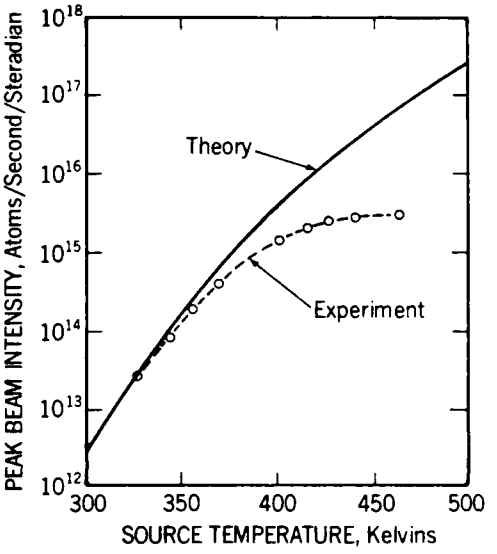


Figure 2. Peak beam intensity as a function of source temperature. The data is for the oven described in the text (r = 1 mm and L = 4.5 cm). The peak beam intensity should depend only on the source temperature.

Figure 3 shows the measured beam profile compared with the predictions of theory for the dark-wall, bright-wall and recirculating ovens. The source temperature in the recirculating-oven theory was adjusted to give the best fit to the experiment and the dark-wall and bright-wall theory then used the same source temperature. The temperature for this best fit was 14 K below the temperature measured adjacent to the heater. This was consistent with the discrepancy between theory and experiment in figure 2. No correction for the detector angular resolution of ~ 0.1° was made although, if included in the theory, this correction would not quite round off the peak enough to bring complete agreement between theory and experiment in the central region.

In an attempt to determine the mechanism for the saturation observed in figure 1, we measured the beam profile with the source at 405 K, a temperature which is well into the saturation region. If the saturation were produced by collisions in the vicinity of the source, then the effective emitting surface would be displaced forward and the collimation would not be as good,

that is, the beam profile would be broadened [4]. Our measurements showed no such broadening, so the saturation must arise from another cause. We will speculate on this effect in the next section.

Discussion and Conclusions

The good agreement between experiment and theory for the beam profile measurements suggests

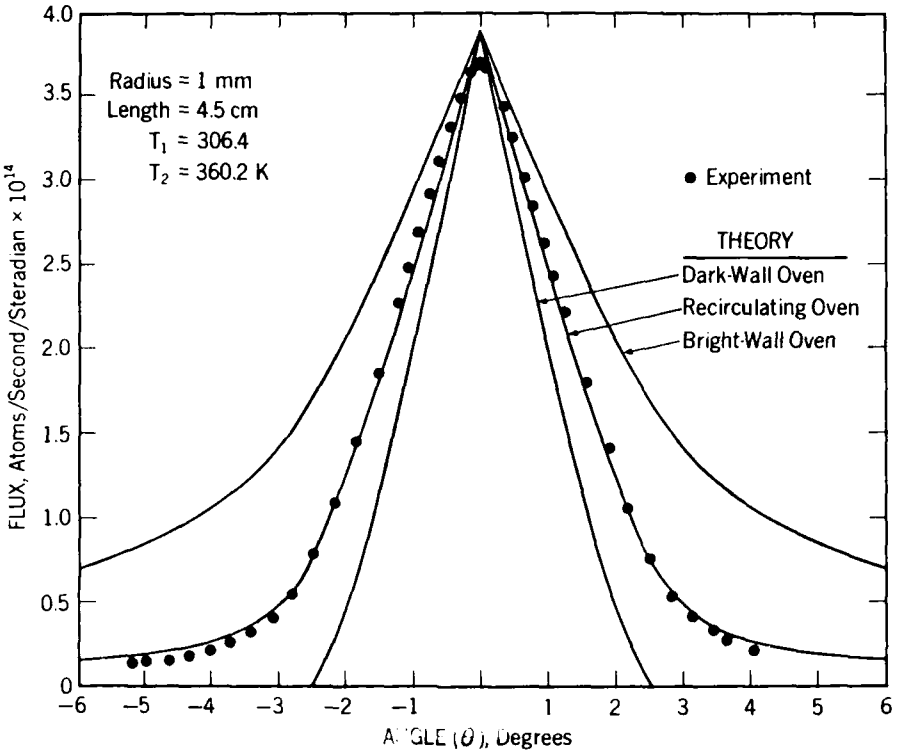


Figure 3. Beam profile for three types of ovens. The experimental points are for the recirculating oven described in the text. The parameters in the upper lefthand corner refer to Figure 1a.

To test the hypothesis that the cesium is bound by capillary action in the porous matrix, the oven was operated vertically (emitting upward and downward) as well as horizontally. Within our experimental resolution of 2% we found no change in peak beam intensity. With the oven pointing downward we would have expected problems with the detector if cesium had spilled onto it. In simple experiments performed in a sealed glass tube, the cesium charge was found to be fully absorbed by the porous matrix, so these results were not unexpected.

The oven described in this report has been operating continuously for more than six months with no appreciable decrease in emitted flux. This is not a particularly significant test of the lifetime of this type of oven, but it lends support to the idea that cesium is effectively expended from the reservoir, not the collimator or source surface. The oven was not designed with maximum lifetime as a major consideration, so, while we will keep it on life test, it will be important to address life testing with another oven.

that the model and assumptions upon which it is based are reasonable. The peak beam intensity as a function of temperature is reasonably consistent with theory in view of how that temperature was measured, so one can also conclude that the experiment and theory are in accord with regard to the magnitude of the peak intensity.

The discrepancy between actual and measured source temperature is most likely due to a modest thermal resistance between the stainless steel tube and the oven reservoir. It is possible that this thermal resistance also plays a role in the "saturation" effect in figure 2. As noted in the previous section, this might appear to be a mean-free-path effect, but the beam-profile measurements do not support that interpretation. This apparent saturation could simply reflect a non-linear dependence of the thermal resistance (between the heater and oven) on temperature. If liquid cesium plays a significant role in heat transport between the stainless steel surface and the reservoir, then vaporization of the cesium at higher temperatures would result in a higher thermal resistance. The "saturation" would then be an artifact reflecting an increasing descre-

pancy between the actual source temperature and the temperature measured adjacent to the heater. Clearly, it will be important to use different heater and sensor arrangements in future tests.

Using the model for the recirculating oven, we have integrated over all angles in the forward hemisphere to obtain an estimate of the total emitted flux. Similar calculations were also performed for the dark-wall and bright-wall ovens. For the conditions of the experiment described in this paper (see figure 3), the total fluxes for the bright-wall, recirculating and dark wall ovens are 3.6×10^{13}, 2.2×10^{13} and 6.2×10^{11} atoms/second, respectively. The total emission from the dark-wall oven is dramatically lower than that of either of the other ovens. This should be the ideal goal of any new oven design. For these conditions the recirculating oven is about 1.6 times as efficient in its' use of cesium as the bright-wall oven. Unfortunately, the total flux from the recirculating oven rises as T_1 rises and at $T_1 = 312$ K the total flux for the two ovens is about equal. For small angles (< 5°), the shape of the collimated beam changes little for this increase in T_1. It is primarily an increase in the large angle flux which produces the increased net flux for the oven. Thus, to take full advantage of the design, it is necessary to hold T_1 near the melting point of cesium (301.5 K). As will be discussed shortly, design modifications might alter this situation.

It is not surprising that the net emitted flux of the recirculating oven crosses over that of the bright-wall oven as the front temperature is increased. One only has to consider the limiting case where T_1 reaches T_2 and this becomes obvious. In this limit the source becomes essentially uncollimated with the tube area emitting flux with a cos(θ) distribution. For comparison with the figures above, the total flux (for $T_1 = T_2 = 360.2$ K) is 1.25×10^{15}, clearly much worse than for the equivalent bright-wall oven. This limiting case is also useful as a test for the integrations and the form of the equations.

The models for the three types of ovens appear to fit experimental beam profiles quite well, but their accuracy in predicting total emission has not been rigorously tested. With respect to total emission, the dark-wall and recirculating oven models would appear to rest on fairly solid assumptions, since no corrections for large angle emission seem necessary. However, as we noted in the modeling section, an end correction (not included in our calculations) is needed in the model for the bright-wall oven and this would affect the large angle emission as well as the total emission. Thus, until the models are verified, the estimates of relative total flux should be considered with care.

Several possible modifications of the simple recirculating oven can be considered. First, the oven might be designed so that the entire porous collimator is held at the front temperature T_1. For low values of T_1, the model predicts that the central beam profile for such an oven would be very close to that of the dark-wall oven (see fig. 3). However, the total integrated flux (using our experimental conditions) would not be significantly smaller than that where the wall temperature varies linearly between T_1 and T_2. This is because the large angle flux dominates the total flux and there is little difference in that flux for the two cases. Where the entire collimator is held at T_1, it might also be possible to fabricate a parallel tube structure. This type of design (which is used in ovens in commercial standards) would be problematic if the collimator had to support a large temperature gradient, since such a collimator would be short and broad, and maintenance of the temperature gradient would most likely require a very large heat flow.

Another design modification, and one that might improve the performance dramatically, involves the combination of a bright-wall and recirculating collimator. We have considered the addition of a section of bright-wall collimator at the front end of the recirculating oven described above. The model suggests that the beam profile and total flux for such a device would be very close to that of the ideal dark-wall oven. Furthermore, the performance would not be degraded as severely by a rise in T_1. In the worst case where $T_1 = T_2$, the profile and total flux would be those determined by the bright-wall part of the collimator. The key concern in the model for this hybrid oven is the validity of the assumption regarding the interface between the recirculating and bright-wall sections of the collimator. We have assumed that the vapor pressure in the bright-wall section varies linearly from zero at the emission end to a value commensurate with T_1 at the interface with the recirculating portion. The impressive potential performance of this hybrid oven clearly warrants further study.

Acknowledgement

The work described in this paper was supported by the U.S. Air Force Space Division. The authors would like to acknowledge the very helpful input of Dr. Fred Walls.

References

1. R. D. Swenumson and U. Even, Rev. Sci. Instr. 52, 559 (1981).
2. Melissa Lambropoulos and S. E. Moody, Rev. Sci. Instr. 48, 131 (1977).
3. Norman F. Ramsey, Molecular Beams, (Clarendon Press, Oxford) 1956.
4. J. A. Giordmaine and T. C. Wang, J. Appl. Phys. 31, 463 (1960).
5. An. N. Nesmeyanov, Vapour Pressure of the Elements, translated and edited by J. I. Carasso, (Academic Press, New York) 1963.

OPTICALLY PUMPED SMALL CESIUM BEAM STANDARDS; A STATUS REPORT*

Amy Derbyshire, R. E. Drullinger, M. Feldman(a), D. J. Glaze, D. Hilliard,
D. A. Howe, L. L. Lewis(b), and J. H. Shirley
National Bureau of Standards
Time and Frequency Division
Boulder, Colorado 80303

and

I. Pascaru and D. Stanciulescu
Frequency Electronics Incorporated
Mitchel Field, NY 11553

Abstract

We report on our project to study and to demonstrate the potential performance achievable in cesium beam frequency standards in which laser driven optical pumping is used for the atomic state selection and state detection in place of the conventional magnetic state selection. The beam tubes used have been derived from commercial devices. In the first unit the only functional change was a simple replacement of state selection magnets with optics. In a second unit, the magnetic shields and c-field have been extended to include the regions of optical pumping.

The short Ramsey cavities resulted in observed resonances which are 1300 Hz wide in agreement with theory for the geometry and detecting transition used. Both devices have demonstrated $\sigma_y(\tau) \leq 10^{-11}\ \tau^{-1/2}$ and are not yet limited by beam shot noise. Systematic errors resulting from resonantly scattered light have been shown to be less than 10^{-12} and one of these devices has been operated as a clock for 45 days.

Introduction

Optical Pumping in Cesium

The possibility of improving the performance of atomic beam frequency standards through the use of state selection and detection by optical means has been a subject of investigation for several years. [1-3] In recent times semiconductor laser diodes which are suitable for laboratory experiments in optical pumping have become commercially available. [4] These lasers operate with sufficient output power in a nearly single longitudinal mode, and sufficiently small intensity and frequency noise to allow some interesting measurements of optically pumped frequency standards. In addition, since the lasers can produce light at 852 nm, they can be used with the D2 transition in atomic cesium. The small size and low cost of laser diodes may allow the production of commercial cesium standards [5] with improved stability characteristics over present designs, or the construction of a laboratory primary frequency standard [6] with greater accuracy than that of existing primary cesium standards.

*Work supported in part by the USAF Space Division and USAF Rome Air Development Center.

(a) Current Address: Lawrence Livermore Laboratories Livermore, California 94550

(b) Current Address: Ball Corporation/Efratom Division, Broomfield, Colorado 80020

Contribution of the National Bureau of Standards, not subject to copyright.

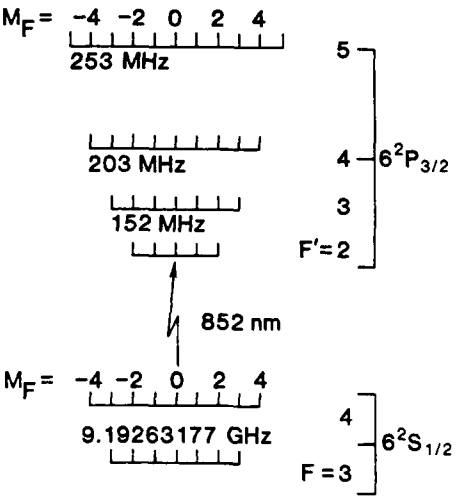


Figure 1. Energy level diagram for cesium optical pumping.

The energy level diagram of Figure 1 shows the relevant optical and microwave transitions for optical pumping in cesium. A number of different pumping schemes are possible, which may prepare cesium atoms in either the $^2S_{1/2}$, F=3 or F=4 hyperfine states. For example, the method used most often in the measurements reported here tuned a laser diode to the $(^2S_{1/2}, F=4) \Leftrightarrow (^2P_{3/2}, F'=4)$ transition (852 nm) from the ground state to the second excited state. In this case the excited atoms decay back to both the F=3 and F=4 ground state levels with roughly equal probability. Continued exposure to the laser light drives the optical transition from the F=4 ground state again until no atoms are left in this state. The result is an increase in the population difference between the $^2S_{1/2}|F,m_F\rangle = |4,0\rangle$ and $|3,0\rangle$ magnetic sublevels. Figure 2 is a schematic representation of this optical pumping method, where Laser 1

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

is the pump laser. Virtually complete optical pumping of the cesium occurs, with 95 to 99% of the cesium atoms placed into the F=3 ground state. This results in approximately 22% of the atoms in the $|F,m_F\rangle = |3,0\rangle$ state, as opposed to about 7% for traditional magnetic state selection. In addition, all velocity classes are included, instead of the 8 to 10% common for magnetic state selection.

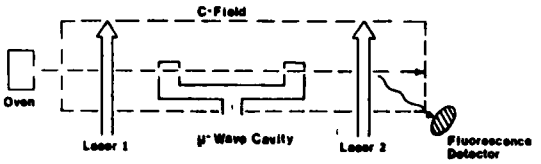


Figure 2. Schematic of an optically pumped cesium beam clock.

More complete pumping into the $|3,0\rangle$ state is possible if two lasers are used for the optical pumping. [7] One laser may be tuned to the $^2S_{1/2}$, $F=3 \leftrightarrow {}^2P_{3/2}$, $F'=3$ transition using π-polarization. This pumps all magnetic sublevels of the F=3 ground state HFS into the F=4 HFS, with the exception of the $m_F=0$ sublevel. A second laser, tuned to the $^2S_{1/2}$, $F=4 \leftrightarrow {}^2P_{3/2}$, $F'=3$ transition pumps atoms back into the F=3 HFS. After many cycles of this process, most atoms should reside in the F=3, $m_F=0$ sublevel. Experimental verification for this process has been reported. [5,8]

Optical Detection

In order to detect the number of cesium atoms which made a transition from the F=3 to the F=4 ground state hyperfine levels after the application of the microwave signal (Figure 2), the same laser used for the single-laser optical pumping described above may be applied to the downstream atoms. The fluorescence collected from the atoms is a direct measure of the number of atoms which have made the microwave transition. In this case, approximately two photons per atom are emitted, placing sever demands on the collection efficiency and input noise specifications of the detection optics. An alternative approach is to use another laser tuned to the $^2S_{1/2}$, $F=4 \leftrightarrow {}^2P_{3/2}$, F=5 transition. With this transition, atomic selection rules require the excited atoms to decay back to the F=4 level, where they may interact again with the laser light. Since more than 100 photons can be emitted per atom, the collection optics need not be very efficient, and yet the net quantum efficiency of detection will be essentially unity.

Experiment

We have constructed and partially tested two beam tubes. Using the fabrication capabilites at Frequency Electronics and making maximum use of standard parts and geometries it is possible to build small, simple, dedicated tubes in which all the geometrical parameters are fixed. This arrangement has the advantage that its easy and one does not subsequently have problems controlling and adjusting a large number of parameters. However, it has the disadvantage that is hard to go back into an all welded package and correct the inevitable design and assembly errors.

The first device was a minimum modification on an existing commercial beam tube and is schematically shown in figure 3. The "A" and "B" magnets, hot wire ionizer, electron multiplier and vac-ion pump were omitted in construction. Instead vacuum flanges were added at both ends for the addition of test ovens, in the "A" and "B" regions for the addition of the optical pumping and detecting optics and in place of the vac-ion pump so a large vacuum pumping station could be fitted to facilitate repeated openings of the system. The magnetic shields, c-field and Ramsey cavity remained unchanged from the standard commercial tube. The ovens used on this device were of the ideal "dark wall" type (see the accompanying paper on recirculating ovens). The ovens were separated from the beam tube by gate valves which could be closed when the beam tube system was opened. The fluorescence collection optics were a combination of a spherical mirror and an aspheric lens followed by a relay lens which refocused the fluorescence onto a photodiode mounted outside the vacuum envelope. The overall collection-detection efficiency was about 5%.

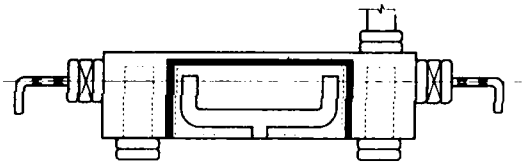


Figure 3. Schematic of OPCS-1.

The second unit differed more radically from a conventional standard as shown in figure 4. The "c-field" and the magnetic shields were extended to include the entire length of the device in an attempt to control magnetic field gradient induced transitions (Majorana transitions). The fluorescence collection optics included a spherical mirror segment facing an elliptical mirror which directly imaged the fluorescence onto a large area photo diode mounted outside the vacuum. The overall collection-detection efficiency was about 35%. As in the first unit, ovens were fitted to both ends. However, unlike the first unit, the ovens were of the standard production type and the total device hard sealed with no vacuum flanges.

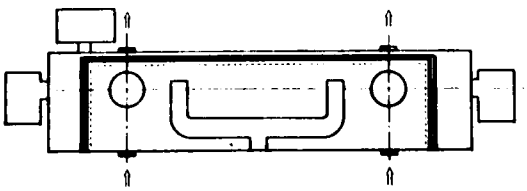


Figure 4. Schematic of OPCS-2.

We have called these two units OPCS-1 and OPCS-2 for "optically pumped cesium standard" numbers 1 and 2. A third device is under construction which will not be functionally different from OPCS-2 except in some of its abilities to generate and study Majorana transitions. It will, however, solve some optical problems associated with the hard seal technology used on OPCS-2 as well as employing a superior magnetic shielding package.

The lasers used were of the "channel stripe planer" (CSP) type with simple cleaved facet cavities. However, in this form they proved somewhat mode unstable and excessively sensitive to feedback from other optical elements in the experiment. The solution to this was the use of a flat mirror located as close to the back facet of the laser as the heat sink would allow (about 2 mm) and mounted on the piezo-electric transducer (PZT). This fed back enough light (about 10^{-6} of the back facet emission) to force mode control. Changing the drive voltage to the PZT caused mode changes about every 100 V with broad regions of mode stability in between. When followed by a Faraday isolator with about 30 dB of isolation, this laser system proved mode stable for indefinite periods. The laser system together with its collimating objective and mode control mirror was temperature controlled.

The frequency of the laser was locked by a two level servo to the optical resonance of the cesium beam itself. The laser drive current was modulated at several hundred hertz. The resulting FM produced an AM on the fluorescence from the atomic beam. Synchronous detection then resulted in an error signal which when integrated was used to correct the laser drive current. With a cross over of some 10 seconds, this error signal was also used to control the temperature of the laser and hence keep the drive current centered about a specified value. It appeared subsequently, that through slight alignment errors or residual first order Doppler shift the two lasers could be locked to slightly different velocity groups of the beam atoms and hence introduce an enhanced sensitivity to laser frequency noise that degraded the clock short term stability. The solution to this was to modulate the pump and detection lasers at different frequencies and detect both modulations in the fluorescence emitted from the detection region. That is, just as with the microwave servo system, modulation of the up stream resonance produces a modulated population which shows up at the down stream detector.

Results

All of the results reported here have been obtained using a single laser for state preparation. The detection has always involved the use of the F=4 → F=5 cycling transition. In this case, the slower atoms contribute more strongly to the observed signal by virtue of their greater interaction time with the detection laser. The result is a slightly narrowed Ramsey resonance from that which would be produced by a purely effusive beam (see Figures 5 and 6). That is, a purely effusive beam would have a velocity distribution given by

$$I(v) = \frac{2I_o}{\alpha^4} v^3 e^{-v^2/\alpha^2}$$

where $\alpha = \sqrt{2kT/m}$. However, when detected via an optical cycling transition, the effective velocity profile has the form

$$I'(v) = \frac{2I_o}{\alpha^4} v^2 e^{-v^2/\alpha^2}$$

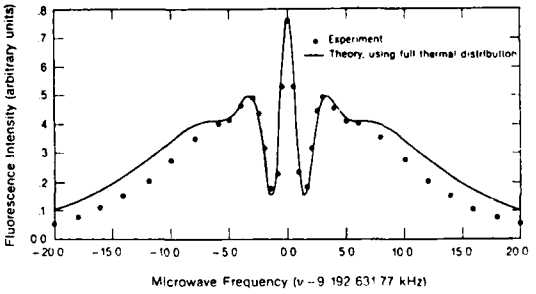


Figure 5. The observed Ramsey resonance fit by a theoretical line shape for a full thermal velocity distribution.

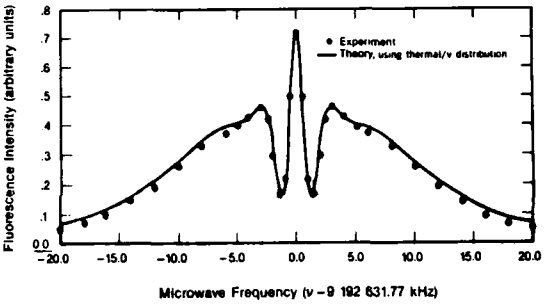


Figure 6. The observed Ramsey resonance fit by a thermal distribution weighted by v^{-1}.

Both OPCS-1 and -2 have been locked up and run as clocks. They both exhibit short term stability $\sigma_y(\tau) \leq 1 \times 10^{-11} \tau^{-1/2}$ for $3 < \tau < 10^4$ s. At longer times the frequency stability of OPCS-1 was dominated by environmental effects which were not carefully controlled in this device. The longer term behavior of OPCS-2 has not been studied.

In OPCS-1 the short term stability was limited by residual FM noise in the diode lasers. The hard seal window technology used in OPCS-2 resulted in a high level of scattered laser light at the fluorescence detector and thus required the use of laser powers well below optical saturation. Also, the lasers used on this device happened to tune to the cesium resonance at relatively low power levels where the laser linewidth and FM noise are comparatively worse than at higher power levels [9]. The result was an initial operating performance $\sigma_y(\tau) > 10^{-10} \tau^{-1/2}$ for $3 < \tau < 10^4$ s. When a dye laser with its superior linewidth and FM noise characteristics was substituted for the diode lasers, the stability of $\sigma_y(\tau) < 10^{-11} \tau^{-1/2}$ quoted above was obtained. The FM noise of the diode lasers was then reduced by servo control to

stable reference cavities and the $\sigma_y(\tau) < 10^{-11} \tau^{-1/2}$ performance was again obtained. In this case the stability was limited by shot noise of the scattered light.

In an attempt to show the potential for sustained operation, OPCS-1 was run uninterrupted for a 45 day period. Also a search was made for frequency offsets in OPCS-1 due to the optical pumping and detection process itself. No frequency shifts at the 10^{-12} level were found. Tests included changing both the pump and detection laser intensities, changing the specific optical transition used for locking the laser frequency, and varying the laser injection current, modulation frequency and amplitude. These results make the possibility of frequency shifts due to light shifts (ac Stark shift) or modulation cross-talk unlikely at the 10^{-12} level. This is consistent with the shift of 3×10^{-14} predicted in the accompanying paper on light shift by J. Shirley.

Conclusions

The optical state preparation process treats all velocity classes equally. As a result, the mean velocity of an optically selected beam is about twice that of a conventional magnetic state selected beam. For the same Ramsey cavity, this results in a microwave linewidth which is about twice as wide. This is the case for the devices studied here. The parent tube has a linewidth of about 800 Hz while the optically pumped derivations of it used here have a linewidth of 1.3 KHz. In spite of this loss of line Q, the increased utilization of atomic beam flux resulted in a substantial overall performance improvement. The conventional parent tube has a short term stability of $\sigma_y(\tau) = 3.5 \times 10^{-11} \tau^{-1/2}$. In the case of OPCS-2 we used the same oven and beam flux and achieved a $\sigma_y(\tau) < 1 \times 10^{-11} \tau^{-1/2}$. However, we are still being limited by phenomena other than the beam shot noise. In fact, the limit for OPCS-2 under the conditions that it has been operated is $\sigma_y(\tau) = 3.5 \times 10^{-12} \tau^{-1/2}$.

Because of experimental problems with the two devices tested so far, we have not studied either the long term stability or the effects on Majorana transition expected from the change in magnetic shielding design used in OPCS-2. The lack of long term stability has prevented a serious test of the theoretically predicted light shift but we have shown the light shift to be below 10^{-12} in OPCS-1. The operation of OPCS-1 as a clock for 45 days shows that there are no problems with the lasers, as operated here, that would prevent long term clock operation.

References

[1] G. Singh, P. DiLavore and C. O. Alley, IEEE J. Quant. Electr. QE-7, 196 (1971).

[2] M. Arditi and J.-L. Picqué, J. Phys. (Paris) 41, L-379 (1980).

[3] L. L. Lewis and M. Feldman, 35th Ann. SFC, 612 (1981).

[4] A. Aiki, M. Nakamura, T. Kuroda, and J. Umeda, Appl. Phys. Lett. 30, 649 (1977).

[5] L. Lewis, Prog. Quant. Electr. 8, 153 (1984).

[6] L. L. Lewis, F. L. Walls, and D. A. Howe, "Prec. Meas. and Fund. Const. II," B. N. Taylor and W. D. Phillips, Eds. Nat'l Bur. Stand. (US), NBS Spec. Publ. 617, p. 25 (1984).

[7] This method was suggested by L. Cutler of Hewlett Packard. See for example H. J. Gerritsen and G. Nienhuis, Appl. Phys. Lett. 26, 347 (1975).

[8] G. Avila, E. De Clerg, M. de Labachelerie and P. Cérez, IEEE Trans Instrum Meas, IM-34, 139 (1985).

[9] Aram Mooradian Phys. Today, 38 no. 5, 43 (1985).

FLUORESCENT LIGHT SHIFT IN OPTICALLY PUMPED CESIUM STANDARDS

Jon Shirley
Time and Frequency Division
National Bureau of Standards
Boulder, Colorado 80303

Abstract

We have calculated the light shift in an optically pumped cesium beam frequency standard caused by fluorescence co-propagating with the atomic beam. Both scalar and tensor contributions are included to give the dependence on light polarization. The results provide design criteria for proposed new standards.

The light shift, or light-induced AC Stark shift, is a potential source of error in proposed optically pumped atomic beam frequency standards that is not present in existing cesium standards. This shift can be quite large (parts in 10^9) while the atoms are directly exposed to a light beam intense enough to produce optical pumping [1]. But in an atomic beam the optical pumping and resonance observation regions are spatially separated. Careful design can minimize scattering of pumping light into the observation region. However, fluorescence by the pumped atoms can still propagate along the direction of the atomic beam into the observation region to produce a light shift. The present investigation was made to establish a quantitative estimate for this fluorescence-induced light shift and its dependence on design parameters.

We first find the light shift as a function of the light intensity bathing the atoms. Later we will relate the intensity to the atomic beam flux and propagation distances.

Brillet [2] has already made a calculation of the scalar [3] part of the light shift. We wish to also include the tensor [3] part to determine the dependence of the shift on polarization of the light. Rather than using the formalism of Mathur et al. [4] we have chosen to begin with a basic formula for the energy shift of a state g [5]:

$$\Delta E_g = \hbar \sum_n V_{ng}^2 \frac{(\omega-\omega_{ng}-\vec{k}\cdot\vec{v})}{(\omega-\omega_{ng}-\vec{k}\cdot\vec{v})^2 + \frac{1}{4}\Gamma_{ng}^2} \qquad (1)$$

Here ω and $\vec{k}$ are the frequency and wave vector of monochromatic exciting light, $\vec{v}$ is the velocity of the atomic absorber, ω_{ng} is the transition frequency between g and an excited state n, Γ_{ng} is the spontaneous emission rate from n to g, and V_{ng}, ω_{ng}, ω, and Γ_{ng} are all in angular frequency units. The state g is either the F=3 or F=4, m_F=0 ground electronic state of cesium, the levels involved in the clock transition. Equation (1) can be derived by second order perturbation theory with (1/2) Γ_{ng} brought in as an imaginary part of the excited state energy [6].

Contribution of the National Bureau of Standards, not subject to copyright.

Using the Wigner-Eckart theorem and angular momentum algebra V_{ng}^2 can be written as a constant times the light intensity times a factor which depends only on the light polarization and the angular momentum quantum numbers of the states n and g. The detuning dependence, averaged over the velocities in the atomic beam, can be expressed in terms of a dimensionless integral function

$$D(\delta,\varepsilon) = \int_0^\infty \frac{(\delta-u)\ 2u^3 e^{-u^2} du}{(\delta-u)^2 + \varepsilon^2}$$

which we have evaluated numerically. The relative light shift of the clock transition (frequency ω_o) can then be written

$$\frac{\Delta E_g}{\hbar\omega_o} = C I_q \sum_n (a_{ng}^q)^2\ D(\delta_{ng},\varepsilon) \qquad (2)$$

where $C = \mu^2/(2\omega_o\varepsilon_o\hbar^2 ck\alpha) = 1.33 \times 10^{-7}\ m^2/W$, I_q is the light intensity having polarization q, a_{ng}^q is the angular momentum factor, $\delta_{ng} = (\omega-\omega_{ng})/k\alpha$ is the detuning variable and $\varepsilon = \frac{1}{2}\Gamma/k\alpha = 0.01$. The Doppler width in a gas with most probable velocity α [7] is $0.83\ k\alpha$. The reduced electric dipole matrix element μ is related to the excited state lifetime by $1/\tau = \Gamma = \mu^2 k^3/3\pi\varepsilon_o\hbar$. For pumping of the D2 transition to the $^2P_{3/2}$ electronic state the sum over n includes those m_F values allowed by the light polarization and all three allowed F values, since the Doppler width is comparable to the hyperfine structure separation. For unpolarized light a summation must also be made over q. Equation (2) holds for monochromatic light co-propagating with the atomic beam, as for example, pumping light scattered along the beam.

For fluorescence emitted by beam atoms the frequency ω also has a Doppler shift and hyperfine components to be averaged over. When this is done we obtain a result similar to Eq. (2) except that the summation includes the different frequencies and polarization of the fluorescence and the integral D is replaced by the dual velocity integral function

$$D(\delta,\varepsilon) = \int_0^\infty D(\delta+u,\varepsilon)\ 2u^3 e^{-u^2} du$$

Its arguments are $\delta_{n'g'ng} = (\omega_{n'g'}-\omega_{ng})/k\alpha$ and $2\varepsilon = \Gamma/k\alpha = 0.02$ where n' and g' are the initial and final states of the fluorescing atom. Evaluation of the summation and dual velocity integral leads to the coefficients in the table below:

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

Light Shift Coefficients		
Fluorescing State F'	Polarization π	σ
2	3.1	6.9
3	-2.6	-1.3
4	-5.6	-5.1
5	1.3	4.8

Each coefficient is to be multiplied by $10^{-8}I$ where the light intensity I is in W/m^2. For unpolarized light (2/3 σ polarization and 1/3 π polarization) our result for F' = 4 is about 1/4 the corresponding result of Brillet [2].

We now estimate the fluorescent light intensity arising from optical pumping. The intensity reaching beam atoms from atoms excited a distance d away is

$$I_q = I_{at}\, N_q P_q \hbar\omega / 4\pi d^2$$

where I_{at} is the atomic beam intensity in atoms per second at the point of excitation, N_q is the mean number of photons of polarization q emitted by each atom, and P_q describes the angular dependence for polarization q. N_q may be greater than one, especially for a cycling transition. If the magnetic field is parallel to the beam, the polarization is all σ. If the field is perpendicular to the beam, the polarization is half σ and half π.

With Ramsey excitation of the clock transition the observed light shift will be predominantly the average shift occurring in the drift region between the microwave cavity ends (see Fig. 1). The required average of $1/d^2$ is $1/\ell_2(\ell_2+\ell_3)$ where ℓ_2 is the distance from the point of excitation to the beginning of the drift region and ℓ_3 is the length of the drift region. The atomic beam intensity is proportional to $1/\ell_1^2$ where ℓ_1 is the distance from the beam source to the point of excitation. The inverse dependence on $\ell_1^2\,\ell_2(\ell_2+\ell_3)$ of the light shift provides a guide for design of new standards. Fluorescence from the detection region is usually negligible because the atomic beam flux is so much smaller there.

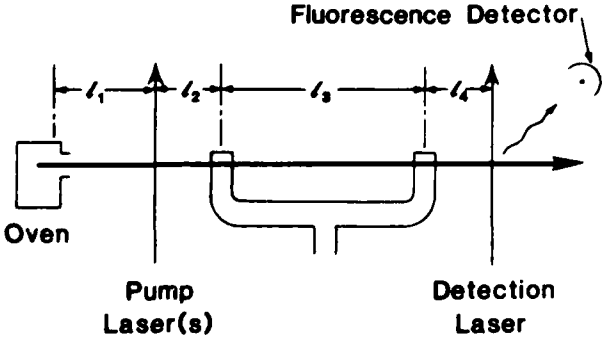


Fig. 1. Schematic of an optically pumped cesium beam standard.

For a typical commercial cesium beam tube we may have beam intensities at the optical pumping position of about 3×10^{11} atoms/sec. Using ℓ_2 = 5 cm, ℓ_3 = 10 cm and N = 3/4 photons per atom for pumping on the F = 3 to F' = 4 transition we obtain an average drift region intensity of 6×10^{-7} W/m^2. With the magnetic field perpendicular to the atomic beam the relative light shift is 3×10^{-14}.

For a proposed primary standard the atomic beam intensity would be about 8×10^{10} atom/sec. Using ℓ_2 = 10 cm, ℓ_3 = 1.4 m and N = 1/2 photon per atom for pumping on the F = 4 to F' = 3 transition we obtain an average fluorescence intensity in the drift region of 5×10^{-9} W/m^2. With the magnetic field parallel to the atomic beam the relative light shift will be only 6×10^{-17}.

On the basis of these estimates we can conclude that the light shift should be included in the error budget of any optically pumped cesium standard. However, it need not be the limiting factor in the overall accuracy of a well-designed standard. In fact, to experimentally observe the fluorescence-induced light shift it may be necessary to introduce an extra light beam very close to the cavity entrance (small ℓ_2) which drives a cycling transition (large N).

References

[1] M. Arditi and J. L Picqué "Application of the Light-Shift Effect to Laser Frequency Stabilization with Reference to a Microwave Frequency Standard," Optics Communications, vol. 15, pp. 317-322, October 1975.

[2] A. Brillet, "Evaluation of the Light Shifts in an Optically Pumped Cesium Beam Frequency Standard," Metrologia, vol. 17, pp. 147-150, 1981.

[3] W. Happer and B. S. Mathur, "Effective Operator Formalism in Optical Pumping," Phys. Rev., vol. 163, pp. 12-25, Nov. 1967.

[4] B. S. Mathur, H. Tang, and W. Happer "Light Shifts in the Alkali Atoms," Phys. Rev., vol. 171, pp. 11-19, July 1968

[5] See for example, V. S. Letoknov and V. P. Chebotaev in Nonlinear Laser Spectroscopy, Springer Series in Optical Sciences, Vol. 4, p. 171, 1977.

[6] C. Cohen-Tannoudji, "Effect of a Non-Resonant Irradiation on Atomic Energy Levels--Application to Light-Shifts in Two-Photon Spectroscopy and to Perturbation of Rydberg States," Metrologia, vol. 13, pp. 161-166, 1977

[7] N. F. Ramsey, Molecular Beams, Oxford: Clarendon Press, 1956, Ch. II. p. 20.

CESIUM BEAM FREQUENCY STANDARD FOR THE PLSS PROGRAM

D. Silvermetz, I. Pascaru and M. Meirs

Frequency Electronics, Inc.
Mitchel Field, New York 11553

Summary

Frequency Electronics has developed a low noise Cesium Standard to be used under dynamic vibration environments. The PLSS (Precision Location Strike System), under development for the USAF by Lockheed Missiles and Space Company, requires a precision and stable frequency reference to perform mission timing and ranging functions. The frequency standard is located in a high-flying aircraft and must operate in a hostile environment (temperature and altitude extremes, high shock and vibration levels, EMI) while maintaining frequency stability of 7×10^{-12}. Low single-sideband phase noise must be maintained under condition of random vibration. This phase noise is less than -110 dBc at 10 Hz, -120 dBc at 100 Hz and -130 dBc at 1000 Hz.

Frequency Electronics developed the Frequency Standard Set, Model FE-5105A for the PLSS Program, and met all of Lockheed's specification requirements. The vibration isolated standard features the FEI cesium beam tube and FEI's low "g" sensitivity SC cut quartz crystal oscillator. The crystal oscillator exhibits an acceleration sensitivity of less than $2X10^{-10}/g$.

The five sets delivered to Lockheed are currently employed in PLSS flight tests.

Introduction

PLSS stands for Precision Location Strike System. It is a multi-mission system developed by Lockheed for the U.S. Air Force. Three high-flying Lockheed aircraft serve as listening platforms to detect short radar transmissions and determine their precise time of arrival at each aircraft. By this means, the location of the radar can be pinpointed and further action initiated.

Each aircraft carries a Cesium Beam Frequency Standard (CBFS) manufactured by Frequency Electronics, Inc. (FEI), to facilitate the precision timing required. The frequency standard shown in Figure 1 operates in an extremely hostile environment of temperature extremes, high shock and vibration and electromagnetic interference, while still maintaining low single-sideband noise and a frequency stability of $7PP10^{12}$ in this high vibration environment.

A simplified block diagram of the system appears in Figure 2. The oven-controlled, voltage-controlled oscillator utilizes a low g sensitivity ($<2X10^{-10}/g$) SC-cut quartz crystal to generate a 5 MHz signal. This signal is multiplied to 200 MHz and fed to a harmonic generator. A synthesized 7.368+ MHz signal is also fed to the harmonic generator. The lower sideband around the 46th harmonic occurs near the hyperfine transition frequency of Cesium 133, namely 9.192+MHz. Any drift of the synthesized signal away from the hyperfine transition frequency causes an error signal to be developed in the cesium beam tube that is used to stabilize the quartz oscillator frequency. The 5 MHz signal is doubled to 10 MHz and distributed to 7 highly isolated amplifiers, for use by the PLSS subsystems.

Major Assemblies

The Frequency Standard Set (Figure 1) hangs down from a baseplate in the TR-1 aircraft. Two major assemblies comprise the set; the suspension structure with distribution amplifier and the CBFS. The former (Figure 3) is permanently mounted to the aircraft baseplate and cabled to the aircraft subsystems. The CBFS is delivered to the aircraft just before a mission in a fully warmed-up state.

Figure 4 shows the Cesium Standard bottom-side up. The upper rear module is a rechargeable battery pack which can keep the cesium standard warmed-up and operating within specifications for 1-1/2 hours while it is in transit between ground support equipment and the aircraft. Modularized construction is used throughout, for ease of maintainability. The middle module is the quartz oscillator and the module in the foreground, the power supply. The flip side is shown in Figure 5, with coverplate removed. It shows the Cesium Beam Resonator Assembly with its hybridized tuned preamplifier, the multiplier module containing four hybrid circuits, the hybridized synthesizer module and the interface module. The six hybrid circuits in this Cesium Standard are designed and manufactured by FEI. The interface module monitors twelve critical functions in the standard and provides normalized voltage to fault logic circuits located in the mating distribution amplifier assembly.

Ground Support Equipment

Between missions, the Cesium Standard is maintained in a ready state in the hot rack (Figure 6). The hot rack, built by FEI, can accommodate up to five cesium standards in the 3 large drawers. Front panels contain status indicators, loop and modulation controls, and diagnostic test meters which permit fault location to the module level. In the event of a primary power failure, a back-up battery unit can supply power for four hours of operation. An FEI phase comparator provides a continuous record of the frequency and phase of a selected standard compared to either a second standard in the hot rack, or an external frequency source. The multiplexer sequentially steps the test input of the phase comparator between the cesium frequency standard inputs. Dwell time is selectable in multiples of 1/2 hour. In addition, the hot rack recharges the batteries of the cesium standards.

Figure 7 provides a closer view of the distribution amplifier. Seven isolated outputs are available at 78 ohm 2-wire balanced connectors. Logic circuits examine the monitor outputs of the cesium standard, as well as the outputs of the seven distribution amplifiers to determine operating status of the FSS. Status is available at the "Fault Out" connector. Controls are provided for switching off the modulation and opening the frequency control loop. Status indicators display the condition of the system and the power source in use.

CH2186-0/85/0000-0024$1.00©1985IEEE

Figure 8 shows the suspension structure with the distribution amplifier removed. The vibration isolators are visible in this view. Looking up at this assembly (Figure 9) provides a better view of the isolators which, in conjunction with the low "g" sensitivity crystal, are so important in providing low noise performance in this high-vibration environment.

Vibration-Induced Phase Noise

The low noise performance can be seen in Figure 10, which is a graph of phase noise surrounding the 10 MHz carrier. The specification limit is shown in the center curve. With the cesium standard mounted in the isolators, the worst case aircraft vibrations produce a vibration input into the standard of $3 \times 10^{-4} G^2/Hz$. With a low "g" sensitivity SC-cut crystal, these conditions produce vibration sidebands as shown in the lower calculated curve.

Two of the curves show actual phase noise measurements on five cesium standards under static conditions and under specified random vibration conditions. Note that this data is for a 10 MHz carrier. For a 5 MHz carrier, 6 dB less noise can be expected.

The upper curve is calculated and shows the sidebands that would result if the vibration isolators were not used. Therefore, both a low "g" sensitivity crystal and the vibration isolators were necessary in this high vibration environment.

Performance

The Frequency Standard Set was qualified for the PLSS program by successfully passing the tests listed in Table 1.

A few of the important performance parameters appear in Table 2.

After conducting performance tests, the qualification unit was subjected to the environmental tests shown in Table 1. The unit operated successfully throughout those tests.

Figures 11 through 18 show the qualification unit in various stages of environmental testing.

Conclusion

Frequency Electronics, Inc. produced five Frequency Standard Sets for the PLSS program, using high performance, in-house fabricated cesium beam tubes, SC-cut quartz crystal oscillators and hi-reliability hybrid circuits. The frequency standards are performing well in PLSS system tests being conducted by Lockheed in Northern California.

TABLE 1

Frequency Standard Model FE-5105A
Qualification Tests

TEST
o Electrical Performance
o Random Vibration
o Shock
o Acceleration
o High-Temperature Exposure
o Thermal Vacuum
o Humidity
o Sand and Dust
o Salt Fog Atmosphere
o Explosive Atmosphere
o Electromagnetic Compatibility

TABLE 2

Frequency Standard Model FE-5105A
Performance

Frequency Accuracy	$\pm 7 \times 10^{-12}$
Reproducibility	$\pm 5 \times 10^{-12}$
Short-Term Stability	
τ = 100 msec	5×10^{-9}
τ = 20 msec	8×10^{-10}
$1 < \tau < 10^5$ sec	$7 \times 10^{-11}/\sqrt{\tau}$
SSB Phase Noise (1Hz BW)	
10 Hz from carrier	-122 dBc
100 Hz from carrier	-133 dBc
1000 Hz from carrier	-142 dBc
Output Isolation	40 dB
Output Amplitude into 78 Ohms	1.45 $\pm$ .25 Vrms
Phase Coherence - 7 Outputs	± 4 Nanoseconds
Settability	$\pm 5 \times 10^{-13}$
Battery Capacity	1.5 Hours
Operating Power	28 Watts

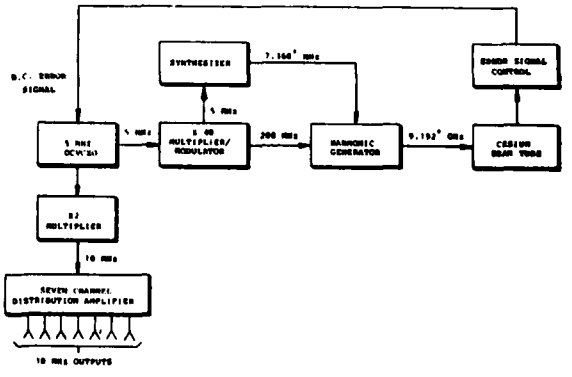


Figure 1. Frequency Standard Set (FSS) Model FE-5105A

Figure 2. Frequency Standard Block Diagram

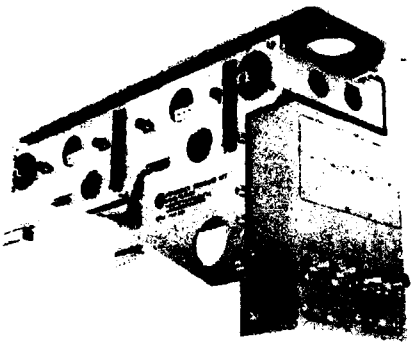


Figure 3. Suspension Structure and Distribution Amplifier

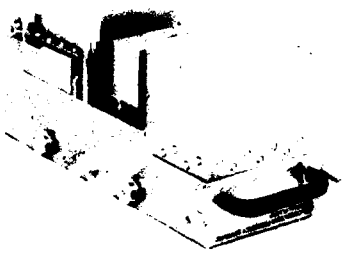


Figure 4. Cesium Beam Frequency Standard (CBFS) Model FE-5460P

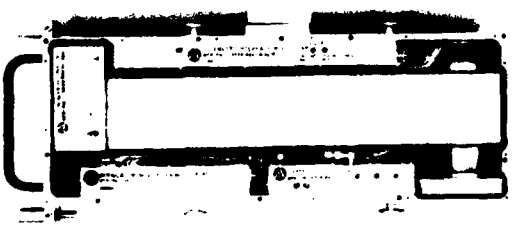


Figure 5. Cesium Beam Frequency Standard, Internal View

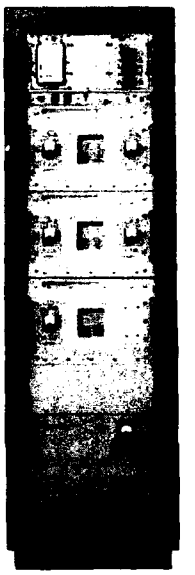


Figure 6. Hot Rack

Figure 7. Distribution Amplifier

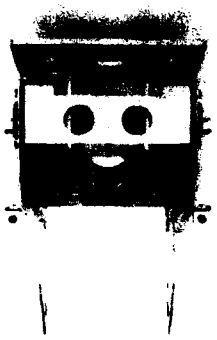


Figure 8. Suspension Structure

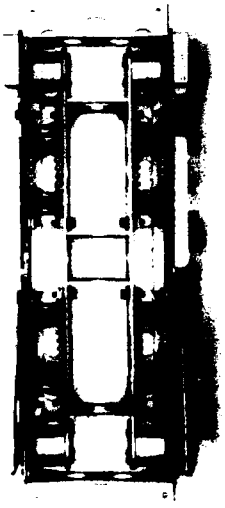


Figure 9. Vibration Isolators

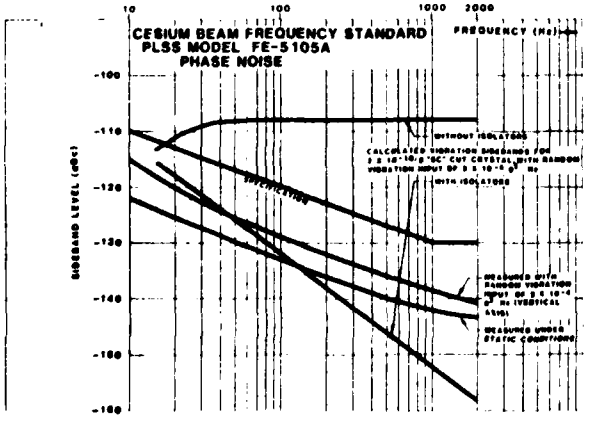


Figure 10. Cesium Beam Frequency Standard Phase Noise

Figure 11. Random Vibration Test

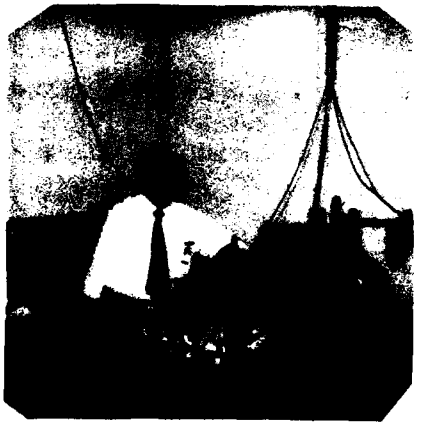


Figure 12. Shock Test With Mass Simulator

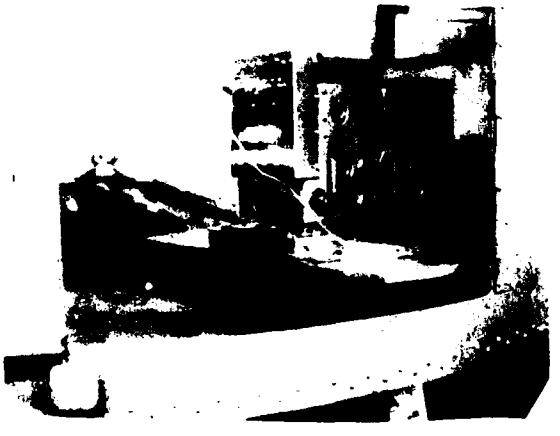


Figure 13. Constant Acceleration Centrifuge

Figure 14. Thermal Vacuum Test

Figure 15. Humidity Test

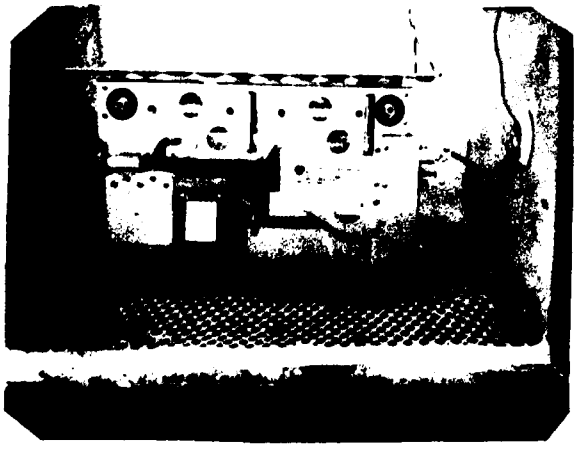


Figure 16. Sand and Dust Test

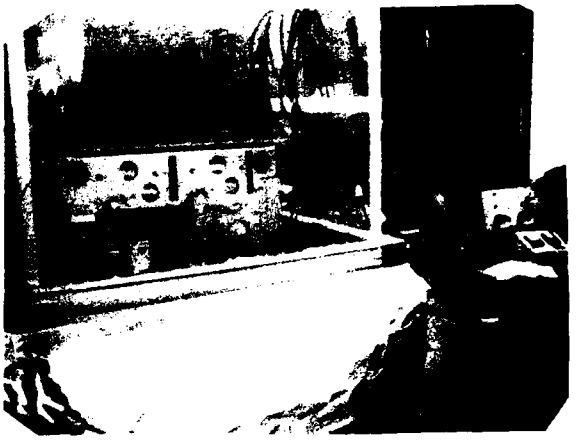


Figure 17. Salt Fog Atmosphere Test

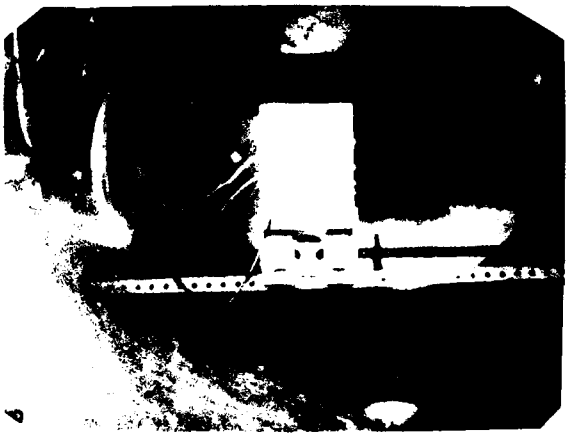


Figure 18. Explosive Atmosphere Test

A RUBIDIUM CLOCK MODEL

Robert P. Frueholz and James C. Camparo
Chemistry and Physics Laboratory
The Aerospace Corporation, P. O. Box 92957, Los Angeles CA 90009

ABSTRACT

In this paper a signal model for the rubidium (Rb) clock is developed. This model combines feedback analysis of the clock's servo-control circuitry with the atomic physics required to describe the processes occurring within the Rb absorption cell. The model allows clock performance, in terms of Allan variance, to be predicted from a number of electronic and physical parameters; and while building on prior studies of Rb standards, incorporates a number of features which make it distinctive.

All previous models of the Rb clock were limited to an analysis of the clock's short term performance, Allan variance averaging times less than 10,000 seconds. However, by explicitly including the effects of discharge lamp intensity fluctuations, which are transformed into output frequency variations via the light shift effect, clock performance can be predicted for averaging times greater than 10,000 seconds. Furthermore, the model is the first which incorporates the influence of an optically thick Rb vapor along with the diffusion of optically pumped atoms to the walls of the absorption cell into the calculation of clock performance. In addition, the model has been developed in sufficient generality to allow its application to diode laser pumped clocks using other alkali-metals.

As part of our validation of the model, it is checked against experimental results from a recent test of an EG&G Rb clock. Agreement between measured and predicted Allan variances for both short and long averaging time periods is excellent.

I. INTRODUCTION

The passive Rb gas cell atomic clock is presently the most often used atomic frequency standard in high-performance systems because of its particular advantages of small size, low weight, low power consumption, and excellent short-term performance. In this standard the Rb^{87} atomic resonance frequency associated with the hyperfine ground state splitting is used as a reference to control the frequency of a quartz crystal oscillator (see Fig. 1 and Ref. 1). Servo control circuitry is used to transfer the stability of the microwave atomic transition to the crystal oscillator, and for times greater than the time constant of the feedback loop the crystal oscillator frequency stability reflects that of the atomic system. The Rb hyperfine resonance is optically detected using the standard technique of optical pumping double resonance, where the source of optical pumping radiation is typically a Rb^{87} discharge lamp. For effective optical pumping the output of the discharge lamp is filtered by transmission through a vapor of Rb^{85} atoms. The filtering may be carried out using a discrete separated filter cell (SFC) as is shown in Fig. 1 or by combining the filter and absorption cells into an integrated filter cell (IFC) [1,2]. In either case the atomic microwave signal appears at a photodiode as a small light intensity modulation component resting on a large DC background. In this paper a non-empirical model of the gas cell frequency standard is developed, with the aim of putting the operation and performance of the gas cell clock on a sound theoretical footing.

In past years a number of other authors have investigated Rb clock performance by developing clock models of varying complexities [1,3-7], and we would be remiss if we did not acknowledge the foundation these studies have provided for our own work. The present model, though, has a number of significant differences from the previous investigations. All previous models were limited to short term performance analysis, Allan variance averaging times less than 10^4s. During this time regime shot noise at the photodetector limits clock performance. The current analysis is the first that is valid for times greater than 10^4s. This was accomplished by explicitly including the effects of discharge lamp intensity fluctuations, which are transferred into output frequency variations via the lightshift effect. Additionally the model is the first which incorporates the influence of an optically thick Rb vapor along with the diffusion of optically pumped atoms to the walls of the absorption cell into the calculation of clock performance. Inclusion of these processes very dramatically effects the axial Rb hyperfine polarization within the cell and how each segment of the cell contributes to the observed hyperfine resonance signal. With the advent of diode laser sources for optical pumping [8] practical gas cell clocks based on alkalies other than Rb, e.g. Cs, may be envisioned. Consequently, the current analysis has been carried out for arbitrary nuclear spin allowing analysis of alternative gas cell clocks.

This paper has been divided into several sections. In the Analysis section the feedback analysis of the servo-control loop along with the physical analyis and computational procedures yielding the Rb hyperfine lineshape are developed. Integration of these components yields the functioning model. In the Model Verification section the validity of the model is established by modelling the performance of a prototype "state of the art" Rb standard [9,10] manufacturered by EG&G, Electronic Components Division for use on NAVSTAR/Global Positioning System satellites. In the final section, implications of this clock model are discussed along with potential avenues for Rb clock performance improvement.

II. ANALYSIS

A. FEEDBACK ANALYSIS OF SERVO CONTROL CIRCUITRY

The function of the servo control circuitry is to transfer the stability of the microwave atomic transition to the crystal oscillator, and the feedback analysis of this system provides a means of understanding how the electronics and the physics package affect clock performance at all analysis times. In Fig. 2 the clock is displayed in a functional block diagram to which standard linear feedback analysis may be applied. The analysis is carried out in the frequency domain and the inputs are the power spectral densities of the various system noise sources. $S_y^{cr}(f)$ and $S_y^{ph}(f)$ are the spectral densities of the fractional frequency fluctuations of the crystal oscillator and the physics package, respectively. [11] Other authors have carried out

CH2186-0/85/0000-0029$1.00©1985IEEE

most of the required feedback analysis previously. [7,12,13] The spectral density of the fractional frequency fluctuations of the entire clock, quartz crystal oscillator locked to the atomic system, $S_y^{Rb}(f)$ is given by

$$S_y^{Rb}(f) = \frac{(f/f_n)^2}{1 + (f/f_n)^2} S_y^{cr}(f) + \frac{1}{1 + (f/f_n)^2} S_y^{ph}(f), \quad (1a)$$

with

$$f_n \equiv \frac{M\, m\, R\, \mu}{2\pi t_1}. \quad (1b)$$

R is the effective impedance of the current to voltage converter/amplifier, M is the frequency multiplier's multiplication factor, and m is the atomic system response relating a microwave frequency detuning Δf to a detected error signal current i,

$$i = m\, \Delta f. \quad (2)$$

Effectively, m is found to be the slope of the discriminator pattern, proportional to the second derivative of the Rb hyperfine lineshape evaluated at the Rb resonance's central frequency. This parameter is obtained through modelling of the physics package (see II.B and C). The VCXO response μ relates a change in control voltage ΔV_{in} to a change in the crystal's output frequency Δf_{VCXO},

$$\Delta f_{VCXO} = \mu\, \Delta V_{in}. \quad (3)$$

t_1 is the time constant of the integrating filter which has a transfer function of the form $1/(2\pi i f t_1)$. Equation 1a indicates that the spectral density of the clock's fractional frequency fluctuations for frequencies above f_n is determined by the crystal oscillator and for frequencies below f_n by the physics package. In the time domain this implies that for times less than the loop attack time ($1/2\pi f_n$) the clock frequency stability is specified by the crystal oscillator while the atomic system (physics package) defines frequency stability at times greater than the attack time.

The principal difference between previous analyses and the present one is the explicit inclusion of the effects of the lamp induced light shift [14,15] of the Rb hyperfine transition frequency. It is clear that lamp intensity fluctuations at frequencies less than f_n are mapped directly into variations of the clock output frequency via the light shift effect. Recent experimental studies have shown that for the EG&G Rb frequency standard these light induced frequency fluctuations specify the long term frequency stability of the standard [16]. Further, these lamp induced frequency fluctuations should set a fundamental limitation to the long term frequency stability of Rb clocks in general. Including the light shift effect, the physics package's spectral density of fractional frequency fluctuations takes the form,

$$S_y^{ph}(f) = \left[K_{LS}^2\, S_y^I(f) + \frac{4e\, i_o}{(f_{Rb}^o\, m)^2}\right], \quad (4)$$

with the lightshift coefficient K_{LS} relating a change in Rb hyperfine resonance clock frequency f_{LS} to a change in light intensity I and the nominal light intensity I_o

$$\Delta f_{LS} / f_{Rb}^o = K_{LS}\, (\Delta I / I_o).$$

$S_y^I(f)$ is the power spectral density of the fractional light intensity ($\Delta I/I_o$) fluctuations. The second term on the right hand side of Eq. 4 is due to shot noise at the photodetector with e the electronic charge, i_o the total lamp induced photocurrent and f_{Rb}^o the Rb hyperfine resonance frequency.

Modelling of the physics package supplies i_o while $S_y^I(f)$ can be experimentally determined from lamp intensity fluctuations recorded as a function of time. While in principle the light shift coefficient can be calculated from fundamental parameters, an experimentally measured value is used in the present analysis. Equation 4 explicitly shows how lamp intensity fluctuations and the light shift can impact clock performance. The importance of minimizing the light shift coefficient as well as the design of lamps yielding extremely stable intensities is apparent.

The spectral density of fractional frequency fluctuations is related to the Allan variance $\sigma_y(\tau)$ via the following integral transformation [11],

$$\sigma_y^2(\tau) = 2 \int_0^\infty S_y^{Rb}(f) \frac{\sin^4(\pi\tau f)}{(\pi\tau f)^2}\, df, \quad (5)$$

with τ the Allan variance frequency averaging time. For the EG&G Rb frequency standard $S_y^I(f)$ was of the form [16]

$$S_y^I(f) = A + \frac{B}{f^2}, \quad (6)$$

the sum of white and random walk contributions. Crystal oscillators display fractional frequency spectral densities of the form [17],

$$S_y^{cr}(f) = \frac{C}{f} + Df + Ef^2. \quad (7)$$

Combining this information with Eq. 1a we observe that the following integrals must be evaluated when obtaining the Allan variance over all averaging times.

$$\Lambda(\tau) = \int_0^\infty \frac{f^\alpha}{1 + (f/f_n)^2} \sin^4(\pi\tau f)\, df \quad (8)$$

$$\alpha = -4, -2, -1, 1, 2$$

To allow rapid evaluation of Allan variances, particularly for averaging times near the loop attack time, these integrals were evaluated analytically using complex contours. The results of these

evaluations are tabulated in Appendix I.

With the evaluation of the integrals the Allan variance may be readily extracted from the modelled clock's spectral density of fractional frequency fluctuations. This spectral density is a function of crystal and loop parameters as well as inputs describing the performance of the physics package. In the following paragraphs the analysis applied to the physics package, yielding the total detected photocurrent and the atomic system response, m, is described.

B. ATOMIC SYSTEM ANALYSIS

Figure 1a shows the basic physics package elements of an optically pumped gas cell atomic frequency standard: the lamp, the filter cell, the resonance cell and the photodetector. In combination, the lamp and filter cell produce a spectral emission that preferentially excites atoms in the resonance cell out of one of the $5^2S_{1/2}$ ground state hyperfine multiplets. For example, the optical emission lines corresponding to transitions from F=2 (i.e. transition numbers 3 and 4 in Fig. 1b) may be effectively blocked by the filter cell, so that only those atoms in the F=1 hyperfine multiplet can absorb light. Since an excited atom can decay from the $5^2P_{3/2}$ or $5^2P_{1/2}$ state into either ground state hyperfine multiplet, the successive action of several optical absorptions and reemissions will lead to a transfer of population from one hyperfine multiplet into the other. Furthermore, since this process of optical pumping depopulates the optically absorbing hyperfine multiplet, the equilibrium light intensity transmitted by the resonance cell and detected by the photodiode will be at a maximum. If microwaves of the appropriate frequency, 6835 MHz, are now applied to the cavity, atoms in the overpopulated hyperfine multiplet will be induced into returning to the optically absorbing hyperfine multiplet. This appropriate frequency is then detected as a decrease in the light intensity transmitted by the resonance cell. Since the Q of the atomic interaction with the microwaves is 10^7-10^8, the decrease in transmitted light intensity can be used to discriminate against very small changes in the 6835 MHz microwave frequency. [1]

From the outline of gas cell frequency standard operation presented above, it is clear that an accurate calculation of the response of the transmitted light intensity to variations in the microwave frequency is a primary concern for a model of clock performance. Typically, this problem is approached in one of three directions. Quite often one attempts to describe the atomic system as a two level atom interacting with the microwaves, [18] an approach that has the advantage that it is intuitive and easy to work with. In particular, when one is concerned with obtaining insight into difficult problems, for example the influence of microwave modulation on the atomic lineshape, [19] a two level atom description is the most reasonable first step. However, this approach is quantitatively incorrect, and can even be in qualitative disagreement with reality under normal frequency standard operating conditions. [20] An alternative approach is to start from first principles, and as much as possible to rigorously calculate collisional interactions, diffusion effects, optical transmission profiles through the resonance cell and the like. [21-24] This program will obviously yield quantitatively correct results, but in light of the number of interrelated physical processes that occur in the gas cell frequency standard, it appears to be an unreasonably complex procedure for predicting clock performance. Furthermore, as more physical processes are incorporated into the model this rigorous approach will soon lose the intuitive nature that makes the two level atom so attractive. Obviously a middle approach, where important physical processes are recognized and treated fairly rigorously, while suitable simplifying approximations are made for less important processes, is the desired approach. In this regard the starting point of the model to be presented below is the generalized Vanier theory of the 0-0 hyperfine lineshape in optically pumped alkali-metal vapors. [6,20,25]

In the generalized Vanier theory all the Zeeman sublevels of an alkali's ground state are considered. However, by assuming that: i) all Zeeman sublevels of a given hyperfine multiplet interact with the light equivalently, ii) all Zeeman sublevels of the ground state have an equal probability of being populated as a result of decay from either the $5^2P_{3/2}$ or $5^2P_{1/2}$ excited states, and iii) the microwaves only induce transitions between the (F=2, m_F=0) and the (1,0) ground state Zeeman sublevels; the density matrix equations describing the evolution of the ground state Zeeman sublevels are considerably simplified. Additionally, relaxation of the ground state Zeeman sublevels is included phenomenologically into the model by considering "dark" longitudinal and transverse relaxation rates γ_1 and γ_2, respectively. Thus, one does not complicate the physical picture by differentiating among uniform, electron randomization and Zeeman relaxation mechanisms. [26,27] In the remainder of this section we will use the generalized Vanier theory to construct a model for the transmitted light intensity as a function of microwave frequency in a possibly optically thick gas cell frequency standard. In this regard our procedure for constructing the model will be to consider each important clock element and physical process individually before incorporating it into the evolving clock model.

1. Lamp and Filter Cell

We assume that the lamp light entering the resonance cell can be decomposed into five parts: four gaussian lineshapes of full width $\Delta\nu_G$ corresponding to transitions 1-4 in Fig. 1b, and a component corresponding to the lamp's buffer gas light which contributes to the shot noise at the photodetector. Defining ε as the ratio of total buffer gas light to total rubidium light, we have for the optical spectrum entering the resonance cell

$$I(\nu) = \sum_{i=1}^{4} I_{io} \left\{ \varepsilon\delta(\nu-\nu_{BG}) + \frac{4\ln 2}{\pi\Delta\nu_G^2} \exp\left(-\frac{4\ln 2\,(\nu-\nu_i)^2}{\Delta\nu_G^2}\right)\right\} \qquad (9)$$

where I_{io} is the total entrance intensity of line i, and ν_i is the center frequency of the i^{th} spectral component. The presence of the Dirac delta function $\delta(\nu-\nu_{BG})$ is a simple means of including lamp buffer gas light into the analysis. The I_{io} are determined by the properties of the lamp and the action of the filter cell, both of which have recently received considerable study. [28-32] Our model should be applicable to both the SFC and IFC Rb clock

designs. This flexibility was included by allowing for an isotopic mixture in the resonance cell, and by maintaining the I_{io} as input parameters to the model (so that the I_{io} can represent the direct spectral emission from either the lamp or the lamp/filter cell combination).

2. Transmitted Light

To take into account the effects of optical pumping on the light transmitted by the resonance cell, we write the Bouger-Lambert law for the spectral components as

$$\frac{\partial I_i(\Delta,z)}{\partial z} = -\eta_\Delta(z)\,[Rb^{87}]\,\sigma_i\,I_i(z) \quad (i=1,2) \tag{10a}$$

$$\frac{\partial I_j(\Delta,z)}{\partial z} = -\{(1-\eta_\Delta(z))\,[Rb^{87}] + (\tfrac{7}{12})\,[Rb^{85}]\}\,\sigma_j I_j(z) \quad (j=3,4). \tag{10b}$$

Equation (10b) reflects the fact that spectral components 3 and 4 are absorbed by not only Rb^{87} atoms but also Rb^{85} atoms, which would be present if a IFC design was employed. In these expressions $\eta_\Delta(z)$ is the fractional population in the F=1 hyperfine multiplet at some axial position z in the resonance cell, and for some detuning Δ between the hyperfine transition frequency and the applied microwave field; $[Rb^{87}]$ and $[Rb^{85}]$ are number densities, and the σ_μ are optical absorption cross sections:

$$\sigma_\mu = 2r_o f_\mu c \sqrt{\pi \ell n 2/(\Delta\nu_G^2 + \Delta\nu_D^2)}, \tag{11}$$

where f_μ is the μ^{th} transition oscillator strength, r_o is the classical electron radius and $\Delta\nu_D$ is the doppler broadened absorption linewidth. [33]

We note that the present model of gas cell clock absorption is basically a one dimensional model; that is in cylindrical coordinates we consider the transmitted intensity at a specific axial position z to be equivalent for all r and θ. In actuality, however, we know that this is not the case in presently manufactured designs, because the microwave field strength in the clock cavity can have both a radial and angular dependency. [34,35] Thus, as applied to presently available gas cell frequency standards the present model must be considered as a first order approximation. It is our belief though that corrections due to the radial and angular distribution of microwave field strength within the clock cavity are within the uncertainties generated by the imperfect knowledge of the input parameters required by the model. Furthermore, advanced gas cell frequency standards will very likely employ dielectrically loaded cavities whose microwave fields have no radial or angular dependencies, [36] or photodetectors that only sample transmitted light intensities at specific values of r and θ. [34] For these advanced frequency standards we expect our model to be somewhat more rigorous.

Integrating Eqs. (10), the intensity of the rubidium light at any axial position z is given by

$$I_i(\Delta,z) = I_{io}\exp(-\overline{\eta_\Delta}(z)\,[Rb^{87}]\,\sigma_i z) \quad (i=1,2) \tag{12a}$$

$$I_j(\Delta,z) = I_{jo}\exp(-\{(1-\overline{\eta_\Delta}(z))\,[Rb^{87}] + (\tfrac{7}{12})\,[Rb^{85}]\}\,\sigma_j z \quad (j=3,4) \tag{12b}$$

where $\overline{\eta_\Delta}(z)$ is the fractional population in the F=1 hyperfine multiplet averaged over the vapor length traversed by the light:

$$\overline{\eta_\Delta}(z) \equiv (\tfrac{1}{z})\int_o^z \eta_\Delta(z)\,dz. \tag{13}$$

(For convenience, at $z=\ell$, $\overline{\eta_\Delta}(\ell)$ will simply be denoted as $\overline{\eta_\Delta}$.) Thus, the quantity of interest for determining clock performance becomes

$$I(\Delta,\ell) = \sum_{i=1}^{2} I_{io}\,[\varepsilon + \exp(-\overline{\eta}_\Delta[Rb^{87}]\,\sigma_i\ell)] + \sum_{j=3}^{4} I_{jo}\,[\varepsilon + \exp(-\{(1-\overline{\eta}_\Delta)\,[Rb^{87}] + (\tfrac{7}{12})\,[Rb^{85}]\}\sigma_j\ell). \tag{14}$$

This equation is deceptively simple in appearance; the transmitted light intensity is just the sum of several exponential functions. In actuality, the $\overline{\eta}_\Delta$ terms appearing in the exponents greatly complicate the situation, since they depend on the $I_i(z)$. Thus, the equation for $I(\Delta,\ell)$ is transcendental, and requires special consideration. It is worth noting that the transcendental nature of Eq. (14) is due solely to the fact that under normal clock operating conditions the resonance cell's alkali vapor is optically thick to the rubidium lamp light. For an optically thin vapor $\overline{\eta_\Delta}$ would be a function of the given $I_{\mu o}$.

3. The Fractional Population η_Δ

In the generalized Vanier theory the fractional population in the lower F=b=I-1/2 hyperfine multiplet, where I is the alkali's nuclear spin and considering diffusion phenomenologically, is given by the expression [20]

$$\eta_\Delta' = \{\frac{g_b(A+\gamma_1)}{g_a(B+\gamma_1) + g_a(A+\gamma_1)}\} \times \{\frac{\Gamma_2^2 + \Delta^2 + (\Gamma_2/\Gamma_{1\alpha})\,\omega_1^2}{\Gamma_2^2 + \Delta^2 + (\Gamma_2/\Gamma_{1\beta})\,\omega_1^2}\}, \tag{15}$$

where

$$\Gamma_2 = \frac{(A+B)}{2} + \gamma_2, \tag{16}$$

$$\Gamma_{1\alpha} = (B+\gamma_1)\ [1 + (\frac{g}{4g_b})\ (\frac{B-A}{A+\gamma_1})]^{-1} \tag{17}$$

and

$$\Gamma_{1\beta} = \{\ (A+\gamma_1)\ (B+\gamma_1)\ [g_a B + g_b A + g\gamma_1]\}/$$
$$\{g[(A+B)/2]^2 + A(B-A) + \gamma_1[g((A+B)/2 + \gamma_1)$$
$$+ g_a B + g_b A]\} \tag{18}$$

In these expressions the subscripts a and b refer to the F quantum numbers of the hyperfine multiplets (e.g. a=I+1/2), ω_1 is the microwave Rabi frequency, and the g's refer to the various degeneracies in the alkali ground state: g is the total ground state degeneracy equal to 2(2I+1), g_a is the degeneracy of the upper F=a hyperfine multiplet equal to 2(I+1), and g_b is the degeneracy of the lower hyperfine multiplet equal to 2I (for Rb^{87} g=8, g_a=5 and g_b=3). A and B are defined as optical photon absorption rates for the F=2 and F=1 hyperfine multiplets, respectively. Due to the variation of optical pumping light intensity with axial position the magnitudes of these quantities depend on z; the calculation of A and B is carried out subsequently. The quantities which depend on axial position A, B, and ω_1 result in η'_Δ also displaying an axial dependence. However due to the inclusion of the effects of atomic axial diffusion the quantity of interest η_Δ has a different axial dependence than η'_Δ, though as shown below there is a relatively simple expression relating them.

Probably the stickiest issue for the generalized Vanier theory to handle is that of atomic diffusion. In uncoated resonance cells after an alkali atom hits the wall, the population density distribution among all of the atom's ground state Zeeman sublevels is randomized. [37] Thus, the measure of hyperfine polarization $\langle \hat{I}\cdot\hat{S}\rangle$ is zero at the resonance cell walls. Rigorously, a term $D\nabla^2\rho$ must be included into the density matrix rate equations in order to describe diffusion, where D is the diffusion coefficient of Rb atoms in the resonance cell's vapor. [38] However, in the generalized Vanier theory diffusion is considered as a phenomenological relaxation mechanism, contributing a term γ_{dif} to both γ_1 and γ_2. If R is taken as a measure of the resonance cell's dimensions, typically the resonance cell's radius, then $\gamma_{dif} = D/R^2$. This procedure for including diffusion into the determination of η_Δ is actually quite reasonable, since if no other relaxation mechanisms were present η_Δ would relax in the dark (ie. in the absence of optical pumping light and microwaves) with a rate of this order of magnitude. [39] However, diffusion is fundamentally different from other "bulk" relaxation mechanisms like Rb-buffer gas and Rb-Rb collisions, since in steady state diffusion leads to a spatial distribution of $\langle \hat{I}\cdot\hat{S}\rangle$; the bulk relaxation mechanisms only effect the magnitude of $\langle \hat{I}\cdot\hat{S}\rangle$ not its spatial distribution.

Minguzzi et al. [40] have shown that the spatial distribution of $\langle \hat{I}\cdot\hat{S}\rangle$ is in general characterized by a sum of diffusion modes. In the case of the gas cell clock, though, where the optical pumping rate is low and the resonance cell is fairly uniformly illuminated, a reasonable approximation is to consider only the first order mode. Thus, we have for the axial distribution of $\langle \hat{I}\cdot\hat{S}\rangle$

$$\langle \hat{I}\cdot\hat{S}\rangle\ (z) = \langle \hat{I}\cdot\hat{S}\rangle_p \sin(\pi z/\ell). \tag{19}$$

Since η_Δ is simply related to $\langle \hat{I}\cdot\hat{S}\rangle$ by the equation [41]

$$\eta_\Delta = \frac{1}{2} - \frac{4\langle \hat{I}\cdot\hat{S}\rangle + 1}{g}, \tag{20}$$

we have for the spatial distribution of η_Δ, considering only the first order axial diffusion mode,

$$\eta_\Delta(z) = \frac{g_b}{g} + (\eta'_\Delta - \frac{g_b}{g})\ \sin(\frac{\pi z}{\ell}). \tag{21}$$

In essence what we have done is to decompose the contributions of diffusion to η_Δ into two parts. In the first part we assume that diffusion has some limiting effect on the magnitude of η_Δ, and that this can be incorporated into the generalized Vanier theory through contributions to γ_1 and γ_2. In the second part we recognize the fact that the final form of the axial distribution of η_Δ is to some degree known, and so a functional form can be superimposed onto the solution η'_Δ. In this way the major physical aspects of diffusion are brought into the model without recourse to cumbersome and obscure expressions.

The bulk relaxation mechanisms which contribute to γ_1 and γ_2 are typically buffer gas and spin exchange collisions. Additionally, for completeness one should include the slow down factor, (6I+1)/(8I+4), into the spin exchange contribution to the transverse relaxation rate γ_2. 25 Thus, defining σ_{BG} and σ_{ex} as the buffer gas relaxation and spin exchange cross sections respectively, [42,43] the phenomenological relaxation rates γ_1 and γ_2 become

$$\gamma_1 = D/R^2 + [BG]\ \bar{v}\sigma_{BG} + [Rb]\bar{v}\sigma_{ex} \tag{22a}$$

and

$$\gamma_2 = D/R^2 + [BG]\bar{v}\sigma_{BG} + (\frac{6I+1}{8I+4})\ [Rb]\bar{v}\sigma_{ex} \tag{22b}$$

where [BG] is the number density of buffer gas molecules or atoms, and $\bar{v}$ is the average relative velocity between colliding species.

Finally, by assuming that the Rb atoms are essentially frozen in place due to the presence of the buffer gas, [35,44,45] one can determine $\eta'_\Delta(z)$ by substitution of the position dependent photon absorption rates, A and B along with the suitable expression describing $\omega_1(z)$ into Eq. (15). For typical gas cell frequency standards employing a TE_{011} or TE_{111} microwave cavity this expression is simply [46]

$$\omega_1(z) = \omega_{1p} \sin\left(\frac{\pi z}{\ell}\right) \tag{23}$$

where ω_{1p} is the peak microwave Rabi frequency in the cavity.

4. Determination of the Dependence of Photon Absorption Rates on Axial Position

To calculate the positional dependencies of the optical photon absorption rates we first assume that the axial optical intensity distribution can be approximated by a Bouger-Lambert law of the form

$$I_i(\zeta,z) = I_{io} \exp\left(-\zeta[Rb^{87}]\,\sigma_i z\right) \tag{24a}$$

$$I_j(\zeta,z) = I_{jo} \exp\left(-\{(1-\zeta)[Rb^{87}] + \left(\frac{7}{12}\right)[Rb^{85}]\}\sigma_j z\right), \tag{24b}$$

where ζ is to be thought of as a resonance cell parameter describing the "global" fractional population in the F=1 hyperfine multiplet as a result of optical pumping. In effect ζ is a zeroth order approximation to $\eta'_\Delta(z)$. For the low optical pumping rates expected from lamps, this approximation should be a reasonable description of reality; since in the limit of negligible optical pumping $\bar{\eta}_\Delta(z)$ in Eqs. (12) is independent of the light intensity, and the standard Bouger-Lambert law is recovered. We note, however, that recent theoretical work has indicated that this approximation may not be generally valid. [22,23] Equations (24) will allow us to calculate the positionally dependent optical photon absorption rates,

$$A(\zeta,z) = \left(\frac{1}{h\nu}\right) \sum_{j=3}^{4} I_j(\zeta,z)\sigma_j \tag{25a}$$

$$B(\zeta,z) = \left(\frac{1}{h\nu}\right) \sum_{i=1}^{2} I_i(\zeta,z)\sigma_i \tag{25b}$$

as functions of ζ.

The task which now needs to be discussed is the determination of ζ. Considering Eqs. (14) and (24), it is clear that any consistent choice of ζ must satisfy the equation

$$\sum_{k=1}^{4} I_k(\zeta,\ell) = \sum_{i=1}^{2} I_{io} \exp\left(-\bar{\eta}_\infty [Rb^{87}]\sigma_i \ell\right) + \sum_{j=3}^{4} I_{jo} \exp\left(-\{(1-\bar{\eta}_\infty)[Rb^{87}] + \left(\frac{7}{12}\right)[Rb^{85}]\}\sigma_j \ell\right), \tag{26}$$

and the simpliest choice of ζ which satisfies this requirement is

$$\zeta = \bar{\eta}_\infty = \frac{1}{\ell}\int_0^\ell \eta_\infty \, dz. \tag{27}$$

Substituting Eqs. (15) and (21) into Eq. (27), our expression for ζ becomes

$$\zeta - \frac{g_b}{g} + \frac{g_b g_a}{\ell g}\int_0^\ell \left\{\frac{[B(\zeta)-A(\zeta)]\sin\left(\frac{\pi z}{\ell}\right)}{g_a[B(\zeta)+\gamma_1] + g_b[A(\zeta)+\gamma_1]}\right\} dz = 0, \tag{28}$$

where we have explicitly noted the fact that A and B depend parametrically on ζ. Using a Newton-Raphson method, [47] Eq. (28) can be solved numerically for ζ. A discussion of the results of this procedure for determining ζ and their consistency will be presented below.

Summarizing, with the determination of ζ, A(z) and B(z) are specified and when substituted into Eq. (15) determine η'_Δ. η_Δ and $\bar{\eta}_\Delta$ follow directly from η'_Δ via Eqs. (21) and (13), respectively. Substituting $\bar{\eta}_\Delta$ into Eq. (14) yields the transmitted light intensity $I(\Delta,\ell)$, and the optically detected hyperfine lineshape, $S(\Delta) = I(\infty,\ell)-I(\Delta,\ell)$.

C. INTEGRATION OF ATOMIC PHYSICS WITH FEEDBACK ANALYSIS

It is a straightforward matter to extract the quantities required by the feedback analysis from the preceeding study of the physics occurring in the absorption cell. If we take κ and A_p as the clock photodiode's sensitivity (amps/W) and area, which is assumed to be identical to the clock cavity's cross-sectional area, the total lamp induced photocurrent is given by,

$$i_o = \kappa A\, I(\infty,\ell). \tag{29}$$

The slope of the discrimination pattern is given by

$$m = \delta_m\, I''(0,\ell) \tag{30}$$

where $2\delta_m$ is the modulation depth of the microwave frequency and $I''(0,\ell)$ is the second derivative of the transmitted light intensity evaluated at the center frequency of the hyperfine resonance. By equating the slope of the discriminator pattern with $\delta_m I''(0,\ell)$ we have made the small-modulation-depth/quasi-static approximation with regards to the effect of the microwave frequency modulation on the signal lineshape. [7]. As long as we confine ourselves to frequency modulation regimes such that $\delta_m/\Gamma_2 \ll 1$ and $\omega_M/\Gamma_2 \gg 1$, where ω_M is the modulation frequency, then the recent work of Audoin et al. [19] would indicate that this approximation is reasonably valid. As will be demonstrated in the following paragraphs the exclusion of higher order modulation effects does not appear to have affected the ability of our model to accurately predict the performance of current generation Rb clocks. It is not clear though, that this would be true under all potential clock operating conditions. Should clock conditions vary significantly from

nominal values used in current standards the ramifications of microwave modulation would have to be reconsidered. However, these effects could be approximated in the present model by using the work of Audoin et al.[19] to determine a correction factor for Eq. (30). In our calculations we take $2\delta_m$ as being equal to the frequency separation of the inflection points of the approximately Lorentzian resonance lineshape and assume that $\omega_M < \delta_m$.

III. MODEL VERIFICATION

In the model of the clock signal lineshape presented above quite a few assumptions were made in order to make the problem tractable. Several of these assumptions are inherent to the generalized Vanier theory, and have been validated by recent experimentation; [20] these can be considered as microscopic assumptions, since they are related to the statistical behavior of individual atoms. The other assumptions in the model (i.e. the functional form of the axial light intensity distribution, and the axial distribution of n_Δ as a result of diffusion) should be considered as macroscopic assumptions, since they are related to the alkali vapor as a whole; and with these we are on less firm ground experimentally. Thus, before proceeding to calculations of clock performance it would be prudent to explore the validity of these assumptions. This will be accomplished by showing that the model in its present form leads to qualitatively correct predictions of the global fractional population ζ, the axial distribution of $\langle \vec{I} \cdot \vec{S} \rangle$ and the signal lineshape as functions of temperature.

The global fractional population ζ is actually a quite useful parameter for understanding gas cell clock performance, since it reflects the degree of optical pumping, and hence the clock signal amplitude, for the resonance cell as a whole. Qualitatively, one expects ζ to depend on temperature in two ways: i) as the resonance cell temperature rises the global optical pumping rate necessarily changes, and ii) the longitudinal relaxation rate is temperature dependent as a result of spin exchange. Thus, for high temperatures one expects ζ to approach its statistical value g_b/g. At low temperatures, when the vapor is optically thin, one should be able to arrive at a relatively simple expression for ζ, since in this case one can assume that A and B are constants independent of axial position. Regarding Eq. (28) in this optically thin regime one has

$$\zeta_{thin} = \frac{g_b}{g}\left\{1 - \frac{2g_a(B-A)}{\pi[g_a(B+\gamma_1) + g_b(A+\gamma_1)]}\right\}, \qquad (31)$$

which is still a temperature dependent quantity because of γ_1.

If we now consider the case of optical pumping out of only the F=b hyperfine multiplet as a specific example, the above arguments would require us to make the following predictions regarding ζ: i) at low resonance cell temperatures ζ should be relatively small; ii) at high temperatures ζ should saturate at 3/8 in the case of Rb; and iii) for intermediate temperatures one should have $\zeta_{correct} > \zeta_{thin}$, since ζ_{thin} increases as a function of temperature because of only one mechanism while $\zeta_{correct}$ increases as a result of two cooperating mechanisms. A sample calculation of ζ over the temperature range 10-120°C is presented in Fig. 3. The qualitative agreement between our expectations and the calculations lends confidence to the interpretation and use of ζ in the clock signal model.

Since the atoms are essentially frozen in place due to the presence of the buffer gas, the maximum change in light intensity per unit length (for (microwaves on and off resonance) occurs in a unique spatial region of the resonance cell determined by the light intensity distribution, the microwave field distribution and diffusion. The ability to determine this unique spatial region is critically important for a model of gas cell clock performance, since the microwave field strength and light intensity in this region have the primary influence on both the clock's Q and signal to noise ratio. Furthermore, variations in this position with microwave power and light intensity are responsible for the so called position shift [44] and inhomogeneous light shift, [45] respectively.

As previously discussed, the present model is a one dimensional model. Thus, if it is to be trusted in its predictions of clock performance, it should be able to predict the qualitatively correct axial distribution of $\langle \vec{I} \cdot \vec{S} \rangle$ as a function of resonance cell temperature. For the case of an optically thin cell and relatively low optical pumping light intensities one would expect a distribution of $\langle \vec{I} \cdot \vec{S} \rangle$ peaked in the center of the resonance cell. As the resonance cell temperature was raised, and the cell became more opaque to the optical pumping light, one would expect the position corresponding to the maximum $\langle \vec{I} \cdot \vec{S} \rangle$ to shift towards the front of the resonance cell. In Fig. 4 we have plotted the model's predictions of $\langle \vec{I} \cdot \vec{S} \rangle$ as a function of axial position using the parameters listed in Table I. We note that the curves are very similar to those of Minguzzi et al., [40] where the effects of diffusion were analyzed more rigorously, and are again in qualitative agreement with our expectations.

Of course the most critical test for our macroscopic assumptions concerns the signal lineshape. Experimentally, it is known that the transmission signal resonance is very well approximated by a Lorentzian, being however somewhat more sharply peaked. [48] Regarding the model, there is no reason a priori to expect a lineshape resembling anything like a Lorentzian; yet if this is not predicted over a reasonable range of operating temperatures, the validity of the model must be questioned. In Figs. 5a and 5b we show the signal lineshape calculated at 20 and 100°C, respectively, using the parameters listed in table I; the dots are model calculations, and the solid curve is a least squares fit of the model results to a Lorentzian. As is readily apparent the model gives excellent agreement in the wings of the lineshape, but near resonance is more sharply peaked as expected. These predictions are in qualitative agreement with the experimental results of Camparo and Frueholz. [48] Thus, our macroscopic assumptions appear to be valid, at least over normal clock operating conditions, allowing the final test of the model: a comparison between predicted and measured Rb gas cell clock performance.

Recently, Riley [9] and Lynch and Riley [10] performed a stability test on a newly designed Rb frequency standard. This test was somewhat unique in that many of the input parameters required by the present model of clock performance were documented.

Thus, these results can be used to establish the accuracy of the present model in predicting clock performance. The parameters required for the calculation are listed in table I, along with reasonable estimates of their values based on the information presented by Riley, or our own measurements on similar components of the Rb clock. The uncertainties presented with some of the more critical parameters represent our best estimates of the inaccuracies associated with extracting the parameters from the information presented by Riley. In particular, the uncertainty associated with the resonance cell temperature represents the difficulty associated with measuring liquid alkali metal temperatures accurately, and the difficulty of accurately converting these liquid temperatures to saturated alkali vapor pressures. The only parameter that could not be determined by the information presented by Riley was the peak Rabi frequency. This was therefore evaluated by matching the calculated signal linewidth with Riley's measured signal linewidth. We note, however, that this parameter can be determined experimentally quite easily by the ARP method of Frueholz and Camparo. [35] The choice of buffer gas was not crucial, and so for convenience nitrogen at 10 torr was used.

Table II and Fig. 6 show the results of our calculations along with uncertainties in the calculated quantities which are associated with the uncertainties in the input parameters. Several points concerning the predicted results are worth noting. First, the predicted Q and discriminator slope agree very well with the measured values, especially since the measured discriminator slope is probably a bit too large; based on the measured Allan variance and DC photocurrent one would have expected to measure a discriminator slope of 268 pA/Hz. Furthermore, the results indicate that the line Q is relatively insensitive to reasonable variations in the global optical pumping rate, but that the signal amplitude is very sensitive to these changes. This is precisely the same conclusion reached by Matsuda et al. in their experimental studies of the signal characteristics of the Rb gas cell clock. [49] Finally, the calculated Allan variance agrees very well with that measured by Lynch and Riley [10] as shown by Fig. 5. Thus, within the uncertainties associated with the input parameters we have demonstrated the validity of a model of gas cell clock performance that is completely independent of measurements of the clock signal's Q or signal to noise ratio.

Vanier et al have carried out extensive analyses of the effect of loop attack time on clock performance. As an additional demonstration of the model capabilities, the Allan variances for two different servo-control loop attack times are calculated. The crystal oscillator is assumed to have a power spectral density of fractional frequency fluctuations of [17]

$$S_y^{cr}(f) = \frac{6 \times 10^{-26}}{f} + 1 \times 10^{-26} f + 2 \times 10^{-28} f^2, \quad (32)$$

while the atomic system behaves as previously calculated. Two attack times, 0.1 and 10s, are considered; the resulting Allan variances are displayed in Fig. 7. These results were calculated quite efficiently due to the analytical evaluation of the intergrals of Eq. (8). As an aside, we note that the present analysis of the hyperfine lineshape, based on the quasi-static approximation, will not be valid at extremely short attack times. At very short attack times the Rb atoms will no longer be able to respond adiabatically to the microwave frequency variation induced by the control loop; furthermore, at short attack times atomic coherence effects should be considered. For typical gas cell frequency standards a reasonable limit to our models validity would be an attack time of 0.01s.

IV. CONCLUSIONS AND SUMMARY

In practice a clock model's usefulness is intimately connected to the realization of a device with optimal performance characteristics for a specific application. Thus, though reasonable accuracy in model predictions is required, equally important attributes of a clock model are its ability to: i) analyze design trade-offs for a specific application, ii) diagnose the clock parameters that limit performance, and iii) predict the avenues for further improvement. Since the present model is inherently numerically faster than its predecessors, as a result of the analytic evaluation of the spectral density integrals and density matrix inversion, design trade-offs can be quickly evaluated. Furthermore, since the model treats the physics of the resonance cell signal in a semi-rigorous yet perspicuous manner, diagnoses of performance limits and predictions for improvements are in terms of physically meaningful and conceptually lucid parameters.

As an example of the present model's ability to predict possible avenues of improvement for the Rb gas cell clock, we have computed the short term performance for a clock with the parameters listed in table I as a function of microwave power, or what is equivalent the peak Rabi frequency in the clock cavity; these results are shown in Fig. 8. As is clear from the figure, the model predicts that a short term stability of $6 \times 10^{-13}/\sqrt{\tau}$ could be achieved by reducing the 6835 MHz microwave power fed into the cavity by 23 db. Of course these predictions must be tempered by the realization that for these Rabi frequencies one might not be in the small-modulation-depth/quasi-state regime with regards to the discriminator slope and that the microwave field strength displays only a longitudinal variation within the absorption cell. However, they do indicate a possible avenue for improvement that in the final analysis must be experimentally verified.

Finally, with a straightforward modification the microwave detuning in Eq. (9) can be made to depend on: local magnetic field strengths, local optical pumping light intensities, the buffer gas pressure, the alkali vapor pressure, etc. The model lineshape would then exhibit the inhomogeneous nature that results in the position shift effect [44] and the inhomogeneous light shift. [45] With the inclusion of these resonant frequency shift terms the model could be used to investigate possible mechanisms of gas cell clock frequency drift, which is a principle impediment to the gas cell frequency standard's more widespread use.

ACKNOWLEDGEMENT

The authors wish to thank L. G. Redekopp of the University of Southern California for aid in evaluating the contour integrals. This work was sponsored by the United States Air Force Space Division under contract No. F04701-83-C-0084.

Appendix I. Results of Complex Contour Integration

When extracting the Allan variance from the spectral density of fractional frequency fluctuations

integrals of the form

$$\Lambda(\tau) = \int_0^\infty \frac{f^\alpha}{1 + (f/f_n)^2} \sin^4(\pi\tau f)\, df \qquad (A1)$$

with

$$\alpha = -4, -2, -1, 1, 2$$

must be evaluated. Evaluation of these integrals may be carried out using complex contour techniques. To facilitate evaluation, $\sin^4(\gamma f)$ was written in terms of complex exponentials. For the even integrands the contour chosen extended from $-R$ to $+R$ along the real axis with a semi-circular indention around origin when required. The contour was closed along the upper half of the circle $|z| = R$ in the complex plane with R ultimately allowed to go to infinity. For odd integrands the contour was composed of three segments. The first extending from $+iR$ to zero along the imaginary axis with semi-circular indentions about singularities. The second from 0 to $+R$ along the real axis followed by a quarter segment of the circle $|z| = R$ connecting the other segments. Again R was allowed to go to infinity. In evaluating the integral for $\alpha = -1$ it was found to be expeditious to first differentiate the integrand with respect to $\gamma(=\pi\tau)$ eliminating the singularity at the origin. The resulting integral was evaluated and then indefinite integration with respect to γ was performed.

The results of this analysis are as follows

$$\alpha = -4: \quad \Lambda(\tau) = \frac{\pi}{4}\left[\frac{4}{3}(\pi\tau)^3 - \frac{\pi\tau}{f_n^2} + \frac{1}{f_n^3}\left(\frac{e^{-4\pi\tau f_n}}{4} - e^{-2\pi\tau f_n} + 3/4\right)\right], \qquad (A2)$$

$$\alpha = -2: \quad \Lambda(\tau) = \frac{\pi^2\tau}{4} - \frac{\pi}{16 f_n}\left[e^{-4\pi\tau f_n} - 4e^{-2\pi\tau f_n} + 3\right], \qquad (A3)$$

$$\alpha = -1: \quad \Lambda(\tau) = \frac{1}{4f_n}\left[F(4f_n\pi\tau) - 2F(2f_n\pi\tau)\right], \qquad (A4a)$$

with

$$F(c\pi\tau) = \frac{e^{c\pi\tau}}{c}\left[Ei(-c\pi\tau) - e^{-c\pi\tau}Ei(c\pi\tau) + \sum_{n=1}^{A}(-1)^n \sum_{k=1}^{n}(-1)^k \frac{(c\pi\tau)^{n-k}}{n\cdot(n-k)!}\right] + \frac{e^{-c\pi\tau}}{c}\left[Ei(c\pi\tau) - e^{c\pi\tau}Ei(-c\pi\tau) + \sum_{n=1}^{B}\sum_{k=1}^{n}\frac{(c\pi\tau)^{n-k}}{n\cdot(n-k)!}\right] - \frac{1}{c}\left[\sum_{n=1}^{A}\frac{1}{n} + \sum_{n=1}^{B}\frac{1}{n}\right], \qquad (A4b)$$

for $c\pi\tau < 20$. $Ei(x)$ is the exponential integral function[50] and A and B are chosen to insure convergence of the respective sums. When $c\pi\tau > 20$ various asymptotic expansions may be employed in deriving $F(c\pi\tau)$ to yield,

$$F(c\pi\tau) = \frac{-2}{c}\left[\ln(c\pi\tau) - \sum_{\substack{k=3 \\ k\ \text{odd}}}^{M}\frac{(k-2)!}{(c\pi\tau)^{k-1}}\right], \qquad (A4c)$$

with M chosen to insure series convergence to desired level of accuracy.

When α equals either +1 or +2 the integral of Eq. (A1) is not finite. One way this can be overcome is by limiting the range of integration to a maxmium frequency. Physically this represents the maximum frequency to which the measurement system responds. An alternative means of addressing this problem, preserving analytic evaluation of the integrals, is to modify the integrand in such a manner that it goes smoothly to zero above a cut-off frequency f_c. This is carried out by alterring Eq. (A1) to give

$$\Lambda(\tau) = \int_0^\infty \frac{f^\alpha}{1 + (f/f_n)^2}\,\frac{1}{1 + (f/f_c)^2}\sin^4(\pi\tau f)df. \qquad (A5)$$

Evaluating Eq. A5 for $\alpha = +1$ and +2 yields,

$$\alpha = +1: \quad \Lambda(\tau) = \frac{(f_c)^2}{16(f_c^2 - f_n^2)}\left[G(4\pi f_c\tau) - G(4\pi f_n\tau) - 4\left\{G(2\pi f_c\tau) - G(2\pi f_n\tau)\right\} + 6\ln(f_c/f_n)\right] \qquad (A6)$$

with $G(x) = e^{-x}\,Ei(x) + e^{x}\,Ei(-x)$,

and

$$\alpha = +2: \quad \Lambda(\tau) = \frac{\pi(f_c)^2}{16(f_c^2 - f_n^2)}\left\{f_c\left(e^{-4\pi f_c\tau} - 4e^{-2\pi f_c\tau} + 3\right) - f_n\left(e^{-4\pi f_n\tau} - 4e^{-2\pi f_n\tau} + 3\right)\right\}. \qquad (A7)$$

Table I. Parameters for Clock Signal Model

Parameter	Value[a]	Uncertainty
Linewidth of Lamp Lines	2 GHz	$\pm$ 0.2 GHz
Temperature	65°C	$\pm$ 5°C
Photodiode Responsivity	0.5 amps/watt	
Ratio of Buffer Gas Light to Rb Resonance Light	0.64	
$P_o(D_1,F=1)$	.029 mW	$\pm$ 5%
$P_o(D_2,F=1)$	.046 mW	
$P_o(D_1,F=2)$	.018 mW	$\mp$ 5% (c)
$P_o(D_2,F=2)$	.029 mW	
Peak Rabi Frequency ω_{1p}	450 Hz	
Buffer Gas	N_2	
Buffer Gas Pressure	10 torr	
Light Shift Coefficient	-3×10^{-10} (b)	$\pm 1 \times 10^{-10}$
Lamp Intensity Spectral Density	$1.2 \times 10^{-9} + \frac{1.3 \times 10^{-13}}{f^2}$ (b)	

(a) W. Riley Ref. 9

(b) C. H. Volk and R. P. Frueholz Ref. 16

(c) Optical powers were varied in the calculations according to the rule that if $P_o(D_1,F=1)$ changed by $\pm$ 5%, then simultaneously $P_o(D_1,F=2)$ changed by $\mp$ 5%.

Table II: Comparison of Predicted Clock Performance with Measured Clock Performance

Quantity	Predicted Valve	Measured Valve
Fractional Population ζ	0.359	
Fraction Population Change: $\bar{\eta}_o - \zeta$	4.3×10^{-3}	
DC Photocurrent	81.8 μA	82 μA
Photocurrent Change: $i_{dc} - i_{\Delta=0}$	$(45.7^{+35\%}_{-30\%})$ nA	
Q:	$(1.63^{+1.2\%}_{-1.8\%}) \times 10^7$	1.6×10^7
Discriminator Pattern Slope: $\delta_m I''(0,\ell)$	(251^{+93}_{-74}) pA/Hz	351 pA/Hz
Short Term Allan Variance: $\sigma_y(\tau)$	$(3.1^{+1.1}_{-0.9}) \times 10^{-12}/\sqrt{\tau}$	$2.8 \times 10^{-12}/\sqrt{\tau}$

REFERENCES

[1] See for example J. M. Andres, D. J. Farmer, and G. T. Inouye, IRE Trans. Mil. Elec. MIL-3, 178 (1959); R. J. Carpenter, E. C. Beaty, P. L. Bender, S. Saito and R. O. Stone, IRE Trans. Instrum. I-9, 132 (1960); M. E. Packard and B. E. Swartz, IRE Trans. Instrum. I-11, 215 91962).

[2] E. Jechart, in the Proceedings of the Twenty-Seventh Annual Symposium on Frequency Control, USAEC Fort Monmouth, NJ (Springfield, Virginia, National Technical Information Service, 1973), p. 387.

[3] J.-P. Gourber, Ann. Radioele. 20, 191 (1965).

[4] R. F. Lacey, A. L. Helgesson and J. H. Holloway, Proc. of IEEE 54, 170 (1966).

[5] H. Fukuyo, K. Iga, N. Kuramochi, I. Matsuda and H. Ishikawa, Bull. Tokyo Inst. Tech. Number 107, 75 (1971).

[6] G. Missout and J. Vanier, Can. J. Phys. 53, 1030 (1975).

[7] J. Vanier and L.-G. Bernier, IEEE Trans. Instrum. Meas. IM-30, 277 (1981).

[8] J. C. Camparo, "The Diode Laser in Atomic Physics", submitted to Contemporary Physics.

[9] W. J. Riley, in Proceedings of the Thirteenth Annual Precise Time and Time Interval (PTTI) Application and Planning Meeting, NASA Conference Publication 2220 (NASA, Greenbelt, MD, 1981), p. 609.

[10] T. J. Lynch and W. J. Riley, in Proceedings of the Fifteenth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, (Naval Research Laboratory, Washington, DC, 1984, available from NTIS, ADA-149-163), p. 269.

[11] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leeson, T. E. McGunigal, J. A. Mullen, Jr., W. L. Smith, R. L. Sydnor, R. F. C. Vessot and G. M. R. Winkler, IEEE Instrum. Meas. IM-20, 105 (1971).

[12] L. S. Cutler and C. L. Searle, Proc. IEEE 54, 136 (1966).

[13] J. Vanier, M. Tetu, and L. G. Bernier, IEEE Trans. Instrum. Meas. IM-28, 188 (1979).

[14] A. Kastler, J. Opt. Soc. Am. 53, 902 (1963).

[15] B. S. Mathur, H. Tang, and W. Happer, Phys. Rev. 171, 11 (1968).

[16] C. H. Volk and R. P. Frueholz, J. Appl. Phys. 57, 980 (1985).

[17] C. Audoin and J. Vanier, J. Phys. E. Sci. Instrum. 9, 697 (1976).

[18] W. E. Bell and A. L. Bloom, Phys. Rev. 107, 1559 (1957).

[19] C. Audoin, J. Viennet, N. Cyr and J. Vanier, in the Proceedings of the Fourteenth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, NASA Conference Publication 2265 (NASA, Greenbelt, MD, 1983), p. 87.

[20] J. C. Camparo and R. P. Frueholz, Phys. Rev. A 31, 1440 (1985).

[21] V. V. Batygin, M. B. Gornyi and B. G. Matisov, Sov. Phys. Tech. Phys. 26, 1314 (1982).

[22] M. B. Gornyi, D. L. Markman and B. G. Matisov, Opt. Spectrosc. 55, 20 (1983).

[23] M. B. Gornyi and B. G. Matisov, Opt. Spectrosc. 56, 380 (1984).

[24] M. B. Gornyi and B. G. Matisov, Sov. Phys. Tech. Phys. 28, 25 (1983).

[25] J. Vanier, Phys. Rev. 168, 129 (1968).

[26] F. A. Franz, Phys. Rev. 141, 105 (1966).

[27] At the present time there appear to be no compelling reasons for considering relaxation more rigorously; see Ref. 20.

[28] T. Tako, Y. Koga, I. Hirano and M. Ohi, Jap. J. Appl. Phys. 14, 1641 (1975).

[29] N. Kuramochi, H. Fukuyo, I. Matsuda and N. Shimoi, Jap. J. Appl. Phys. 15, 949 (1976).

[30] N. Kuramochi, T. Matsuo, I. Matsuda and H. Fukuyo, Jap. J. Appl. Phys. 16, 673 (1977).

[31] N. Kuramochi, I. Matsuda and H. Fukuyo, J. Opt. Soc. Am. 68, 1087 (1978).

[32] J. Vanier, R. Kunski, P. Pavlin, M. Tetu and N. Cyr, Can. J. Phys. 60, 1396 (1982).

[33] Note that for an isotopically mixed resonance cell we assume negligible optical pumping of the Rb^{85} isotope. Considering the low light levels involved, this is probably a reasonable first order approximation.

[34] P. Tremblay, N. Cyr and M. Tetu, submitted to Can. J. Phys.

[35] R. P. Frueholz and J. C. Camparo, J. Appl. Phys. 57, 704 (1985).

[36] H. E. Williams, T. M. Kwon and T. McClelland, in the Proceedings of the 37th Annual Frequency Control Symposium, (IEEE Press, New York, 1983), p. 12.

[37] W. Franzen, Phys. Rev. 115, 850 (1959).

[38] W. Happer, Rev. Mod. Phys. 44, 169 (1972).

[39] F. A. Franz, Phys. Rev. A 6, 1921 (1972).

[40] P. Minguzzi, F. Strumia and P. Violino, Nuovo Cimento 46B, 145 (1966).

[41] L. C. Balling, Optical Pumping, in Advances in Quantum Electronics Vol. 3 (Academic Press, London, 1975), pg. 1.

[42] F. A. Franz and C. H. Volk, Phys. Rev. A 14, 1711 (1976).

[43] N. W. Ressler, R. H. Sands and T. E. Stark, Phys. Rev. 184, 102 (1969).

[44] A Risley and G. Busca, in the Proceedings of the 32 Annual Frequency Control Symposium (Electronics Industries Association, Washington, D.C., 1978), p. 506.

[45] J. C. Camparo, R. P. Frueholz and C. H. Volk, Phys. Rev. A 27, 1914 (1983).

[46] J. D. Jackson, Classical Electrodynamics (Wiley, New York, 1962), p. 256.

[47] R. H. Pennington, Introductory Computer Methods and Numerical Analysis (Collier-MacMillan Canada, Toronto, 1970).

[48] J. C. Camparo and R. P. Frueholz, Phys. Rev. A 30 803 (1984).

[49] I. Matsuda, N. Kuramochi, N. Shiomi and H. Fukuyo, Jap. J. Appl. Phys. 16, 391 (1977).

[50] Handbook of Mathematical Functions, edited by M. Abramowitz and I. A. Stegun (Dover Pulbications, Inc., 1965) pp. 227-251.

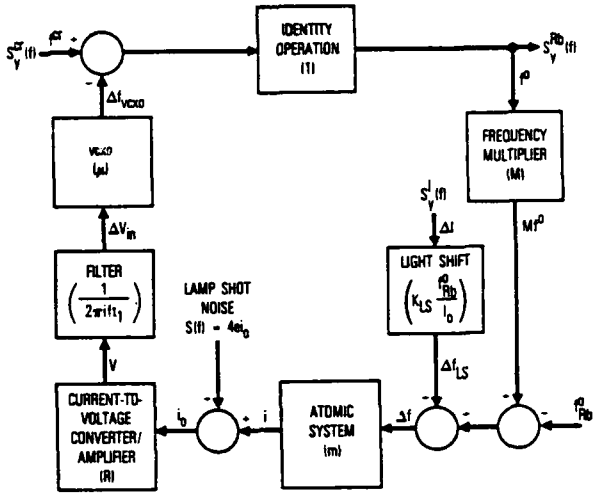


Figure 2. Block diagram of the gas cell frequency standard employed in servo-loop feedback analysis. The transfer function of each element is in parentheses.

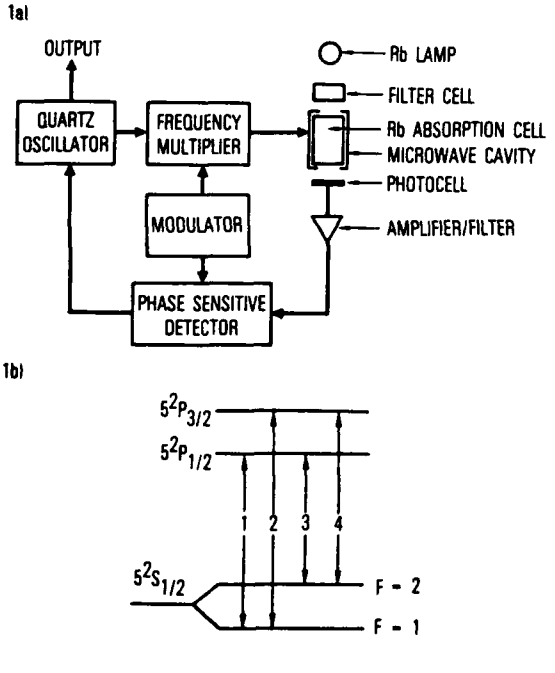


Figure 1. a) Schematic diagram of a passive rubidium gas cell frequency standard.

b) Energy level diagram of Rb^{87} showing the radiative transitions of interest.

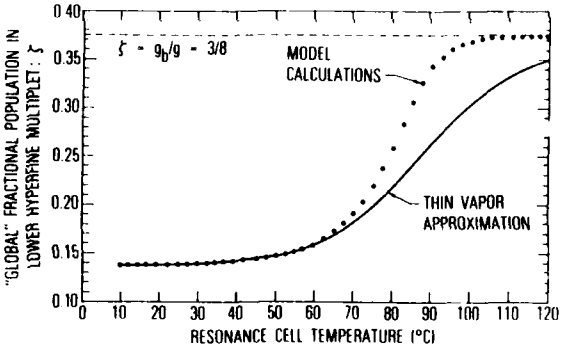


Figure 3. Global fractional population ζ in the F=b=I-1/2 hyperfine multiplet as a function of resonance cell temperature for the case of optical pumping out of only this multiplet. The solid line is a thin vapor approximation where the spin exchange rate is the only temperature dependent quantity; dots correspond to the actual case where the optical thickness of the vapor must also be considered. In the limit of no optical pumping the equilibrium value of ζ is 3/8.

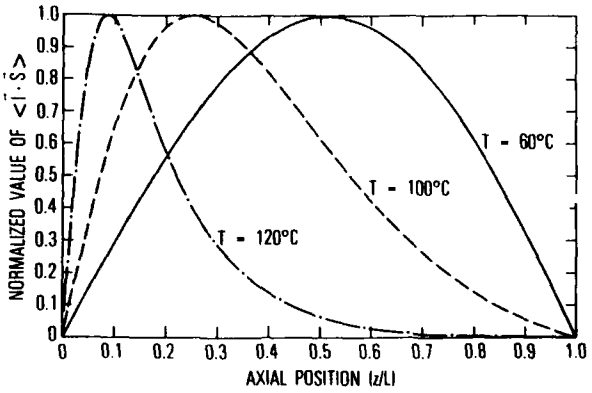


Figure 4. The measure of hyperfine polarization $\langle \vec{I} \cdot \vec{S} \rangle$ has an axial spatial distribution that is temperature dependent. For relatively low resonance cell temperatures the spatial distribution is primarily determined by diffusion. As the resonance cell's alkali-metal vapor becomes progressively more optically thick, efficient optical pumping becomes confined to regions closer to the entrance window of the resonance cell.

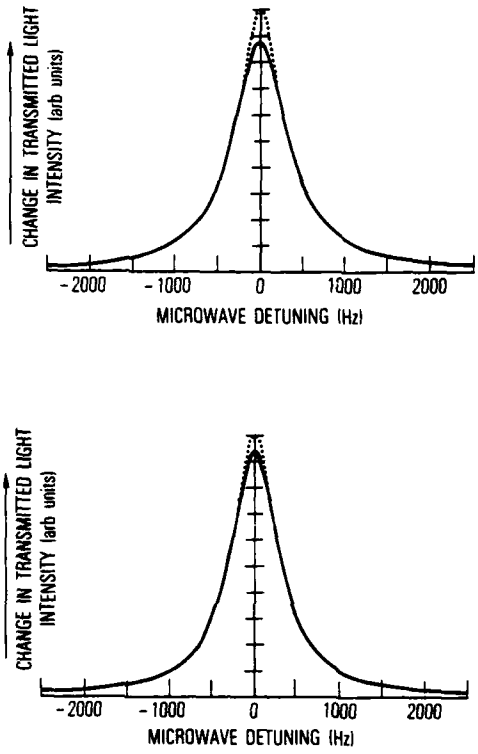


Figure 5. Optically detected hyperfine transition lineshape: a) T=20°C, b) T=100°C. The dots are the calculated change in transmitted light intensity as a function of microwave frequency; the solid curve is a least squares Lorentzian fit. Note that at reasonable resonance cell temperatures the lineshapes are well approximated by a Lorentzian; they are however somewhat more sharply peaked.

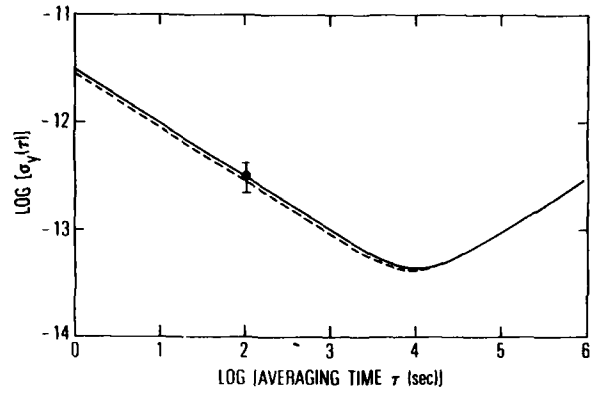


Figure 6. Frequency stability of the EG&G Rb gas cell frequency standard in terms of Allan variance $\sigma_y(\tau)$. The dashed line is the standard's experimental performance while the solid line is the predicted performance yielded by the model. The error bar represents the uncertainty in the standard's predicted performance due to uncertainties in the input parameters. For averaging times less than 10^4s the standard's performance is limited by shot noise at the photodetector. When the averaging time is greater than 10^4s clock performance is limited by discharge lamp intensity fluctuations transferred into frequency fluctuations through the light shift effect.

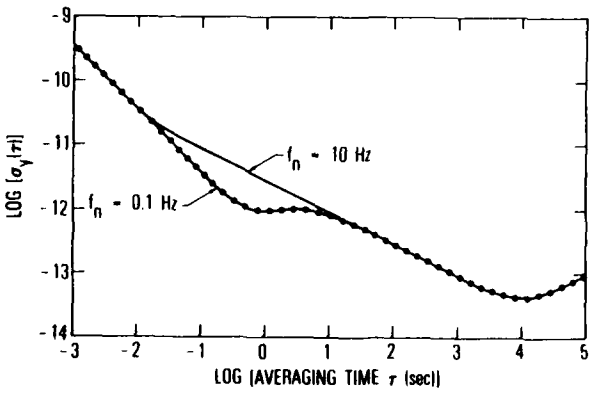


Figure 7. Calculated Allan variance $\sigma_y(\tau)$ for passive Rb^{87} gas cell standard for two control loop attack times. Crystal oscillator performance as specified in Ref. 17. The atomic system performance is that of the EG&G prototype standard.

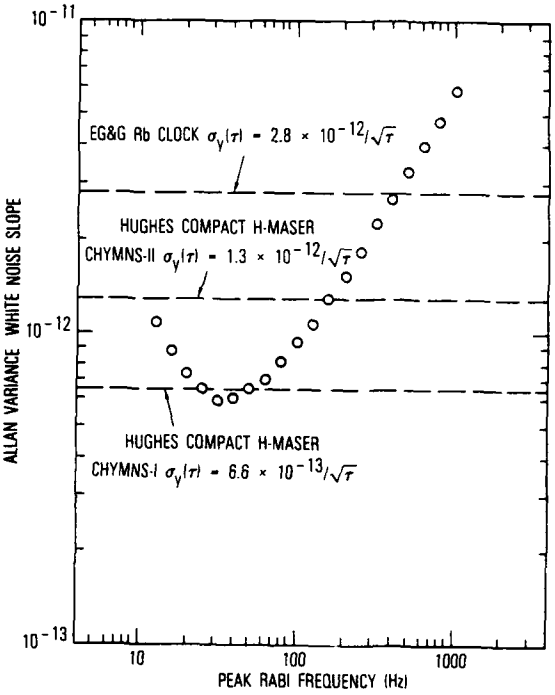


Figure 8. To demonstrate the manner in which the model could be used as a design tool, we have plotted the short term Allan variance slope for the parameters of Table I as a function of microwave Rabi frequency, or equivalently, microwave power fed into the cavity assuming a Q=100. These calculations would indicate roughly a factor of 5 improvement in the short term performance by decreasing the microwave power by ~23 db.

A Highly Stabilized Semiconductor Laser and Its Application to Optically Pumped Rb Atomic Clock

Motoichi Ohtsu, Minoru Hashimoto, and Hidetaka Ozawa

International Cooperation Center for Science and Technology,
Tokyo Institute of Technology,
2 - 12 - 1, O-okayama, Meguro-ku, Tokyo 152, Japan

Abstract

Frequency stability of 1×10^{-12} at $\tau = 100$ s was obtained for 0.8 μm AlGaAs laser by using spectral lines of Rb vapor as frequency references. It was confirmed that this value of the stability was as high as the value limited by spontaneous emission noise. Through an analysis based on a semiclassical Langevin's equation, it was estimated that the stability can be improved to as high as $1.7 \times 10^{-14} \tau^{-1/2}$. Spectral linewidth reduction was also tried to improve the coherence of the semiconductor laser. A novel technique, i.e., electrical feedback, was proposed for this reduction instead of using a conventional technique of optical feedback. The linewidth was stably reduced by this technique. The minimum value obtained was 330 kHz for an InGaAsP laser at 1.5 μm, which was fifteen times narrower than that of a free-running laser. It was estimated that the linewidth can be ultimately reduced to a value less than 1 kHz by this technique. Experiments on optical pumping for Rb atomic clock were carried out by using the highly stabilized semiconductor laser mentioned above. As the first step, experiments on saturated absorption spectroscopy of ^{87}Rb - D_2 lines were carried out. Eleven lines, including cross-resonance lines, were clearly observed. As the next step, double resonance signal was obtained by laser optical pumping. The microwave frequency shift by the laser frequency and power were measured. The microwave frequency stability was also evaluated. Furthermore, a comment on the spectral lifetime of semiconductor laser for Rb atomic clock was given.

I. Introduction

The spectral properties of semiconductor lasers have been recently improved as a result of the demands of the optical communication industry. These lasers can be used as reliable light sources for coherent optical communication and coherent optical measurements. For these purposes, we have improved their frequency stabilities, and carried out their spectral linewidth measurements. A part of these works has been reported in this symposium[1]. In the present paper, recent progresses on this study and its application to Rb atomic clock are reported.

II. Frequency Stabilization of Semiconductor Lasers

Frequency stabilization of a 0.8 μm AlGaAs laser was carried out by using a stable Fabry-Perot interferometer, absorption spectral lines of H_2O and ^{85}Rb as frequency references[2] - [4]. The injection current of the laser was controlled for stabilization by employing a PID servo-control circuit. Figure 1 summarizes the experimental results[5]. When the ^{85}Rb - D_2 line was used as a frequency reference, the highest frequency stability of $\sigma_y(\tau) = 1.4 \times 10^{-12}$ was obtained at $\tau = 100$ s. Figure 2 shows the limit of the frequency stability estimated through an analysis based on a semiclassical Langevin's equation[5]. Comparison between Figs. 1 and 2 shows that experimental results have already approached to the value limited by spontaneous emission. Quite recently, Saito, et. al., pointed out that the electrical feedback may control the quantum FM noise of 0.8 μm AlGaAs lasers, and reduce the FM noise to a value less than that limited by spontaneous emission[6]. If this result is applied to the present case, the frequency stability can be improved to the value limited by noise of the detector in the feedback loop. This value is given by the curve G of Fig. 2,

CH2186-0/85/0000-0043$1.00©1985IEEE

i.e., the stability can be improved as high as $\sigma_y(\tau) = 1.7 \times 10^{-14}\ \tau^{-1/2}$.
Figure 3 shows the result of frequency stabilization by using a ^{87}Rb - D_2 line as a frequency reference, which was recently obtained for developing a ^{87}Rb atomic clock. The laser was installed in a small vacuum chamber, and the fluctuations of the temperature and injection current were reduced as low as 1×10^{-4} K and 0.6 nA/$\sqrt{Hz}$, respectively. Two kinds of ^{87}Rb absorption cells were used, i.e., with buffer gases and without buffer gases. For the ^{87}Rb absorption cell without buffer gases, saturated absorption spectral lines as well as linear absorption spectral lines were measured, and were used as frequency references for stabilization. For the ^{87}Rb absorption cell with buffer gases (Ar/N_2 = 1.65, total pressure ; 43 Torr), linear absorption spectral lines were used as a frequency reference. In all of these cases, frequency stability as high as $\sigma_y(\tau) = 1 \times 10^{-12}$ at τ = 100 s was obtained. Higher frequency stability can be expected by improving the servo-controlling circuits. For 1×10^{-2} s < τ < 1×10^{1} s, the stability of the curve D is slightly higher than others because of higher frequency discrimination of the frequency reference by saturated absorption line.
Figure 4 shows the deterioration in the power stability observed when the laser frequency was stabilized by controlling the injection current. This is due to that the power was disturbed by the change in the injection current for frequency stabilization. Since the deterioration in power stability will reduce the detection sensitivity of double resonance signal in ^{87}Rb atomic clock, simultaneous stabilization of the power should be required by controlling, e.g., temperature. Simultaneous power stabilization is now in progress.

III. Linewidth Reduction of Semiconductor Lasers by Electrical Feedback

It has been reported that the linewidth of a free-running semiconductor laser was larger than several mega herz[7]. However, if this laser is used for coherent optical communication or coherent optical measurements, the linewidth should be narrower than 1 MHz. Several techniques have been proposed to reduce the linewidth for these applications. One of them is to increase the cavity Q factor by using an external mirror. This has been called an optical feedback technique, and it makes use of the injection of reflected light into the laser from an external mirror. The linewidth has been reduced to a value as narrow as 1 kHz by this technique[8]. However, this technique presents several problems. One of them is that the linewidth can be temporally affected by phase fluctuations of the reflected light induced by the mechanical vibration of the external mirror. Furthermore, it is essentially required to considerably increase the size of the laser cavity in this technique, which sacrifices such an advantageous property of the semiconductor laser as its small size.
To overcome these difficulties, we have proposed a simpler and more stable technique, i.e., an electrical feedback to reduce the linewidth by controlling the injection current[9]. Saito, et. al.[6] have also pointed out that the electrical feedback can reduce its linewidth to a value smaller than the one given by the modified Schawlow - Townes formula[10]. This makes the electrical feedback a more promising technique to realize a stable and ultranarrow linewidth laser. Figure 5 shows the experimental apparatus. In this experiment, a distributed feedback (DFB) - type InGaAsP laser at 1.5 μm was used to get a single longitudinal mode oscillation for a wide range of the injection current. However, this technique can be applied also for 0.8 μm AlGaAs lasers. FM noise of the laser was detected by using a compact Fabry - Perot interferometer of 10 mm length as a frequency discriminator. The output signal from a Ge - APD, which is proportional to FM noise, was fed back to the injection current after amplified by a video amplifier with 100 MHz bandwidth. A delayed self - heterodyne technique was employed for linewidth measurements[11]. Figure 6 shows the experimental results. The minimum value obtained was 330 kHz, which is 15 times narrower than that of a free - running laser. The spectral line shape showed none of the temporal fluctuations which have sometimes been observed in the optical feedback technique[8]. Figure 7 shows the minimum attainable linewidth, where R_{FP} represents the reflectance of the mirrors of the Fabry - Perot interferometer. For this estimation, it was assumed that the linewidth can be reduced to a value limited by the noise of the detector which is installed in the initial stage in the feedback loop, as was pointed out by Saito, et. al.[6]. From Fig. 7, it can be concluded that the linewidth can be ultimately reduced to a value less than 1 kHz when $R_{FP} > 0.9$.

IV. High Resolution Spectroscopy of ^{87}Rb

Highly stabilized semiconductor laser described in II and III can be used for high resolution spectroscopy and optical pumping of atomic clocks. In this section, experimental results of high resolution spectroscopy of ^{87}Rb - D_2 lines are presented. Figure 8 shows a popular energy levels of ^{87}Rb atoms, in which each optical transition is assigned (o ~ t). Two kinds of ^{87}Rb absorption cells, employed in II, were also used here at room temperature. Figure 9 shows the linear absorption spectral shapes observed by both of the absorption cells. By comparing these figures, it was found that the frequency of F = 1 line for

the cell with buffer gases was located 260 MHz lower than that of the cell without buffer gases, i.e., this transition suffered the pressure shift.

Figure 10 shows saturated absorption spectral shapes observed by the cell without buffer gases. Five lines for F = 1 and six lines for F = 2 were clearly resolved, which were assigned to be the saturated absorption and cross - resonance lines. Least-square fitted curves are also shown in Fig. 10, which was derived by using a model given by Nakayama[12]. These curves fit well with those of the experimental results, and the linewidth of these spectral lines were estimated as 40 MHz through this fitting. Further calculations are now in progress by employing a more detailed model which includes also the dependence of the line shapes on the polarization of the laser light[13].

Figures 11 and 12 shows the dependences of the signal strength and linewidth of two cross - resonance lines (s - t, p - q) on the laser power density, where the cross - sectional area of the laser beam was about 0.1 cm^2. A saturation due to the laser power can be clearly seen in these figures.

V. Application to ^{87}Rb Atomic Clock

The laser frequency was locked at the center of a linear absorption spectral line of F = 1 of the ^{87}Rb absorption cell with buffer gases. The frequency stability and power stability of the laser have been given in Figs. 3 and 4. Fig. 13 shows the derivative of a double resonance signal of ^{87}Rb obtained by using this stabilized laser as a pumping source. Figure 14 shows the dependences of the linewidth and S/N value of this signal on the laser power density. In Fig. 14, the linewidth decreases for the power density range of larger than about 100 $\mu W/cm^2$.One of the possible reasons may be due to anomalous line narrowing, which has been recently predicted[14].

Figures 15 and 16 show the shifts of the stabilized microwave frequency due to the laser frequency and power density. Center of the dispersive curve of Fig. 15 was selected as zero point of the axes of this figure, i.e., frequency shifts of the microwave $\Delta\nu_M$ and laser $\Delta\nu_L$ represent the shifts from this point. Figure 17 shows the slopes of the curves of Figure 15 at $\Delta\nu_L = \Delta\nu_M = 0$. Figure 18 shows the frequency stability of the microwave of the ^{87}Rb atomic clock. In this measurement, the laser power density incident into the atomic clock and $\Delta\nu_L$ were fixed at 88.5 $\mu W/cm^2$ and 0, respectively. The curve A represents the contribution of the FM noise of the frequency stabilized laser estimated by using the results of Figs. 3 and 17. The curve B also represents this contribution obtained by assuming that the laser frequency stability is improved as high as the ultimate value given by the curve G in Fig. 2. AM noise does not give direct contributions to this stability as long as the laser frequency is fixed at $\Delta\nu_L = 0$ because the power shift of Fig. 16 is zero at $\Delta\nu_L = 0$. However, since the AM noise could reduce the S/N value of the double resonance signal detection, this will limit the frequency stability of the atomic clock. Quantitative estimation of this effect is now in progress. Experimental results in this figure show that the stability of the laser-pumped ^{87}Rb atomic clock obtained in this preliminary exkperiment was already as high as that of a conventional atomic clock, and is almost equal to the value reported by L. Lewis[15] Further improvements of this stability can be expected by reducing the noise from the photodetector and servo - control circuit.

VI. Spectral Lifetime of Semiconductor Lasers

When a semiconductor laser is used for ^{87}Rb atomic clock , its spectral lifetime should be long enough. That is, the wavelength of a free-running should stay at the resonance wavelength of ^{87}Rb - D_2 line at 780.0 nm for at least more than several years. However, since the wavelengths of commercially available AlGaAs lasers distribute in a wide range of between 760 - 800 nm, it is not easy to find a laser with the wavelength accurately coincident with that of ^{87}Rb - D_2 line even though the wavelength tuning can be performed by widely varying the temperature. If the lasers are operated at the room temperature for practical use, the probability of finding appropriate lasers at 780.0 nm among commercially available lasers are only between 10 - 40 %. Even though an appropriate laser can be found, it often shows the long - term variation of the wavelength. By these reasons, the spectral lifetime, i.e., the time period in which the laser wavelength stays at that of ^{87}Rb - D_2 line, is rather limited. This spectral lifetime is a limiting factor to the performances of a laser - pumped ^{87}Rb atomic clock. However, the detailed investigation has not yet been carried out. The discussion in this section gives a comment on this point.

Figure 19 shows a experimental results of the variation of the range of injection current for stable oscillation of each longitudinal mode, which represents the lifetime of the modes. It is seen that the lifetime of the mode A, oscillated with a lower injection current just above the threshold current, is rather long. On the other hand, those of the modes with higher injection current (modes B $\sim$ E)are quite short. From this result, it may be concluded that it is safer to use the laser with a lower injection current to maintain the spectral lifetime long enough. Furthermore, the variation of the range of these injection currents is not gradual but stepwise, which may induce such a catastrophic phenomenon that a laser-pumped ^{87}Rb atomic clock

suddenly dies. The phenomenon in Fig. 19 is completely different from a popularly observed mode hopping[16]. This could be explained by a temporal decrease of the thermal resistance due to an oxidation of the In bonding layer or by thermal effects due to nonradiative recombinations of carriers near the facets, which has been pointed out also by Fabre and Guen[16]. By this decrease in self-heating, the wavelength change of each mode shows blue shift, which is shown by Fig. 20. Figure 20 shows the variation of the injection current required to tune wavelength of a longitudinal mode to that of the optical transition from F = 1 of ^{87}Rb - D_2 lines. Increase in this injection current means that the laser actually suffers a blue shift. It should be pointed out that this shift is also stepwise. Average of the shift given by this figure was about + 40 MHz/hour. From the discussion presented so far, it may be concluded that the thermal effect is a dominant factor to limit the spectral lifetime of semiconductor lasers.

Further reduction of the thermal resistance can be expected by improving the design of laser structures. We are now trying to fabricate an improved laser for this purpose in cooperation with laser fabricating group, and a prototype of these lasers have been already fabricated[18].

VII. Summary

Recent progress in frequency stabilization and linewidth reduction in semiconductor lasers were presented. From these results, it may be concluded that the semiconductor lasers have a possibility of becoming an ultrahigh coherent light sources by applying electrical feedback technique. These lasers were used for high resolution spectroscopy of ^{87}Rb, and eleven saturated absorption lines were well resolved. The laser-pumped ^{87}Rb atomic clock was constructed, and its stable operation was confirmed. Furthermore, several comments on spectral lifetimes of the laser were given to develop a reliable light source for the atomic clock.

Acknowledgements

We would like to express our gratitudes to Messrs. Kasai, Hashi, Chiba (Fujitsu Co. Ltd.), Prof. Kano (Univ. of Elefctro-Comm.), Dr. Nakayama (Kyoto Inst. Tech.), Prof. Oura, Drs. Kuramochi, and Teramachi (Tokyo Inst. Tech.), for their valuable discussions.

References

[1] M. Ohtsu, H. Tsuchida, and T. Tako, "Frequency Stabilization of AlGaAs Lasers", 36th Annual Frequency Control Symposium, pp. 327 - 337, 1982, Philadelphia

[2] H. Tsuchida, M. Ohtsu, and T. Tako, "Improvements in the Short-Term Frequency Stability of AlGaAs DH Laser", Trans. IECE Japan, vol. E65, pp.65 - 66, 1982

[3] H. Tsuchida, M. Ohtsu, and T. Tako, "Frequency Stabilization of AlGaAs Semiconductor Laser to the Absorption Line of Water Vapor", Jpn. J. Appl. Phys., vol. 21, pp. L1 - L3, 1982

[4] H. Tsuchida, M. Ohtsu, T. Tako, N. Kuramochi, and N. Oura, "Frequency Stabilization of AlGaAs Semiconductor Laser Based on the ^{85}Rb - D_2 Line", Jpn. J. Appl. Phys., vol. 21, pp. L561 - L563, 1982

[5] M. Ohtsu, H. Fukada, T. Tako, and H. Tsuchida, "Estimation of the Ultimate Frequency Stability of Semiconductor Lasers", Jpn. J. Appl. Phys., vol. 22, pp. 1157 - 1166, 1983

[6] S. Saito, O. Nilsson, and Y. Yamamoto, "Frequency modulation noise and linewidth reduction in a semiconductor laser by means of negative frequency feedback technique", Appl. Phys. Lett., vol. 46, pp. 3 - 5, 1985

[7] Y. Yamamoto, S. Saito, and T. Mukai, "AM and FM Quantum Noise in Semiconductor Lasers. Part II : Comparison of Theoretical and Experimental Results for AlGaAs Lasers", IEEE J. Quantum Electron., vol. QE - 19, pp. 47 - 57, 1983

[8] R. Wyatt, "Line-narrowed lasers for coherent optical fiber systems", Proceedings of Conference on Optical Fiber Communication, TUP2, February 11 - 13, 1985, San Diego, California, USA

[9] M. Ohtsu and S. Kotajima, "Linewidth Reduction of a 1.5 μm InGaAsP Laser by Electrical Feedback", Jpn. J. Appl. Phys., vol. 24, pp. L256 - L258, 1985

[10] C. H. Henry, "Theory of the Linewidth of Semiconductor Lasers", IEEE J. Quantum Electron., vol. QE - 18, pp.259 - 264, 1982

[11] T. Okoshi, K. Kikuchi, and A. Nakayama, "Novel method for high resolution measurement of laser output spectrum", Electron. Lett., vol. 16, pp. 630 - 631, 1980

[12] S. Nakayama, "Theoretical Analysis of Rb and Cs D_2 Lines in Saturation Spectroscopy with Optical Pumping", Jpn. J. Appl. Phys., vol. 23, pp. 879 - 883, 1984

[13] S. Nakayama, "theoretical Analysis of Rb and Cs D_2 Lines in Doppler- Free Spectroscopic Techniques with Optical Pumping", Jpn. J. Appl. Phys., vol. 24, pp. 1 - 7, 1985

[14] J. 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

[15] L. L. Lewis and M. Feldman, "Optical Pumping by Lasers in Atomic Frequency Standards", 35th Annual Frequency Control Symmposium, pp. 612 - 624, 1981

[16] M. Ohtsu, Y. Otsuka, and Y. Teramachi,

"Precise measurements and computer simulation of mode-hopping phenomena in semiconductor lasers", Appl. Phys. Lett., vol. 46, pp. 108 - 110, 1985

[17] F. Favre and D. Le Guen, "Emission frequency stability in single - mode - fibre optical feedback controlled semiconductor lasers", Electron. Lett., vol. 19, pp. 663 - 665, 1983

[18] K. Kishino, S. Kinoshita, S. Konno, and T. Tako, "Selective Meltbacked Substrate Inner - Stripe AlGaAs/GaAs Lasers Operated under Room Temperature CW Condition", Jpn. J. Appl. Phys., vol. 22, pp. L473 - L475, 1983

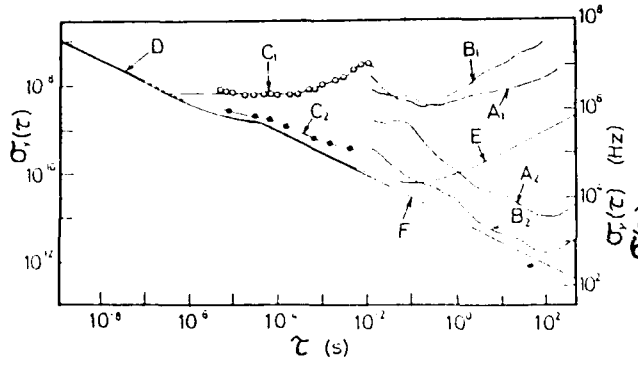


Fig. 1 Summary of the experimental results of frequency stabilization[5].
A_1, B_1, C_1, D : Free-running.
A_2 : Stabilized by H_2O.
B_2 : Stabilized by ^{85}Rb - D_2.
C_2: Stabilized by a rigid Fabry - Perot interferometer.
E : Theoretical limit for the free - running laser (curve F of Fig. 2).
F : Theoretical limit given by spontaneous emission (curve H of Fig. 2)

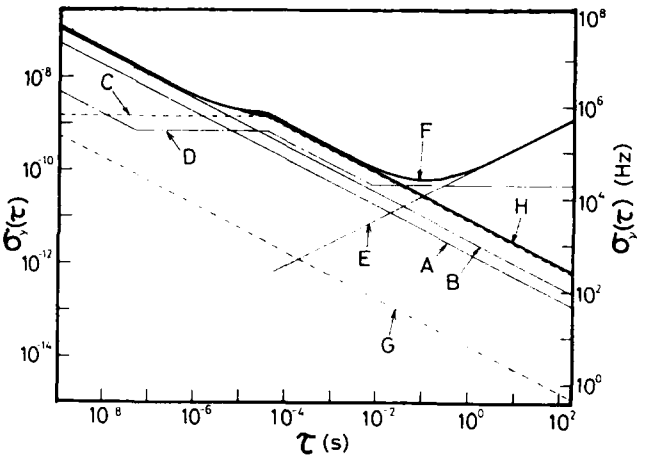


Fig. 2 Calculated results of the frequency stability of a 0.8 μm AlGaAs laser[5].
A : Spontaneous emission noise.
B : Carrier noise.
C : Current noise.
D : Current source noise.
E : Temperature noise.
F : Free-running laser.
G: Detector noise limited value for the stabilized laser.
H : Spontaneous emission noise limited value for the stabilized laser.

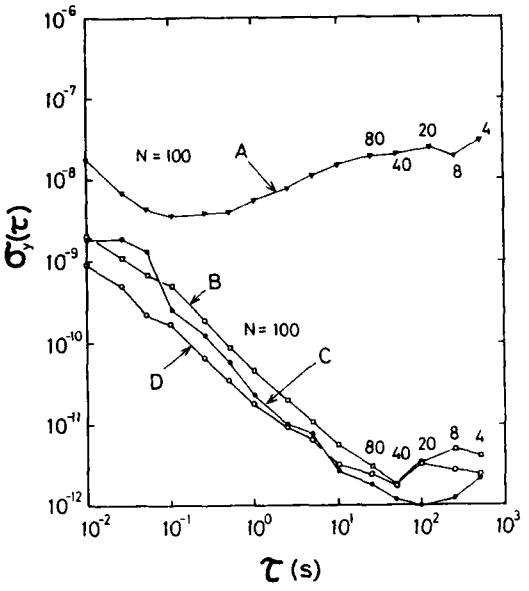


Fig. 3 Results of the frequency stabilization of an AlGaAs laser by using the ^{87}Rb - D_2 line as a frequency reference.
A : Free-running.
B : ^{87}Rb with buffer gases, linear absorption.
C : ^{87}Rb without buffer gases, linear absorption.
D : ^{87}Rb without buffer gases, saturated absorption.

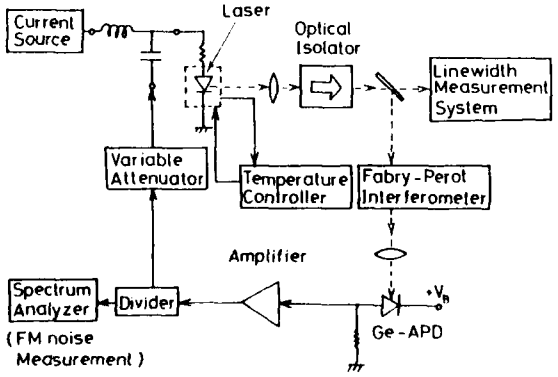


Fig. 5 Experimental apparatus for linewidth reduction of the laser

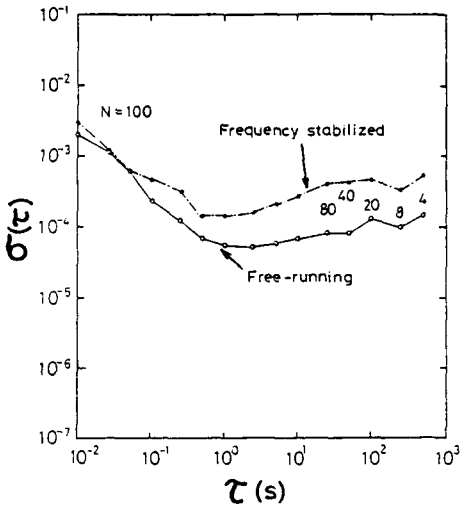


Fig. 4 Stabilities of the laser power when the laser frequency is in free-running and stabilized conditions.

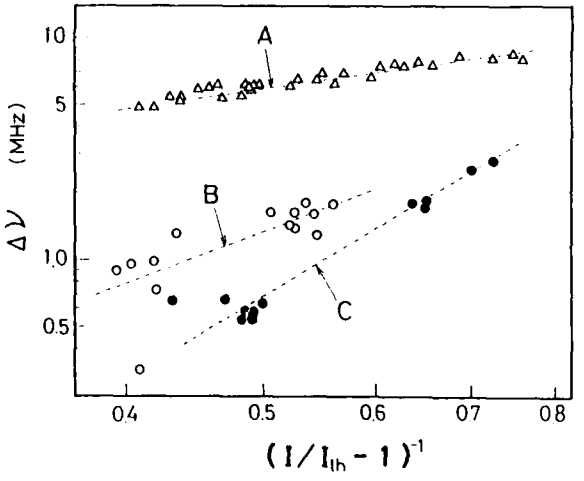


Fig. 6 Experimental results of the linewidth reduction of the laser. I/I_{th} represents the injection current normalized to its threshold value. A : Free-running laser. B, C : Under feedback condition with R_{FP} = 0.9 and 0.95, respectively. R_{FP} represents the reflectance of the Fabry - Perot interferometer in Fig. 5.

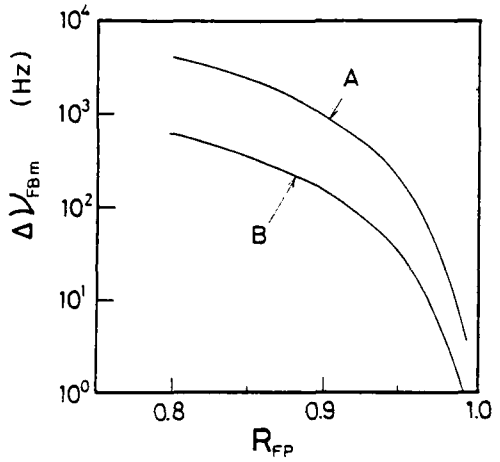


Fig. 7 Estimated minimum attainable linewidth limited by the detector noise in the feedback loop. The curves A and B represent the results when a Ge - APD and Ge - PIN photodiode were used as detectors in the feedback loop, respectively.

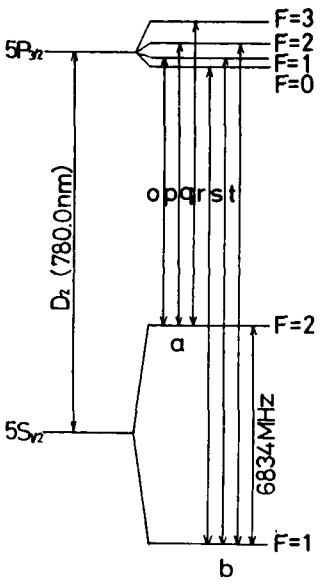


Fig. 8 Energy level diagram of ^{87}Rb atoms.

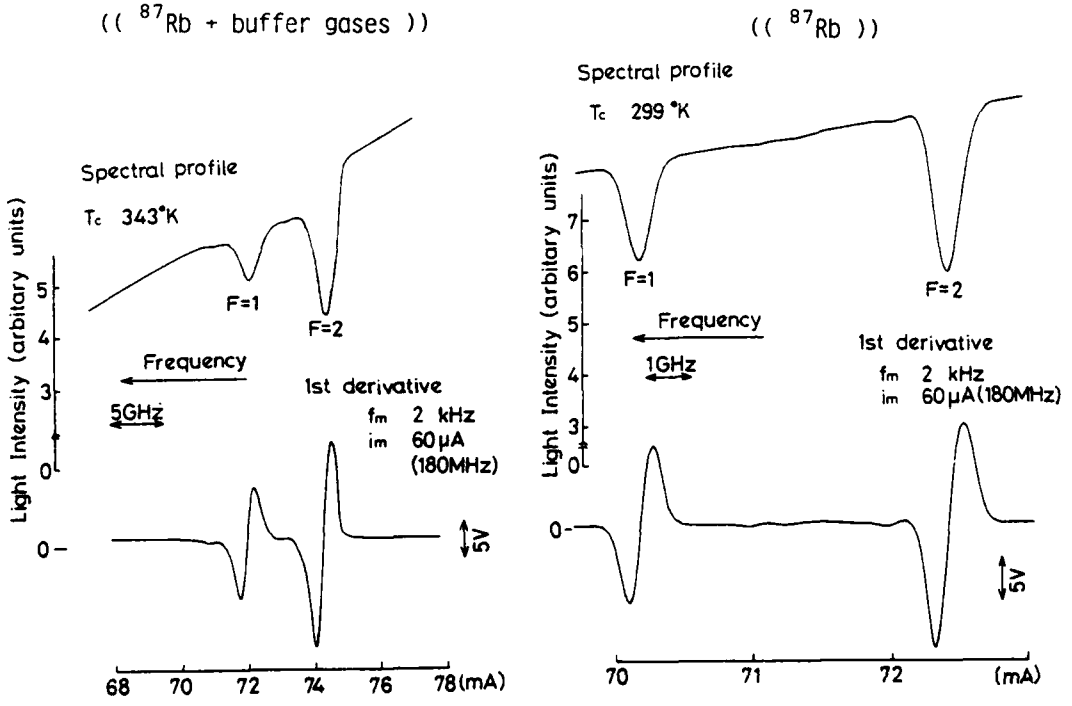


Fig. 9 Linear absorption spectral shapes observed by using ^{87}Rb cells with and without buffer gases.

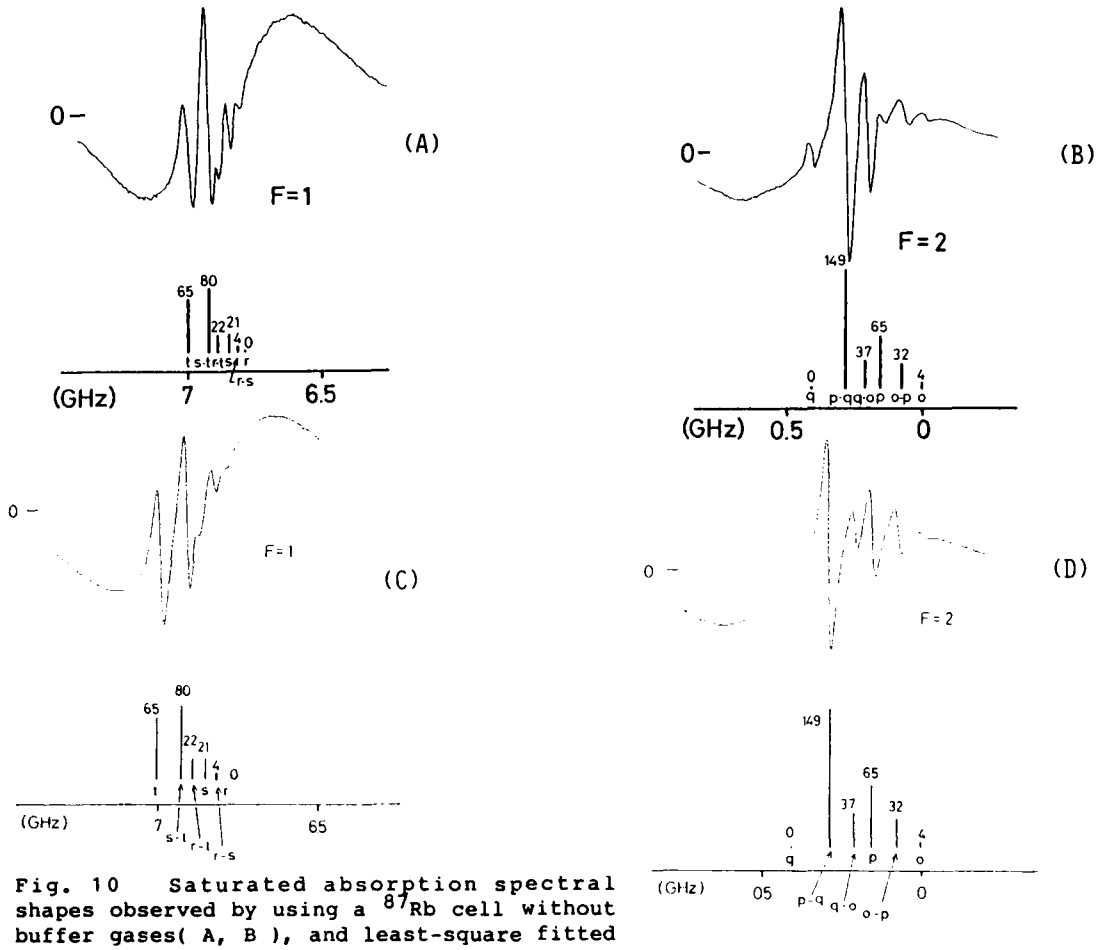


Fig. 10 Saturated absorption spectral shapes observed by using a ^{87}Rb cell without buffer gases(A, B), and least-square fitted curves (C, D).

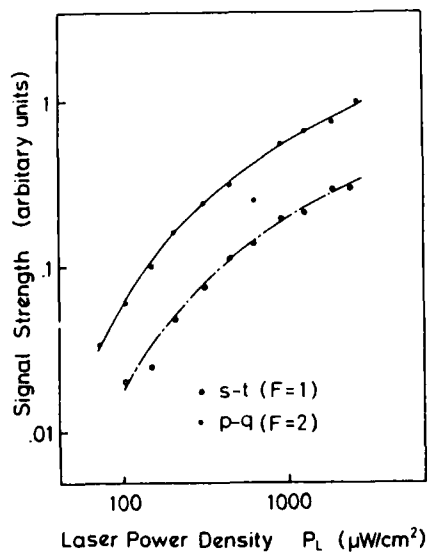


Fig. 11 Dependence of the signal strengths of two cross-resonance lines (s - t, p - q) on the laser power density.

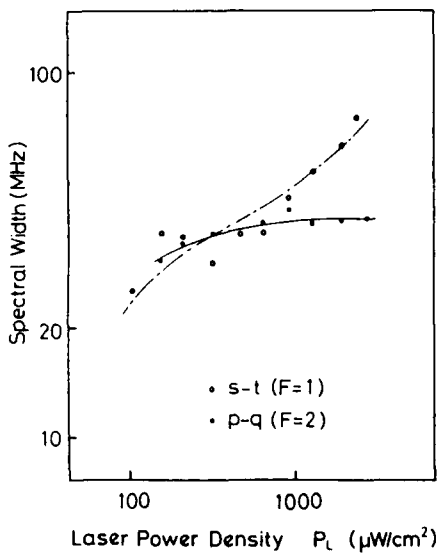


Fig. 12 Dependence of the linewidths of two cross-resonance lines (s - t, p - q) on the laser power density.

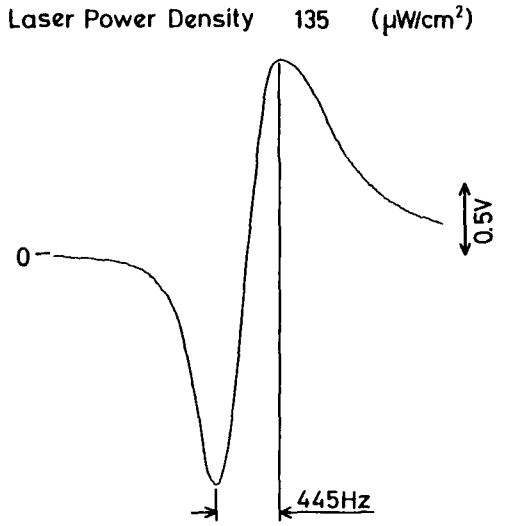
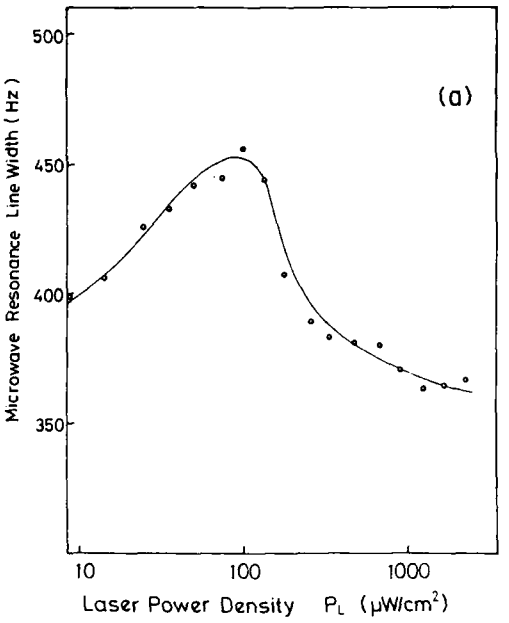


Fig. 13 Derivative of the double resonance signal of ^{87}Rb obtained by using the stabilized laser as a pumping source.



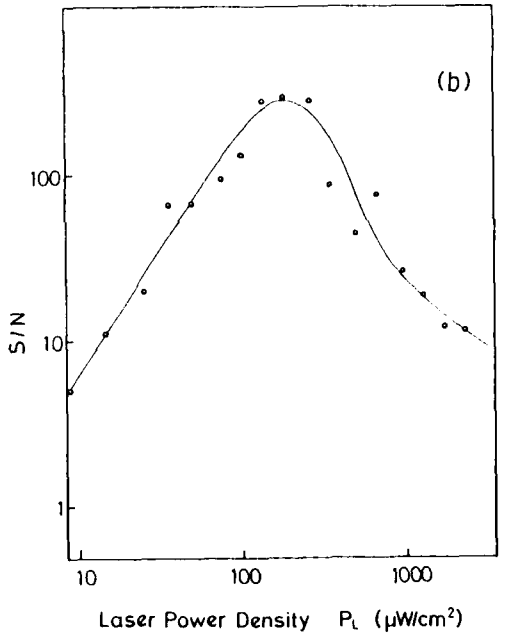


Fig. 14 Dependences of the linewidth (a) and S/N value (b) of the double resonance signal on the laser power density.

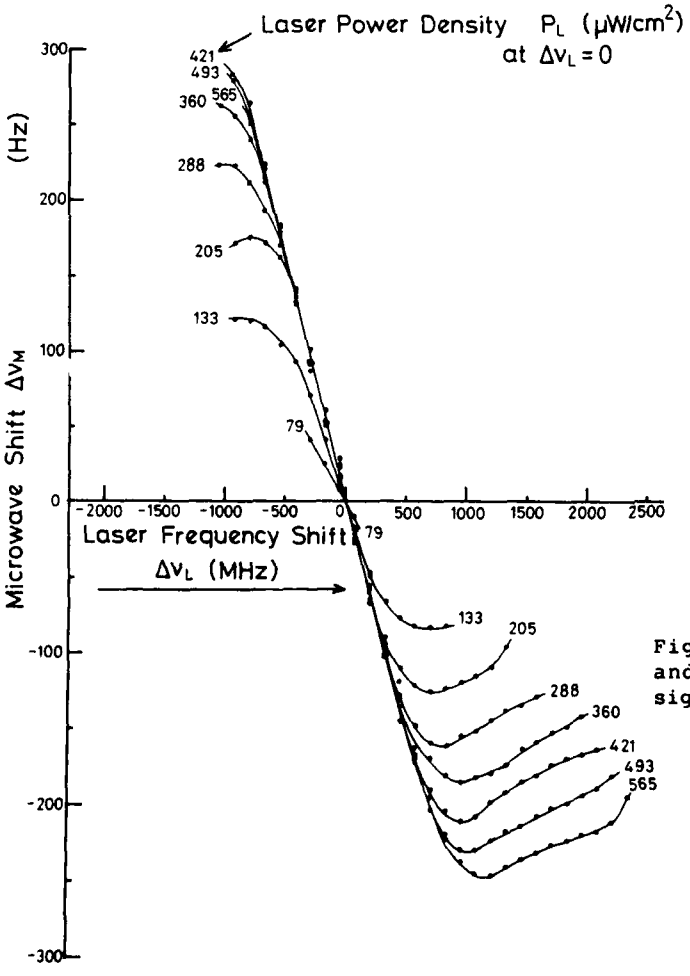


Fig. 15 Shift of the stabilized microwave frequency due to the shift of the laser frequency.

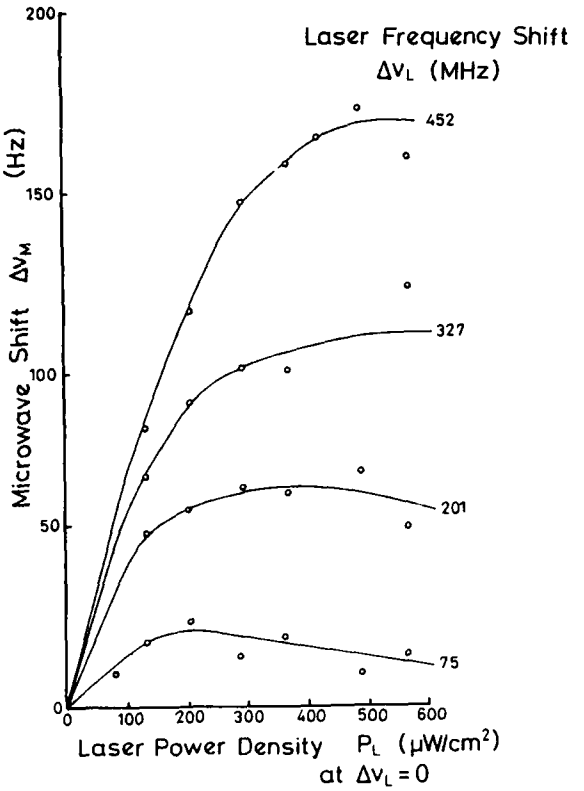


Fig. 16 Shift of the stabilized microwave frequency due to the laser power density.

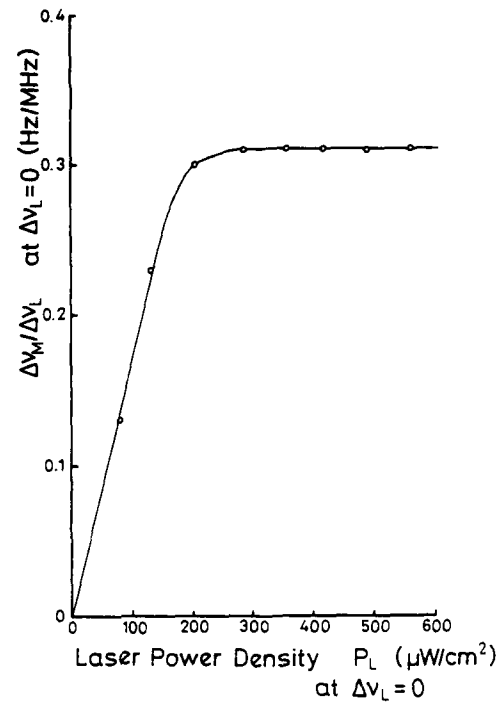


Fig. 17 Slope of the curves of Fig. 15 at $\Delta\nu_L = 0$.

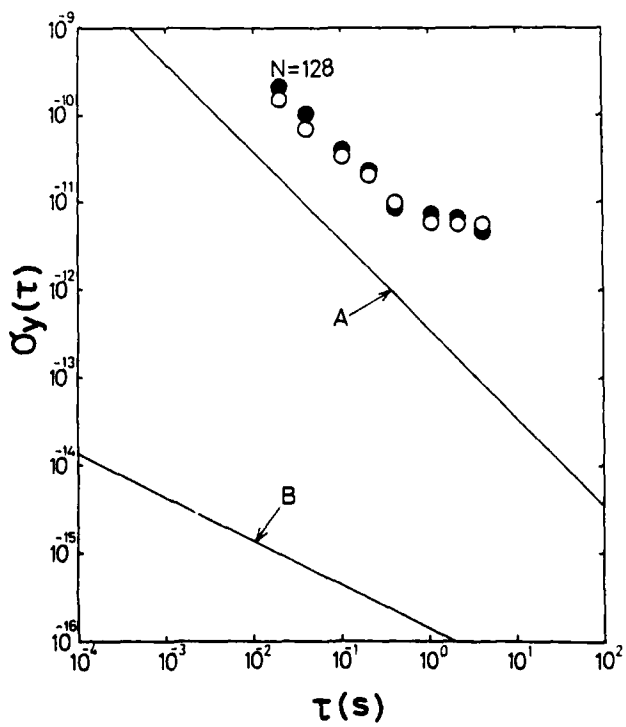


Fig. 18 Frequency stability of the microwave.
A : Contribution of the FM noise of the frequency stabilized laser estimated by using the curve B of Fig. 3 and Fig. 17.
B : Contribution of the FM noise of the laser obtained by using the curve G of Fig. 2 and Fig. 17.
○ : Experimental results of the laser-excited ^{87}Rb atomic clock.
● : Experimental results of the conventional ^{87}Rb atomic clock.

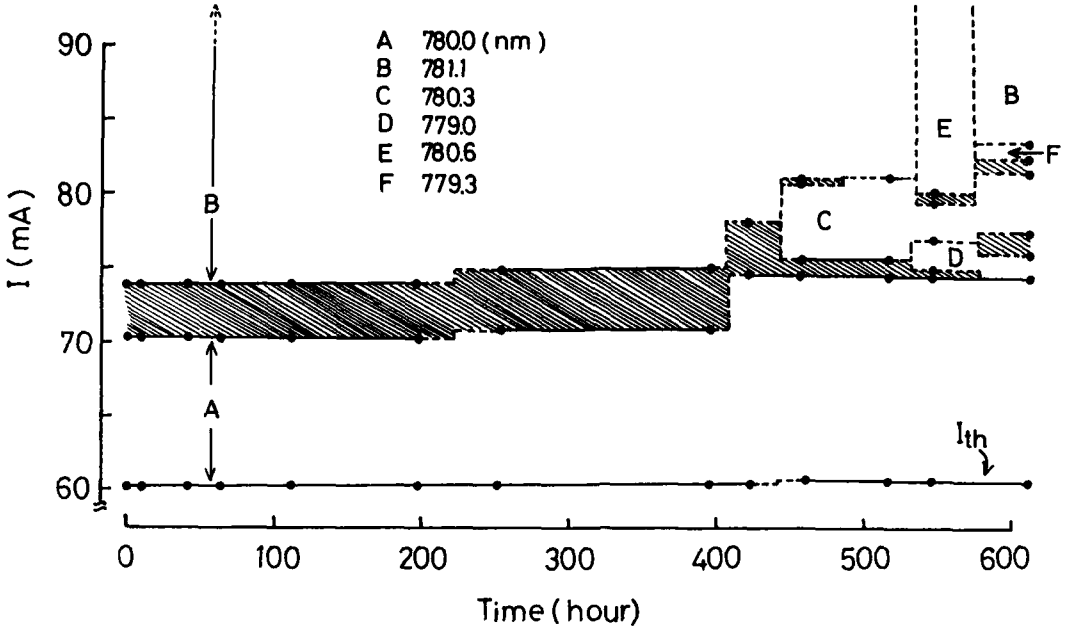


Fig. 19 Spectral lifetimes of longitudinal modes (A - F) of the laser. Hatched area represents the area of multi-longitudinal mode oscillation.

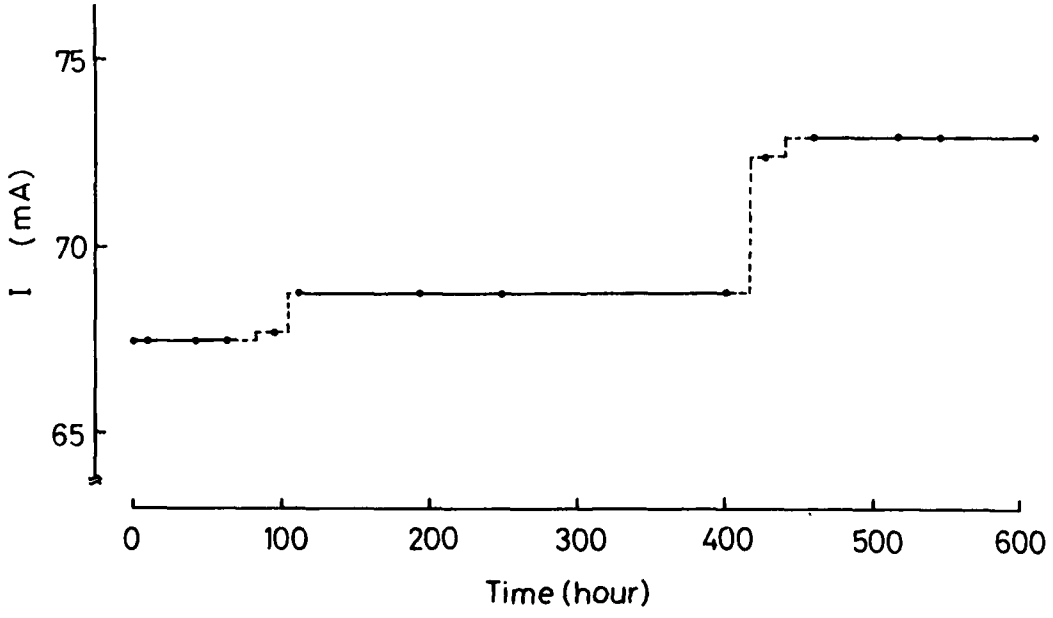


Fig. 20 Variation of the injection current required to tune the wavelength of a longitudinal mode to that of the optical transition from F = 1 of ^{87}Rb - D_2 lines.

AN ULTRA-MINIATURE RUBIDIUM FREQUENCY STANDARD

Kazuharu Chiba and Toshio Hashi

FUJITSU LIMITED
1015 Kamikodanaka, Nakahara-ku, Kawasaki 211, Japan

Abstract

An ultra-miniature, general-purpose rubidium gas cell frequency standard has been designed and investigated experimentally, with the aim of developing one of the smallest atomic frequency standards yet developed. This paper describes the details of the design concept and the experimental results obtained so far.

This rubidium frequency standard embodies new electrical and structural design features; an ultra-miniature physics package including a dielectric-loaded cylindrical TE_{111} microwave cavity, an IT cut crystal unit for the slave oscillator, advanced servo electronics using sample-hold technology, and a simplified, fabrication-oriented structure, in addition to the use of an integrated filter-resonance cell.

As a result of these features, the rubidium frequency standard has an extremely small volume (0.4 liter) and light weight (0.6 kg) which allow the standard to be easily installed in a variety of systems.

Some important experiments have been made to confirm the validity of the design concept. In spite of the substantial reduction in size, excellent quantities, by which the performance of the rubidium frequency standard is determined, have been obtained. From these data, excellent performance of the standard can be expected together with compactness and light weight.

Introduction

Atomic frequency standards have already found many applications in the field and in systems involving precision frequencies. FUJITSU has been producing rubidium frequency standards in several models for about fifteen years. Since our miniature, high-performance rubidium frequency standard, designated 5407A, was successfully developed in 1981[1], a large quantity of rubidium standards have been rapidly introduced for practical frequency sources in a wide range of applications, such as communications, broadcasting, and navigation systems in Japan. Recent applications of rubidium standards include precision clock sources for synchronization of local area networks (LAN), because expansion of the LAN area via global networks requires plesiochronous operation of LAN's located at long distances.

Our continuous activity in this field has been directed toward a volume equivalent to or less than that of a high stability crystal oscillator and toward a high performance level, which are generally incompatible requirments, to meet the ever-increasing demands for precision frequency standards and to expand applications of rubidium standards by developing a more attractive standard.

In order to achieve a compact design, which is one of the precondition to realize an attractive model, it was indispensable to reduce the size of the physics package, which occupies a major part of the volume. For this purpose, excitation mode of a microwave cavity and loading dielectric inside the cavity were investigated, in addition to the use of a two-cell scheme already developed and used in conventional models.

Together with this approach, electronic circuits were also redesigned to achieve a low parts count, simplified, and cost saving design without sacrificing performance.

In the following, the design and the experimental results of the new rubidium frequency standard obtained so far are discussed in detail.

Design

Figure 1 is a simplified block diagram of the tentative model. This system consists of the physics package and three printed circuit boards:

1. The physics package, which serves as a very high-Q frequency discriminator to produce an error signal corresponding to the difference between the input microwave frequency and the intrinsic resonance frequency.

2. The frequency multiplier, which generates a VHF signal from the output of the slave oscillator mounted on the same printed circuit board.

CH2186-0/85/0000-0054$1.00©1985IEEE

3. The frequency synthesizer, which produces a fractional frequency necessary for synthesizing the interrogation frequency.

4. The low frequency circuit including the servo control circuit which produces the control voltage for steering the slave oscillator by analysing the error signal from the physics package, a voltage regulator, and a temperature control circuit for the cavity in the physics package.

The mechanical structure was simplified for ease of construction. All four assemblies are mounted on a base plate which is a major part of the structure, and are electrically connected by only a few cables and wires. This simplified structure is expected to greatly facilitate assembly and testing of these circuits.

Physics package

A microwave cavity containing the resonance cell accounts for a large part of the volume of the physics package. In order to realize a compact design of the physics package, a reduction in size of the microwave cavity itself is needed. As one of the means, we considered modification of the excitation mode in the cavity and loading dielectric entirely into the remaining space inside the cavity. We studied two excitation modes, TE_{011} conventionally used and TE_{111}, both of which have relatively simple distribution of electromagnetic field.

The resonance length, ℓ, of the cylindrical cavity without the cell and holes for light transmission is given by

$$TE_{011}: \quad \ell = \frac{\lambda}{2\sqrt{1 - (\frac{\lambda}{1.64a})^2}} ,$$

$$TE_{111}: \quad \ell = \frac{\lambda}{2\sqrt{1 - (\frac{\lambda}{3.41a})^2}} ,$$

where

$$\lambda = \frac{\lambda_0}{\sqrt{\epsilon_r \mu_r}} , \quad \mu_r = 1$$

$$\lambda_0 = \frac{c}{f\sqrt{\epsilon_0 \mu_0}} = \text{wavelength in free space,}$$

a = radius of the cavity.

Figure 2 shows the relationship between a and ℓ of the cavity with variation of ϵ_r. From this relationship, a cavity with a capacity of 20 cc is expected to be realized by selecting dielectric loaded in the cavity, specifically by loading either dielectric with ϵ_r of 4 (quartz glass) for TE_{011}, or dielectric with ϵ_r of 2.1 (teflon) for TE_{111}.

Next, our concern was directed toward the size of the resonance cell, because degradation of signal-to-noise ratio (SNR) of the resonance signal and of long-term frequency stability may be caused by reduction in size of the resonance cell. Hence, we decided to use the TE_{111} mode cavity loaded with teflon because the use of the cavity allows a relatively larger resonance cell than that in the TE_{011}, and teflon used as the dielectric is easy to process by machine and is electrically stable. Aluminum was selected as the cavity material, because of its light weight, high electric conductivity for microwave cavities, and high heat conductivity for temperature ovens.

The experiments indicated that the tuning frequency of the cavity was adjustable by means of both a movable plate and a tuning screw located on the plate over a range of approximately 400 MHz, with a loaded quality factor of about 300.

Thus, a 20 cc cavity has been achieved without sacrificing the quality factor. The resultant size of the physics package has been reduced to 46 mm in diameter by 60 mm in length (100 cc), which is 1/5 of our conventional package, including a rubidium lamp exciter and a single magnetic shield of mu-metal. Figure 3 is an exploded view of the physics package.

Electronics section

(1) Frequency multiplier

The frequency multiplier includes the slave oscillator and a multiplier chain together on a printed circuit board.

A feature of this section is the use of an IT cut, small size, 10 MHz crystal unit for the slave oscillator. An IT cut crystal can be designed so that its temperature coefficient is nearly zero in ambient temperature range of 70°C to 80°C, which approximately coincides with the operating temperature range of the resonance cell. The use of the IT cut crystal allows the crystal to be mounted on the cavity block for temperature control and eliminates a separate temperature oven for the crystal.

The multiplier chain, in spite of its compactness, synthesizes a high power (+26 dBm into 50 ohms), low noise VHF (90 MHz) signal multiplying 3 x 3 times the output frequency of the slave oscillator, together with phase-modulation at low frequency at the 30 MHz stage.

(2) Frequency synthesizer

This section uses a sampling phase-locked loop composed of a miniature packaged HIC VCXO (2 cc), newly developed by FUJITSU, and programmable counters. The use of the VCXO enables the compact design of the circuit. The circuit also permits adjustment of the standard's output frequency over a range of $> 1 \times 10^{-7}$ with

incremental steps of about 2×10^{-9} by changing presettings of the programmable counters.

(3) Servo control circuit

The servo control circuit of our conventional standard uses a active filter, which needs a number of bulky capacitors as its element, to extract efficiently only a fundamental component from the resonance signal.

The new circuit uses the combination of a synchronous filter and a synchronous D.C. restorer, which is based on sample-hold technology and serves as a filter having a very sharp frequency selectivity. Figure 4 is the simplified block diagram of this circuit and the waveform at each port.

An output current of the solar diode is converted to a voltage by a preamplifier using a virtual ground point of an operational amplifier. In the integrated circuit for sample-hold, the converted voltage signal is sampled by sampling pulses whose frequency is twice that of the modulation signal and held, resulting in a rectangular waveform shown at (c) in Figure 4. The synchronous D.C. restorer clamps only the low level of the signal in phase and the high level of the signal out of phase by means of its synchronous switching operation. Since the phase of the resultant signal is different by π as shown at (d) in Figure 4, depending on whether the input microwave frequency is higher or lower than the reference atomic frequency, efficient synchronous detection with respect to the modulation signal is achieved.

Both analysis and experiments of a total frequency control system suggest that it has satisfactory frequency discrimination sensitivity and sufficient loop gain.

The preamplifier configuration, whose input impedance is extremely low, contributes to the low noise detection of the solar diode output[2], because the configuration derives a short-circuit current of the solar diode without forward bias voltage which is the source of shot noise. The use of the combination of the synchronous filter and the synchronous D.C. restorer enables significant reduction in parts count in this circuit, and the circuit is suitable for large scale integration.

Packaging

Figure 5 shows the exterior view of the rubidium frequency standard. The 10 MHz output is on the rear panel. Lock or loss of lock is indicated on the front panel. C-field adjustment is accessible from the front panel.

Experimental results

Some important preliminary experiments were made to confirm the validity of the design concept.

Short-term frequency stability

Short-term frequency stability has now become to be the most important quality to evaluate the performance of an atomic frequency standard.

It has already been verified through previous work that the short-term stability ($\sigma_y(\tau)$) of a passive atomic frequency standard depends on the SNR of the resonance signal and the line Q_l of the atomic resonance in the τ's range where $\sigma_y(\tau)$ is determined by white FM noise. This relationship is expressed as[3]:

$$\sigma_y(\tau) = \frac{1}{\sqrt{2}\,\pi\, Q_l\ (SNR)} \cdot \tau^{-1/2}$$

where $$SNR = \sqrt{\frac{\text{Signal power at } \Delta\nu_\mu = LW/2}{\text{Noise power/Hz}}},$$

$$Q_l = \frac{\nu_\mu}{LW},$$

ν_μ = atomic resonance frequency,

LW = line-width of atomic resonance.

The SNR of the resonance signal of 63 dB/$\sqrt{\text{Hz}}$, which is shown in Figure 6, and the line-width, LW, of 570 Hz at 6834 MHz have been successfully obtained at a reasonable light intensity of the rubidium lamp. According to the value obtained, excellent short-term frequency stability of $1.3 \times 10^{-11}\,\tau^{-1/2}$ is expected.

Long-term frequency stability

The long-term frequency stability of the rubidium gas cell frequency standard mainly depends on that of the resonance cell. In order to evaluate the effect of the reduction in size of the resonance cell, the frequency change of the resonance cell for the new standard has been measured by mounting the cell into our existing rubidium standard.

Figure 7 shows the long-term frequency stability plots of the resonance cell measured over about four months. No noticeable drift is exhibited and its aging rate is better than 1×10^{-11}/month, because of high-vacuum processing techniques already established for the production of highly stable resonance cells. From the data and the established cell production technology, the long-term frequency stability of the standard is also expected to be typically less than 1×10^{-11}/month.

Environmental frequency stability

In environmental frequency stabilities, temperature dependence of the rubidium standard will be major concern for practical use. It

mostly depends on the temperature dependence of the physics package, in particular that of both resonance and lamp cells. The temperature coefficients of both cells have been measured to estimate the temperature dependence of the rubidium frequency standard.

Figure 8 shows the temperature characteristics of the cells. These coefficients slightly change, compared with that of cells for the conventional two-cell scheme. The temperature coefficients, however, can be easily optimized by changing the ratio of buffer gas mixture and total gas pressure in the resonance cell.

Technical Data

A summary of the tentative technical data on the standard is shown in Table 1.

Conclusion

The details of the design of an extremely small, new rubidium frequency standard have been presented as well as the results of some preliminary experiments.

Many electrical and structural features have been incorporated into the design resulting in extreme compactness and light weight.

Preliminary experiments have verified that the critical quantities by which the performance of the rubidium standard is determined are quite satisfactry to realize the smallest, high performance rubidium frequency standard.

On the basis of the results, continuous effort will be made to establish the technology for the development of this extremely small, new rubidium frequency standard.

Acknowledgement

The authors would like to thank Messrs. T. Kojima, M. Yokouchi of FUJITSU LIMITED for their guidance and encouragement throughout the work.

References

[1] T. Hashi, et al., "A Miniature, High-Performance Rubidium Frequency Standard", Proc. 35th Annual Symposium on Frequency Control, pp.646-650, 1981.

[2] K. Chiba, et al., "Low Noise Detection of Atomic Resonance in Rubidium Frequency Standard", National Convention Record of IECE of Japan (in Japanese), No. 2712, 1984.

[3] T. Hashi, et al., "Analysis, Design and Performance of Cesium Beam Frequency Standard", FUJITSU Scientific and Technical J., Vol. 20, No. 4, 1984.

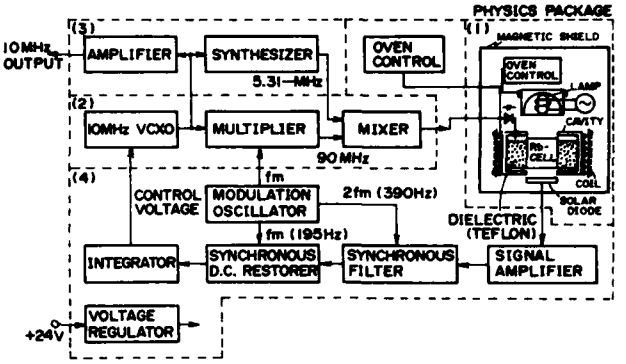


Fig. 1 Block diagram of rubidium frequency standard

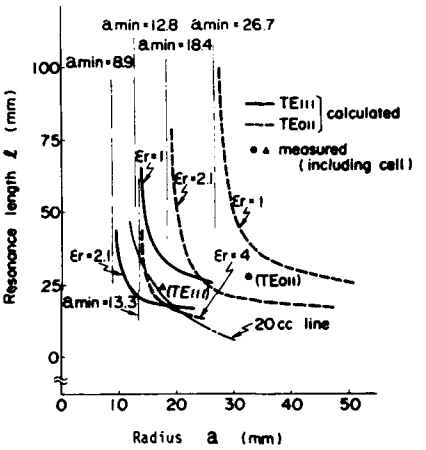


Fig. 2 Relationship between a and ℓ of cavity

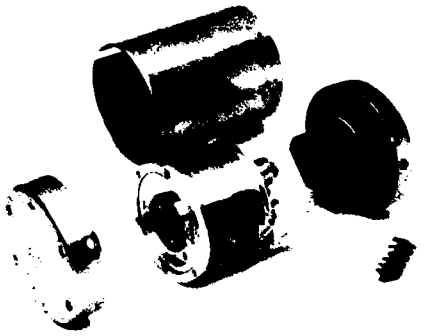


Fig. 3 Exploded view of physics package

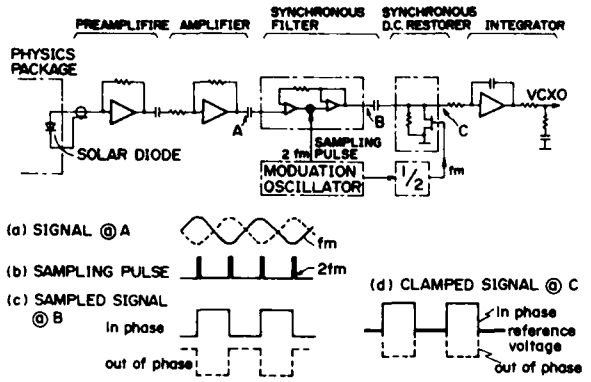


Fig. 4 Block diagram of servo control circuit

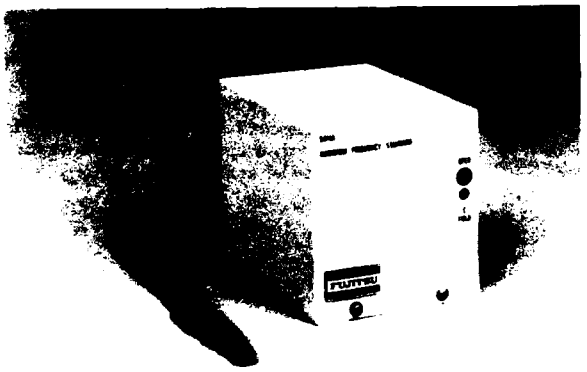


Fig. 5 Exterior view of rubidium frequency standard

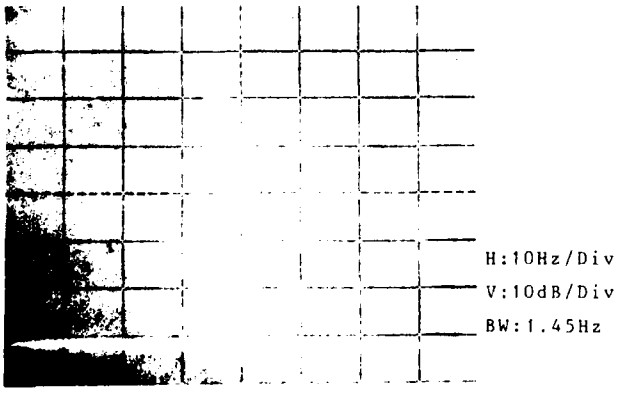


Fig. 6 SNR of resonance signal

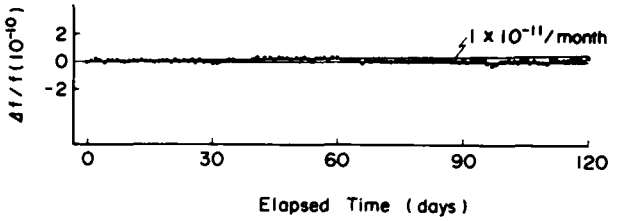


Fig. 7 Long-term frequency stability of resonance cell

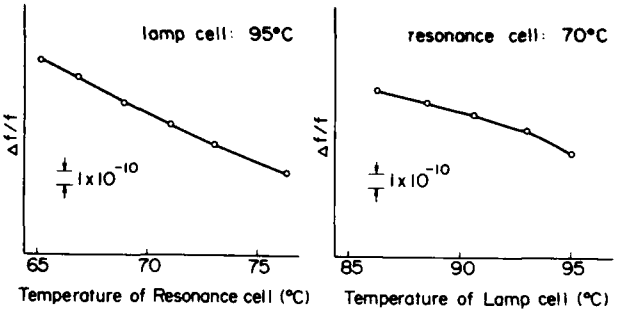


Fig. 8 Temperature dependence of cells

Table 1. Tentative technical data on rubidium frequency standard

Output
- Frequency: 10 MHz
- Level: 0.5 Vrms into 50 ohms
- Connector: OSM

Input
- Voltage: +24 V
- Power: 12 W (steady at 25°C)
  28 W (during warmup)

Dimmensions: 65 mm (2.6") wide, 70 mm (2.8") high, 90 mm (3.5") deep

Weight: 0.6 kg (1.4 lbs)

EVALUATION OF THE RUBIDIUM ATOMIC FREQUENCY STANDARD DEVELOPED IN INDIA

G.M.Saxena, A.Chatterjee(Mrs), D.S.Sachdeva and B.S.Mathur

Time and Frequency Section

National Physical Laboratory, Hillside Road,

New Delhi-110012,India.

Abstract

A laboratory model Rubidium atomic frequency standard has been developed at N.P.L., New Delhi, India, in collaboration with Indian Institute of Technology,Kanpur. In order to evaluate frequency stability of the indigenously developed standard we have studied effect of various critical parameters like modulation level, temperature of the oven of the optical package, efficiency and noise behaviour of Silicon Photovoltaic cell, the detector etc.on the S/N of hyperfine resonance signal and its linewidth. We have tried both phase modulation of frequency multiplier chain and frequency modulation of synthesized frequency (5.314 MHz) and corresponding results on resonance signal and linewidth are reported. After optimization of various parameters the frequency stability of the standard has been computed with respect to commercial Cs frequency standard in time domain. The short term and medium term frequency stability of our Rb standard is of the order of 1×10^{-11}. Efforts for miniaturization and improving the frequency stability are being made.

Introduction

We report in this paper indigenous development of Rubidium Atomic frequency standard in the National Physical Laboratory, New Delhi, India. It is a secondary frequency standard but far less complex compared to cesium and Hydrogen frequency standards. In view of its wide ranging utility and relatively simple mechanism, the development programme of frequency standards at N.P.L., India has been initiated with Rubidium Atomic Frequency Standard Fig.1.

The principle[1] on which Rubidium Atomic Frequency Standard works is based on phase locking the VCXO (voltage controlled oscillator) to the atomic transitions between the Rb^{87} ground state hyperfine levels F=1, $m_F = 0$ and F=2, $m_F = 0$, Fig.2. These levels have second order dependence on the magnetic field.

Description of the Frequency Standard

The block diagram of the frequency standard is shown in the Fig.3. The heart of the Rubidium atomic frequency standard is the optical package or the frequency reference assembly. It will be discussed latter on in detail. The resonance r.f. magnetic field is derived from a 5 MHz VCXO after miltiplication and synthesis of its frequency. The 5 MHz is phase modulated at 137 Hz and multiplied to 60 MHz in one channel while in the other 5.314 MHz,synthesized frequency, is referenced by VCXO.

The signals from both the channels are fed to properely biased SRD mounted just inside the microwave cavity. Here 114th harmonic of 60 MHz signal is mixed with the synthesized 5.314 MHz signal to produce the resonance frequency 6.834 GHz for inducing the hyperfine transitions between two ground state sublevels of Rb^{87} isotope. The scheme of phase modulation of 60 MHz produces a sinusoidal scan of the exciting frequency in the optical package with peak deviation kept very small. The second harmonic 274 Hz appears at the resonance and fundamental 137 Hz when off the resonance. The magnitude and phase of 137 Hz produce necessary error signal after the synchronous detector and integrator and it is used to correct 5 MHz VCXO.

The optical package Fig.4 of the frequency standard consists of conventional Rb^{87} lamp, Rb^{85} filter cell and Rb^{87} absorption cell enclosed inside a microwave cavity operating in TE_{011} mode and having on its rear end a tuning plunger for fine tuning. A lense and solar cell combination is mounted on the tuning plunger. The optical components are maintained at the same temperature using non-inductive heater winding. For better temperature control another heater winding is used and it also provides the necessary temperature gradient between Rb^{87} reservoir and absorption cell. The optical package is magnetically shielded using two layers of μ-metal sheets. For applying necessary small magnetic field along the light axis a solenoid is used. The necessary inputs for the lamp excitation,

CH2186-0/85/0000-0059$1.00©1985IEEE

temperature stabilization and magnetic field are controlled by sophisticated electronics circuits.

Observations

To evaluate the performance of the Rb atomic frequency standard we have studied the effect of following factors on the linewidth and S/N of the resonance signal.

(i) Level of 137 Hz phase modulating frequency.

(ii) Frequency modulation of 5.314 MHz.

(iii)Characteristics of the solar cell.

(i) Level of 137 Hz Phase Modulating Frequency : We have studied how the level of 137 Hz the phase modulating frequency affected the linewidth of the resonance signal. We observe that under optimised conditions a level of 100 mV pp of 137 Hz modulating frequency corresponds to linewidth of 350 Hz of microwave transition. On increasing the modulation level to 200 mV pp the linewidth goes up to 600 Hz and for 250 mV pp levellinewidth is 1 KHz Fig.5. The linewidth broadens to 1.5 KHz for the level of 500 mV pp of the modulating frequency. After a careful study of the strength and linewidth of the resonance signal we found that a level of 100 mV pp of modulating frequency is optimum.

(ii) Frequency Modulation of Synthesizer Frequency (5.314 MHz): We have also tried the scheme of frequency modulation of the synthesizer frequency, which in the present case is 5.314 MHz. We observed that for 500 Hz, 1 KHz and 1.6 KHz FM levels of the modulating frequency the linewidths of the order of 475 Hz, 710 Hz and 950 Hz have been obtained respectively Fig.6. This scheme of frequency modulation could be used with the advantage that the 60 MHz, the multiplied frequency, of higher spectral purity is obtainable. It is because higher Q circuit in the multiplier chain could be used resulting ultimately in higher S/N of the resonance signal.

(iii) Characteristics of the Solar Cell :

In the Rubidium Atomic Frequency Standard a Solar cell is used for detecting the hyperfine transitions. We have undertaken a study[2] to estimate the weightage of characteristics of the Solar Cell, the detector, on the performance of the frequency standard. The detector, on the performance of the frequency standard. The detector is simply a silicon photovoltaic cell. Its role is crucial as the correction signal for offsetting the aging of the crystal is derived through it. Its characteristics like :

(a) dependence of output on the temperature,

(b) quantum efficiency and spectral response, and

(c) noise behaviour,

are important parameters governing the frequency stability of the frequency standard. The effects of these characteristics we have studied and report in the paper.

(a) Dependence of Output on the Temperature: The Solar cell output falls with increase in temperature. We have observed that Solar Cell d.c. output falls from 105 mV at 60°C to 40 mV at 80°C i.e. over a span on 20°C the solar cell output changed by 65 mV. As the Rb^{85} filter cell yields good results near 70°C we have maintained the operating temperature of the optical components at 70°C. At this temperature the d.c. output is 72 mV and 137 Hz resonance signal is also strong. We have, also observed that S/N of the d.c. output of the Solar Cell is a function of the temperature and it falls with increase in temperature. The resonance signal strength falls off on either side of the operating temperature Fig.7. It is because at lower temperature gain in the solar cell's d.c. output is more than off set by the fall in the filtering efficiency of Rb^{85} filter cell. While at higher temperature the efficiency of the filter cell increases upto certain temperature but it is accompanied by considerable fall in the efficiency of the photocell.

(b) Quantum Effiency and Spectral Response: The Solar cell detector monitors very minute changes in the intensity of the light falling on it near and at the resonance. It is important that it has high quantum effiency and good spectral response in the near infrared region. The quantum effiency and the response time depend on the width of the depletion region. A solar cell with large depletion width is desirable.

(c) Noise Behaviour : A Solar Cell is beset with the following types of noise processes :

(i) Photon Noise : This type of noise arises due to statistical variation in the number of photons incident on the solar cell and it also depends on the nature of the radiation. For longer wavelength the photon noise reduces as a result in present case the noise is little less than what is expected in the visible region. Any fluctuation in the Rb^{87} lamp output adds to this type of

noise. It is necessary to keep the photon noise low, for this lamp should be highly stable.

(ii) Johnson Noise : The thermal fluctuation is the basic cause of the Johnson noise. To keep the Johnson noise low we used two ovens each having temperature constancy better than 0.1°C. A better temperature control of the optical package may reduce Johnson noise to some extent.

(iii) Generation-Recombination Noise : This kind of noise originates from the fluctuation in the number of thermally generated carrier and depends on their lifetime. The generation-recombination noise can be reduced using better fabrication techniques.

Besides, these fundamental noise processes there are two other kinds of noise namely modulation and contact noise. Both are similar in nature with a difference that former is generated at the surface while latter originates at the metal-semiconductor contacts. The basis cause of these noise processes is dislocation or imperfection in the semiconductors. These kinds of noise could also be reduced with better techniques of fabrications. For application in Rb frequency standard a careful selection of solar cell should be made in order to keep low these types of noise.

Frequency Stability Measurements in Time Domain

The frequency stability of the indigenously developed Rubidium Atomic Frequency Standard has been measured against a commercial HP Cesium Atomic beam standard by comparing their relative phases. Under optimized conditions, 5 MHz signals from indigenously developed Rb frequency Standard and Commercial Cs frequency standard have been compared for several days using a phase recorder. The two sample variance $\sigma_y(2,\tau)$ for different sample times have been calculated. We have estimated that medium term frequency stability of the indigenously developed Rubidium frequency standard is of the order of 1×10^{-11} for a sample time of 1.4×10^{4} sec. Fig.8. Efforts are being made to make more elaborate frequency stability measurements stretched over longer sample time.

Conclusion

In this paper an attempt has been made to analyse the effect of various parameters on the frequency stability of the Rubidium atomic frequency standard. We have specifically studied the effect of level of phase modulation and frequency modulation of the exciting microwave frequency and Solar Cell characteristics on the linewidth and S/N of the resonance signal. The study on the solar cell characteristics has revealed that, beside, several critical factors which are well documented and have bearing on the frequency stability, the quantum effiency, response time, spectral response and noise behaviour of the Solar cell are no less important parameters. It is observed that better quantum effiency and noise behaviour of the solar cell may improve the frequency stability of the Rubidium Atomic Frequency Standard.

References

1. M.E.Packard and B.E.Swartz,"The Optically Pumped Rubidium Vapour Frequency Standard", IRE Transaction on Instrumentation, Vol. I-11, p.p. 215-223, December 1962.

2. G.M.Saxena, Mrs.A.Chatterjee, D.S. Sachdeva and B.S.Mathur, "Application of Photovoltaic Cell in Rubidium Atomic Frequency Standard", Photovoltaic Materials and Devices, New Delhi: Wiley Eastern Limited Publication, May, 1985.

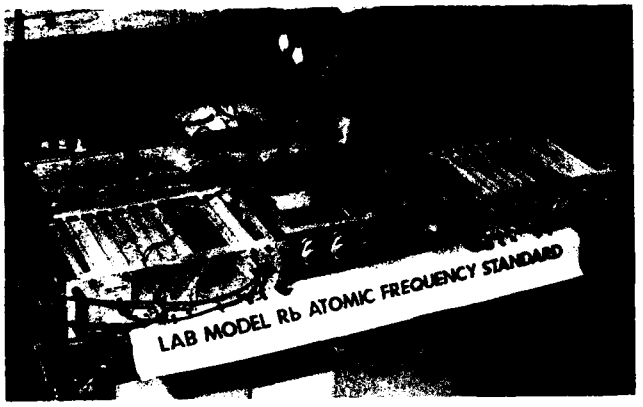
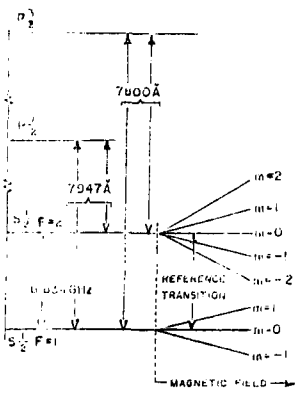


Fig. 1



Rb^{87} GROUND STATE ENERGY LEVEL DIAGRAM

Fig. 2

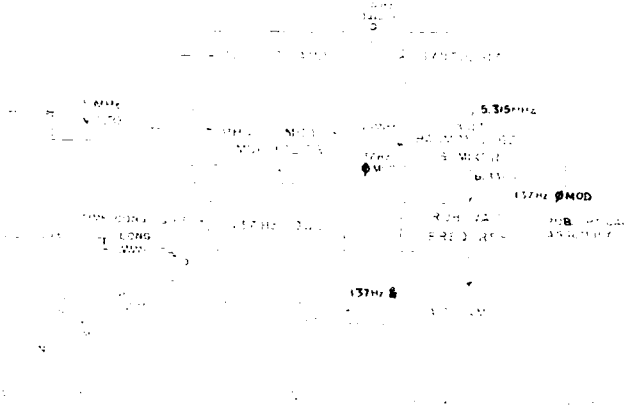


Fig. 3

Fig. 4

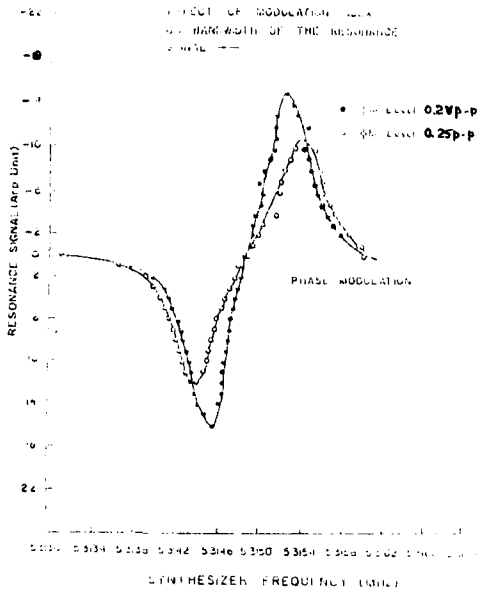


Fig. 5

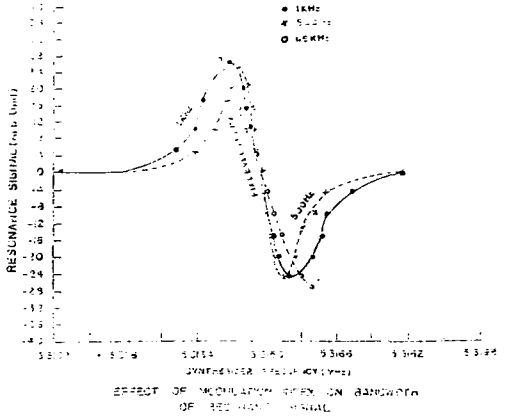
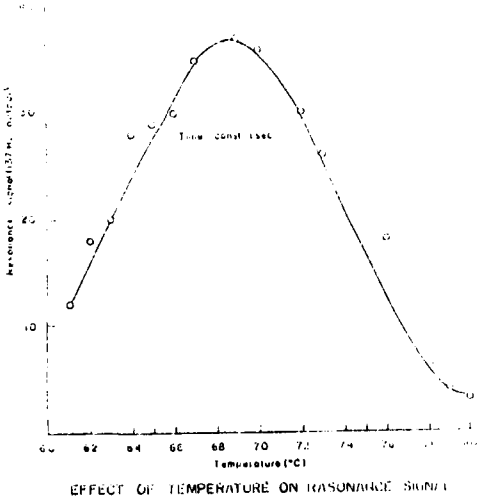
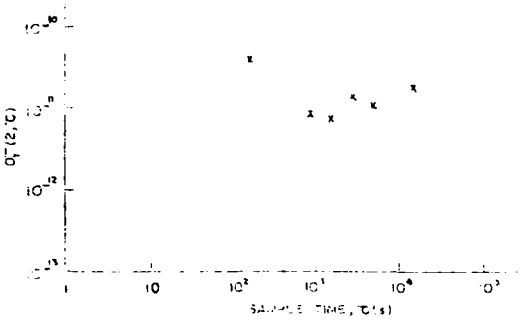


Fig. 6



EFFECT OF TEMPERATURE ON RESONANCE SIGNAL

Fig. 7



FREQUENCY STABILITY CHARACTERIZATION OF Rb-FREQUENCY STANDARD

Fig. 8

EXPERIMENTAL RESULTS ON A FREQUENCY STANDARD BASED ON A RUBIDIUM 87 MASER

by

M.TETU, R.BROUSSEAU, N.CYR, A.MICHAUD, P.TREMBLAY, B.VILLENEUVE

Laboratoire de Recherche sur les Oscillateurs et Systèmes
Département de génie électrique
Université Laval, Québec, G1K 7P4, Canada

Abstract: This paper gives some experimental results related to the study and the development of an atomic frequency standard using a rubidium 87 maser as the reference oscillator. The main features of the maser design are a sealed quartz coated bulb, a mixture of N_2-CH_4 buffer gas, an atmospheric pressure proof enclosure, a threefold oven and a four-layer magnetic shield. The characteristics of these features are presented. The design of a modular low noise coherent receiver needed to phaselock the quartz oscillator is also discussed and its phase noise contribution, measured. The transfer of frequency stability between the maser and the quartz oscillator is studied experimentally for short averaging times. A preliminary measurement of the mid term frequency stability is also obtained and an estimated overall performance of the frequency standard is given.

Introduction

At its very beginning the rubidium 87 maser was expected to become a highly stable frequency standard for short averaging times [1]. However many difficulties were encountered in its realization and the interest in its development faded out for some time. The second generation which incorporated a sealed quartz bulb containing the active vapor proved to be a satisfactory approach [2] and the short term frequency stability demonstrated was impressive [3]. On the other hand the mid term and long term frequency stability was somehow discouraging. More works were devoted to study some aspects of this problem and a better understanding of the origin of these instabilities came about [4],[5].

Recently a thorough study of the fundamental limits of the short term frequency stability [6],[7] and the evaluation of the systematic effects on the maser frequency [8] have shown that this type of maser oscillator could be much more stable than what is observed today. Following these expectations a new device has been constructed and a frequency standard based on this device is under development. We report in this paper a summary of the experimental results obtained regarding the behavior of the maser itself and some preliminary measurements of the frequency stability of this standard.

Maser design (MRb2)

Maser description

The rubidium maser consists of an optically pumped ^{87}Rb vapor in interaction with a buffer gas. The pumping light is emitted by an electrodeless lamp filled with ^{87}Rb vapor and 2 Torr of Kr. This light is filtered through a cell containing a vapor of ^{85}Rb and 50 Torr of Ar. The buffer gas used is a mixture of N_2 and CH_4 under a proportion given in the next section.

The active medium is contained in a sealed quartz bulb whose inner surface is coated with Paraflint [9]. The total weight of the bulb is less than 72 g. This bulb is placed in a right cylindrical quartz cavity operated in the mode TE_{021} and tuned at 6.835 GHz. Probably due to the low amount of quartz in the cavity and the presence of the Paraflint coating, the loaded cavity Q of this set up is found to be 42 000 at 55 °C with a coupling coefficient of 0.2. In the earlier design maser the loaded cavity Q was as low as 23 000 in equivalent operating conditions. Such a high Q system ensures much easier criteria for oscillation. The measured thermal coefficient of the cavity frequency is -2.2 kHz/°C.

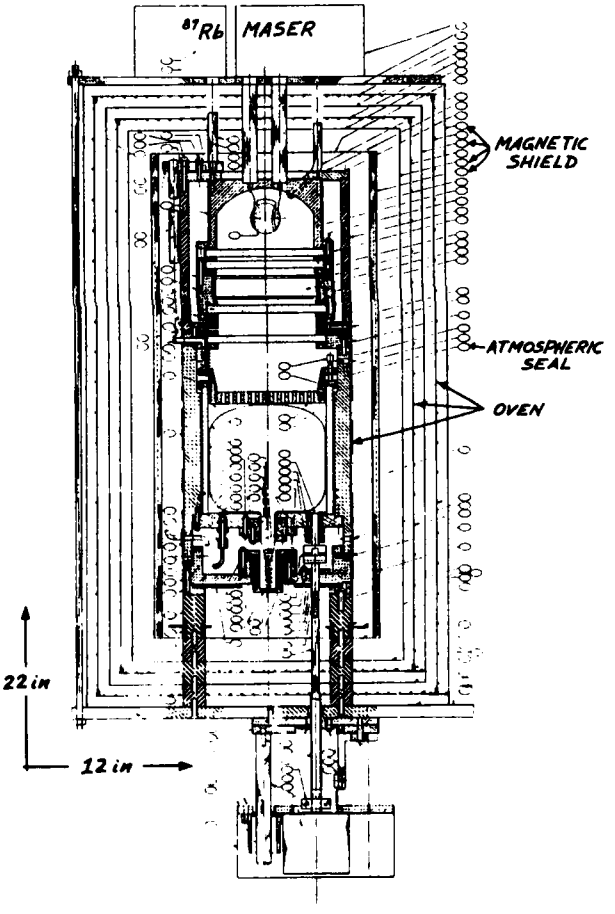


Fig. 1 Layout of the rubidium maser (MRb2).

A layout of the maser realized is shown in figure 1. The quartz bulb and the microwave cavity are contained into a vacuum tight enclosure to reduce the atmospheric pressure influence. The fine tuning is done with a moving rod driven by a stepping motor. The lamp

CH2186-0/85/0000-0064$1.00©1985IEEE

bulb is also in a vacuum tight enclosure. The temperature of the lamp enclosure, the filter holder and the cavity elements are regulated separately. Two other cylindrical ovens are used to regulate the maser temperature. The magnetic shielding effect is ensured by a fourfold cylindrical setup. The overall maser is a right cylinder of 12 in. diameter by 22 in. long.

Buffer gas mixture

Buffer gas is used in rubidium maser to reduce the Doppler effect, to increase the active volume and to improve the optical pumping efficiency. Unfortunately the buffer gas shifts the resonant frequency of the atoms by an amount which is proportional to the volume density and to the temperature of this gas. For a fixed volume cell, the buffer gas shift has been expressed phenomenologically by the relation [10]:

$$\Delta\nu_{BG}(T) = P_S[\beta+\delta(T-T_o)+\Gamma(T-T_o)^2], \quad (1)$$

where P_S is the buffer gas pressure at sealing temperature (25°C) and T_o is a nominal temperature (usually 60°C), β is the pressure coefficient, δ is the first order thermal coefficient and Γ is the second order thermal coefficient.

The value of the pressure and temperature coefficients for molecular nitrogen (N_2) and methane (CH_4) interacting with ^{87}Rb were measured and are given in Table 1.

	N_2	CH_4
β (Hz/Torr)	556.4 ± .1	-510 ± 1.5
δ (Hz/Torr/°C)	.572 ± .002	-.605 ± .002
Γ (Hz/Torr/°C^2)	-.00175 ± .0001	-.00001 ± .00006

Table 1: Measured values for the pressure and temperature coefficients of N_2 and CH_4 with ^{87}Rb.

As it is shown, the values of the first-order temperature coefficient for the two gases are close but of opposite sign so a compensation is possible. The second-order temperature coefficients are very small and play a role only when full thermal compensation is accomplished.

In the actual maser, a mixture in the ratio

$$r = \frac{P_{CH4}}{P_{N2}} = 0.465 \quad (2)$$

is used at a total pressure of 12.5 Torr, measured at 25 °C. This total pressure ensures an adequate maser gain [11]. The residual first-order thermal coefficient is then:

$$\delta = \frac{\delta_{N2} + r\,\delta_{CH4}}{1 + r} = 0.20 \quad (3)$$

The total first order thermal coefficient becomes 2.5 Hz/°C.

Maser operating conditions

Light shift and cavity pulling: A thorough study of the optical spectrum emitted by the prescribed setup has been conducted on Fabry-Perot to find the effects of the temperature of operation for the lamp reflector and the filter holder [12]. Self-reversal of the lamp [13] was observed for a reflector temperature of 70 °C which is a low value when compared to the 105 °C needed with the earlier lamp design. So it turns out that the pumping light module could be operated at much lower temperatures than expected, relaxing the constraint on the temperature gradients to be maintained.

Measurement of the maser frequency as a function of the cavity tuning frequency (cavity pulling) has been done for various light intensities (lamp oscillator voltages) in order to find the conditions at which the maser frequency is almost independent of the light variations when the cavity is tuned to the maser frequency. A typical result of this measurement is given in figure 2. It is seen that such a setting can exist after a right choice of the maser parameters. Also worth mentioning, the relatively low and close value for the temperature of the lamp reflector, the filter and the cavity. These measurements show also that the maser line Q varies from 5.5 x 10^7 to 7.5 x 10^7 accordingly to the light intensity and the cavity temperature.

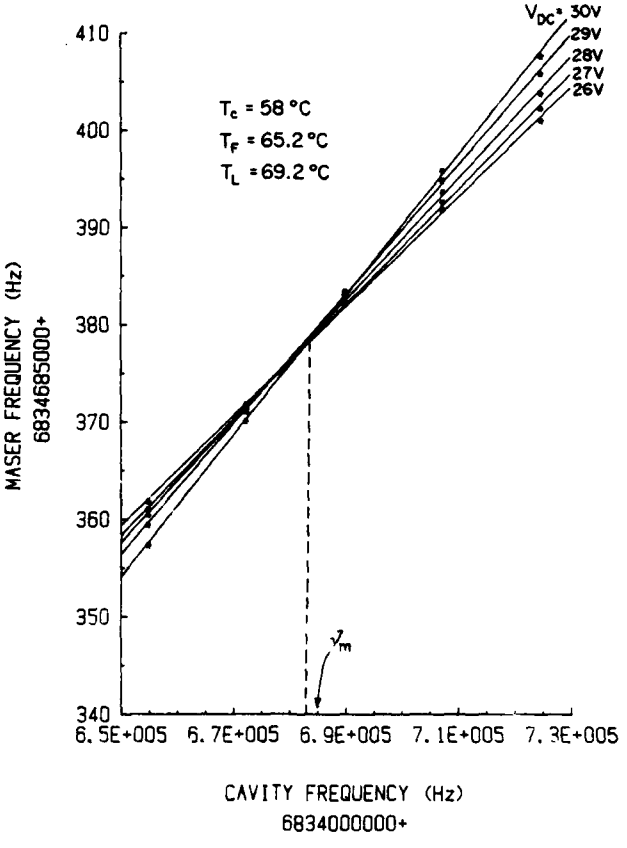


Fig. 2 Variation of the maser frequency as a function of the cavity frequency for various light intensities (lamp oscillator d.c. voltages).

Maser frequency thermal coefficient: From an ensemble of measurements as the one shown in figure 2 but obtained with different cavity temperatures, it is possible to evaluate the residual maser frequency thermal coefficient and typical results are given in

figure 3. This graph shows the maser frequency thermal coefficient as a function of the lamp oscillator d.c. voltage setting (light intensity) and for various cavity frequencies. It is seen that typical value for this coefficient is 1 Hz/°C. When the thermal variation of the total buffer gas shift (+2.50 Hz/°C) and the cavity pulling effect (-1.14 Hz/°C) are taken into account, we see that they justify the results obtained and that the temperature dependence of other frequency shifts, if they exist, are small.

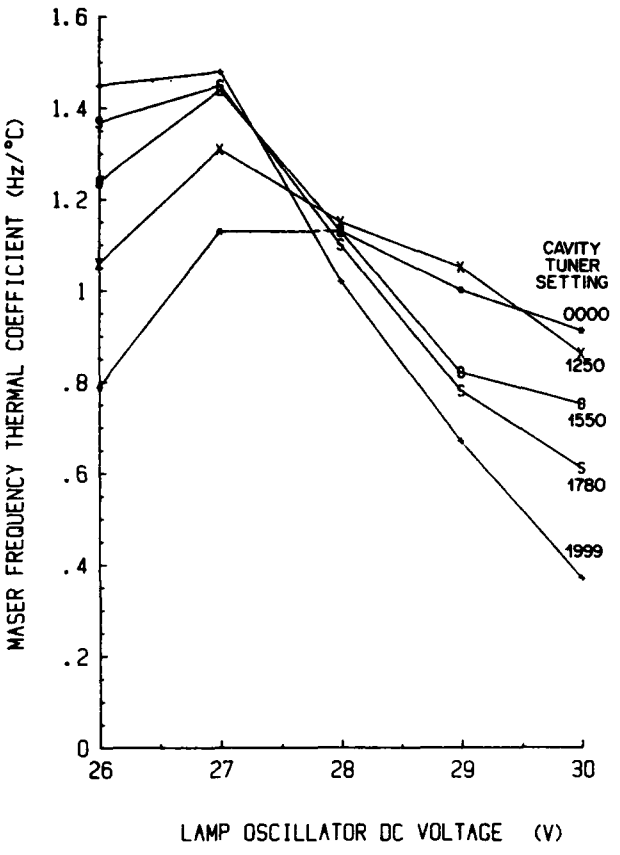


Fig. 3 Maser frequency thermal coefficient as a function of the lamp oscillator d.c. voltage (light intensity) for various cavity frequency.

An important feature of this measurement is that there exist experimental conditions where the maser frequency thermal coefficient is almost independent of the variations of the light intensity and frequency cavity tuning. Moreover, it appears possible to bring the residual absolute thermal coefficient to a very low value by a proper compensation between the cavity pulling effect and the buffer gas shift. However the adjustment of all these experimental conditions involves many tedious measurements.

Atmospheric pressure shift: Variations of the atmospheric pressure can alter the maser frequency. We have measured a relative maser frequency shift of -1.85 x 10^{-11}/Torr for a fully exposed maser while operated in standard conditions. With the cavity and the maser bulb placed in a sealed vacuum tight enclosure, this shift has been reduced to -4.57 x 10^{-12}/Torr and with the light bulb placed also in a sealed enclosure the shift has been reduced further to a value smaller than -4 x 10^{-13}/Torr. The origin of the contribution from the lamp is not known yet but the effect coming from the cavity is attributed to the cavity pulling, resulting from the variations of the cavity frequency with atmospheric pressure.

Magnetic shielding: The rubidium maser sensitivity to magnetic field variations is expressed as:

$$\Delta\nu_B = 5.74 \times 10^{10} B^2 , \quad (4)$$

where B is the magnetic field induction experienced by the atoms. A variation of the induction, dB, will shift the maser frequency by an amount given by:

$$d\Delta\nu_B = 11.48 \times 10^{10} B dB . \quad (5)$$

To reduce the effect of external magnetic field variations, a magnetic shield is necessary. Four concentric right cylinders with end caps, made out of a sheet of moly-permalloy are used. The dimensions of the cylinders are given in Table 2.

Cylinder no.	1	2	3	4
Radius (in.)	3.75	4.5	5.25	6
Length (in.)	16	17.5	19	22
Thickness (in.)	0.031	0.031	0.031	0.031

Table 2: Dimensions of the magnetic shield cylinders

The shielding factor as defined in [14], is evaluated to be 132 dB for the longitudinal variations and 137 dB for the transversal variations if operated in standard environment.

The rubidium maser here described is a compact device which has proven to be very reliable over the two years that it has been under study. The maser power is fairly high (10^{-9} W) due principally to the high loaded cavity Q factor and the efficient optical pumping scheme. This signal power plus the fact that low noise front end microwave amplifiers are now commercially available make possible the realization of a very stable frequency standard for short averaging times.

Coherent receiver design

Modular design

A coherent receiver (coherent transponder) is needed to phaselock a quartz oscillator to the maser signal. This allow the generation of standard (cardinal) frequencies based on the maser transition frequency [15]. The receiver design to be used with the rubidium maser is shown schematically in figure 4. It includes a low noise microwave amplifier (Miteq, model AMF-2S-6769-6367) located as close as possible to the maser output port and a four-stage down-converter. The amplifier and its input isolator are placed in a temperature stabilized box. The intermediate frequencies of the downconverter are 34.7 MHz, 4.7 MHz and 311 kHz and each local oscillator frequency is generated from the phase locked oscillator. The phase comparison is done at the level of 311 kHz with the use of a commercial synthesizer (Wavetek-Rockland, model 5100). The pilot oscillator is a 10 MHz quartz oscillator (Hewlett-Packard, model 10811B) and the standard output frequencies are 1, 5, 10 and 100 MHz.

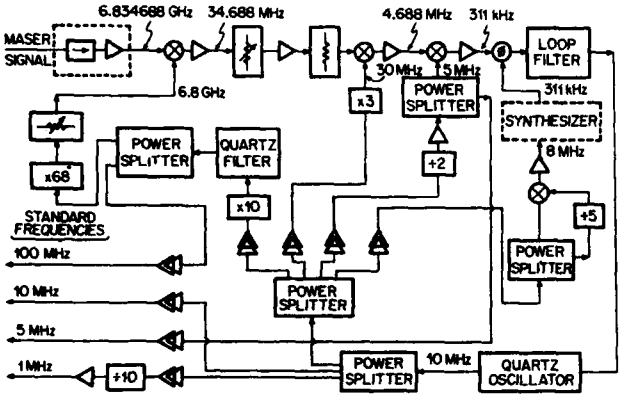


Fig. 4 Schematic representation of the coherent receiver used to phaselock a 10 MHz quartz oscillator to a ^{87}Rb maser.

Phase noise contribution

Down-converter: The noise contribution from the down-converter section to the phase of the signal, transmitted through it, has been measured with the setup shown in figure 5. The same high frequency signal is sent to two identical converters in which the local frequencies are generated from the same free running quartz oscillator. Two low frequency signals are produced at the output port of the phase detectors and the value of their frequency can be adjusted with the synthesizers multiplication factor. If the two frequencies are set to the same value, the phase noise contribution can be measured either in the frequency domain by a direct phase comparison within a phase detector followed by a low frequency spectrum analyzer or in the time domain with a time interval counter in a setup similar to the dual mixer one. Another possible approach to get a measurement in the time domain, is to set the two frequencies at slightly different values and measure the frequency stability of the beat frequency. These three methods were used in our experiment and they gave consistent results.

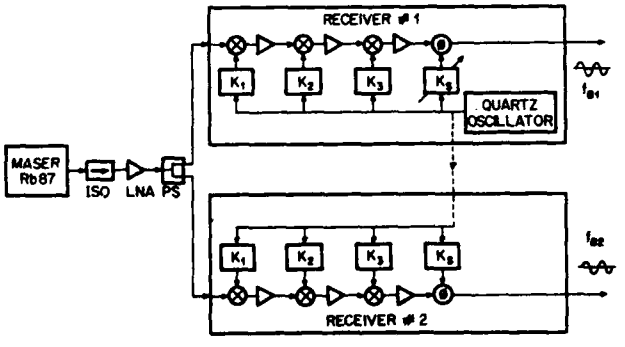


Fig. 5 Measurement setup used for the determination of the receiver phase noise contribution.

We give in figure 6, the results of the measurement in the frequency domain expressed as the spectral density of the phase fluctuations. These results show, unambiguously, the presence of flicker phase noise at a level:

$$S_{\varphi,r}(f) = 2.7 \times 10^{-6}\ f^{-1} \quad [rad^2/Hz] \qquad (6)$$

for a frequency range between .1 Hz and 30 Hz. The main source for this noise has not been identified in the present work. However it is supposed to be produced in the frequency multipliers [16].

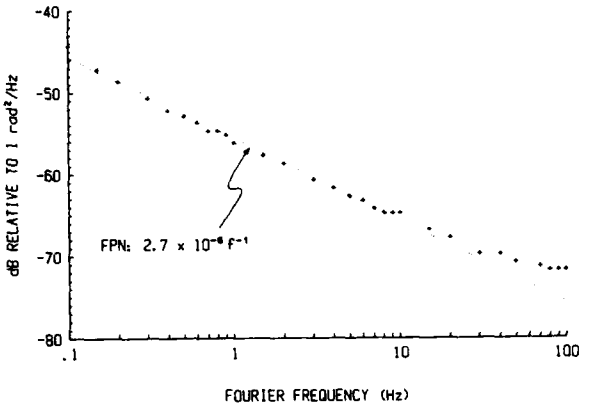


Fig. 6 Spectral density of the phase noise generated within the down-converter section of the receiver.

The translation of this result to a time domain measurement can now be done, taking properly into account the filter used to limit the signal R.F. spectrum [17]. If a low-pass rectangular filter with a cutoff frequency at 200 Hz is used and the beat frequency is set at 100 Hz, the two sample standard deviation will be:

$$\sigma(\tau) = 6.6 \times 10^{-14}\ [\ 7 + \ln \tau\ ]^{1/2}\ \tau^{-1}\ , \qquad (7)$$

which gives a value of 1.7×10^{-13} at 1 s averaging time. This value has been confirmed by experimental verification.

Additive noise: The thermal noise generated in the various elements of the receiver will add a white phase noise contribution proportional to the equivalent noise temperature. The gain and the noise temperature of the microwave amplifier were measured to be 22.2 dB and 208 K respectively. The gain of the down-converter was set at 69.1 dB and its noise temperature measured to be 3220 K. The overall receiver gain and noise temperature are then respectively 91.3 dB and 227 K. The single sided spectral density of the resulting white noise is given by the relation:

$$N_o = k\ T\ G\ R \qquad [V^2/Hz] \qquad (8)$$

where k is Boltzmann's constant, T is the equivalent noise temperature of the receiver (227 K) and G its gain (1.35×10^9), and R is the characteristic impedance of the system (50 ohms).

The single sided spectral density of the phase fluctuations is obtained through the relation [17]:

$$S_{\varphi,a} = \frac{N_o}{A^2}\ . \qquad (9)$$

Since the receiver gain is adjusted to deliver a signal at the output of the phase comparator having an amplitude of 0.5 V, the contribution of the additive noise to the phase spectral density is evaluated to 8.45×10^{-10} rad^2/Hz.

The contribution of the additive noise to the two sample standard deviation is given by [17]:

$$\sigma(\tau) = \left[\frac{3\, N_o}{4\, \pi^2 \nu_o^2\, A^2}\, F_h \right]^{1/2} \tau^{-1} \qquad (10)$$

where F_h is the equivalent limiting bandwidth of the measuring system. Evaluated for the present case, equation (10) gives the following contribution:

$$\sigma(\tau) = 1.2 \times 10^{-15}\, F_h^{1/2}\, \tau^{-1} , \qquad (11)$$

Discrete modulation: A close look at figure 6 shows the presence of a 60 Hz modulation at a level of -56.8 dB. Such a modulation will bring to a time domain measurement a contribution given by the relation [18]:

$$\sigma(\tau) = \frac{\Delta\nu_o}{\nu_o} \frac{(\sin \pi f_m \tau)^2}{\pi f_m \tau} . \qquad (12)$$

$\Delta\nu_o$ is the frequency excursion resulting from the modulation, which is in the present case 0.12 Hz, and f_m is the modulation frequency. The envelope of equation (12) will contribute to a time domain measurement by an amount given by:

$$\sigma(\tau) = 9.3 \times 10^{-14}\, \tau^{-1} . \qquad (13)$$

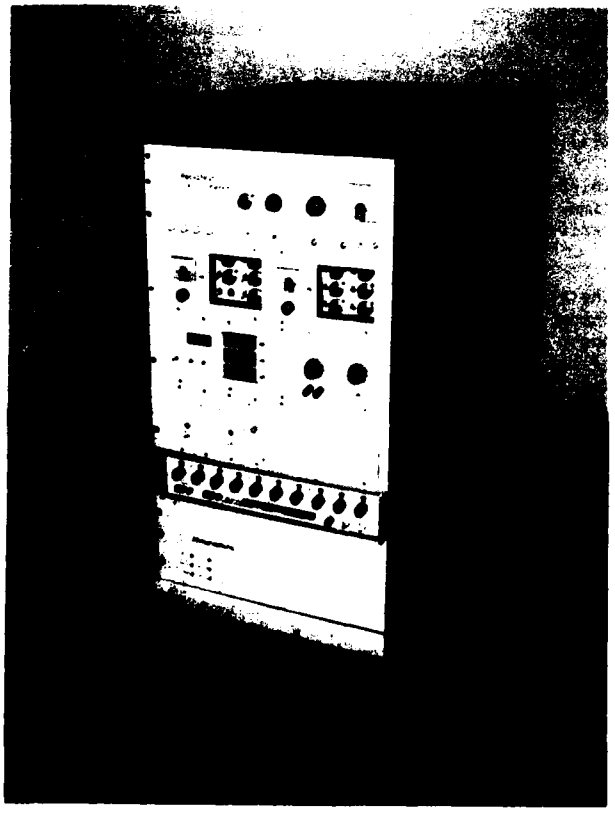


Fig. 7 a) Photograph of the frequency standard developed with a rubidium maser as the reference oscillator.

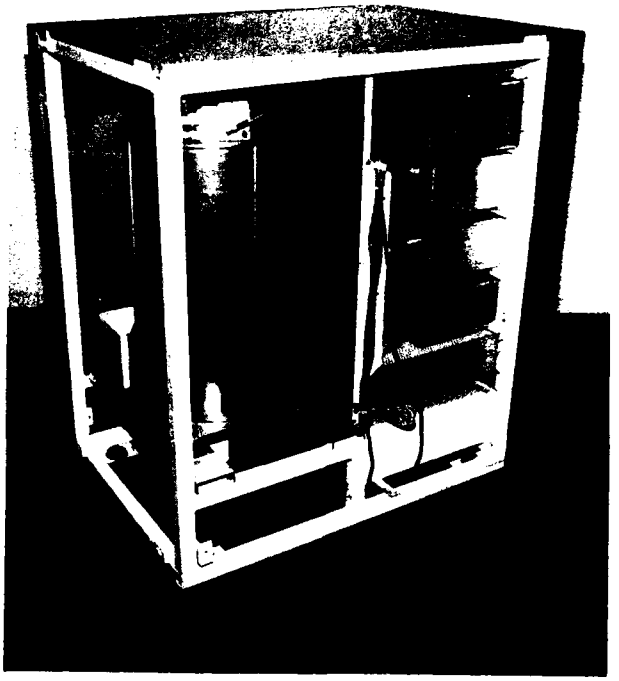


Fig. 7 b) Same standard with side panel removed, giving a view of the volume occupied by the maser.

From all these measurements, we conclude that the coherent receiver will contribute to the phase noise of the signal transmitted through it, a quantity which is, in terms of the two sample standard deviation, approximately 2×10^{-13} at 1 s averaging time. The main type of noise contributed is a flicker phase noise and, as it will appear in the following section, the present receiver design might become a limiting factor in the transfer of short term frequency stability.

Phase-locked loop parameters: The loop filter (see figure 4) is made out of a low-pass filter with a 15 kHz cutoff frequency followed by a compensated integrator. It can be shown that the overall system is equivalent to a second order, type 2, phase-locked loop [15],[19]. In such a PLL, both the natural frequency and the damping can be adjusted to the desired values. In the present receiver the damping is maintained to a value of 0.7 and the natural frequency can be selected between 10, 100 and 1000 Hz. In the following study on the frequency stability transfer, the last value has been selected.

Transfer of frequency stability

Rubidium maser frequency standard

The rubidium maser, its control electronics and the coherent receiver are integrated in a single unit shown in figures 7a and 7b. Although no important effort has been paid to reduce the size to a minimum, the unit is 2 feet wide by 3 feet high by 3 feet long. We consider in the following that the frequency standard is the quartz oscillator phase-locked to the rubidium maser. As demonstrated through theoretical

work, the short term frequency stability of the standard is expected to be that of the quartz oscillator while the long term stability should be that of the maser thus improving the long term frequency stability of the quartz oscillator. However in the present case, the maser is a more stable oscillator for short averaging times and it could also improve slightly the short term frequency stability of the quartz oscillator [20].

Rubidium maser short term frequency stability

The rubidium maser short term frequency stability has been measured by comparison with another rubidium maser of an earlier design (RbM1) [21]. The measurement is done by beating together the signal of each maser at the input of an A.M. receiver and counting the frequency of the beat signal generated through an envelope detector [3].

The results obtained for the two sample standard deviation are indicated in figure 8 by the asterisks. These values are obtained with a measurement setup using a bandpass filter centered at 3 kHz and having an equivalent bandwidth (F_h) of 750 Hz. We deduce from this measurement that the short term frequency stability is limited by a white phase noise contribution expressed as:

$$\sigma(\tau) = 5.5 \times 10^{-14}\ \tau^{-1} \qquad (14)$$

for averaging times shorter than 0.8 s. For longer averaging times, the frequency instabilities start to increase and are attributed to the comparison maser, RbM1.

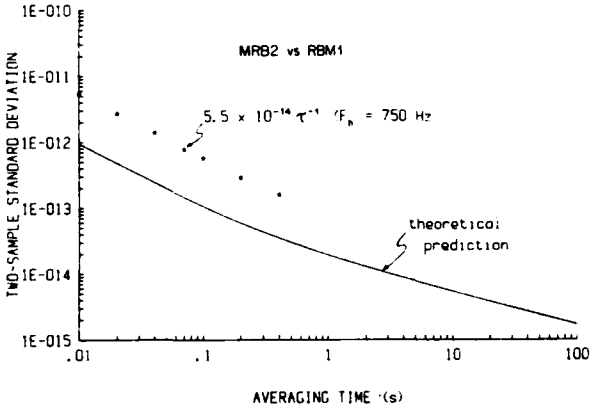


Fig. 8 Measurement of the short term frequency stability of the rubidium maser. The asterisks are experimental data and the solid line is the theoretical evaluation.

When compared to the theoretical evaluation done with the same maser and measurement setup parameters, and indicated by the solid line of figure 8, we see that this impressive result is still a factor 6 higher than expected. This discrepancy is not resolved yet but there are good reasons to believe that the measured data belong to the comparison maser or are the result of the additive noise generated in the envelope detector. More measurements are needed to obtain a definite answer and we will consider at this point that the observed data are the upper limit the short term frequency stability of the new maser MRb2.

Transfer of short term frequency stability

The short term frequency stability of the free running quartz oscillator has been measured by comparison between two similar oscillators in a dual mixer operated at 100 MHz. The resolution of the measurement setup has been tested to be at a level of $1 \times 10^{-13}\ \tau^{-1}$ in a 1.8 kHz equivalent bandwidth filter. The results obtained are given by the crosses in figure 9. The two sample standard deviation can be expressed as:

$$\sigma(\tau) = [2.3 \times 10^{-25}\ \tau^{-2} + 9 \times 10^{-24}\ \tau^{0}]^{1/2}\ , \qquad (15)$$

while the measurement setup equivalent bandwidth is 1.8kHz.

When each quartz oscillator is phase-locked to one of the rubidium masers used in the measurement of figure 8, the results indicated by the circles in figure 9 are obtained. It is observed that the flicker frequency noise of the quartz oscillator is corrected and that the instabilities are brought to the level of the white phase noise contribution ($4.8 \times 10^{-13}\ \tau^{-1}$ with F_h = 1.8 kHz). Although the results are interesting, they are not representative of the maser itself. To achieve this goal, a quartz oscillator with lower white phase noise is necessary and eventually a lower flicker phase noise receiver.

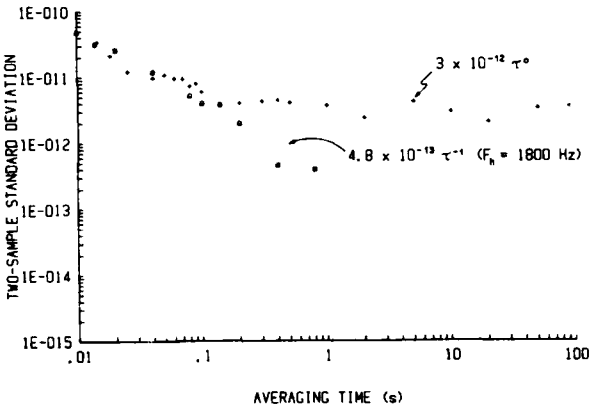


Fig. 9 Measurement of the transfer of frequency stability for short averaging times. The crosses represent the stability of the free running quartz oscillator and the circles are the stability of the same oscillator when phase-locked to the maser.

Mid term frequency stability

A preliminary evaluation of the mid term frequency stability of the standard has been conducted by comparing the frequency of the 100 MHz output signal to another 100 MHz signal, derived from a frequency standard based on an hydrogen maser. The two 100 MHz frequencies were offset by 50 Hz, mixed, and low pass filtered with an equivalent filter bandwidth of 150 Hz. The results obtained are given in figure 10. We consider this measurement as preliminary because we have verified, from measurements on the H-maser standard, that the white frequency noise contribution observed for averaging times shorter than 200 s belongs to the H-maser standard in its actual operating

conditions. Moreover, since the H-maser was not under controlled temperature, it is not proved that the long term instabilities belong to the Rb-maser standard. Anyhow, since a value as low as 6 x 10^{-14} is observed at 250 s averaging time, we are encouraged by this preliminary result. We shall then consider that the frequency instabilities observed for averaging times longer than 250 s represent the worst case value for the present rubidium maser frequency standard.

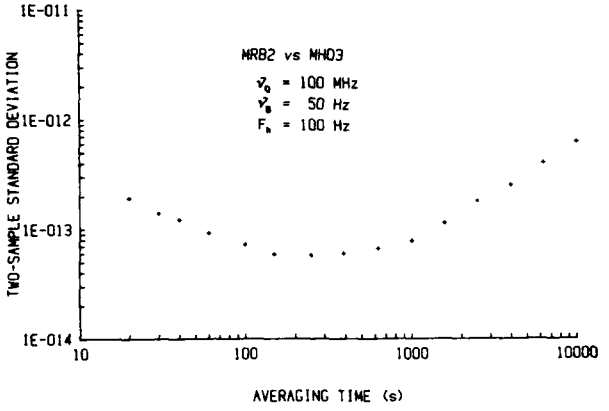


Fig. 10 Measurement of the mid term frequency stability of the Rb-maser standard when compared to an H-maser standard.

Overall performance

The overall performance of the rubidium maser frequency standard is depicted in figure 11. The solid line represent the frequency stability of the quartz oscillator phase-locked to the maser signal. The short term frequency stability can be expressed as 1 x 10^{-14} $F_h^{1/2}$ τ^{-1} and can be improved if a more stable quartz oscillator becomes available. For longer averaging times, the frequency stability of the standard is characteristic of the maser and reaches a level of 6 x 10^{-14} at 250 s. The dashed line indicates the worst case frequency stability of the rubidium maser itself. The short term value can be written as 2 x 10^{-15} $F_h^{1/2}$ τ^{-1}. Between 0.8 s and 200 s, the frequency stability cannot be specified since no other sufficiently stable frequency standard was on hand at the time of the measurement.

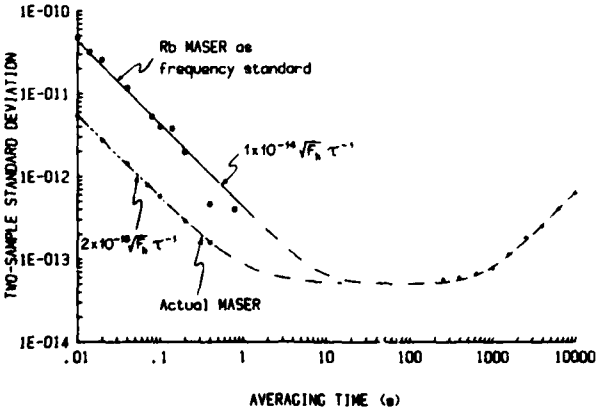


Fig. 11 Overall performance of the Rb-maser frequency standard.

Conclusion

The results of this study show that the behavior of a rubidium maser is now more under control and that its realization can be made easier. They also prove that the rubidium maser is effectively a very stable oscillator for short averaging times and that its long term frequency stability can be increased. The barrier of 1 x 10^{-13} has been broken for both short and mid term averaging times. The transfer of this maser performance to a pilot oscillator is possible but the actual coherent receiver design might become a limiting element. A detailed study of the long term frequency stability is needed before we specify the full performance of the actual maser design (MRb2).

Acknowledgements: The authors wish to thank MM Roger Blier, Yvon Chalifour, Claude Dubé and Julien Morasse for their skillful contribution during various phases of this project. This work has been financed through grants from CRSNG/NSERC, Ottawa, Canada and Fonds FCAC, Province of Québec, Canada.

References

[1] P. Davidovits and R. Novick, "The Optically Pumped Rubidium Maser", Proc. IEEE, Vol. 54, No. 2, pp.155-170, February 1966.

[2] J. Vanier, "Relaxation in Rubidium-87 and the Rubidium Maser", Phys. Rev., Vol. 168, No. 1, pp. 129-149, April 1968.

[3] M. Têtu, G. Busca, and J. Vanier, "Short-term Frequency Stability of the Rb^{87} Maser", IEEE Trans. Instr. Meas., Vol. IM-22, No. 3, pp. 250-257, September 1973.

[4] J. Vanier, D. H. Nguyen, G. Busca, and M. Têtu, "Study of the effect of light shifts and buffer gas shifts on the hyperfine transition of ^{87}Rb: influence of the long term frequency stability of the rubidium maser", Can. J. Phys., Vol. 57, No. 9, pp.1380-1387, December 1979.

[5] G. Busca, R. Brousseau and J. Vanier, "Long-Term Frequency Stability of the Rb^{87} Maser", IEEE Trans. Instr. Meas., Vol IM-24, No 4, pp.291-296, December 1975.

[6] P. Tremblay, "Etudes sur le maser au rubidium 87: Comportement en présence d'une boucle de réaction", M. Sc. Thesis, Université Laval, Génie électrique, Septembre 1982.

[7] M. Têtu, P. Tremblay, P. Lesage, and P.Petit, "Experimental Study of the Frequency Stability of a Maser Oscillator Operated with an External Feedback Loop", IEEE Trans. Instr. Meas., Vol. IM-32, No. 3. pp. 410-413, September 1983.

[8] M. Têtu, P. Tremblay, D. Bonnier and J. Vanier, "Short term Frequency Stability and Systematic Effects on the Rubidium 87 Maser Oscillator Frequency", Proc. 36th Annual Symposium on Frequency Control, Philadelphia, U.S.A., pp. 340-347, June 1982.

[9] Paraflint is a trade name of a hydrocarbon of the polyethylene type sold by Moore and Munger, Inc., Stamford, CT, U.S.A.

[10] J. Vanier, R. Kunski, N. Cyr, J. Y. Savard, and M. Têtu, "On hyperfine frequency shifts caused by buffer gases: Application to the optically pumped passive rubidium frequency standard", J. Appl. Phys., Vol. 53, No. 8, pp. 5387-5391, August 1982.

[11] W. A. Stern and R. Novick, "Further results on the rubidium-87 maser frequency standard", Proc. of the 26th Annual Symposium on Frequency Standard, Atlantic City, U.S.A., pp. 223-224, May 1972.

[12] G. Missout, R. Vaillancourt, M. Têtu, and J. Vanier, "Etudes de Filtres Isotopiques pour les Masers au Rubidium," Rev. Phys. Appl., Tome 6, pp. 307-311, September 1971.

[13] R. Bernheim, "Optical Pumping", W. A. Benjamin, New York, U.S.A., 1965.

[14] R. Kunski and J. Vanier, "On magnetic shielding in hydrogen masers", J. Phys. E: Sci. Instrum., Vol. 15, pp. 1207-1209, 1982.

[15] J. Vanier and M. Têtu, "Phase-Locked Loops Used with Masers", IEEE Trans. Communications, Vol Com-30, No. 10, pp. 2355-2361, October 1982.

[16] V. F. Kroupa, "Frequency Stability: Fundamental and Measurement", IEEE Press, 1983, Part VI

[17] P. Tremblay and M. Têtu, "Frequency Stability Characterization from the Filtered Signal of a Precision Oscillator", Proc. of the 39th Annual Symposium on Frequency Control, Philadelphia, U.S.A., this issue, 1985.

[18] J. Rutman, "Characterization of Phase and Frequency Instabilities in Precision Frequency Sources: Fifteen Years of Progress", Proc. IEEE, Vol. 66, pp. 1048-1075, september 1978.

[19] F. M. Gardner, "Phaselock Techniques", 1966, J. Wiley and Sons, New York, U.S.A.

[20] J. Vanier, M. Têtu, and L.-G. Bernier, "Transfer of Frequency Stability from an Atomic Frequency Reference to a Quartz-Crystal Oscillator", IEEE Trans. Instr. Meas., Vol. IM-28, No. 3, September 1979.

[21] G. Busca, J. Racine, and J. Vanier, "A Compact Rb^{87} Maser", Proc. 28th Annual Symposium on Frequency Control, Atlantic City, U.S.A., pp. 344-347, May 1974.

MEASUREMENT AND INTERPRETATION OF HYDROGEN MASER QUALITY PARAMETERS

Edward M. Mattison,* Wei Shen,*† and Robert F.C. Vessot*

*Smithsonian Astrophysical Observatory
Cambridge, Massachusetts 02138, U.S.A.
†Shaanxi Observatory
Academia Sinica
The People's Republic of China

INTRODUCTION

Two properties of interest in characterizing the operation of the hydrogen maser are (a) the effectiveness of the state selecting magnet and (b) the relaxation rates of hydrogen atoms in the storage bulb. A measurement of the maser's line Q as a function of output power is capable of yielding information about these properties. This measurement provides an estimate of the maser quality parameter "q" and of the relaxation rate "γ_t", from which other relaxation rates can be inferred. (Methods of measuring q by varying the cavity Q using external gain[1] and by applying inhomogeneous magnetic fields[2] have been described in the literature.) We derive a relationship between q, line Q, and maser power, and then discuss the application of this technique.

THEORY

Q_ℓ, the maser's line Q, is related to the total transverse relaxation rate γ_2 of the hydrogen atoms' magnetic moment:

$$Q_\ell = \frac{\omega}{2\gamma_2} \tag{1}$$

where $\gamma_2 = \gamma_d + \gamma_2'] + \gamma_{2se}$. Here γ_d, the total rate at which atoms leave the storage bulb volume, is given by

$$\gamma_d = \gamma_b + \gamma_r \tag{2}$$

where γ_b is the rate of escape through the bulb entrance aperture and γ_r is the rate of recombination into molecular hydrogen on the bulb wall. γ_2' is the density-independent part of γ_2 exclusive of γ_d, and γ_{2se} is the part of γ_2 due to hydrogen-hydrogen spin exchange. γ_{2se} is related to the hydrogen density n in the bulb by

$$\gamma_{2se} = \frac{1}{2} n \bar{v} \sigma \tag{3}$$

where σ is the spin exchange cross section[3,4] and $\bar{v}$ is the mean relative velocity of atoms in the bulb. The density is related in turn to the hydrogen beam flux by[3]

$$n = \frac{\dot{}_{tot}}{\gamma_d V_b} \tag{4}$$

where I_{tot} is the total flux of hydrogen atoms entering the bulb and V_b is the bulb volume. Using Eqs. 1,2, and 4 one can relate the line Q to I,the flux of atoms in the upper maser state (F=1, m_F=0):

$$Q_\ell^{-1} = \frac{2}{\omega}\left[\gamma_d+\gamma_2' + q\frac{I}{I_{th}}\gamma_t\right] \tag{5}$$

where

$$I_{th} = \frac{hV_c\gamma_t^2}{4\pi\mu_o^2 Q_c \eta} \tag{6}$$

and the quality parameter q is defined by

$$q = \left[\frac{\sigma\bar{v}h}{8\pi\mu_o^2}\right]\frac{\gamma_t}{\gamma_d}\left[\frac{V_c}{\eta V_b}\right]\left(\frac{1}{Q_c}\right)\frac{I_{tot}}{I} \tag{7}$$

Here h is Planck's constant, μ_o is the Bohr magneton, V_c is the volume of the resonant cavity, Q_c is the loaded cavity Q, η is the bulb magnetic filling factor[3], and $\omega=2\pi f$, where f is the maser oscillation frequency. The quantities in square brackets can be calculated for a given cavity and bulb geometry, and Q_c can be measured. The relaxation rate γ_t is given by

$$\gamma_t = \left[(\gamma_d + \gamma_1')(\gamma_d + \gamma_2')\right]^{1/2} \tag{8}$$

γ_1' is the density-independent part of the longitudinal relaxation rate exclusive of γ_d. Eq. (5) shows that Q_ℓ^{-1} is proportional to the flux I. While I is not a directly measurable quantity, it can be related to the power P radiated by the masing atoms:[3]

$$\frac{P}{P_c} = -2q^2\left(\frac{I}{I_{th}}\right)^2 + (1-cq)\frac{I}{I_{th}} - 1 \tag{9}$$

where

$$P_c = \hbar\omega I_{th}/2$$

and

$$c = \left[\frac{\gamma_d+\gamma_1'}{\gamma_d+\gamma_2'}\right]^{1/2} + 2\left[\frac{\gamma_d+\gamma_2'}{\gamma_d+\gamma_1'}\right]^{1/2}$$

Note that P is related to the power P_o output from the maser cavity by

$$P = P_o\frac{1+\beta}{\beta}$$

where β is the cavity coupling factor[5,6]. P is a quadratic function of I and thus of Q_ℓ^{-1}. The parameter q is generally less than 0.1 in most masers, and I/I_{th} is of the order of 2 to 3; thus the first term of Eq. (9) is smaller than the other terms and may be ignored for a linear approximation. Algebraic manipulation of Eqs. 5-9 then gives the following linear relationship between Q_ℓ^{-1} and P:

$$Q_\ell^{-1} = \frac{2}{\omega}(\gamma_d+\gamma_2')\left[1 + \frac{q}{1-cq}\frac{\gamma_t}{\gamma_d+\gamma_2'}\right] + A\left(\frac{q}{1-cq}\right)\frac{1}{\gamma_t}P \tag{10}$$

where

$$A = \frac{16\pi\mu_o^2 Q_c \eta}{\omega^2 h^2 V_c}.$$

If we assume that $\gamma_1' = \gamma_2' \equiv \gamma'$, then

$$\gamma_t = \gamma_d + \gamma', \tag{11}$$

$$c = 3, \tag{12}$$

and

$$Q_\ell^{-1} = b + mP \tag{13}$$

with $b = \frac{2}{\omega}\gamma_t\left[1 + \frac{q}{1-3q}\right]$ and $m = A\left[\frac{q}{1-3q}\right]\frac{1}{\gamma_t}$ (14)

The slope m and intercept b of Eq. 13 are each function of two unknowns, γ_t and q. Thus by measuring Q_ℓ^{-1} as a function of P and calculating the slope and intercept of the straight line fit to the data, one can obtain q and γ_t. Fig. 1 shows measured values of Q_ℓ^{-1} vs. P for two masers.

Since q is generally much less than 1 (its theoretical maximum value is 0.172[3]), Eq. (14) shows that γ_t is primarily determined by b, which is the extrapolation of the curve of Q_ℓ^{-1} vs. P to zero power, and is thus essentially independent of the measurement of P. q, on the other hand, is directly determined by the slope m, and is thus sensitive to the calibration of maser output power and the measurement of the cavity coupling factor β.

CH2186-0/85/0000-0072$1.00©1985IEEE

APPLICATIONS

With the values of q and γ_t thus determined, and the bracketed quantities in the expression for q measured or calculated, Eq. 7 yields the product of γ_d^{-1} and I_{tot}/I, but not the two quantities separately. However, a strategy to determine them is available. If γ_d is made large, the ratio of γ_t/γ_d approaches unity and I_{tot}/I can be calculated. γ_d can be increased either by increasing the recombination rate in the bulb (for example by suspending an object in the bulb by a fine thread) or by increasing the effective escape rate from the bulb (by placing a thin plastic or metal collar in the bulb's collimator). As γ_d is increased in stages by such a method, the q decreases to a minimum value at which $\gamma_t/\gamma_d \approx 1$, and I_{tot}/I can be calculated. Since the state selection efficiency is independent of the bulb's properties, I_{tot}/I has the same value for other conditions of the bulb, allowing γ_d, and thus γ_r, to be determined for the original (unaltered) bulb.

The measurement of q and γ_t can be used to assess the effect of changes in the magnetic state selector with or without determining γ_d. For a particular bulb geometry and wall coating, q varies only with I_{tot}/I. Then if the state selection efficiency is changed, q changes proportionately. This approach is particularly useful in measuring the effectiveness of techniques intended to eliminate the ($F=$ $m_F=1$) state from the hydrogen beam.[7,8]

An example of this approach is shown in the data of Table 1. Measurements were made on a maser with a deteriorating bulb coating. (A low line Q and good magnetic field properties pointed to a bad coating.)

Table 1.

	Before Recoating	After Recoating
$Q_\ell(0)$	1.21×10^9	2.94×10^9
Q_c	44526	32086
β	0.222	0.171
q	0.79	0.09
γ_t	3.35	1.35 sec-1
γ_t/γ_d	1 (assumed)	1.07
I_{tot}/I	3.50	3.50 (assumed)
γ_d	3.35	1.26
γ_b	0.87	0.87 sec-1
γ_r	2.48	0.39 sec-1

For the calculations of Table 1 we used $\sigma=2.33\times10^{-15}cm^2$[4], $\bar{v} = 3.58\times10^5$cm/sec, $V_c=15184cm^3$, $V_b=2893cm^3$, and $\eta=2.14$. Assuming $\gamma_t/g_d = 1$ gives $I_{tot}/I = 3.5$ and $\gamma_d = 3.35$ sec^{-1}. Since the calculated bulb escape rate is $\gamma_b = 0.87$ sec^{-1}, this gives $\gamma_r = 2.48$ sec^{-1}. After the bulb was recoated, $\gamma_t = 1.35$ sec^{-1}. Assuming that I_{tot}/I remained constant, $\gamma_t/\gamma_d = 1.07$ $\gamma_t = 0.39$ sec^{-1}. The calculated value of γ_t/γ_d is very close to unity even after the bulb was recoated and a high line Q was measured, indicating that the coating had a long storage time. This shows that recombination dominates wall relaxation due to dephasing collisions with the bulb wall, and that $\gamma_t/\gamma_d \approx 1$ for most conditions. (For strongly inhomogeneous magnetic fields, magnetic relaxation may cause $\gamma_t/\gamma_d > 1$.) It should be noted that the calculated value of I_{tot}/I is much greater than the theoretical value of 2. This indicates that the calculated values of q are substantially in error. q is sensitive to measurements of power and coupling factor. (We suspect that the mismatch between the maser's resonant cavity and its output isolator may have lead to an error in these measurements.) Thus a single measurement of q can be considered to represent a calibration of β and power for the system under test, so that the ratios of subsequent measurements of q and γ_t yield valid conclusions about changes in relaxation rates or state selection conditions.

REFERENCES

1. Audoin, Claude, Michel Desaintfuscien, and Jean-Pierre Schermann, "Mesure du paramètre q d'un maser à hydrogène," C. R. Acad. Sc. Paris, **264**, 698 (1967).

2. Vanier, J. and R.F.C. Vessot, "Relaxation in the level F=1 of the ground state of hydrogen: application to the hydrogen maser," IEEE J. of Quantum Electronics, **QE-2**, 391 (1966).

3. Kleppner, D., 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**, 972 (1965).

4. Allison, A.C., "Spin-change frequency shifts in H-H collisions," Phys. Rev. **5**, 2695 (1972).

5. Ginzton, E.L., Microwave Measurements, Ch. 9. McGraw-Hill, 1957.

6. Ginzton, E.L., "Microwave Q measurements in the presence of coupling losses," IRE Trans. on Microwave Theory and Techniques, p. 383 (1958).

7. Audoin, Claude, Michel Desaintfuscien, Pierre Petit, and Jean-Pierre Schumann, "Design of a double focalization in a hydrogen maser," IEEE Trans. on Instr. and Meas., **IM-17**, 351 (1968).

8. Urabe, S., K. Nakagiri, Y. Ohta, M. Kobayashi, and Y. Saburi, "Majorana effect on atomic frequency standards," IEEE Trans. on Instr. and Meas., **IM-29**, 304 (1980).

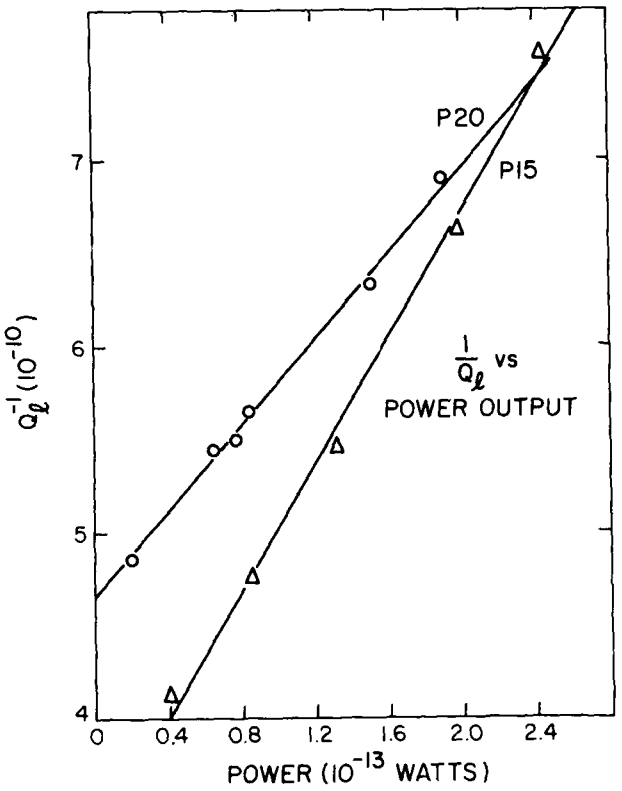


Fig. 1. Q_ℓ^{-1} as a function of maser power.

PROPERTIES OF LOW-EXPANSION MATERIALS FOR HYDROGEN MASER CAVITIES

Edward M. Mattison and Robert F. C. Vessot
Smithsonian Astrophysical Observatory
Cambridge, Massachusetts 02178
and
Stephen F. Jacobs
Optical Sciences Center
University of Arizona
Tucson, Arizona

ABSTRACT

We have measured thermal and mechanical properties of four low-expansion materials that are candidates for use in hydrogen maser resonant cavities: Cer-Vit® C-101 and C-126, ULE®, and Zerodur®. Thermal expansivities were measured at room temperature and after heating the samples to 300°C and to 700°C. The approximate expansivities, $(1/L)(dL/dT)$, in units of 10^{-8} $^{o}C^{-1}$, are: Cer-Vit C126, -40; Cer-Vit C101, -13; Zerodur, -12 to -6; and ULE, +5. Zerodur's expansivity decreased in magnitude by 50% after heating to 700°C, while the expansivities of ULE and C101 remained substantially unchanged.

Measurements of the effects of grinding rectangular samples of the the materials show that grinding produced surface stresses of 4 to 7 N/cm. Etching with hydrofluoric acid relieved essentially all of this stress.

Samples were coated with fired-on silver films. The initial internal stress, σ, in the films was approximately $1.5x10^{3}$ N/cm^{2}. The rate of change of stress with temperature was $d\sigma/dT=-43 N cm^{-2} {}^{o}C^{-1}$. The samples did not return completely to their initial shapes after thermal cycling by 10°C. Within the experimental uncertainty, we observed no consistent change of shape of the samples over a two-year span due to relaxation of the internal film stress.

INTRODUCTION

Chief among the systematic perturbations that can affect the frequency of the hydrogen maser are variations in the resonance frequency of the maser's microwave cavity. A change in cavity frequency of 0.1 Hz, which corresponds to a change in cavity length of roughly $1x10^{-8}$cm, shifts the maser output frequency by roughly 2 parts in 10^{15}. Since maser frequency stabilities of several parts in 10^{16} and better have been achieved, such a shift is readily observable; thus it is desirable to maintain the cavity dimensions constant to the order of Angstroms.

Because dimensional stability is important, maser cavities are usually constructed of mechanically stable, low thermal expansion dielectric materials coated with a conductive layer of silver. A material commonly used for maser cavities (Cer-Vit C-101) is no longer available, making it important to evaluate the characteristics of alternative cavity materials. In addition, it is desirable to understand how surface stress in the material and in its silver coating affects the cavity's dimensions. In the present study we have measured the thermal expansion coefficients of four candidate cavity materials and have observed the behavior of the materials when ground, acid etched, and coated with silver.

The materials examined are ULE®, Code 7971, Corning Glass Works, Corning, NY; Cer-Vit® C-101 and C-126, Owens-Illinois, Inc., Toledo, Ohio; and Zerodur®, Schott Optical Glass Inc., Duryea, PA.

ULE is an amorphous single-phase silica glass composed of titanium silicate and silicon dioxide. Cer-Vit and Zerodur are glass-ceramics, in which a transparent crystalline phase is distributed throughout a vitreous (glassy) material. Because the crystalline and vitreous phases have opposite expansion coefficients, the compositions of these materials can be adjusted to produce a net expansivity close to zero over a selected temperature range. Cer-Vit C-101 is a transparent material, cast in large ingots, that must be cut and ground to shape; it has been used for the cavities of SAO's ground masers. Cer-Vit C-126 is an opaque white material that can be cast in the shape of the finished product, and has been used for the cavity of the SAO/NASA Redshift space maser.[1] Neither material is currently produced by Owens-Illinois, which has sold the rights for manufacturing Cer-Vit to Schott Optical Glass, Inc. Selected mechanical properties of these materials are given in Table 1.[2-5]

Table 1. MECHANICAL PROPERTIES OF CAVITY MATERIALS

Material	Density (g/cm^3)	Young's Modulus ($10^6 N/cm^2$)	Poisson Ratio	Knoop Hardness* (kg/mm^2)
ULE	2.21	6.75	0.17	459
Zerodur	2.53	9.06	0.24	
Cervit C-101	2.50	9.24	0.25	540
Cervit C-126	2.51	8.48	0.26	

* 100 gram loading

THERMAL EXPANSIVITY

The thermal expansivities (linear coefficients of thermal expansion) of the materials were measured under three conditions: (i) after fabrication of the test samples; (ii) after the samples were heated to 300°C; and (iii) after the samples were reheated to 700°C, to simulate the heat treatment the material undergoes when it is coated with silver. (The C-126 sample was not heated.)

For the expansivity measurement, a cylindrical sample with a hole bored along the symmetry axis forms the etalon of a confocal Fabry-Perot interferometer. A HeNe laser mode-matched to the etalon is electronically servoed to follow the resonant frequency of the etalon. The servoed laser beam is beaten with that of a frequency-stabilized laser. The change in resonant frequency caused by a temperature-induced change in etalon length is measured by a frequency counter that measures the change in beat frequency.

One sample of each of the four materials was ground to shape and etched with hydrofluoric acid to relieve the surface stress created by grinding[6]. The expansivity, α, of each sample was measured near room temperature. Because C-126 was unknown to us, its expansivity was measured at temperatures between 100 and 400°K. No further heat treatments and measurements were performed on C-126. The other three samples (ULE, Zerodur, and C-101) were heated to 300°C over a period of approximately 1 hour, held at 300°C for approximately 10 minutes, and cooled to room temperature over approximately 2 hours, after which their expansivities were measured at room temperature. The

CH2186-0/85/0000-0075$1.00©1985IEEE

samples were then heated to 700°C over approximately 4 hours, held for 20 minutes, cooled over 6 hours, and measured at room temperature.

The expansivity of C-126 is plotted in Fig. 1. The data for the other materials are given in Fig. 2. α_{C-126} is approximately $-40 \times 10^{-8} °C^{-1}$ at room temperature, which is 3 to 7 times that of the other materials. For this reason C-126 is not desirable for maser cavities, and the measurements after heat treatment were not performed. As seen in Fig. 1, $\alpha_{C-126}=0$ at T=162°C.

Of the other materials, ULE has the lowest expansivity (in magnitude), approximately $5.5 \times 10^{-8} °C^{-1}$ at room temperature. α_{ULE} is also the most constant with measurement temperature and after heat treatment.

α_{C-101} is the largest expansivity in magnitude, approximately $-14 \times 10^{-8} °C^{-1}$ at room temperature, and becomes rapidly more negative with small temperature increases. Heat treatment reduces the magnitude of α_{C-101} slightly -- by about 6.3% for heat treatment to 700°C.

The initial expansivity of Zerodur, approximately $-12.2 \times 10^{-8} °C^{-1}$, was comparable to α_{C-101}. Heating to 300°C and then to 700°C, however, changed $\alpha_{Zerodur}$ by about 50%, reducing its magnitude to $-6 \times 10^{-8} °C^{-1}$. This substantial change indicates that Zerodur's expansivity can be reduced by proper heat treatment, but raises questions about the stability and predictability of the material.

SURFACE STRESS DUE TO GRINDING

Method of Measurement

The shape of the resonant cavity can be affected by stress acting parallel to the surface of the cavity material. Change in this stress can alter the cavity's shape and thus change the maser's application frequency. Two major sources of surface stress are grinding of the cavity parts, and application of a conductive metallic coating to the inside surface. To determine the amount of initial stress and of stress change from these sources, we measured the bending of rectangular samples of cavity materials. Samples of ULE, Zerodur, C-101, and C-126 were formed into rectangular plates 10.1 x 1.9 x 6.4 cm (4 x 3/4 x 1/4 inch), optically polished on the two broad (4 x .75 inch) faces. Plate bending was measured by photographing interference fringes formed by the plates relative to a non-contacting optical flat illuminated by a HeNe laser. The fringe patterns were digitized and transferred to a computer programmed to eliminate the tilt of the plate relative to the flat. The program fit a set of polynominals to the data and calculated the amount of deflection, δ, of the bent plate, expressed in wavelengths of 6328Å light (see Fig. 3).

Measurement Precision

The deflection measurement process consisted of photographing the fringes, digitizing the photographs, and calculating the amount of deflection. To test the repeatability of the measurement process, 13 photographs of two plates in their initial polished state, were taken during one day. To ensure measurements independence, the samples were removed from the interferometer and replaced between photographs. The standard deviation of the measured deflections, calculated by pooling the variances for the two plates, is 0.040 λ.

Effect of Grinding and Etching

The polished samples were treated as follows: One face of each plate was ground to simulate the grinding process used on maser cavities; the unground surfaces were masked with wax; and the ground face was etched with hydrofluoric acid to relieve the stress of grinding. Four plates were photographed and digitized in their initial states after being ground and after being acid etched. The data are given in Table 2. The average increase in deflection due to grinding is $(1.62 \pm .14)\lambda$ (mean ± standard error).

Table 2. EFFECT OF GRINDING AND ETCHING

	Deflection δ (wavelengths)				
	A	B		C	
PLATE	INITIAL	GROUND	B-A	ETCHED	C-A
11	1.405	3.214	1.809	1.303	-.022
12	1.029	2.553	1.524	.956	-.073
29	1.775	3.396	1.621	1.894	+.119
30	1.380	2.903	1.523	1.274	-.109
			Mean = 1.619±.14		Mean = .021±.10

Grinding is known[6] to create microscopic cracks in the the ground surface that produce an internal compressive stress, curving the plate away from the ground surface. Acid etching tends to relieve this stress. The last column of Table 2 shows that etching the ground plates returns the plates to their initial deflections to within $(.02 \pm .10)\lambda$. Thus essentially all of the initial stress is relieved.

The initial grinding-induced stress can be calculated from the measurements as follows. If a stressed thin film of thickness d is applied to a plate of thickness D, length L, Young's modulus E, and Poisson's ratio p, the radius of curvature r produced in the plate is related to the stress σ (force per unit area) in the film by [7,8]

$$\sigma = \frac{ED^2}{6(1-p)rd} = \frac{4ED^2\delta}{3(1-p)L^2d} \qquad (1)$$

where δ is the plate deflection. If the region near the ground surface is regarded as a thin stressed film, the surface stress σ in the ground surface (force per unit width) is

$$\sigma_s = \frac{4\,ED^2\delta}{3(1-p)L^2} \qquad (2)$$

Table 3 gives the surface stresses calculated using Eq. 2, the values of E and p from Table 1, and the measured deflections of Table 2.

Table 3. SURFACE STRESS PRODUCED BY GRINDING

PLATE	MATERIAL	SURFACE STRESS (N/cm)
11	C-101	7.35
12	C-101	6.19
29	ULE	4.35
30	ULE	4.08

These results show that lower surface stress was induced in the ULE plates than in the Cer-Vit samples. This is probably due to less crack propagation in ULE, which may be related to ULE's relative softness, shown by the Knoop hardness values given in Table 1.

STRESS DUE TO SILVER COATING

After being ground and etched, ten of the sample plates were coated on their ground faces with the silver film used at SAO for producing conductive cavity coatings.[9] This ink consists of finely divided sil-

ver mixed with a small amount of glass-forming oxides (frit) suspended in a liquid vehicle. The ink is sprayed on the vitreous substrate, dried, and fired at approximately 700°C. At the firing temperature the frit fuses the silver particles to each other and to the substrate; in addition, the silver tends to sinter together. As the unit cools, the silver contracts more rapidly than the substrate, setting up a tensile stress in the coating that tends to bend the substrate concave toward the coated side. Table 4 gives the deflections of the ten plates in their initial states and after being silver coated.[10] The differences range from 2λ to 11λ. In the preceding section we have seen that the average difference in deflection between the initial and etched states is a fraction of a wavelength. Thus, we can calculate the film stress by comparing the coated deflection with the initial deflection rather than with the etched deflection, which is unavailable for eight of the plates.

Using Equation 1 and the film thicknesses given in Table 4, we obtain the stresses shown in the last column of Table 4. The stresses on the films on C-126 are significantly less than those on other substrates. C-126 when initially cast is a clear glassy material with a greenish hue. After heat treatment it becomes opaque white and appears to be inhomogeneous, displaying light and dark rings at cut surfaces. The material was obtained for this experiment in its white state. It may have changed properties or shape when the silver was fired, resulting in a lower apparent film stress.

The average stress in the silver films applied to the non-C126 plates is $(1.52 \pm 0.3) \times 10^3 N/cm^2$.

Table 4. STRESS DUE TO SILVER COATING

Plate	Mat'l	Deflection δ (λ) Init	Coated	Init - Coated	Coating Depth 10^{-3}inch	Stress $10^3 N/cm^2$
4	Zerodur	1.22	-7.41	8.63	6.25	1.12
7	C-101	1.66	-7.84	9.50	4.88	1.64
9	C-101	1.11	-6.09	7.20	4.20	1.44
11	C-101	1.41	-0.85	2.26	1.00	1.90
15	C-126	0.77	-2.99	3.76	5.05	0.58
16	C-126	-0.55	-3.50	2.98	4.95	0.47
21	C-126	0.78	-1.09	1.87	2.08	0.70
24	ULE	2.16	-8.78	10.94	4.37	1.39
28	ULE	1.51	2.03		1.95	
30	ULE	1.38	-1.68	3.06	1.05	1.62
					Mean (excl C-126):	1.52 ±0.3

rmal Expansion and Hysteresis

The deflections of the coated plates were measured at 23°C and 33°C to determine the effect on the film of thermal expansion. The average rate of change of deflection with temperature, $d\delta/dT$, is given in Table 5. ($d\delta/dT$ for plate 28 is positive because its deflection increases with temperature, indicating that the plate is convex toward its coated face.)

Eq. 1 predicts that the change in deflection is proportional to film thickness d, assuming that the thermally-induced change in surface stress is a constant. We tested this by assuming a power-law relationship, $d\sigma/dT = kd^m$, and regressing log $|d\sigma/dT|$ against log d. We find m = 1.013, indicating that the relationship between $d\delta/dT$ and d is very close to linear, as predicted. The change in surface stress is obtained by differentiating Eq. 1 with respect to temperature:

Table 5. THERMAL EXPANSION OF COATING

Plate	Coating Thickness (10^{-3} inch)	$d\delta/dT$ (λ/°C)	$d\sigma/dT$ (N cm^{-2} $°C^{-1}$)
4	6.25	-.353	-46.0
7	4.88	-.362	-62.4
9	4.20	-.210	-42.0
11	1.00	-.048	-40.4
15	5.05	-.210	-32.5
16	4.95	-.241	-38.1
21	2.08	-.060	-22.6
24	4.37	-.485	-61.6
28	1.95	+.133	-37.9
30	1.05	-.094	-49.7
		Mean	-43.3±12

$$\frac{d\sigma}{dT} = \frac{4}{3} \frac{ED^2}{(1-P)L^2 d} \frac{d\delta}{dT} \qquad (3)$$

The average value is

$$\left\langle \frac{d\sigma}{dT} \right\rangle = (-43.3 \pm 12)\ N cm^{-2} {}^{o}C^{-1} \qquad (4)$$

The standard error in Eq. 4 indicates a good deal of scatter. This is due partly to thermally induced changes in the plates between measurements. The temperature of the experiment was controlled to approximately ±0.5°C, leading to nonrepeatability in the deflection measurements on the order of 0.2 to 0.5 wavelengths. In addition, there appears to be hysteresis in the plate curvature following warming and cooling. The coated plates were photographed at 23°C, 33°C, and again at 23°C, with overnight equilibrations at the three conditions. As shown in Table 6, heating to 33° reduced the initial 23°C deflection (marked 23A), while returning to 23°C increased the deflections (marked 23B) to values that were consistently smaller than their initial values. (For reasons explained previously[10], plate 28 runs contrary to this behavior: heating increases the deflection, and returning to 23°C leaves $\delta(23B) > \delta(23A)$.)

Table 6. THERMALLY INDUCED HYSTERESIS

Plate	Deflection δ (λ) 23A	33	23B	|23A-23B| / |33 -23A|	ΔT(°C)
4	7.91	4.14	7.35	0.15	1.6
7	9.63	5.10	7.84	0.40	4.9
9	7.06	4.05	6.01	0.35	5.0
11	0.99	0.36	0.90	0.14	1.9
15	3.75	0.70	2.92	0.27	4.0
16	4.47	0.87	3.63	0.23	3.5
21	1.65	0.37	1.13	0.41	8.7
24	9.21	4.15	8.78	0.08	0.9
28	1.57	3.13	2.03	0.29	3.5
30	1.90	0.76	1.64	0.23	2.8

The fifth column of Table 6 gives the fractional hysteresis as a proportion of the initial change from 23°C to 33°C; with one exception the values are on the order of 15-40%. The last column of Table 6 gives the difference in ambient temperature, ΔT, that would be necessary to account for the deflection difference $\delta(23B) - \delta(23A)$. It is calculated by:

$$\Delta T = \frac{|\delta(23B) - \delta(23A)|}{d\delta / dT} \qquad (5)$$

where $d\delta/dT$ is obtained from Table 5. Most of the values of ΔT are considerably larger than the actual temperature uncertainty in the experiment, indicating that at least part of the hysteresis effect is due to actual changes in the materials. This would account

in part for the scatter in the deflection measurements of coated samples, which is much greater than the scatter observed in the repeatability measurements.

Temporal Behavior of Samples

The silver coating clearly exerts a tensile stress on the cavity material. If this stress relaxed as a function of time, it could change the frequency of the cavity and thus of the maser. The cavity elements most susceptible to change of coating stress are the flat cavity endplates. The coating stress would tend to make them concave, and as the stress relaxed the endplates would tend to flatten, raising the cavity resonance frequency.

To measure the temporal stability of the coated samples, fringe photographs of various samples were made over periods up to two years. For each plate the average rate of change of deflection was calculated by means of a least squares fit of deflection against time. For plates 4, 7, and 28 the scatter in the data is large enough to make a straight-line fit meaningless. The coefficient of determination, R^2, was calculated for plates with more than two data points. R^2 measures the proportion of total variation about the average deflection that is due to the regression. (R^2=1 implies a perfect fit of the data to the regression line, and R near unity indicates a slope that differs significantly from zero.) The results of the regression analysis are given in Table 7.

The most significant regressions are those for plates 6 and 8, both of which were uncoated. The smaller degree of scatter for these data is probably due to the uncoated plates' relative immunity to temperature. For the coated plates, however, Table 7 shows that no significant systematic variation of deflection is discernible within the scatter of the data.

Table 7. TIME VARIATION OF DEFLECTION

	-------23 °C------			-------33 °C-------		
PLATE	dδ/dT (λ/mo)	R^2	POINTS	dδ/dT (λ/mo)	R^2	POINTS
4	----------		5	----------		3
6*	.0092	.68	7			
7	----------		4	-.015	---	2
8*	.021	.95	8			
9	.032	.32	7	.030	---	2
11	-.021	.40	12	-.013	---	2
15	.043	---	2	.088	---	2
16	-.017	---	2	.105	---	2
21	-.073	---	2	-.046	---	2
24	-.144	.59	3	.037	.85	3
28	----------		3	.004	---	2
29*	.014	55	3			
30	-.036	.51	12	-.022	---	2

*=uncoated

ACKNOWLEDGEMENT

This work was sponsored by the Jet Propulsion Laboratory, Pasadena, California, under contract 954938.

REFERENCES

1. Vessot, R.F.C., et. al., "Test of relativistic gravitation with a space-borne hydrogen maser," Phys. Rev. Lett. 45, 2081-2084 (1980).

2. ULE titanium silicate, code 7971, brochure LEM-ULE-2/72 GLP, Corning Glass Works, Corning, New York (1972).

3. Zerodur transparent glass ceramics, Schott Optical Glass Inc., Duryea, Pennsylvania (undated).

4. Cer-Vit material for reflective optics, brochure RO-3 3-69, Owens-Illinois, Inc., Toledo, Ohio (1969).

5. Typical properties of Cer-Vit material C-126, Owens-Illinois, Inc., Toledo, Ohio (undated).

6. Muffoletto, C. Verne, "Reflective and refractive scattering of ultraviolet radiation caused by state-of-the-art optical grinding and polishing techniques," in Reflecting optics for synchrotron radiation, Soc. Photo-optical Instrumentation Engineers, vol. 315, pp. 85-88 (1981)

7. Campbell, David S., "Mechanical properties of thin films," in Handbook of thin film technology, L. I. Maissel and R. Glang, eds., McGraw-Hill, pp. 12-3--12-50 (1970).

8. Glang, R., R. A. Holmwood, and R. L. Rosenfeld, "Determination of stress in films on single crystal silicon substrates," Rev. Sci. Inst. 36, 7-10 (1965).

9. Silver ink 421, Engelhard Industries Division, Engelhard Minerals and Chemicals Corporation, East Newark, NJ.

10. The signs of the "initial" and "coated" deflections in Table 4 indicate the direction of curvature of the plates: a positive deflection indicates a plate concave toward the observed (polished) side, while a negative deflection indicates convexity toward the polished side. All of the plates were initially observed toward their concave faces (except plate 16, which was almost flat and had a degree of twist) and were then ground and etched on their convex faces. All of the coated deflections except 28's are negative, showing that the silver films caused the plates to reverse their curvatures and to bend concave toward the silvered faces, making the polished faces convex. Plate 28 is anomalous in having a positive deflection after coating, implying that it is concave toward its polished face. This would result if it had initially been more concave toward the polished face and the coating on the convex face had partially reduced its curvature. This picture is consistent with the fact that plate 28's deflection increases with temperature rather than decreasing, but means that the observed initial deflection (given in Table 4) is not the same as the curvature prior to coating. Thus it is not possible to calculate a coating-induced deflection change for this plate.

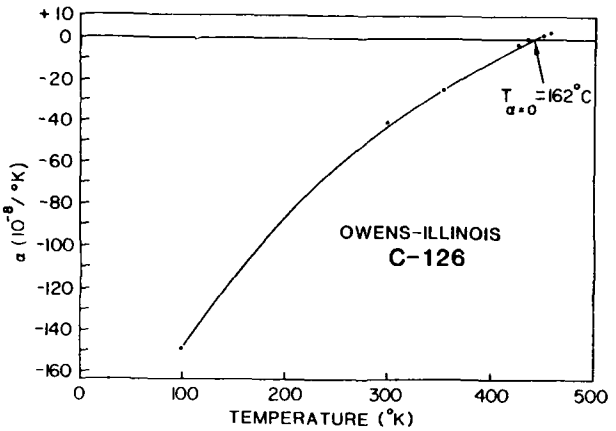


Fig. 1. Expansivity of Cer-Vit C-126

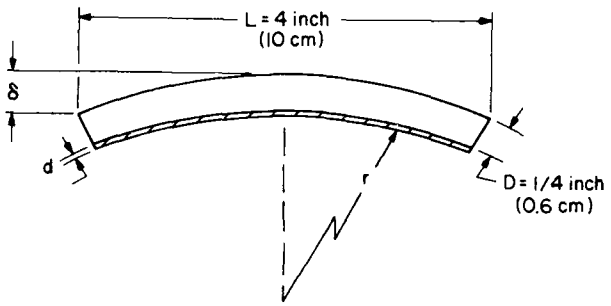


Fig. 3. Diagram of sample plate for surface stress measurements

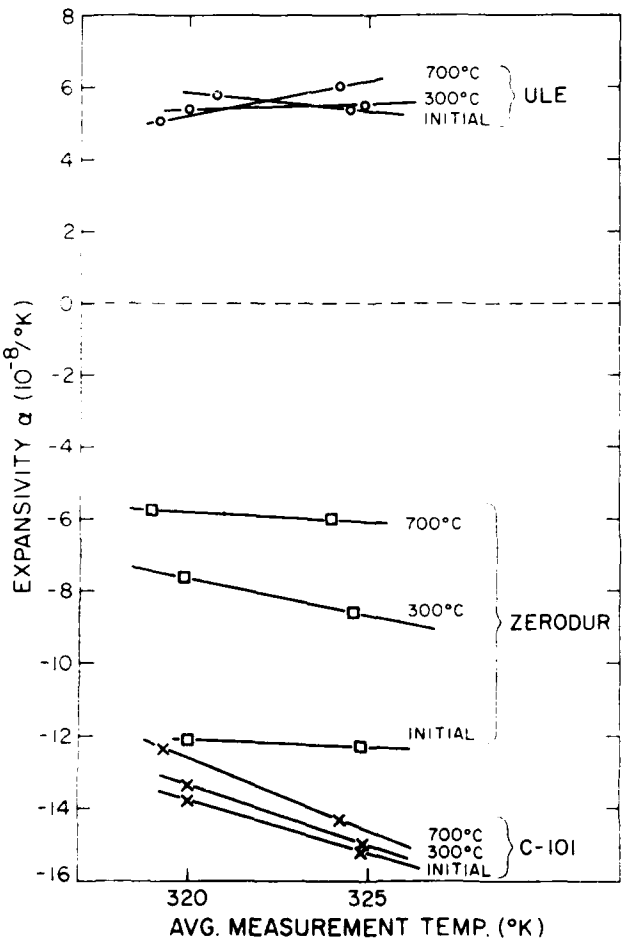


Fig. 2. Expansivities of ULE, Zerodur, and Cer-Vit C-101 as functions of heat treatment

DESIGN FOR A SUBCOMPACT Q-ENHANCED ACTIVE MASER*

Robert R. Hayes and Harry T.M. Wang
Hughes Research Laboratories
Malibu, CA 90265

Abstract

A design is presented for a compact hydrogen maser considerably smaller than those previously developed by this laboratory. The design uses cavity loading to reduce the diameter and length of the microwave cavity to 3 in., and Q-enhancement to achieve maser oscillation. The smaller cavity size leads to commensurate reductions in the size of several other components, particularly the magnetic shields, with the result that the assembled physics package (microwave cavity, vacuum chamber, dissociator, state selector and four nested magnetic shields) will fit within a 5.5-in. by 5.5-in. by 14-in. envelope. Significant improvements have also been made in the size and power consumption of the maser electronics. New, smaller components have allowed a complete microwave front-end (a 1.42-GHz amplifier, an active mixer, an active phase shifter and attenuator, a 5 to 1400-MHz frequency multiplier, several dividers and combiners) and a critical part of the cavity-control electronics to be housed in a 5.5-in. by 5.5-in. by 2-in. package. Further reductions in system size have been made possible by the commercial availability of integrated-circuit number controlled oscillators (NCOs), which have permitted the development of small, efficient frequency synthesizers for the receiver and cavity-control system. The maser which will result when all of these components are combined should prove very attractive for applications in which high stability, small package size and low weight are all critical requirements.

Introduction

The increasing sophistication of space-borne systems, with their attendant need for highly accurate and extremely stable frequency sources, has led to the development of several compact hydrogen masers, some of which operate in a passive (frequency selective amplifier) mode,[1] and others, such as those developed by this laboratory, which operate in an active (sustained oscillation) mode. Both approaches employ compact cavities, and both are considerably smaller and lighter than conventional, full size masers. The technology has matured considerably over the last several years, so much so that one of these systems, a slightly modified version of an active system developed earlier by this laboratory, is now being redesigned and qualified for use in long-term satellite missions.

The design to be described here represents the next logical step in further reducing the size of this particular type of active compact maser. The primary size determining element, the microwave cavity, was first reduced in size using well-understood loading techniques. A theoretical analysis was then performed to determine the optimal values of several adjustable parameters, and to predict the ultimate stability of the device. The results of these analyses, together with some preliminary drawings and schematics of the planned implementation, are presented in this paper.

* This work has been supported by the U.S. Naval Research Laboratory under contract N00014-83-C-2023.

Fabrication of this maser is in progress, so that measured stability data is not available at this time. Nevertheless, it was felt that the concept and future availability of a supersmall maser was of sufficient interest to members of the frequency standard community, especially those involved with system planning, to warrant a brief presentation of the design now, with a more complete report, including performance data, to be given at a future meeting.

Microwave Cavity

The microwave cavity[2] is a 3-in. diameter by 3-in. long copper cylinder, the resonant frequency of which is lowered by means of either two or four "electrodes" which are attached to the storage bottle. The low-frequency equivalent circuit of these electrodes consists of several capacitors formed by the gaps betweeen the electrodes in series with a one-turn inductor.[3] The loading electrodes used on the earlier versions developed by this laboratory were formed from copper sheet stock and incorporated radial fins to provide a lumped capacitance at the gaps (Figure 1). It was found experimentally that these fins were not necessary, that the required capacitance could be obtained with a structure having no fins and smaller gaps, and that one could consequently use thin foils, metallic paints and electroplate in place of the thicker sheet stock used heretofore. The only constraint on the thicknesses of these electrodes is that they be greater than several skin depths (δ_{Cu} = 1.7 μm). The current design will use silver plated electrodes.

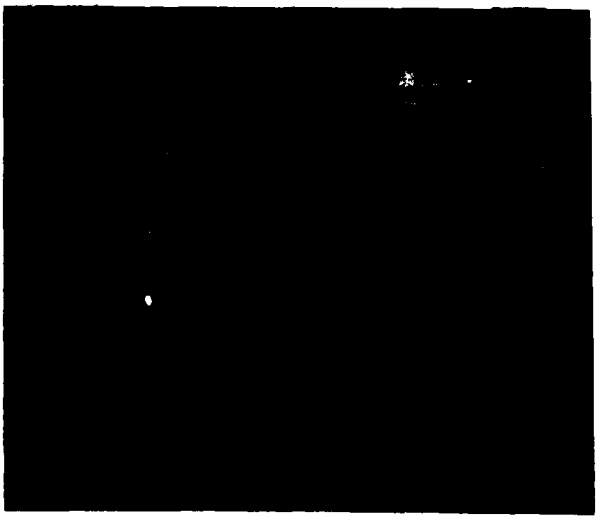


Figure 1. Prototypical compact cavity showing older "finned" electrode construction.

CH2186-0/85/0000-0080$1.00©1985IEEE

The field patterns of the cavity microwave fields have been plotted using a perturbational mapping technique. The patterns for the three L-band modes are shown in Figure 2. One sees that the electric field for the maser mode (topmost figure) is concentrated in the vicinity of the gaps and that the magnetic field is distributed uniformly throughout the bulb volume, in agreement with the simple L-C model. The outer cylinder is no longer the primary frequency determining element and functions mainly as an RF shield. The two other modes shown in Figure 2 are due to the coaxial nature of the electrode-outer cylinder arrangement, and can be troublesome when Q-enhancement is used. It has been found, however, that proper coupling can effectively suppress these modes.

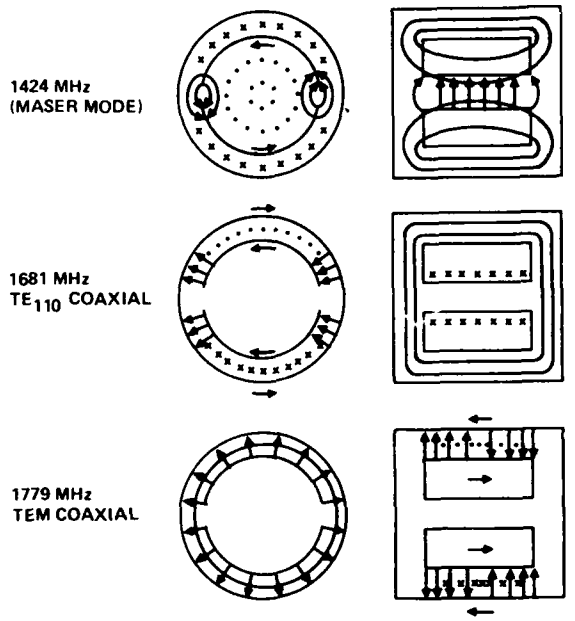


Figure 2. Mode structure for the 3-in. compact cavity. The arrowed lines denote the electric field.

The unit, shown in Figure 3, will have an overall length of 11.5 in. This version, while suitable for laboratory use, would probably not survive the stresses of a space launch. A ruggedised version which employs a two-point mount has been designed and is presently being fabricated. The modifications involved in changing to this type of mount increase the overall physics package length to 14 in. Included in this length is a space for mounting the critical front-end electronics, a modification which places the electronics and RF cables within a thermally-controlled

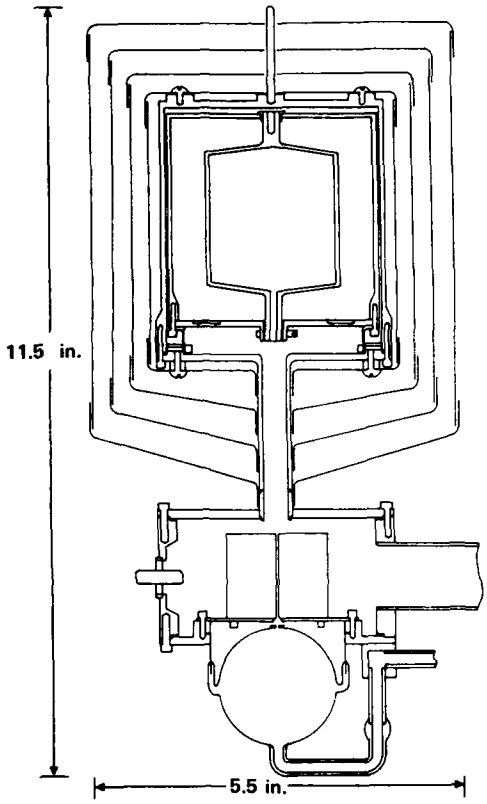


Figure 3. Layout of physics package.

Mechanical Design

The mechanical design is quite similar to that of earlier models,[4] with some interesting exceptions. The cavity is mounted to the vacuum vessel with a leaf spring arrangement which allows the cavity and vacuum chamber to "breathe" radially while preventing any lateral movement; the net effect is a decoupling of the microwave cavity from environmentally induced dimensional changes of the vacuum vessel. An indium gasket is used to make the vacuum seal between the end plate and cylindrical section of the titanium vacuum vessel, and a stainless-to-titanium transition is used to attach the vacuum can throat to the stainless steel plenum chamber. RF cabling is thermally shunted to the bottom of the vacuum can by means of integral clamps. The cables are then channeled down the beam tube and brought out through the walls of the plenum chamber, thus avoiding the need for non-magnetic connectors. The dissociator and quadrupole state selector are of the same type used on earlier models.

environment, and which should lead to significant improvements in maser stability. Both designs use four nested magnetic shields. The longitudinal shielding efficiency has been calculated using the algorithm developed by Gubser, Wolf and Cox,[5] and was found to be 300,000 for 14 mil shield material.

Electronic Design

Q-enhancement allows the threshold condition for oscillation to be met for microwave cavities having relatively low coupled Qs. Generally speaking, enhancement is needed when the natural cavity Q is lower than 25,000. For the small cavity described here, with a Q of 4600, enhancement is essential. To compensate for the fluctuations in cavity frequency which can result from phase variations in the Q-enhancement feedback loop, some form of cavity stabilisation is mandatory. Stabilisation corrects not only electronically-induced shifts, but also variations

due to thermal expansion, aging, and anything else that shifts the natural frequency of the microwave cavity. It is cavity stabilization that makes the use of all-metal cavities in a maser feasible.

The Q-enhancement and cavity-control system used, shown schematically in Figure 4, has been described in detail elsewhere.[4] The reduction in module size achieved in this particular design has been made possible by the availability of smaller and smaller components. The first module, for example, consists of a microwave amplifier, a 10-dB coupler, an image-reject mixer and an IF amplifier, all packaged in a box 4.5 in. by 1.75 in. by 0.75 in. The unit, which draws 145 mA at 15 V, is comparable both in size and power dissipation to previous versions of the amplifier alone. The module's microwave amplifier has a noise figure of 1.1 dB, a value also considerably better than the 2.5 to 3.0 dB of earlier versions. Smaller multipliers, too, are now available. The 5-MHz to 1.4-GHz multiplier is packaged in two 2.2-in. by 1.3-in. by 0.5-in. boxes, and dissipates 2 W, a figure substantially lower than the 10 W of older models. The bulky mechanical attenuator and phase shifter used in the feedback loop have been replaced by 1-in. by 1-in. by 0.15-in. analog controlled devices. The attenuator has less than a 3° phase variation over its operating range, and the phase shifter has less than ±0.1 dB variation in attenuation, so that independent level and phase adjustments can be made within the Q-enhancement loop, with each having a negligible effect on the other.

Significant reductions in the size and power consumption of the cavity control and receiver synthesizers have been achieved by using commercially available integrated circuit number controlled oscillators (NCOs). An operating version of the 5751-Hz receiver synthesizer which uses one of these NCOs has a 15 μHz resolution per least significant bit, and a power consumption of 1.3 W (as compared with 12 W for older versions). Other modifications include a more compact single transistor driver for the dissociator, and a 3-in. by 4-in. by 1-in. high voltage module for the 2 liter/second ion pump. The smaller size of the physics unit should also simplify the thermal control system by reducing the number and required capacity of the various heaters.

Performance Predictions

The phase perturbing effect of the enhanced thermal noise in the microwave cavity limits the ultimate stability that can be attained by a cavity-stabilized, Q-enhanced oscillating maser. The theoretical stability of such a maser has been derived by Lesage and Audoin,[5] and is given by

$$\sigma^2 = \frac{kT}{2Q_a^2 P\tau}\,\frac{Q_e}{Q_o}\left[(1+\beta_1+\beta_2) + \frac{(N-1)}{4\beta_2}(1+\beta_1+\beta_2-\frac{Q_o}{Q_e})^2\right]$$

where σ is the two-sample variance, Q_e and Q_o are the enhanced and intrinsic cavity Qs, Q_a is the atomic Q, β_1 and β_2 are the cavity coupling parameters, N is the noise figure of the first amplifier, P is the power delivered to the cavity by the atoms, T is the cavity and amplifier temperature, and τ is the measurement interval. The relations between power, atomic Q, beam flux and Q_e are given in the seminal paper by Kleppner et al.[7] Using those relations, one can express σ in terms of atomic flux, β_1, β_2, Q_e, and the escape time

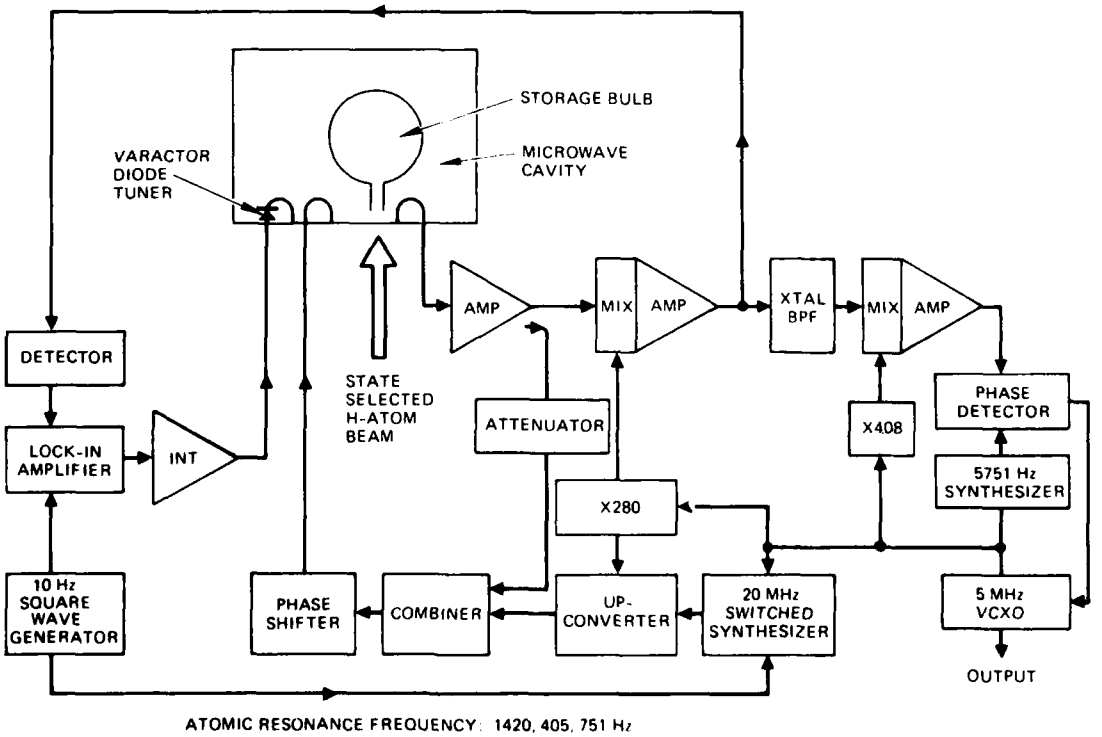


Figure 4. Block diagram of oscillating compact hydrogen maser.

from the bulb, T_b. Those values which will minimize σ are then found using a combination of analytic and numerical techniques.

The results of such a minimization procedure are shown in Figure 5. For a particular flux (I) and for the (current) design parameters given in Table I, one obtains an optimum value for σ. The optimum values of Q_e, β (β_1=β_2) and T_b are also functions of I, although only the last varies appreciably. Note that, in theory at least, one can always improve maser stability by going to higher flux levels, providing that the bottle escape time is simultaneously reduced. The resulting increase in signal is always greater than the increase in noise caused by the decreased lifetime, although an asymptote is eventually reached where the gain of one is just offset by the loss of the other.

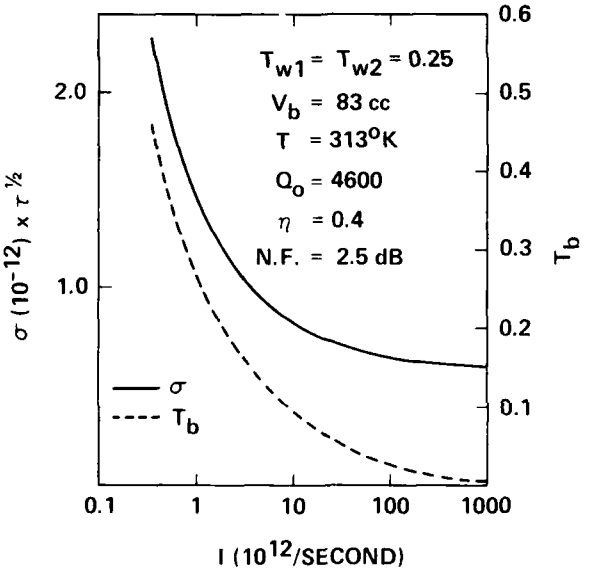


Figure 5. Minimum value of σ achievable for a given atomic flux.

Two sets of values have been listed in Table I, one labeled "current", the other "achievable". Current values correspond to existing RL systems or prototypes, while achievable are those that can be achieved by improvements in design, procurement, or technique. Of particular interest are the differences between the current and achievable wall collision lifetimes (T_{1w}, T_{2w}). The achievable values have been obtained by extrapolating the lifetimes measured by Berg[8] for larger bottles down to the 2-in. size of the compact maser and the current by extrapolating between measured values of in-house systems. The implication is that either the teflon emulsion used, FEP-120, is considerably different from that used by Berg, or that the teflon-coating procedures used at RL could stand some improvement. However, even with these less than optimal coatings, one can obtain theoretically competitive stabilities.

Table II lists the optimum parameter values for an atomic flux of 10^{12}/s. Table III is a comparison of theoretically predicted and measured values for earlier compact masers, CHYMNS I and II, with the predictions for the new design (CHYMNS III). Sizewise, CHYMNS I has a 6-in. dia. by 6-in. cavity, and a 4-in. dia. by 4.5-in. storage bulb, while CHYMNS II has a 4.1-in. dia. by 6-in. cavity and a 2.2-in. dia. by 5-in. storage bulb. The comparison in Table III is somewhat unfair, in that the operating flux for CHYMNS II is less than that used for the other two: for equal fluxes, the stability of CHYMNS II should be a factor of two better than CHYMNS III. It can be seen, incidentally, that the experimental values approach the theoretical, which implies that the major source of short term instability is indeed the enhanced thermal noise in the cavity. More significant, however, is the fact that the theoretical penalty paid in going to this very compact design will not be severe.

Table I. Design Parameters

	CURRENT	ACHIEVABLE
OVERALL LENGTH	11.5 in.	
OUTER SHIELD DIA.	5.5 in.	
CAVITY I.D.	3.0 in.	
CAVITY I. LENGTH	3.0 in.	
BULB VOLUME	83 cc	100 cc
Q_o	4600	
η	0.40	0.50
T_{2w}	0.25*	1.0
T_{1w}	0.25*	0.5
T_b	0.27	
AMPLIFIER N.F.	2.5 dB	1.5 dB

*PREDICTED

Table II. Optimum Values of σ,β, Q_e and T_b for an Atomic Flux of 10^{12} per Second

PARAMETERS	σ (1 s)	I (10^{12}/s)	β	Q_e	I_{min} (10^{12}/s)	T_b
CURRENT	14.65×10^{-13}	1.0	0.24	58,000	0.278	0.27
ACHIEVABLE	5.87×10^{-13}	1.0	0.17	29,000	0.217	0.30

Table III. Theoretical and Measured Values of σ

	CHYMNS I	CHYMNS II	CHYMNS III
T_b	0.5	0.5	0.27
I(10^{12}/s)	1.2*	0.7*	1.0
σ THEORETICAL (1s)	3.8×10^{-13}	10.3×10^{-13}	14.7×10^{-13}
σ MEASURED (1s)	6.6×10^{-13}	13×10^{-13}	?

*ESTIMATED

Conclusions

A subcompact maser is being constructed that is approximately half the size of existing compact masers. Several improvements have been made relative to earlier designs. One, an order-of-magnitude reduction in the dissipated power of several key electronic modules, should considerably enhance the long-term reliability of the system, while others, such as the improved magnetic shielding and better thermal control, should improve the overall stability. The small size, low weight, low power consumption and excellent stability performance of this maser will provide system designers with an additional option in the choice of high stability frequency sources for modern navigation and communication systems.

References

1. F.W. Walls and K.B. Persson, "A new miniaturized passive hydrogen maser," Proc. 38th An. Freq. Control Symp., p. 416, 1984.

2. H.T.M. Wang, J.B. Lewis, and S.B. Crampton, "Compact cavity for hydrogen frequency standard," Proc. 33rd Ann. Freq. Control Symp., p. 543, 1979.

3. H.E. Peters, "Small, very small, and extremely small hydrogen masers," Proc. 32nd An. Symp. on Freq. Control, p. 469, 1978.

4. H.T.M. Wang, "Characteristics of oscillating compact hydrogen masers," Proc. 36th An. Symp. on Freq. Control, p. 249, 1982.

5. D.U. Gubser, S.A. Wolf, and J.E. Cox, "Shielding of longitudinal magnetic fields with thin, closely spaced, concentric cylinders of high permeability material," Rev. Sci. Inst. 50, 751 (1979).

6. P. Lesage and C. Audoin, "Frequency stability of an oscillating maser: analysis of the effect of an external feedback loop," IEEE Trans. IM-30, 182 (1981).

7. 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, A972 (1965).

8. H.C. Berg, "Spin exchange and surface relaxation in the atomic hydrogen maser," Phys. Rev. 137, A1621 (1965).

DOPPLER-FREE TWO-PHOTON LASER SPECTROSCOPY OF HgII*

J. C. Bergquist, D. J. Wineland, Wayne M. Itano, Hamid Hemmati**,
H.-U. Daniel†, and G. Leuchs††

Time and Frequency Division
National Bureau of Standards
Boulder, Colorado 80303

Abstract

The Doppler-free, two-photon $5d^{10}\ 6s\ ^2S_{1/2}$-$5d^9\ 6s^2\ ^2D_{5/2}$ transition in singly ionized Hg, attractive as an optical frequency standard, has been observed for the first time. A few $^{198}Hg^+$ ions were confined in a radio-frequency (rf) trap and the two-photon transition was detected by monitoring the change in the fluorescence light scattered by the ions from a laser beam tuned to the first resonance transition at 194 nm. The radiative lifetime of the $^2D_{5/2}$ state and the absolute wavenumber of the two-photon transition were measured to be 0.090(15) s and 17 757.152(3) cm^{-1} respectively.

Introduction

Microwave or optical transitions of laser cooled ions that are confined in electromagnetic traps offer the basis for frequency standards of high stability and accuracy.[1-5] The advantages of such devices are numerous: very long interrogation times and, therefore, high transition line Q's can be achieved; fractional frequency perturbations that are due to the trapping fields can be held below 10^{-15}; collisions with background gas and cell walls can be largely avoided; Doppler shifts are directly reduced by trapping and cooling and, finally, nearly unit detection efficiency of transitions to metastable states is possible so that the signal to noise ratio need be limited only by the statistical fluctuations in the number of ions that make the transition.[5] Details of ion traps and laser cooling have been published elsewhere.[1-3] A particularly attractive candidate for an optical frequency standard is the two-photon allowed (or single photon, electric quadrupole allowed) $5d^{10}\ 6s\ ^2S_{1/2}$-$5d^9\ 6s^2\ ^2D_{5/2}$ Hg^+ transition near 563 nm.[4,5] The lifetime of the $^2D_{5/2}$ state, expected to decay by emission of electric quadrupole radiation, is calculated to be of order 0.1 s.[5] This gives a potential optical line Q of about 7×10^{14}. Here we describe the first results of our investigation of the two-photon transition in $^{198}Hg^+$ (which is free of hyperfine structure) stored in a miniature rf trap.

*Work of the U.S. Government; not subject to U.S. copyright.
**Present address: Allied Bendix Aerospace Corporation, Columbia, MD
†Present address: Springer Verlag, Heidelberg, West Germany
††Heisenberg Fellow of the Deutsche Forschungsgemeinschaft.
Present address: Max Planck Institute für Quantenoptik, Garching, West Germany

Experiment

Our trap is similar to the small radio frequency traps used in the ion cooling experiments that were conducted at Heidelberg University on Ba^+,[7] and at the University of Washington on Mg^+.[8] A cross section of the trap electrodes is shown in Fig. 1. We note that, although the inner surfaces of our trap electrodes were machined with simple conical cuts, the trap dimensions were chosen to make the fourth and sixth order anharmonic contributions to the potential vanish.[9] The rf drive frequency was 21 MHz with a voltage amplitude, $V_o \leq 1$ kV. The background pressure, excluding deliberately added mercury and helium, was $\leq 10^{-7}$ Pa (133 Pa ≡ 1 torr). After loading 50-200 mercury ions, the mercury vapor was frozen out in a liquid nitrogen cold trap and the vacuum vessel was back-filled with He to the order of 10^{-3} -10^{-2} Pa. This was sufficient to collisionally cool the trapped Hg^+ to near room temperature as verified by the Doppler width of the $^2S_{1/2}$-$^2P_{1/2}$ resonance line near 194 nm.

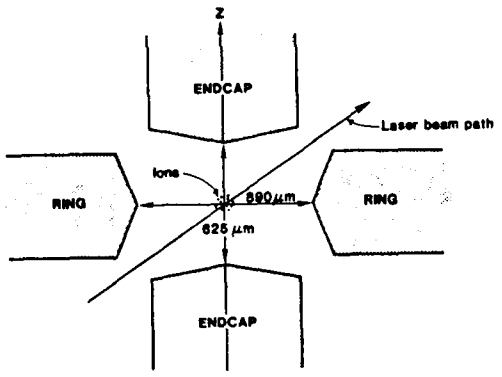


Fig. 1. Schematic showing cross section view of trap electrodes. The electrodes are figures of revolution about the z-axis and are made from molybdenum.

Optical double-resonance[10,11] was used to detect the two-photon transition. About 5 μW of narrowband cw sum-frequency-generated radiation near 194 nm[12] was tuned to the $6s\ ^2S_{1/2}$ to $6p\ ^2P_{1/2}$ first resonance transition and was directed diagonally through the trap (between the ring electrode and the end caps). The fluorescence light scattered by the ions was detected at right angles to the 194 nm beam with an overall detection efficiency of about 10^{-4}. Typically, our signal level was 2-10 x 10^3 counts/s and the signal to

 "US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

background ratio was better than 10/1. When the ions were driven by the radiation from a 563 nm cw ring dye laser out of the $^2S_{1/2}$ ground state into the $^2D_{5/2}$ metastable state, there was a decrease in the 194 nm fluorescence corresponding to the number of ions in the D state.

The dye laser beam also was directed diagonally through the trap; the axes of the dye laser beam, the 194 nm beam and the collection optics were mutually perpendicular. A near-concentric standing wave cavity was placed around the trap in order to enhance the power of the 563 nm radiation and to better ensure nearly equal intensity counter-propagating beams. The cavity was positioned so that its waist ($w_o \cong 25$-30 μm) was located near the center of the cloud of trapped ions. The power buildup factor was approximately 50, giving nearly 5 W of circulating power for typical input power levels of 100 mW. The ring dye laser linewidth in these preliminary experiments was of the order of 300 kHz. The frequency of the laser was offset locked and precisely scanned with respect to a second dye laser locked to a hyperfine component in the Doppler-free, saturated absorption spectrum of $^{127}I_2$.

Results

A typical resonance curve and simplified energy level diagram is shown in Fig. 2. The full scanwidth is 4 MHz. The electric field vector of the 563 nm laser radiation is nearly parallel to a small applied

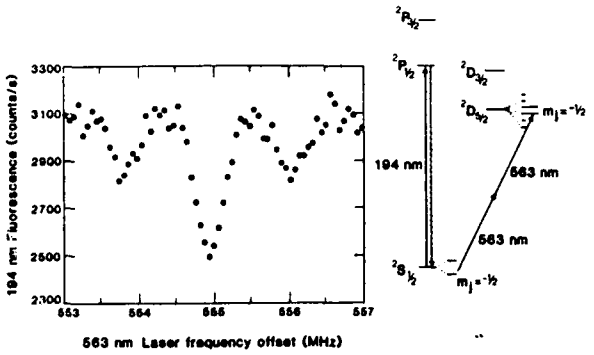


Fig. 2. Two-photon $^2S_{1/2}$-$^2D_{5/2}$ transition in $^{198}Hg^+$. AM sidebands caused by the harmonic secular motion of the ions are visible in this scan. The frequency scan is 4 MHz at the fundamental laser frequency ($\lambda \cong 563$ nm). The depth of the central component is about 25% of full scale. The integration time is 2 s/point. In the inset is a simplified energy level diagram of $^{198}Hg^+$, depicting the levels of interest.

magnetic field of approximately 11.6×10^{-4} T (11.6 G) which differentially Zeeman splits the ground and excited states. The selection rule for the two-photon transition for this polarization is $\Delta m_J = 0$, and, thus, only two components are observed, separated by approximately 13 MHz (approximately 6.5 MHz at the dye laser frequency). In Fig. 2, we scan over only one of these Zeeman components ($m_J = -1/2 \leftrightarrow m_J = -1/2$) but see sideband structure. This structure is due to amplitude modulation (AM) of the 563 nm laser intensity due to the secular motion of the ions in the rf trap.[13] To our knowledge, this is the first observation of secular motion sidebands at optical frequencies. The depth of the central feature in Fig. 2 is nearly 25% of full scale implying that we have nearly saturated the two-photon transition. The linewidth is about 420 kHz, and is determined in nearly equal parts by the laser linewidth of about 320 kHz and the power broadening by the 194 nm radiation of nearly 270 kHz. When the 194 nm radiation is chopped, the two-photon linewidth drops to approximately 320 kHz.

We have experimentally measured the radiative lifetime of the $5d^9\ 6s^2\ ^2D_{5/2}$ state to be 0.090(15) s in good agreement with the calculated lifetime of 0.105 s.[6] Again in this measurement, the ground state population was monitored by measuring the laser induced fluorescence of the 194 nm transition. The radiation from the dye laser near 563 nm was tuned to resonance with the two-photon transition and chopped on and off. During the time that the laser radiation was on, it drove 10-20% of the ion population into the D state. The time constant for the atomic system to relax during the radiation-off period could be determined from the exponential return of the 194 nm fluorescence to steady state. An example is shown in Fig. 3. The relaxation rate was measured over a range of He pressures differing by a factor of four. The reported radiative lifetime is the result of an extrapolation to zero pressure of a linear least-squares fit to the data. The pressure-induced decay rate was determined poorly, but amounted to only about 25% of the radiative decay rate at the highest pressure (about 6×10^{-2} Pa).

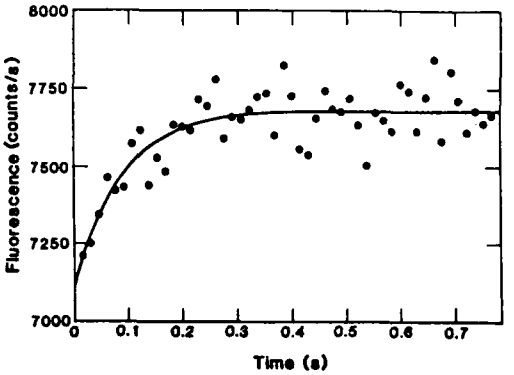


Fig. 3. Measurement of the radiative lifetime of the $5d^9\ 6s^2\ ^2D_{5/2}$ state in Hg^+. The dots are the experimental points indicating the return of the atomic system to steady state after chopping off the radiation from the 563 nm dye laser that drives the ions from the ground state to the metastable D state by two-photon absorption. The solid line is a least squares fit of an exponential to the experimental data. The vertical axis is the counts per second obtained by measuring the laser induced fluorescence from the continuously driven $^2P_{1/2}$-$^2S_{1/2}$ transition. The horizontal axis shows time after turning off the 563 nm laser radiation.

We have also measured the absolute wavenumber of the $^2S_{1/2}$-$^2D_{5/2}$ two-photon transition by measuring the frequency difference between the two photon resonance and the "t" hyperfine component of the nearby R(33) line of the 21-1 band in $^{127}I_2$ (line #1220 in the iodine atlas).[14,15] The two-photon transition in $^{198}Hg^+$ lies 551(2) MHz to the red of this component. From this we determine the wavenumber of the two-photon transition to be 17 757.152(3) cm^{-1}.

In the near future, we anticipate narrowing the 563 nm laser linewidth to the order of a few kHz and studying various systematic effects including pressure broadening and shifts, power broadening, and light shifts. Ultimately, we would like to narrow the laser linewidth to a value near that imposed by the natural lifetime of the D state, and to drive the two-photon (or single photon, electric quadrupole) transition on a single, laser-cooled ion.

Acknowledgments

We gratefully acknowledge the support of the US Air Force Office of Scientific Research and the US Office of Naval Research. We wish to thank R. Blatt (University of Hamburg) for technical assistance in the construction of the rf trap and H. Layer (NBS) for supplying the ^{199}Hg. We take particular pleasure in acknowledging the help of W. Martin, J. Reader and C. Sansonetti (NBS) who provided the correct wavelength for the $^2D_{5/2}$-$^2S_{1/2}$ transition. We also wish to thank D. Huestis (SRI) and our colleague R. Drullinger for many useful conversations and insights.

References

[1] H. G. Dehmelt, Adv. At. Mol. Phys. 3, 53 (1967); H. G. Dehmelt, Adv. At. Mol. Phys. 5, 109 (1969); H. G. Dehmelt, IEEE Trans. Instrum. Meas. IM-31, 83 (1982).

[2] D. J. Wineland, W. M. Itano, and R. S. Van Dyck, Jr., Adv. At. Mol. Phys. 19, 135 (1983).

[3] D. J. Wineland, W. M. Itano, J. C. Bergquist, J. J. Bollinger, and J. D. Prestage, in "Atomic Physics 9", R. S. Van Dyck, Jr. and E. N. Fortson, eds. (World Scientific Publ. Co., Singapore, 1984) p. 3.

[4] P. L. Bender, J. L. Hall, R. H. Garstang, F. M. Pichanick, W. W. Smith, R. L. Barger, and J. B. West, Bull. Am. Phys. Soc. 21, 599 (1976).

[5] D. J. Wineland, J. C. Bergquist, R. E. Drullinger, H. Hemmati, W. M. Itano, and F. L. Walls, J. Phys. (Orsay, Fr.) 42 C8-307 (1981); D. J. Wineland, W. M. Itano, J. C. Bergquist, and F. L. Walls, Proc. of 35th Annu. Symp. on Freq. Control, (1981) p. 602 (copies available from Electronic Industries Assoc., 2001 Eye St., NY, Washington, DC 20006).

[6] R. H. Garstang, J. Research NBS 68A, 61 (1964).

[7] W. Neuhauser, M. Hohenstatt, P. Toschek, and H. G. Dehmelt, Phys. Rev. Lett. 41, 233 (1978).

[8] W. Nagourney, G. Janik, and H. G. Dehmelt, Proc. Natl. Acad. Sci. USA 80, 643 (1983).

[9] E. C. Beaty, to be published.

[10] D. J. Wineland, J. C. Bergquist, W. M. Itano, and R. E. Drullinger, Opt. Lett. 5, 245 (1980).

[11] H. G. Dehmelt, in "Advances in Laser Spectroscopy", F. T. Arecchi, F. Strumia, and H. Walther, eds., (Plenum, New York, 1983) p. 153.

[12] H. Hemmati, J. C. Bergquist, and W. M. Itano, Opt. Lett. 8, 73 (1983).

[13] H. A. Schuessler, Appl. Phys. Lett. 18, 117 (1971); F. G. Major and J. L. Duchene, J. Phys. (Orsay) 36, 953 (1975); H. S. Lakkaraju and H. A. Schuessler, J. Appl. Phys. 53, 3967 (1982); M. Jardino, F. Plumelle, and M. Desaintfuscien, in "Laser Spectroscopy VI", H. P. Weber and W. Lüthy, eds. (Springer-Verlag, NY, 1983) p. 173; L. S. Cutler, R. P. Giffard and M. D. McGuire, Appl. Phys. B36, 137 (1985).

[14] S. Gerstenkorn and P. Luc, Atlas du Spectre d'absorption de la Molecule de L'iode Entre 14,800-20,000 cm^{-1} (Editions du C.N.R.S., 15, quai Anatole-France, 75700 Paris, 1978).

[15] S. Gerstenkorn and P. Luc, Rev. Phys. Appl. 14, 791 (1979).

39th Annual Frequency Control Symposium - 1985

PERFORMANCE OF A LASER-INDUCED RESONANCE RAMAN CLOCK

P.R. Hemmer*, G.P. Ontai**, A. Rosenberg and S. Ezekiel

Research Laboratory of Electronics

Massachusetts Institute of Technology

Cambridge, MA 02139

Abstract

We have stabilized a microwave oscillator using a laser induced resonance Raman transition in a sodium atomic beam and have achieved a fractional frequency (clock) stability of 1.5×10^{-11} for a 1000 sec averaging time. This stability, which is near the shot noise limit for the present setup, compares favorably with the fractional frequency stabilities of commercial cesium clocks at 1000 sec averaging times when differences in transit time and transition frequency are taken into consideration. These preliminary results are very encouraging and were made possible because the major sources of frequency error have been identified and greatly reduced. Remaining sources of long term frequency shifts are still under investigation.

I. Introduction and Background

We have been investigating the performance of a clock based on a laser induced resonance Raman transition in an atomic beam to determine the feasibility of such a scheme and to demonstrate any possible advantages over conventional microwave excited clocks. Although we have been conducting our experiments so far using a sodium atomic beam and dye lasers, this Raman technique is also applicable to a cesium atomic beam employing semiconductor laser excitation and may lead to the development of smaller, lighter and less expensive portable cesium beam clocks.

The stimulated resonance Raman interaction is illustrated schematically in Fig. 1 (a). Briefly, Raman transitions are induced between states 1 and 3 using two laser fields, at frequencies ω_1 and ω_2, simultaneously resonant with the intermediate state 2. Earlier studies show that, for copropagating laser fields interacting with an atomic beam at right angles, the Raman linewidth is determined by the widths of states 1 and 3 only. [1,2] State 2 greatly enhances the transition probability but does not contribute to the linewidth. Thus, for long lived states 1 and 3, the Raman linewidth becomes transit time limited, just as for a direct microwave excitation.

To obtain a very small transit time linewidth we use Ramsey's method of separated oscillatory fields, as illustrated in Fig. 1 (b), in analogy with conventional microwave techniques. [3] In separated field excitation, the atom-field superposition states, excited in zones A and B by the two-photon Raman interaction, interfere quantum mechanically. [3] The resulting interference fringes have frequency spacings which are characteristic of the transit time between interaction zones.

* Rome Air Development Center, Hanscom AFB, MA 01731

** Captain, U.S. Army

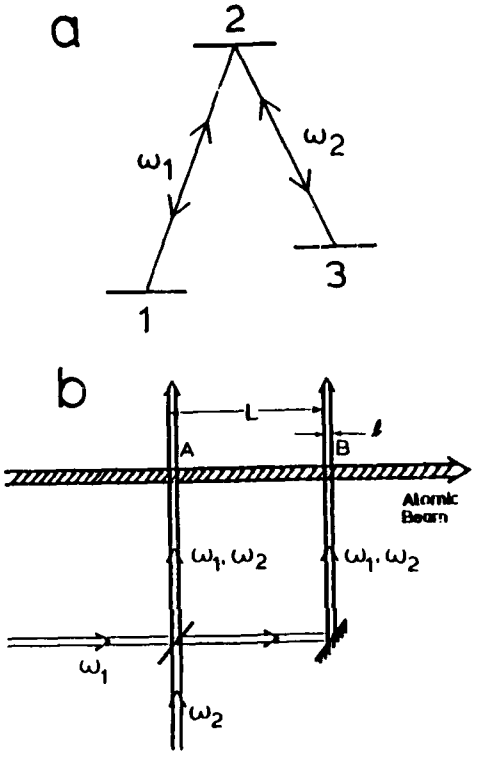


Fig. 1

(a) Schematic of laser induced resonance Raman interaction.
(b) Schematic of separated field excitation.

II. Experimental Setup

The experimental setup used to demonstrate clock applications is schematically illustrated in Fig. 2. [4] The laser at frequency ω_1 is obtained from a single mode dye laser locked to the sodium D_1 transition, at 590 nm., using fluorescence from the atomic beam, as shown in Fig. 2. The laser field at frequency ω_2 is generated directly from that at ω_1 by an acousto optic frequency shifter (A/O), driven with a quartz stabilized microwave oscillator near the 1772 MHz sodium hyperfine transition frequency. This greatly reduces the effects of laser jitter by correlating the frequency jitters of ω_1 and ω_2 so as to produce a highly stable difference frequency. [2] After leaving the A/O, the laser beams at ω_1 and ω_2 are combined in a single mode optical fiber, as shown, before exciting the atomic beam at the two Ramsey separated interaction zones, labeled A and B in Fig. 2.

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

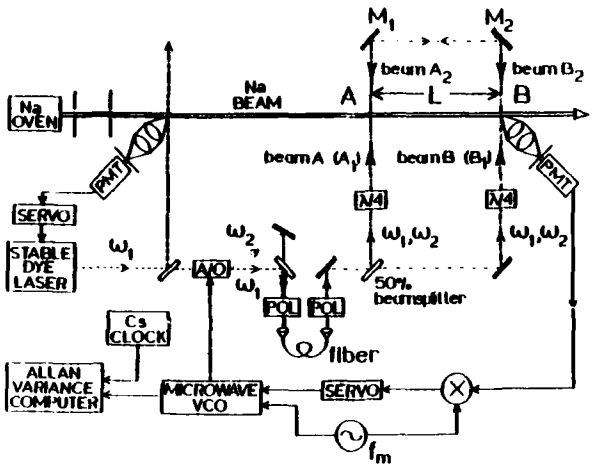


Fig. 2

Schematic of experimental Raman clock setup.

Typical Raman/Ramsey fringe lineshapes obtained using a 15 cm interaction zone separation appear in Fig. 3 (a). These fringes are observed by monitoring the fluorescence induced in interaction zone B while scanning the microwave oscillator frequency with ω_1 locked on resonance with the D_1 transition. The central fringe in Fig. 3 (a) has a width of about 2.6 kHz (FWHM) which is consistent with transit time.

To stabilize the frequency of a microwave oscillator to the central fringe in Fig. 3 (a), a discriminant is needed. This discriminant, which is shown in Fig. 3 (b), is obtained by frequency modulating the microwave source at a rate f_m = 610 Hz and demodulating the zone B fluorescence signal with a lock-in amplifier. The output of the lock-in amplifier is then used in a feedback loop to hold the microwave oscillator frequency at the central zero of the discriminant. The stability of this oscillator is measured by comparing it with a commercial cesium (or rubidium clock).

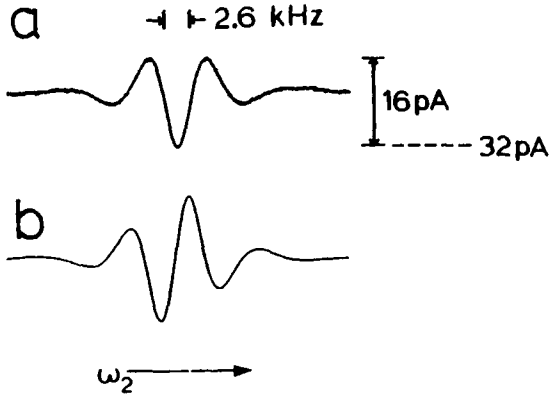


Fig. 3

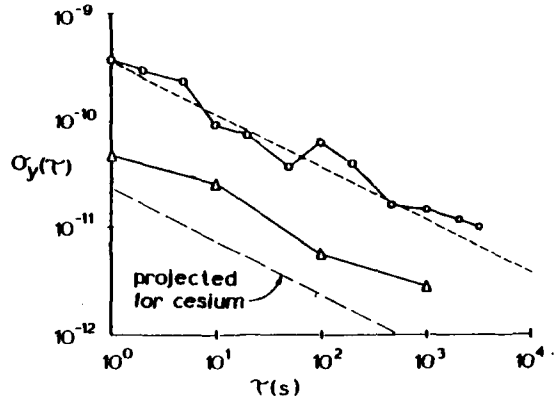
(a) Typical Raman/Ramsey fringe lineshape for a 15 cm interaction zone separation. Photomultiplier photocathode current levels as shown.

(b) Discriminant of (a) obtained using frequency modulation.

III. Clock Performance

Figure 4 shows a plot of the measured fractional frequency stability of the stabilized microwave oscillator, $\sigma_y(\tau)$, as a function of averaging time, τ. For $\tau \approx 1000$ sec, the stability is about 1.5×10^{-11}. The dashed line superimposed on the data in Fig. 4 has the slope predicted for shot noise limited stability. The data in this plot are very close to the predicted shot noise limit.

In order to compare the preliminary stability data of Fig. 4 with commercially available cesium clocks it is necessary to include a factor of about 16 to account for the higher transition frequency and longer typical transit time of cesium. The lower dashed line in the figure is the projected stability expected if cesium were used in place of sodium in our present experimental setup with everything else being equal. It is just the upper dashed line divided by 16. The triangles are the specifications of a portable H-P cesium clock ($L \approx$ 7 cm) and are included for comparison. [5] As can be seen, the projected preliminary sodium results compare favorably with commercial cesium clocks out to averaging times of about 1000 sec.



Δ Portable HP cesium clock

Fig. 4

Plot of fractional frequency deviation $\sigma_y(\tau)$ vs. averaging time τ for stabilized microwave oscillator.

IV. Frequency Error Sources

For $\tau > 1000$ sec we observe departure from shot noise limited stability. At present, potential sources of frequency error that can cause long term frequency drifts in the Raman process are under study, both experimentally and theoretically.

One source of frequency error in the Raman scheme is caused by changes in the relative optical paths to each of the two Ramsey interaction zones. These optical path effects are analogous to cavity phase shifts in conventional microwave excitation and, just as for microwave, are greatly reduced in the Raman setup by using standing wave excitation. In addition, phase shifts can also arise in the Raman process if the polarizations of the laser fields at ω_1 and ω_2 are not identical. However, such phase shifts are greatly reduced in the present setup by directing the laser beams at both frequencies through a common linear polarizer before exciting the atomic beam.

Frequency errors can also be introduced due to the Doppler effect if the ω_1 and ω_2 lasers are not exactly copropagating. To ensure that both laser beams are copropagating they are coupled into the same single mode optical fiber. Additional frequency errors can arise when both laser fields are off-resonance (by the same amount) with their respective components of the D_1 transition. To reduce these errors, optical pumping techniques have proved useful.

Other important error sources include external magnetic field effects, atomic beam misalignments, laser beam misalignments (both frequencies together), fluorescence background slope effects, fluctuations in the relative intensities of the two laser fields, and frequency pulling caused by the presence of other atomic levels near the intermediate state 2.

In order to achieve the stability results shown in Fig. 4 it was necessary to study all of the error sources listed above and to develop the experimental techniques needed to control the important ones. Clearly, more work is needed and efforts are currently in progress to identify and control sources of long term frequency drift for averaging times longer than 1000 sec.

V. Future Applications

As mentioned earlier, one of the more promising applications of the Raman technique is to develop a cesium atomic beam Raman clock, using semiconductor lasers. Fig. 5 illustrates the basic differences between a Raman clock and a conventional microwave clock. As shown in Fig. 5 (a), a conventional cesium clock consists of a cesium atomic beam, a microwave cavity, state selection magnets and a hot wire detector. These components require precise alignment and greatly increase the complexity of the atomic beam apparatus. In contrast, a Raman cesium clock, Fig. 5 (b), consists of a very simple atomic beam, excited by semiconductor laser light and employing optical detection. The bulky and heavy microwave cavity and state selection magnets are no longer required and all alignments can be made using optics external to the vacuum system. [6]

We are also considering the attractive possibility of extending the resonance Raman technique into the mm-wave region of the spectrum. At these much higher transition frequencies it may be possible to achieve better clock stabilities and many of the experimental problems, associated with exciting mm-wave transitions in an atomic beam, could be avoided. Finally, the Raman technique can also be readily applied to slowed or trapped atoms, possibly without greatly increasing the complexity of the experimental setup, since many of these techniques already make extensive use of resonant laser light.

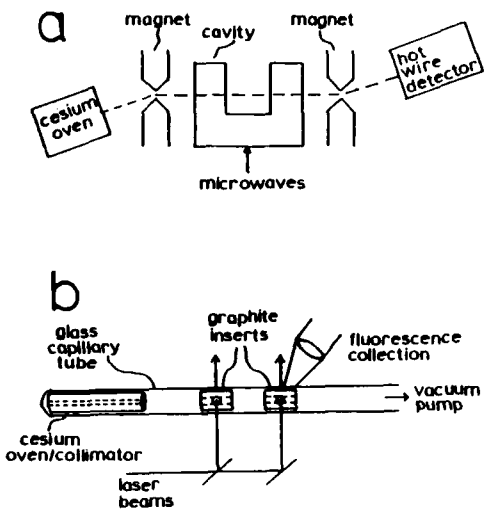


Fig. 5

(a) Schematic of conventional cesium clock.
(b) Schematic of planned Raman cesium clock.

Acknowledgements

This work was supported by Rome Air Development Center, the Joint Services Electronics Program, and the National Science Foundation.

References

[1] J.E. Thomas, S. Ezekiel, C.C. Leiby Jr., R.H. Picard and C.R. Willis, "Ultrahigh-resolution spectroscopy and frequency standards in the microwave and farfrared regions using optical lasers," Opt. Lett. 6, 298 (1981) and references therein.

[2] J.E. Thomas, P.R. Hemmer, S. Ezekiel, C.C. Leiby Jr., R.H. Picard and C.R. Willis, "Observation of Ramsey fringes using a stimulated resonance Raman transition in a sodium atomic beam," Phys. Rev. Lett. 48, 867 (1982).

[3] N.F. Ramsey, Molecular Beams (Oxford U. Press, London, 1963).

[4] P.R. Hemmer, S. Ezekiel and C.C. Leiby Jr., "Stabilization of a microwave oscillator using a resonance Raman transition in a sodium beam," Opt. Lett. 8, 440 (1983).

[5] Hewlett-Packard portable cesium beam clock.

[6] Optically pumped cesium clocks, which are currently under development, also employ semiconductor laser light to replace the state selection magnets and hot wire detector. However, optically pumped cesium clocks still require the microwave cavity.

ERRORS IN SERVO SYSTEMS USING SINUSOIDAL FREQUENCY (PHASE) MODULATION*

Fred L. Walls
Time and Frequency Division
National Bureau of Standards
Boulder, Colorado 80303

Abstract

This paper reviews the errors in determining the center of a resonance line which are due to residual imperfections in practical electronic systems using sinusoidial frequency or phase modulation. In particular the effects of residual amplitude modulation, baseline distortion, and harmonic distortion in the modulation process and the demodulator are qualitatively analyzed for a Lorentzian line in the limit of small modulation index. This permits one to easily calculate analytically the frequency offsets as a function of modulation index and the transfer function of the fundamental and various harmonics of the modulation frequency. Using this model one can easily formulate accurate tests for experimentally measuring the frequency errors in practical servo systems, even if the original assumptions about small modulation index and a pure Lorentzian line are not exactly fulfilled.

Introduction

Many systems use sinusoidal frequency or phase modulation of a probe frequency in order to find the center of a resonance line. The purpose of this paper is to review the residual imperfections which occur in practical systems and the subsequent errors in determining line center. In particular the effects of residual amplitude modulation, baseline distortion, and harmonic distortion in the modulation and the demodulation process are qualitatively analyzed for a Lorentzian line in the limit of small modulation index, this permits one to easily calculate analytically the frequency offsets as a function of modulation index and the transfer function of the fundamental and various harmonics of the modulation frequency. Based on this model one can then compare the relative susceptability of various servo configurations to residual electronic imperfections. Additionally one can easily formulate accurate tests for experimentally measuring the frequency errors in practical servo systems, even if the original assumptions about small modulation index and a pure Lorentzian line are not exactly fulfilled.

Model of a Resonance Line and Error Signal

One of the most common, methods for determining the center of a resonance line with high precision is to sinusoidally modulate the frequency (or phase) of the probe and detect the phase of the resulting amplitude modulated signal at the fundamental of the modulating signal. The general scheme is shown in Figure 1.

The various subsystems and their effect on errors in determining the center of the resonance will be analyzed in later sections.

*Contribution of the National Bureau of Standards, not subject to copyright.

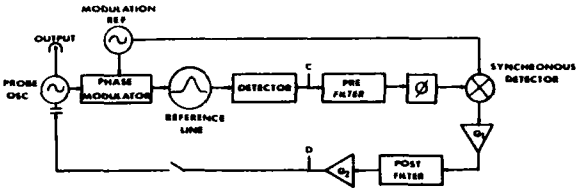


Fig. 1. Block diagram of a sinusoidally modulated probe oscillator which can be locked to the center of a reference line.

Curve a of Figure 2 shows a typical resonance line that would be observed at point C of Figure 1 as a function of slowly sweeping the frequency of the probe (without modulation) across the resonance. Curve b of Figure 2 is the derivative of curve a.

For the moment, let's assume that the probe output is a sinewave with a spectral width very narrow compared to the width of the resonance shown in Figure 2, curve a. If the center of the probe is at the point A, then the output signal increases as the frequency of the VCO is increased, and at point B, the signal decreases as the probe frequency increases. If the frequency of the probe is swept back and forth (FM), then the signal has both a dc and an ac component. If the deviation of the FM is small compared to the half linewidth W, then the demodulated and filtered output of the synchronous detector (measured at Point D of Figure 1) fairly accurately reproduces the derivative of curve a. Note that in curve b, the point of zero signal, which also has the steepest slope, nominally occurs at the center of the resonance line. This curve is referred to as a frequency discriminator curve. The signal at point D can be used to steer the probe frequency because near line center we now have a dc signal proportional to the frequency error between the probe frequency and the center of the resonance.

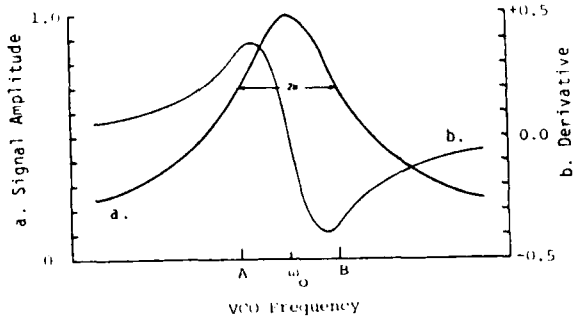


Fig. 2. Reference line (a) and its first derivative (b).

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

Now let's examine this process in a little more detail. More generally, assume that we have a symmetric Lorentzian line superimposed on a sloping and curved background. Then:

$$\text{Signal Amp} = \left(\frac{\gamma^2}{\gamma^2+(\omega-\omega_o)^2}\right)^{1/2} + K_1(\omega-\omega_o) + K_2(\omega-\omega_o)^2 \qquad (1)$$

where K_1 and K_2 are the first two coefficients of a Taylor expansion of the background about line center, $\gamma = \pi W$ is the half angular linewidth, ω is the instantaneous angular frequency of the probe, and ω_o is the true center of the resonance.

Real frequency or phase modulators have small non-linearities and therefore generate small components of modulation at multiples of the modulation frequency. Also the modulation reference generally has some higher harmonic components as well. Therefore let's assume that the modulated signal is of the form

$$\omega = \omega_1 + B\cos \Omega t - K_3 \sin 2\Omega t + K_5 \cos 2\Omega t + K_5 \sin 3\Omega t \qquad (2)$$

where Ω is the modulation frequency.

The effects of distortion in the reference and the modulation process are contained in coefficients K_3, K_5 and K_6. This model assumes that the residual modulation at $\Omega/2$, Ω, 2Ω, etc., due to spurious signals on the probe control line is small compared to that imposed by the modulator. This places a heavy burden on the postfilter (see Figure 1) especially in servos using a square wave reference for the demodulation. The modulation process can and usually does cause some amplitude modulation, therefore another term, $K_4 \cos\Omega t$, needs to be added to equation 1. Substituting for ω in equation 1 and adding the K_4 term yields equation 3. It has been assumed that the amplitude modulation is in phase with the frequency modulation, which yields the maximum offset.

$$\text{Signal Amp} = \left[1 + \frac{1}{\gamma^2}\left[(\omega_1-\omega_o) + B\cos\Omega t - K_3\sin 2\Omega t + K_5\cos 2\Omega rt + K_6\sin 3\Omega t\right]\right]^{-1/2} + K_1[\omega_1 - \omega_2) + B\cos\Omega t] + K_2[\omega_1 - \omega_o) + B\cos\Omega t]^2 + K_4\cos\Omega t \qquad (3)$$

Near line center, with the coefficients K_3, K_4, K_5, K_6 very small compared to γ, and the modulation amplitude B only slightly smaller than γ, the denominator can be expanded using the approximation

$$\frac{1}{1+\delta} = 1 - \frac{\delta}{2} + \frac{3}{8}\delta^2 - \frac{15}{48}\delta^3 \ldots \text{for } \delta<1 \qquad (4)$$

Signal Amplitude =

$$-\frac{B^2}{4\gamma^2}\cos 2\Omega t\,(1+\delta') \qquad (5a)$$

$$-\frac{B}{\gamma^2}\,\Delta\omega\cos\Omega t\;(1+\delta'') \qquad (5b)$$

$$+\frac{B}{2\gamma^2}K_3(\sin\Omega t+\sin 3\Omega t)(1+\delta'') \qquad (5c)$$

$$-\frac{B}{\gamma^2}K_5(\cos\Omega t+\cos 3\Omega t)(1+\delta'') \qquad (5d)$$

$$-\frac{B}{\gamma^2}K_6(\sin 2\Omega t+\sin 4\Omega t)(1+\delta'') \qquad (5e)$$

$$+K_1 B\cos\Omega t \qquad (5f)$$

$$+K_2\left(2\Delta\omega B\cos\Omega t+\frac{B^2}{2}\cos 2\Omega t\right) \qquad (5g)$$

$$+K_4\cos\Omega t \qquad (5h)$$

where $\delta' = -\frac{3}{4}\frac{B}{\gamma^2}\cos^2\Omega t + .267\frac{B^6}{\gamma^6}\cos^6\Omega t \ldots$

$$\delta'' = +\frac{3}{2}\frac{B^2}{\delta^2}\cos^2\Omega t - 2.491\frac{B^6}{\delta^6}\cos^6\Omega t .$$

$\Delta\omega \equiv \omega_1 - \omega_o$

Dc terms, and terms involving the product of two or more small coefficients, eg. K_3K_5, have been dropped.

Note that δ' and δ'' are even power series of $(B/\gamma)^2\cos^2\Omega t$ and could have been given in terms of a Bessel function. They contain mixtures of cos $2\Omega t$, cos $4\Omega t$, cos $6\Omega t$ etc. and have a finite value averaged over a period of the modulation frequency Ω.

Term 5a contains a dc contribution plus even harmonics of Ω (mostly 2nd) due to sweeping over the line profile. Expanding term 5a yields

$$5a = 1 - \frac{B^2}{2\gamma^2}\left(1 - .375\frac{B^2}{\gamma^2} + .16\frac{B^6}{\gamma^6}\ldots\right) - 1/4\,\frac{B^2}{\gamma^2}\cos 2\Omega t\left(1 - 3/4\frac{B^2}{\gamma^2} + .267\frac{B^2}{\gamma^6}\ldots.\right) - 1/4\,\frac{B^2}{\gamma^2}\cos 4\Omega t\,(.129\;\frac{B^2}{\gamma^6}\ldots) \qquad (6)$$

(6)

Term 5b contains the desired error signal proportional to the frequency error $\Delta\omega$. Its harmonic content is odd with contributions at Ωt, $3\Omega t$, $5\Omega t$, etc. coming from the expansion of δ''. Note that dependence of this term on $(B/\gamma)^2$ is the same as for the unwanted error terms 5c through 5e.

$$5b = \frac{-\Delta\omega B}{\gamma^2}\left[\cos \Omega t \left(1 + \frac{9}{8}\frac{B^2}{\gamma^2}\ldots\right) + \frac{3}{8}\frac{B^2}{\gamma^2}\cos 3\Omega t(1+\ldots)\right] \qquad (7)$$

Fundamental Sinewave Demodulation

The most common types of demodulators used to recover the error signal displayed in Equation 5 are the sinewave demodulator and the squarewave demodulator. The primary distinction between the two is the type of reference. The reference can be at the frequency of modulation or at a higher harmonic - typically the third.

The first type to be considered is the fundamental sinewave demodulator. The detector of Figure 1 is assumed to be linear. This is very important as nonlinearities can cause intermodulation between the various terms of Equation 4 yielding large errors. These type of errors will not be analyzed here. The function of the prefilter is to filter noise and spurious signals from the detected signal by narrowing the bandwidth. Of particular importance is the reduction of the signals at $2\Omega t$, $3\Omega t$, $4\Omega t$, etc. In addition to the potentials errors originating from terms 5c, 5d, 5e, and 5f of equation 4, the demodulator should be operated at the highest possible level so as to minimize the relative effects of dc offsets in the demodulator output.

Before filtering, the signals at $2\Omega t$, $4\Omega t$, etc. generally far exceed the noise near line center ($\Delta\omega = 0$) and therefore limit the useful dynamic range of the demodulator if not attenuated. Assume that the prefilter attenuates the signal at $2\Omega t$ by K_{12}, and by K_{13} at $3\Omega t$, etc.

The reference signal is further assumed to be the same as that used in the modulator, phase shifted by ϕ, where ϕ is due to various delays in the electronics and can be a function of the environment-especially temperature. For $\phi << 1$, $\cos(\Omega t+\phi)$ can be approximated as $\cos(\Omega t) + \phi\sin(\Omega t)$ yielding

$$\underset{8a}{\text{Ref}} = \underset{8b}{\cos \Omega t + \phi\sin \Omega t} - \underset{8c}{K_7 \sin 2\Omega t} + \underset{8d}{K_8 \cos 2\Omega t} + \underset{8e}{K_9 \sin 3\Omega t} \qquad (8)$$

Mathematically, the effect of the demodulator is to multiply the signal of equation 5 by the reference signal given in equation 8 [1].

The servo acts to force the output of the demodulator towards zero. The actual error depends on the servo gain. If the dc servo gain, G_1 G_2, is sufficiently large, one can assume that the demodulator output is zero. For simplicity the modulator output has been averaged over three full periods of the modulation frequency Ω and the desired frequency error term, 9a, set opposite to the spurious error terms in equation 9.

The effective dc offset [1] of the demodulator is represented by K_{dc} and is a function of the total ac gain of the system, G_{Ac}.

$$\underset{9a}{\Delta\omega (1+\delta')} = \underset{9b}{\gamma^2 L_1} + \underset{9c}{2L_2\Delta\omega \gamma^2} + \underset{9d}{1/2K_3\phi(1+\delta")} + \underset{9e}{K_4 \frac{\gamma^2}{B}} + \underset{9f}{1/2K_5(1+\delta")} + \underset{9g}{1/2(K_{12})[K_\gamma B)(1-\delta")} + \underset{9h}{K_2 B\gamma^2 k_g} + \underset{9i}{K_5K_7)1+\delta")} \underset{9j}{\frac{2K_{dc}\gamma^2}{B}} \qquad (9)$$

Term 9b is due to the linear component of the background slope and is selected out of the error signal by 8a. This error is just the ratio of the background slope to the slope of the derivative multiplied by the angular half bandwidth. In cases where this effect is exceptionally large and/or unmanageable, a third derivative lock can be used at the expense of signal to noise. See later discussion.

Term 9c is also selected out of the error signal by 8a and causes no frequency error by itself, however in the presence of other error terms it effectively modifies the angular half width γ. This effect is usually small and can be ignored.

Term 9d selected out of the error signal by 8b, is due to the out-of-phase component of the second harmonic distortion in the phase modulator ($\sin 2\Omega t$), the effect of this term can be reduced considerably by making ϕ small. Values of ϕ between .01 and .1 are generally easy to achieve and maintain.

Term 9e is selected out of the error signal by 8a and is due to the fractional amplitude modulation, K_4, at $\cos \Omega t$. Since for most systems $\gamma/B = 1$ the error is approximately K_4 multiplied by the half angular bandwidth γ. This can be a major limitation in some systems.

Term 9f, selected out of the error signal by 8a, is due to the mixing of the in-phase component of the modulator harmonic distortion ($\cos 2\Omega t$) with the fundamental of modulation by the resonance. This can be seen from the expansion of the cross products in the denominator of eq. 3. Because of this there is no method to suppress it other than by making K_5 small. Note that the offset is just 1/2 the amplitude of the in phase 2nd harmonic distortion.

Term 9g is selected out of the error signal by K_8 and is due to the 2nd harmonic generation from sweeping back and forth across the resonance. Near line center the $\cos 2\Omega t$ error signals usually dominate all other error signals. By making K_{12} small one can greatly reduce the susceptability to

2nd harmonic distortion in the demodulator and permit G_{AC} to be increased to the largest value consistent with the noise in the bandwidth of the demodulator.

Terms 9h and 9i are second order small and can be neglected in this approximation.

Term 9j is due to the dc offset in the demodulator. Usually K_{dc} is independent of level for small signal levels but at some point the errors grow exponentially with signal level. By making K_{12} very small, one can increase the signal gain to the point that the noise around frequency Ω in a bandwidth determined by the prefilter is just below the maximum level for the demodulator. This and $B/\gamma = 1$ minimizes the effect of K_{dc}.

Thus, for systems where K_{12} is small the most important error terms for sinewave demodulation at the fundamental are

$$\Delta\omega(1+\delta'') = \gamma^2 K_1 + 1/2 K_3\phi(1+\delta'') + K_4 \frac{\gamma^2}{B} + 1/2\, K_5(1+\delta'') + 2K_{dc}\frac{\gamma}{B} \qquad (10)$$

Fundamental Square Wave Demodulation

For many systems it is easier to implement a squarewave demodulator than it is to use a sinewave demodulator and K_{dc} is often much smaller. In this instance the reference signal of equation 8 is replaced by:

$$\text{Ref} = \cos\Omega t + \frac{1}{3}\cos 3\Omega t + \phi\sin\Omega t + (\frac{\phi}{3} + K_6)\sin 3\Omega t - K_7\sin 2\Omega t + K_8\cos 2\Omega t \qquad (11)$$

where the terms varying as sin $6\Omega t$ or cos $6\Omega t$ have been omitted.

It is easily shown that the frequency offset errors of the closed-loop system are functionally very similar as those derived in eq. 10 above.

$$\Delta\omega(1+\delta'') = \gamma^2 K_1 + \frac{5}{6} K_3\phi(1+\delta'') + K_4\frac{\gamma^2}{B} \frac{5}{6} K_5(1+\delta'') + 2K_{dc}\frac{\gamma^2}{B} \qquad (12)$$

Third Harmonic Demodulation

In some cases the background slope is so large and/or unstable that if is advantageous to use a 3rd harmonic reference to the demodulator. Assume it is of the form

Ref = cos $3\Omega t$ + K_{11} cos $6\Omega t$...

In this case the significant frequency errors are given by

$$\Delta\omega(1+\delta'') = \frac{4}{3} K_3\phi(1+\delta'') + \frac{4}{3} K_5(1+\delta'') + \frac{16}{3} K_{dc}\frac{\gamma^2}{B} \qquad (13)$$

where the 6th order terms have been neglected. Although the sensitivity to sloping background and amplitude modulation is virtually gone, the signal is generally also reduced by a factor of 2 or 3 which increases the relative importance of 2nd harmonic distortion in the modulator and dc offset in the demodulator. K_{12} should be kept small in order to maximize G_{ac} and thereby reduce the effect of K_{dc}.

Tests for Servo Errors

Errors generated from the K_1 coefficient have the same functional dependence on modulation width as the desired signal and are therefore difficult to separate in a fundamental demodulation system. Therefore, one generally has to measure the background slope separately and calculate the offset. One could also compare the frequency of line center for a fundamental and a 3rd harmonic demodulation system. In cases where a hersey structure is present, one can compare the frequency of line center when locked to pairs of successive lobes.

Errors generated from K_5, K_6, and K_7 are generally small and can be neglected.

Errors generated from K_3 can be separated out from the other terms by varying the phase shift ϕ. For most implementations, K_3 varies as B^2. Modeling of the modulator can also be helpful.

The errors associated with K_4 can best be determined by measuring the fractional amplitude modulation at Ω on the probe signal. The phase chosen for the K_4 term is the most likely and has the largest error. K_4 depends on the modulation width B.

The errors associated with K_5 are best illuminated by varying the modulation width B. A plot of frequency change vs. B^2 for $\phi = 0$ yields K_5 while the difference between that curve and the one obtained with $\phi = 0.2$ can be used to determine K_3. K_5 can also be determined from a careful characterization of the phase modulator.

The errors associated with K_8 are unique to the fundamental demodulator systems and can be illuminated by varying K_{12}. For K_{12} small this error can be totally neglected.

The errors originating from K_{dc} are best separated out by varying the ac gain. Varying the dc gain only changes the loop attack time (bandwidth) and should have no effect on these offsets [1]. Another technique for illuminating K_{dc} generated errors is to vary the ac gain with no modulation on the probe and measure the dc error signal.

Summary

A simple model of a resonance system probed by a sinusoidally modulated probe signal has been treated to expose the first order errors in determining line center due to imperfections in the electronics. Although this approach does not easily produce rigorous values for the frequency errors, in that it does not take into account saturation etc., it does yield the correct functional dependence of the errors on modulation index, ac gain, etc. This permits one to compare the offsets in determining line center using various servo configurations. As we've shown, in any servo system with a fundamental demodulator reference, the most serious frequency errors originate from sloping background, 2nd harmonic distortion in the frequency modulation, amplitude modulation on the probe signal, and dc offsets in the demodulator. Servo systems utilizing the 3rd harmonic of the modulation as a demodulator reference are generally not sensitive to baseline tilt or amplitude modulation on the probe, but have increased sensitivity to 2nd harmonic distortion in the modulator, and to dc offsets in the demodulator. With the functional dependence outlined here it's relatively easy to design sensitive tests of these offsets even if the original assumption about a pure Lorentzian line and small modulation index are not exactly fulfilled.

Acknowledgments

The author is very grateful to Dr. J. C. Bergquist and Andrea DeMarchi for fruitful discussions and to the Naval Research Laboratories for financial support.

References

[1] F. M. Gardner, Phaselock Techniques (John Wiley, New York, 1966).

1/f FREQUENCY FLUCTUATIONS IN ACOUSTIC AND OTHER STABLE OSCILLATORS

T.E. Parker

Raytheon Research Division
131 Spring Street
Lexington, MA 02173

ABSTRACT

Flicker, or 1/f, noise is a universal phenomenon that occurs in quartz acoustic oscillators, as well as many other types of oscillators. As more data is acquired on acoustic oscillators, some questions regarding the influence of design, fabrication, and environmental factors on 1/f noise can now begin to be answered. The 1/f noise has been measured on over 70 SAW resonators, including one and two port devices, covering the range of 187 MHz to 984 MHz. There is an approximate +/- 10 dB spread in noise levels among "identical" SAW devices and it has been confirmed that in nearly all cases, the SAW device is a greater source of 1/f noise than most commercial silicon bipolar amplifiers. Also, a strong dependence of 1/f noise on unloaded Q has been observed ($\sim 1/Q_u^4$), but the dependence on loaded Q is weak. The data for one port resonators suggests that they may tend to be noisier than two port devices. Environmental factors have relatively little effect on 1/f noise.

When data from the literature for bulk acoustic wave (BAW) oscillators is included with the SAW data, the $1/Q_u^4$ dependence is even more dramatic. It is clear that the 1/f noise process is essentially the same in BAW and SAW oscillators. The lower 1/f noise levels of most BAW oscillators is due primarily to their higher Q. Data will also be presented for 1/f noise levels on acoustic resonators using materials other than quartz, and for other non-acoustic type oscillators.

INTRODUCTION

Flicker, or 1/f, noise is a universal phenomenon that occurs in most stable oscillators as well as many other physical systems. Quartz acoustic oscillators are no exception, and 1/f noise has been observed in both surface acoustic wave (SAW)[1,2] and bulk acoustic wave (BAW) devices.[3,4] The ultimate source of this noise is of course not yet understood, but other less fundamental questions also remain to be answered. For example, how does the 1/f noise vary with frequency, or Q? Do design parameters such as one port versus two port make a difference? Do fabrication or environmental factors influence the 1/f noise? Is the noise process the same in SAW and BAW devices? Now that techniques for measuring 1/f noise are well established, the body of data is increasing rapidly and some of these questions can now begin to be answered.

This paper reports on the results of 1/f noise measurements on over 70 SAW resonators. Also discussed are published and measured results from BAW resonators as well as 1/f noise levels from other stable oscillator technologies.

1/f Noise in SAW Resonators

A 1/f noise power spectral density of frequency fluctuations is known to exist in SAW resonators over the offset frequency range of approximately 10^{-2} Hz to 10^4 Hz.[5] Recently this noise level was measured on 56 vacuum sealed SAW resonators (fabricated on 40 degree rotated Y cut quartz) covering the frequency range of 187 MHz to 984 MHz. This corresponds to values of unloaded Q (no electrical loading) from over 40,000 to approximately 7,000. Nearly all of these devices exhibited values of Q_u that were greater than 75% of the material limit. The noise measurements were made by operating the SAW devices in oscillators and measuring the phase noise of the oscillators over the frequency range of 1 Hz to 300 Hz. Figure 1 shows a typical plot of single sideband phase noise, $\mathcal{L}(f)$, on a SAW oscillator. (Above ~300 Hz the noise from the reference synthesizer dominates.) From this data an average value for the level at 1 Hz was determined. The value of the power spectral density of

CH2186-0/85/0000-0097$1.00©1985IEEE

frequency fluctuations, $S_{\Delta F}(f=1Hz)$, was calculated from $\mathcal{L}(f=1Hz)$ [in dBc/Hz] by using the relation

$$S_{\Delta F}(f) = (2x10^{\mathcal{L}(f)/10})/f^2 \qquad (1)$$

Figure 2 shows the measured values of $S_{\Delta F}(f)$ at 1 Hz plotted as a function of unloaded Q, Q_u, for the 56 SAW resonators. The values of $S_{\Delta F}(f = 1 Hz)$ serve as a measure of the magnitude of the 1/f noise. The solid data points represent two port resonators and the open data points represent one port resonators.

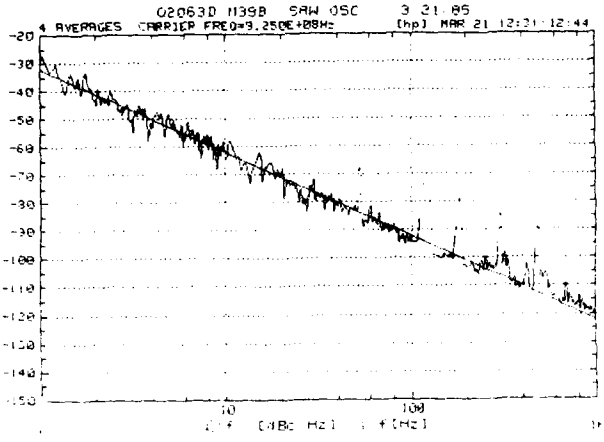


Figure 1. Typical plot of single sideband phase noise as a function of offset frequency for a SAW resonator oscillator.

One Port vs Two Port

The data in Fig. 2 shows that there is approximately +/- 10 dB of scatter for devices with the same unloaded Q. This is similar to what was observed by Elliott and Bray.[2] Though there is considerable scatter, it is clear that there is not much difference in the noise levels of the one port devices as compared to the two port devices. In fact, the one port resonators have a slightly higher (~4 dB) average noise level. The electronic circuits used for the one port resonators were somewhat different from those used for the two port devices so it was desirable to verify that the circuits were not influencing the observed noise levels. This was easy to do since two port devices could

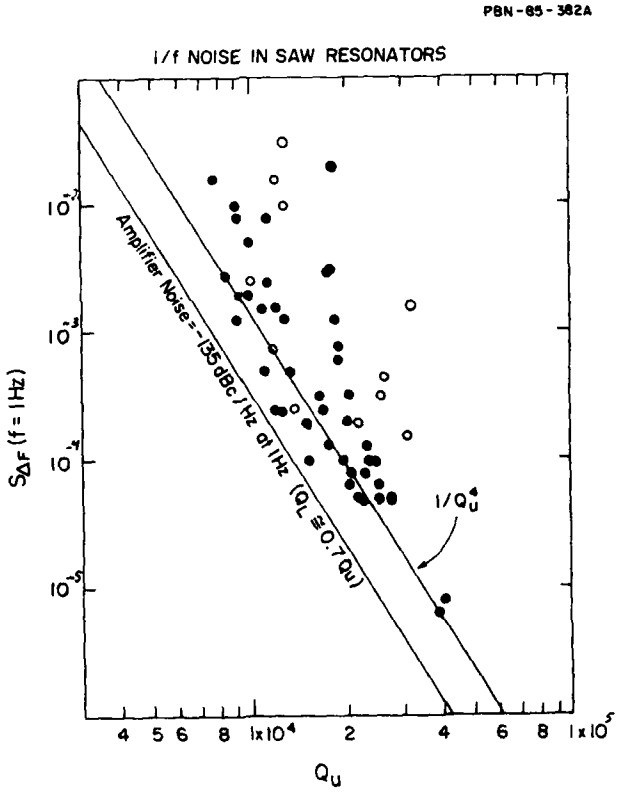


Figure 2. 1/f noise in SAW resonator oscillators as a function of unloaded Q. Open data points represent one port resonators and solid data points represent two port resonators.

be operated in the one port circuits (but not vice versa). Such measurements with a low noise two port device confirmed that the one port measurements were not being affected by the oscillator electronics.

Dependence on Loaded and Unloaded Q

Another feature of the data in Fig. 2 is the strong dependence of 1/f frequency fluctuations on the unloaded Q. This dependence is close to $1/Q_u^4$. In light of this, the apparent slightly higher noise levels of the one port devices are surprising since the one port resonators had a higher average Q, due to the need for fewer fingers to give the same insertion loss. The unloaded Q, Q_u, was chosen for the independent variable in Fig. 2 rather than the loaded Q, Q_L, because it has been observed that the dependence of 1/f noise on Q_L is weak.[6] This is illustrated in Fig. 3. Here the loaded Q was varied on four resonators by adjusting the impedance match-

ing circuits. The accompanying change in SAW insertion loss was compensated for by changing the gain of the loop amplifier by the addition or removal of attenuation. As can be seen in Fig. 3, the dependence of flicker noise on loaded Q in SAW resonators is much weaker then the dependence on unloaded Q.

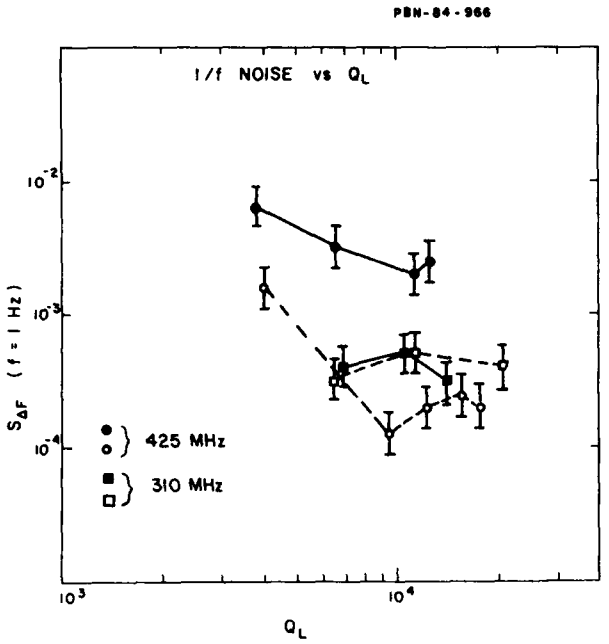


Figure 3. 1/f noise level at 1 Hz of four SAW resonators as a function of loaded Q.

It should be pointed out that, because the devices in Fig. 2 are operating near the material limit for unloaded Q, there is a strong relationship between resonator frequency, F_0, and Q_u. This relationship occurs because the $Q_u \times F_0$ product for material limited acoustic resonators is a constant. (Approximately 1×10^{13} for ST cut quartz.) The variable $1/F_0$ could be substituted for Q_u in Fig. 2 and the data would look essentially unchanged except for a scale factor. To test the dependence on Q_u, flicker noise measurements were also made on 27 ~400 MHz SAW resonators in which a metal pad of either aluminum or gold was added to the acoustic path to lower the value of Q_u.[7] The results of these measurements seemed to confirm that Q_u was the important parameter rather than F_0, since a strong dependence on Q_u was also observed.

However, some recent results with air loading have cast some doubts on this conclusion. In these recent experiments, the 1/f noise was measured on a 425 MHz SAW resonator that was contained in a chamber in which the temperature and pressure could be varied. By varying the pressure of a nitrogen atmosphere from 10^{-6} Torr to 35 PSI the value of Q_u could be changed by a factor of two. The data from this experiment, and the metal loading results, are shown in Fig. 4. The strong dependence of the 1/f noise on Q_u for the metal loaded resonators is clearly seen, but the dependence on Q_u for the air loaded device is weak, and may be even less than $1/Q_u^2$. Therefore, it must be concluded that for devices operating well below the material limit, the means by which Q_u is decreased is important to the level of 1/f noise. However, for material limited devices, the 1/f frequency fluctuations can be thought of as varying as $1/Q_u^4$ or F_0^4. Though the Q_u dependence is not as simple as first thought, the data in Fig. 4 clearly demonstrates that metal films play a major role in determining 1/f noise levels.

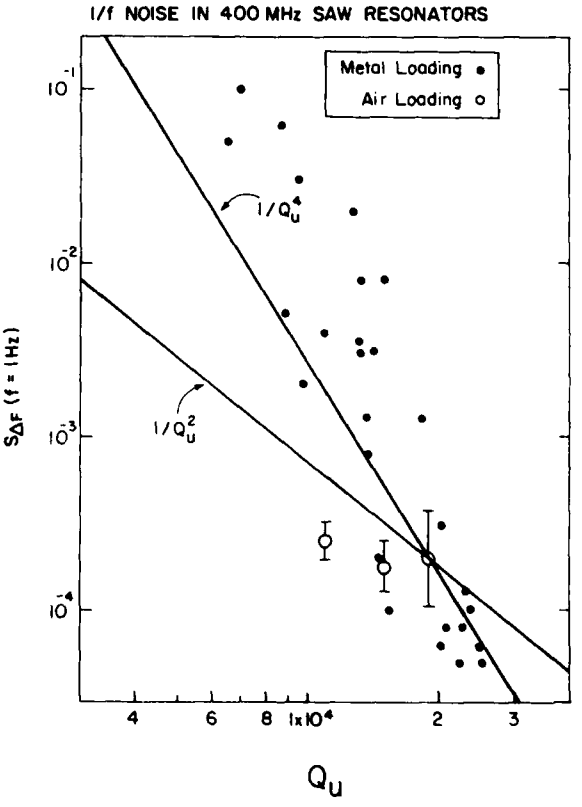


Figure 4. 1/f noise in 400 MHz SAW resonator oscillators as a function of unloaded Q.

Other Environmental Factors

In addition to the pressure dependence of the 1/f noise on the 425 MHz resonator, two other environmental factors were investigated. The flicker noise was measured at -5°C, 40°C, and 90°C, and was found to be constant to within +/- 1 dB. Also, it was observed that exposure to UV light (350 to 380 nm) at 62 microwatts/cm^2 and white light at 20 mW/cm^2 had no effect on the noise level.

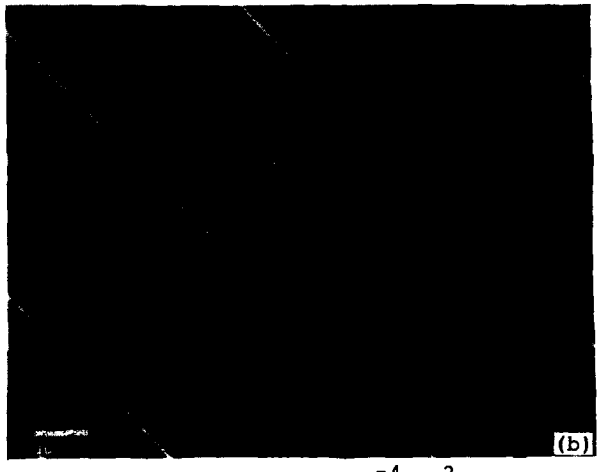
Correlation With SEM Data

As part of this investigation of 1/f noise in SAW devices, some of the resonators for which the noise had been measured were examined with a scanning electron microscope (SEM) to determine if there was any correlation between the physical appearance of the surface of the SAW device and the observed 1/f noise level. At this time nine devices have been examined. These include five 416 MHz resonators that have just completed a two year aging test and four two port resonators in the range of 650 MHz to 800 MHz. The bare quartz surface is essentially featureless and has provided very little information. However, the transducer is quite visible and can have considerable structure.

At this time no correlation has been observed between 1/f noise and the number of open fingers or other defects in the metallization. Also there is no apparent correlation between 1/f noise and the quality of the finger definition. This is illustrated in Fig. 5 which shows transducer fingers from two 416 MHz resonators. For the device in Fig. 5a the lift-off was poor and the finger edges were very ragged. In Fig. 5b a very clean lift-off was achieved. The 1/f noise levels for these two devices are also shown in Fig. 5, and as can be seen, they are very similar.

Another parameter that has been examined for correlation with 1/f noise is the width to space ratio of the transducer metallization. The width to space ratio is defined as the ratio of the finger width to the distance between fingers. For an ideal transducer

$S_{\Delta F}(f=1\ Hz) = 2 \times 10^{-4}\ Hz^2/Hz$



$S_{\Delta F}(f=1\ Hz) = 1 \times 10^{-4}\ Hz^2/Hz$

Figure 5. (a) Ragged finger edges and (b) clean finger edges.

this ratio should be one (i.e., finger width equals finger space), but in practice fabrication variables generally result in fat fingers, particularly at high frequencies. Figure 6 shows a plot of observed 1/f noise levels as a function of width to space ratio for the nine devices that have been examined with the SEM. The data points for the five 416 MHz resonators are plotted as open circles, while the data from the other four resonators are plotted as solid circles. There does appear to be a correlation between width to space ratio and 1/f noise, with the larger values of w/s giving higher noise levels. The data in Fig. 6 is very intriguing, but results from more devices will be required before a convincing case can be made. These measurements are now in progress.

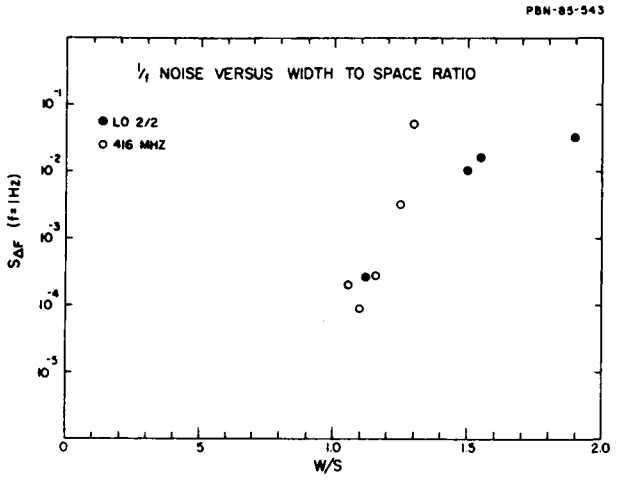


Figure 6. 1/f noise level as a function of width to space ratio.

1/f Noise in Amplifiers

When making noise measurements on oscillators it is important to be sure which components are the dominant source of the excess noise. The amplifiers used in the oscillator circuits for these measurements were commercial wide band silicon bipolar devices which have a very low 1/f noise level. By interchanging SAW devices and amplifiers it was easy to confirm that nearly all of the SAW resonators in Fig. 2 were the dominant source of the flicker noise. However, the 1/f noise levels of several typical amplifiers were also measured directly. These measurements were done both with the amplifiers terminated in 50 ohms and also with the amplifiers terminated with an impedance similar to what they would see in the oscillator loop. In both cases the open loop single sideband phase noise at 1 Hz, $\mathcal{L}'(f=1Hz)$, was ~ -135 dBc/Hz over the full frequency range of the amplifiers.

By using the definition of group delay

$$\tau = \Delta\phi/\Delta\omega_0 \qquad (2)$$

and the relation between loaded Q and group delay, τ,

$$Q_L = \omega_0\tau/2 \qquad (3)$$

it can be shown that, in an oscillator, the power spectral density of frequency fluctuations, $S_{\Delta F}(f)$, caused by open loop phase fluctuations with power spectral density $S_\phi'(f)$ in the loop amplifier is

$$S_{\Delta F}(f) = (F_0/2Q_L)^2 S_\phi'(f) \qquad (4)$$

In these expressions $\omega_0 = 2\pi F_0$, where F_0 is the resonant frequency and ϕ is the phase shift through the SAW device. Note that

$$\mathcal{L}'(f) = 10\ \mathrm{Log}[S_\phi'(f)/2]. \qquad (5)$$

For a constant value of $S_\phi'(f)$ [or $\mathcal{L}'(f)$], equation 4 predicts a $1/Q_L^2$ dependence for $S_{\Delta F}(f)$, which clearly was not present in Fig. 3. This is further evidence that 1/f amplifier phase noise is not dominant in these SAW oscillators.

Using the relation

$$Q_u \times F_0 = C \qquad (6)$$

and assuming

$$Q_L \cong 0.7\ Q_u \qquad (7)$$

equation 4 becomes

$$S_{\Delta F}(f) \cong (C^2/2Q_u^4) S_\phi'(f). \qquad (8)$$

We now see that a constant $S_\phi'(f)$ will give a $1/Q_u^4$ dependence to the 1/f frequency fluctuations of acoustic oscillators operating near the material limit for Q_u. However, when the measured value of the amplifier phase noise [$\mathcal{L}'(f = 1Hz) = -135dBc$] is used in equations 5 and 8, the calculated oscillator noise levels are found to be too low. This is illustrated in Fig. 2 by the line repesenting the amplifier noise. It is now clear that the SAW resonators are the dominant source of the 1/f noise in nearly all cases. Furthermore, Eq. 8 and the $1/Q_u^4$ dependence of the data in Fig. 2 also show that the average open loop phase noise of the SAW resonators, $S_\phi'(f)$, [or $\mathcal{L}'(f)$] is a constant value which is independent of F_0 or Q_u. From the data in Fig. 2, the average value of $\mathcal{L}'(f = 1\ Hz)$ for the SAW resonators is calculated to be -124 dBc/Hz. This value is 11 dB higher than the

amplifier noise and is in close agreement with the results of Elliott and Bray.[2]

It should be noted that not all amplifiers will exhibit the same low 1/f noise level that was present in the amplifiers used in this investigation. In particular, GaAs FET amplifiers have a 1/f noise level that can be as much as 20 dB higher than silicon bipolar amplifiers.

Possible Causes of 1/f Noise in SAW Resonators

There are two potential fundamental types of noise in a SAW resonator that can result in frequency fluctuations in an oscillator. One is a variation of the resonant frequency of the acoustic device, and the second is fluctuations in the phase relationship between the electrical signal (at the pins of the package) and the acoustic energy stored in the resonator. In an open loop measurement, both will result in fluctuations of the transmission phase shift. However, the observed phase fluctuations due to changes in the resonant frequency will depend on the loaded Q, which determines the phase slope. Conversely, fluctuations in the phase relationship may very well be largely independent of the loaded Q. In a closed loop, the situation is reversed. Here the oscillator frequency will track the resonator frequency independent of loaded Q, but fluctuations in the phase relationship will result in oscillator frequency variations that are dependent on loaded Q, as given in Eq.(4). At this time the experimental evidence favors fluctuations in resonant frequency because of the weak dependence of oscillator noise on loaded Q as shown in Fig. 3.

For cavity resonators, the resonant frequency is determined by the equation

$$F_0 = NV/L \qquad (9)$$

where N is a large integer or half integer (~390 for these resonators), V is the average SAW velocity, and L is the effective cavity length. Any parameter that affects V or L can cause a change in resonant frequency. Table 1 is a list of possible sources of frequency fluctuations. These sources are divided into two basic catagories, (1) those that affect velocity and, (2) those that affect cavity length. The ones that affect velocity are subdivided into those that affect the quartz directly, and those that affect the transducer. The transducer parameters are divided into mechanical and electrical phenomena.

Table 1. Sources of frequency fluctuations.

Velocity			Effective Cavity Length
Quartz	Transducer		
	Mechanical	Electrical	
Temperature	Temperature	Finger Resistance	Temperature
Stress	Stress	Interelectrode Resistance	Strain
Defects	Defects	Capacitance	Defects
Damage	Acoustic Attenuation		
Piezoelectric Constant	Adsorption and Desorption		
Acoustic Attenuation			
Adsorption and Desorption			
Surface Conductivity			

There are many possible sources of 1/f noise, but comments can be made about some of them. First of all, any fluctuation that influences the entire substrate in the same fashion as static or dynamic temperture stability, or vibration sensitivity is not a likely source since the fourth power dependence on Q_u (or F_0) of 1/f noise cannot be explained. Temperature and vibration sensitivity are constant with respect to fractional frequency changes. Therefore, temperature and stress related phenomena that affect the entire substrate are probably ruled out. This also includes fluctuations due to time dependent changes in bulk defects or damage. Fluctuations in acoustic attenuation can also cause frequency fluctuations,[4] but the predicted Q dependence for this process is $1/Q_u^4$ for fractional frequency fluctuations, $S_y(f)$, rather than absolute frequency fluctuations, $S_{\Delta F}(f)$, as has been observed here. Surface related phenomena such as surface damage, surface temperature fluctuations, or adsorption and desorption of contaminants would of course be frequency dependent, but again it

is not clear how a Q to the fourth power dependence could be obtained. Furthermore, the 1/f noise is not strongly affected by surface perturbations caused by illumination or pressure changes.

The transducer metallization can also influence the average SAW velocity through both it's mechanical and electrical properties. Though the transducer makes only a small perturbation on the velocity, the fractional frequency fluctuations that are observed are also very small (~3 to 10 parts in 10^{11}). Also, the physical properties of the transducer metal are not nearly as stable as the quartz subtrate. The strong dependence of the 1/f noise on the amount of metal added to reduce the unloaded Q, as seen in Fig. 4, is an indication that the transducer metal plays an important role in 1/f noise. This is further supported by the preliminary data in Fig. 6 which shows a dependence on width to space ratio.

Changes in the physical length of the cavity would also cause frequency fluctuations. Though this seems unlikely to be a source of noise, it must be noted that a change in frequency of 3×10^{-11} corresponds to a change in physical length for a typical SAW resonator of only .0004 angstrom. It is not hard to believe that one or two interstitial atoms could cause this much perturbation.

After looking at the list in Table 1, one begins to wonder if the difficulty in understanding 1/f noise does not stem from the possibility that there may be many simultaneous sources of noise.

1/f Noise in SAW and BAW Resonators

Measured levels of 1/f noise in quartz BAW resonators are now available in the literature,[3,4,8-12] and when this data is combined with the SAW data, the $1/Q_u^4$ dependence becomes even more dramatic. This is shown in Fig. 7. All of the BAW data are taken from the literature as indicated in the figure with the exception of the data point for the 466 MHz device. This device was measured in the author's laboratory using the same type of amplifier as was used with the

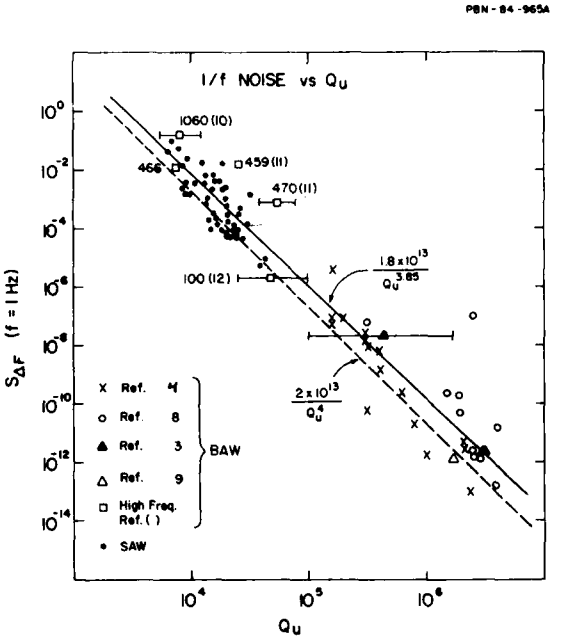


Figure 7. 1/f noise levels of quartz SAW and BAW resonators as a function of unloaded Q.

SAW devices. The frequencies of the BAW devices at or above 100 MHz are indicated beside each data point. The BAW data points with no indicated frequencies fall in the range of 1 to 25 MHz. The presence of error bars indicates that the value of Q_u was not available and had to be estimated.

A least mean square fit to the data in Fig. 7 gives a slope of $Q_u^{-3.85}$ +/- 0.11. The standard deviation of the data is +/- a factor of 8, or +/- 9 dB. The data clearly indicates that within +/- 10 dB, the 1/f noise process is the same in SAW and BAW resonators.

An empirical relation for estimating the level of 1/f noise in material limited SAW and BAW oscillators can be obtained from the data in Fig. 7.

$$S_{\Delta F}(f) \cong (2 \times 10^{13}/Q_u^4)(1/f) \qquad (10)$$

Equation 10 gives a value that is near the low end of the range of data in Fig. 7 and therefore, represents an approximate lower limit for 1/f frequency fluctuations in quartz acoustic resonators. In general it is not unusual to obtain devices with relatively

high noise levels, but devices that are more than 10 dB lower than the value calculated from equation 10 are very rare. Equation 10 can also be written in terms of F_0 (in MHz).

$$S_{\Delta F}(f) \cong [2\times10^{-15}(F_0{}^4)](1/f) \qquad (11)$$

or can be expressed as fractional frequency fluctuations, $S_y(f)$.

$$S_y(f) \cong (2\times10^{-13}/Q_u{}^2)(1/f) \qquad (12)$$

The $1/Q_u{}^4$ (or $F_0{}^4$) dependence of the 1/f noise in quartz acoustic resonators means that when a low frequency quartz oscillator is multiplied to a higher frequency, the degradation in close-to-carrier noise is less than that obtained by operating a quartz oscillator directly at the higher frequency. When multiplying to a higher frequency the phase noise is increased by the square of the multiplication factor. By operating directly at the higher frequency, the close-to-carrier noise is increased by the frequency ratio to the fourth power. This explains why the close-to-carrier noise of a high quality frequency synthesizer is lower than that of a free running high frequency SAW or BAW oscillator.

Excess Noise in Other Stable Oscillators

Information on 1/f noise levels is also available in the literature for other types of resonator stablized oscillators. These results are shown in Fig. 8. The data discussed earlier for SAW and BAW quartz acoustic resonators is shown as solid circles, while data for other types of oscillators are shown as corresponding letters. The data for the two 373 MHz berlinite SAW resonators was measured by the author, but all of the other information was taken from the literature as indicated in the figure. Noise data at 1 Hz was not always available in the literature so the existing data was extrapolated to 1 Hz. In a few cases it was also not totally clear that the excess noise was 1/f in nature.

Considering the variety of resonator technologies present in this data, it is surprising that the excess noise levels do not deviate greately from the $1/Q_u{}^4$ depen-

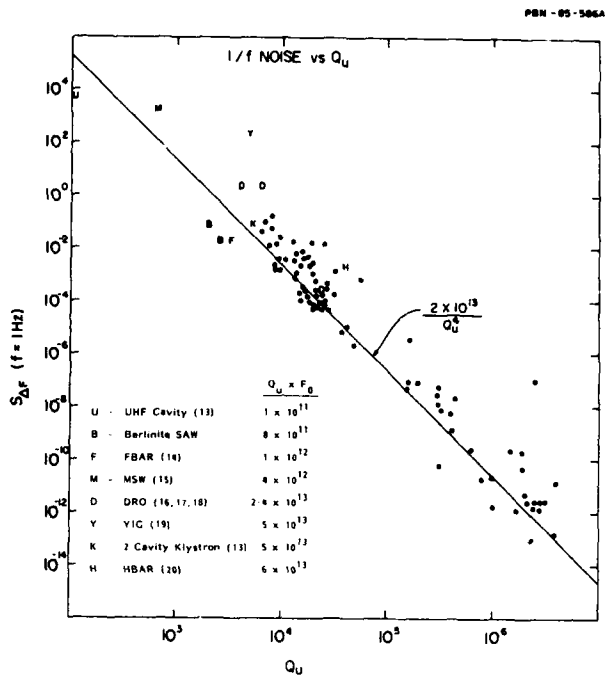


Figure 8. 1/f noise levels of quartz SAW and BAW resonators (solid circles) and other stable oscillators.

dence of the quartz oscillators. However, there is a tendency for the resonators with low Q_uF_0 products (U,B and F) to lie below the quartz line, while the devices with high Q_uF_0 products (D, K and H) fall above the line. This is what you would expect from Eq. 8, except that the deviation from the quartz line is not always as large as predicted. Not unexpectedly, the magnetically tunable resonators (MSW and YIG) have high noise levels even though the data plotted is for the lowest levels reported. Devices that are capable of being tuned over large frequency ranges are most certainly more vulnerable to externally induced fluctuations than devices that cannot be pulled far.

In general it is not clear for the non-quartz devices in Fig. 8 whether the resonant devices or the amplifiers are the dominant source of the excess noise. The magnetic devices are probably dominant, and it is certain that the berlinite resonators have high noise levels ($\mathcal{L}(f = 1\ Hz) = -120$ dBc/Hz). For the other oscillators it is not known which component dominates. However, if the major source of open loop phase noise is constant with oscillator frequency (as with

commercial wide band amplifiers) then Eq. 8 gives a second order dependence on the Q_uF_0 product. Therefore, a device with a high Q_uF_0 product lies above the quartz line and devices with a low Q_uF_0 products lie below the quartz line. This has the appearance of indicating that high Q technologies give higher noise levels, but at equivalent frequencies the higher Q resonators give lower noise.

Acknowledgements

The author would like to thank Gordon Jackson for valuable assistance in obtaining much of the SAW data and J. Lang, E. Sabatino and J. Columbus for fabricating the devices. The author would also like to thank Gary Montress and Joe Callerame for valuable discussions and encouragement.

REFERENCES

1. T.E. Parker, "1/f Phase Noise in Quartz Delay Lines and Resonators", Proc. 1979 IEEE Ultrasonics Symp., p. 878 (1979).

2. S.S. Elliott and R.C. Bray, "Direct Phase Noise Measurements of SAW Resonators", Proc. 1984 IEEE Ultrasonics Symp., p. 180 (1984).

3. F.L. Walls and A.E. Wainwright, "Measurement of the Short Term Stability of Quartz Crystal Resonators and the Implications for Crystal Oscillator Design and Applications", IEEE Trans. Instrum. and Meas., IM-24, p. 15 (1975).

4. 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, First Theoretical Interpretation", Proc. 35th Annual Symp. on Frequency Control, p. 476 (1981).

5. T.E. Parker, "Random and Systematic Contributions to Long Term Frequency Stability In SAW Oscillators", Proc. 1983 IEEE Ultrasonics Symp., p. 257 (1983).

6. T.E. Parker, "1/f Frequency Fluctuations in Quartz Acoustic Resonators", Applied Physics Letters, 46, p. 246 (1985).

7. J. Day, G. Jackson and T.E. Parker, "Bias Controlled Frequency Trimming Of SAW Devices in a DC O_2 Plasma", Proc. 38th Annual Symp. on Frequency Control, p. 310 (1984).

8. J.J. Gagnepain, M. Oliver and F.L. Walls, "Excess Noise in Quartz Crystal Resonators", Proc. 37th Annual Symp. on Frequency Control, p. 218 (1983).

9. Guaranteed noise performance of Model FTS 1050A/004 Quartz Frequency Standard (5 MHz) manufactured by Frequency and Time Systems Inc., 34 Tozer Rd., Beverly, MA 01915.

10. J.P. Aubry, "Quartz and $LiTaO_3$ VHF Resonators for Direct Frequency Generation in the GHz Range", Proc. 1983 IEEE Ultrasonics Symp., p. 487 (1983).

11. L. Bidart and J. Chauvin, "Direct Frequency Crystal Oscillators", Proc. 35th Annual Symp. on Frequency Control, p. 365 (1981).

12. Guaranteed noise performance of Model 1830025 High Frequency Crystal Oscillator (100 MHz) manufactured by Piezo Systems, 100 K St., Carlisle, PA 17013.

13. D.B. Leeson, "Short Term Stable Microwave Sources", Microwave Journal, 13, p. 59 (June 1970).

14. M.M. Driscoll, S.V. Krishnaswamy, R.A. Moore and J.R. Szedon, "UHF Film Resonator Evaluation and Resonator-Controlled Oscillator Design Using Computer-Aided Design Techniques", Proc. 1984 IEEE Ultrasonics Symp., p. 411 (1984).

15. J.P. Castera, P. Hartemann and J.M. Dupont, "A Tunable Magnetostatic Volume Wave Oscillator", Proc. 1983 IEEE MTT Symp., p. 318 (1983).

16. G.D. Alley and H.C. Wang, "An Ultra-Low Noise Microwave Synthesizer", IEEE Trans. on Microwave Theory and Tech, MTT-27, p. 969 (1979).

17. M. Purnell, "The Dielectric Resonator Oscillator-A New Class Of Microwave Signal Source", Microwave Journal, 24, No. 11, p. 103 (1981).

18. M.J. Bianchini, J.B. Cole, R. DiBiase, Z. Galani, R.W. Laton and R.C. Waterman,Jr., "A Single-Resonator GaAs FET Oscillator with Noise Degeneration", Proc 1984 IEEE MTT Symp., p. 270 (1984).

19. Ernst Schloeman, Raytheon Research Div., private communication.

20. J.T. Haynes, H. Salvo, R.A. Moore and B.R. McAvoy, "Low Phase Noise Multiple Frequency Microwave Source", Proc. 37th Annual Symp. on Frequency Control, p. 87 (1983).

39th Annual Frequency Control Symposium - 1985

TIME SCALE STABILITIES BASED ON TIME AND FREQUENCY KALMAN FILTERS

James A. Barnes
Austron, Inc.

and

David W. Allan
National Bureau of Standards
Boulder, Colorado

Abstract

The use of Kalman filters to generate time scales has been well documented in the literature. The typical "Time" Kalman model used for the commercial cesium beam standards is the superposition of white noise frequency modulation (FM) and random walk noise FM. These processes are considered to be continuous and usually sampled at regular intervals. The sample data are the differences in clock readings between a reference clock and each of the other clocks in the time scale system. A Kalman filter can estimate both the time and frequency corrections for each clock in the scale.

There are, however, other options for time scale operation. One can integrate the frequency correction element of the Kalman state vector for a clock in the scale to obtain a new and different time scale. Also, one can re-cast the entire Kalman model in terms of frequency rather than time. The input (measurement) data for this Kalman filter, then, is exactly the first (time) difference of the data discussed above for the "Time" Kalman. As one might expect, however, each of these options provides different performance in the resulting time scale. While this has been pointed out before, the present paper details the various scale performances between measurements and provides an insight into the different performances based on computer simulation studies. For example, the "Time" Kalman filter displays discrete steps in the time corrections where the "Frequency" Kalman filters are continuous (being the integral of a bounded process). Depending on whether one is most interested in minimizing the RMS time error or minimizing the Allan Variance, one chooses the one time scale over the other.

In a more fundamental sense, FREQUENCY is the basic quantity which is measured in the laboratory while TIME is subject to many conventions and exhibits unbounded errors. In fact, as realized today, time is a defined quantity (dependent upon algorithms, definitions and procedures) and not intrinsic to the atomic clocks used to generate time. For these reasons the frequency Kalman algorithms should be used for the realization of primary time scales since it is frequency, not time, which has a physical basis.

I. Introduction

The literature contains several papers on the use of Kalman Filters in the establishment of time and frequency standards [1,2,3,4]. The typical "time" Kalman model is based on commercial cesium beam frequency standards. Specifically the models are the superposition of white noise frequency modulation (FM) and random walk noise FM.

While several researchers have reported the existence of linear frequency drifts, typically the establishment of a statistically significant drift requires measurements extending over an appreciable fraction of the clock's life expectancy. This paper ignores linear frequency drift. For times shorter than 100 seconds or so other model elements become important, but we ignore them also since most time scales sample data at longer intervals.

II. The Time-Kalman Model

Jones and Tryon [1,2] have shown that the individual Kalman noise models can be written in the form:

$$\begin{pmatrix} X(t+\tau) \\ Y(t+\tau) \end{pmatrix} = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} + \begin{pmatrix} \varepsilon\,(t+\tau) \\ \eta\,(t+\tau) \end{pmatrix}$$

where X(t) and Y(t) are the continuous time and frequency (approximately) errors of a clock, and $\varepsilon(t)$ and $\eta(t)$ are independent (band limited) white noises. Jones and Tryon also pointed out that there are actually correlations between these noise terms which can become significant if the time interval, τ, becomes too large. For the clocks considered here, these correlations can be ignored. The idea here is that X(t) and Y(t) are continuous random processes which can be sampled at arbitrary intervals, t and $t+\tau$. (For a more complete discussion see Jones and Tryon [1,2].)

The covariance matrix of the driving noise terms is [1]

$$Q(\tau) = \begin{pmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix} \tau$$

for sufficiently small τ.

The state vector for an ensemble of M clocks is obtained by appending the two state elements of each clock into a column vector of length 2M. The state transition matrix is a 2M by 2M square matrix consisting of the 2 by 2 blocks for the state transition matrices along the main diagonal, with off-diagonal blocks being zeros. Similarly, the

 "US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

ensemble covariance matrix of the driving noise terms is a 2M by 2M square matrix formed from the M 2 by 2 individual blocks, with the off-diagonal blocks being zeros.

The measurements consist of time differences between clock #1 (the reference clock) and each of the other clocks in the ensemble. Although each "measurement" in the simulation studies was treated separately, an equivalent procedure [5] would be to define the H(t)-matrix in the form:

$$H(t) = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & . & . \\ & & & & & & & & \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & . & . \\ & & & & & & & & \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & . & . \\ & & & & & & & & \\ . & . & . & . & . & . & . & . & . \\ & & & & & & & & \\ . & . & . & . & . & . & . & . & . \end{pmatrix}$$

which is an M-1 by 2M matrix for the M-1 independent measurements. All primary time scales have one more unknown than they have independent measurements since absolute time accuracy is impossible.

III. Estimated State Vector for the Time-Kalman Filter

We assume that every τ seconds a complete measurement of M-1 clock comparisons is performed. These comparisons allow an up-date of the estimated Kalman state vector. Of course, the time and frequency errors are thought to evolve continuously even if they were only observed every τ seconds. Kalman theory allows one to estimate (i.e., forecast) these subsequent states even though the comparisons between clocks were not performed continuously. Figure 1 displays the estimated Kalman state vector for clock #1 and the "actual" clock #1 error for a simulated time scale.

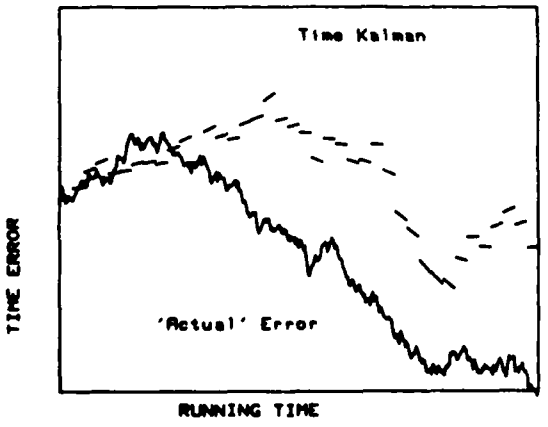


FIG. 1, SIMULATED CLOCK ERROR AND KALMAN TIME STATE

The simulated time scale consists of five "clocks" perturbed by various levels of white noise FM and random walk noise FM. The clocks with the lowest white FM had the highest random walk FM, and conversely. The reference clock, clock #1, had the lowest level of white FM so that it would be the most stable clock between measurements. Figure 2 displays the theoretical Allan variances for the simulated clock data. Of course, it is impossible to generate continuous noise on a digital computer, but the noise was generated every 1/10th τ spacing (ten points per clock per measurement). Hence, the Allan Variances extend to 0.1τ units on the plots.

Each measurement cycle provides new information for the estimated Kalman state vector, and hence abrupt

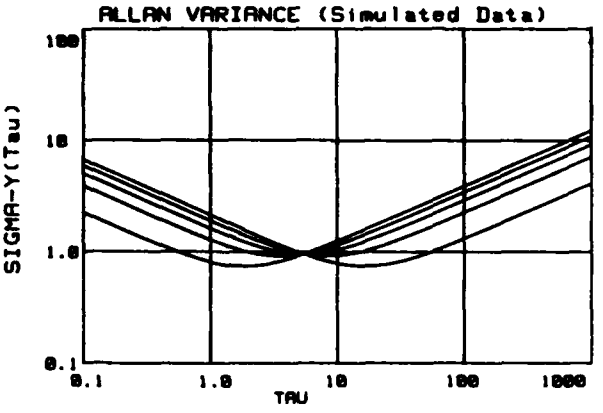


FIG. 2, THEORETICAL ALLAN VARIANCES OF CLOCKS

changes occur (see Fig. 1). Forecasts for the state vector prior to the next measurement are obtained from the frequency element of the estimated state vector. Since the forecast frequency elements of the state vector are constant between measurements, the forecasts in Fig. 1 are segments of straight lines. Of course, the simulation procedure allows one to observe the individual "clock" error, uncontaminated by the instabilities of the other clocks. The absolute time scale error for Fig. 1 is just the difference between the two curves. Clearly, for forecasts following the clock comparisons, the other clocks have essentially no affect on short term variations, and individual clock instabilities are whatever they are. Hence for best time scale resolution, the best reference clock is the most stable clock in short term.

It is not surprising that the abrupt changes in the estimated time error cause observable effects in the Allan Variances (see Fig. 3). Comparing Figs. 1 and 3 reveals that in short term clock #1 by itself is about 8 dB better than the Time-Kalman time scale. This analysis subsumes that the time scale should be based on all of the available data from the clock intercomparisons as provided by the Kalman Filter and its extrapolation up to the next measurement. In effect, this analysis assumes that TIME errors should be minimized regardless of the consequences on FREQUENCY stability.

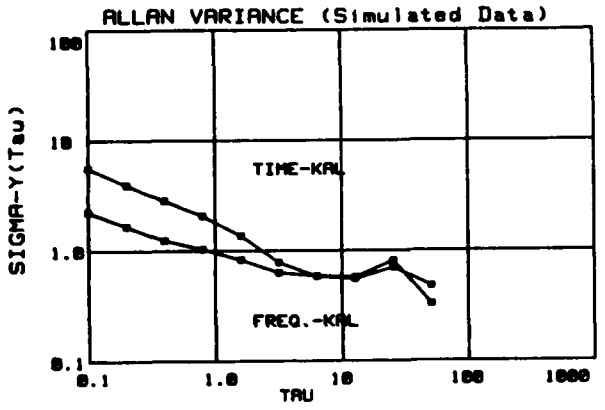


FIG. 3, ALLAN VARIANCES OF TIME SCALES

IV. Time Kalman vs. Frequency Kalman

There is a very important, fundamental and philosophical distinction to be made between time and frequency as it relates to atomic time scales and atomic clocks. Frequency is the fundamental quantity in atomic clocks. It can be well modeled and described by quantum mechanical processes, probability theory, and is the intrinsic natural process upon which the SI second is based as given by the definition for the cesium 133 atom. The energy state selection technique, signal-to-noise and other relevant electronic servo conditions, for example, determine the statistical properties and the averaging time necessary to measure a particular quantum state with a certain level of precision. In addition, because of the good models that have been developed one can evaluate the systematic uncertainties and develop a list of uncertainties associated with the realization of an absolute frequency per some natural resonance phenomenon in an atom. Because of the above, the General Conference of Weights and Measures (CGPM) in 1967 developed the current definition for the second -- as mentioned above. Independently different laboratories can realize the SI second without communication with one another strictly based on fundamental physical processes. When these clocks are compared they should agree within the uncertainty error budgets associated with each experimental determination of the SI second. This is currently the case between the four fundamental primary frequency standards which contribute to the determination of the rate of TAI. These standards are located at NBS, NRC, PTB and RRL, and are independent and agree to within their accuracies (less than or equal to about 1 part in 10^{13}).

In contrast it is impossible to do the same with time. The time from a time scale is typically dependent on three things: 1) an arbitrary origin (date) from whence seconds are accumulated; 2) the particular algorithm used to average a set of clocks used for this accumulation of seconds; and 3) a periodic frequency or rate calibration with either a primary standard or with a commercial standard that may be used in the composition of the scale. Therefore, the time so generated is an artifact of some arbitrary epoch (beginning point), plus some arbitrary algorithm and plus the integrated time errors from the frequency calibration inaccuracies. Whereas independent frequency standards will agree within some error budget, the time difference between independent time scales will be unbounded. We can speak of the natural resonance frequency of quantum mechanical transition, but the natural time of same has no meaning!

Therefore, as one applies Kalman filter theory given these philosophical and physical differences, the idea of optimizing the Kalman parameters around some natural resonance frequency has both intuitive and strong physical meaning; in contrast setting optimum Kalman parameters to minimize the time error is an artifact, is artificial and has no analogous physical meaning. Since the frequency Kalman gives better short-term stability for an arbitrary set of clocks and is apparently comparable to the time Kalman in long term, and since the frequency Kalman has a more sound physical basis the frequency Kalman would appear to be the better approach in the generation of atomic time scales. To date, no one is using this approach. The National Bureau of Standards is in the development process evaluating such an algorithm and doing experimentation to compare it with the current methods of generating time. Should the experiment corroborate the theory and above argument, the plans are to change the official algorithm for the generation of atomic time, TA(NBS), at the National Bureau of Standards to a frequency Kalman approach.

If a person's task were to synchronize a secondary clock to some primary clock, then the Time-Kalman approach would be better [6]. If frequency stability were the principal concern, then the Frequency-Kalman approach may be better.

V. Alternatives

Including the "TIME" Kalman filter discussed above, this paper considers four alternatives:

1. Time-Kalman, discussed above.

2. Linear slewing of the scale time from forecast to forecast as obtained using the Time-Kalman.

3. Forming a new time scale as the (discrete) integral of the frequency elements obtained from the Time-Kalman.

4. A complete re-casting of the Kalman model in terms of frequency rather than time. The underlying model of white FM and random walk FM would still be retained. The time scale, then, would be realized by integration of the estimated frequency elements of the new Frequency-Kalman filter. The measurement data would be exactly the first difference of the same time data used in the Time-Kalman (divided by τ).

Figure 1 depicts the type of instabilities one should expect from option 1, the simple Time-Kalman. Figure 4 uses the same data as for Fig. 1, but slews the correction to clock #1 from the

forecast made at the previous measurement to the forecast for the next measurement, τ seconds in advance. That is, after completing a current measurement, one can forecast the time element of the clock #1 state vector for the next measurement, τ seconds hence. Knowing what the last forecast was for the current point allows one to calculate a slewing rate to apply to the clock #1 data to reach the new correction value.

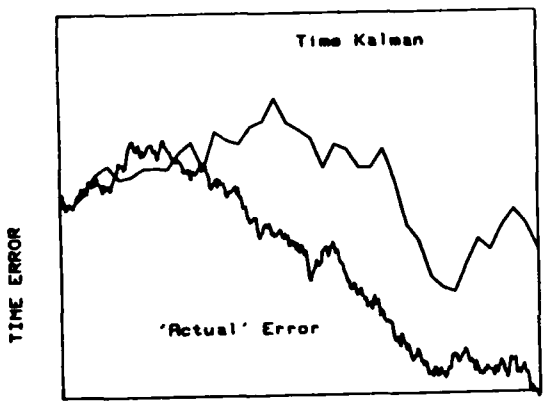


FIG. 4

SIMULATED CLOCK ERROR SLEWED TO NEXT FORECAST (Option 2)

The state vector for the ensemble of clocks includes the frequency element in addition to the time element discussed in the first two options. One can sum up these frequency corrections (and multiply by τ) for the clock with the best short-term stability (clock #1, for the current example). Figure 5 displays the results of these calculations for the same exact data as used for Figs. 1 and 4. Note that the Time-Kalman data and computations are all the same for options 1, 2, and 3. The differences are in what one does with the various elements of the state vector and the transition matrix, in making forecasts.

Option 4, above, is a more substantive change. Starting with the same clock model of white FM and random walk FM, one can show that a Frequency-Kalman model can be written in the form:

$$\begin{pmatrix} Y(t+\tau) \\ Z(t+\tau) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Y(t) \\ Z(t) \end{pmatrix} + \begin{pmatrix} \epsilon(t) \\ n(t) \end{pmatrix}$$

where Y(t) is the frequency element and Z(t) is a dummy variable needed for the model but without any obvious interpretation. Z(t) is essential for the generation of the random walk FM component.

The H(t)-matrix is the same as before since the first element is now frequency rather than time.

The measurement input data is now frequency rather than time. Each of the M-1 measurements is of the form:

$$W_n(t) = (X_n(t+\tau) - X_n(t)) / \tau$$

where $W_n(t)$ is the FREQUENCY measurement for the n-th clock deduced from the exact same data as used in the Time-Kalman. The Q(t) matrix is also unchanged except the factor τ is replaced by $1/\tau$.

As in option 3, option 4 uses the discrete summation of the frequency forecast of the state vector (multiplied by τ) to obtain an estimate of the clock's time error. Figure 6 displays the results of the Frequency-Kalman and integrator using the first difference of the same input data as in Figs. 1, 4, and 5. Figure 7 is an overlay on Fig. 2 of the frequency stability obtained using options 1 and 4 (Time-Kalman and Frequency-Kalman).

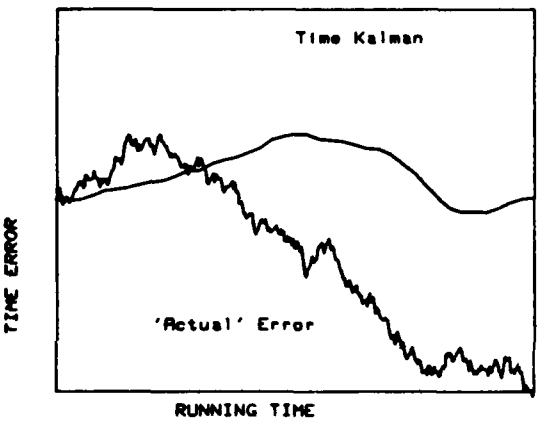


FIG. 5, SIMULATED CLOCK ERROR INTEGRATED FREQUENCY (Option 3)

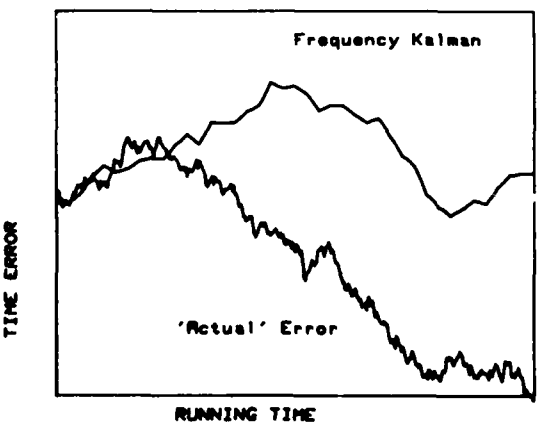


FIG. 6, SIMULATED CLOCK ERROR FOR FREQUENCY KALMAN (OPTION 4)

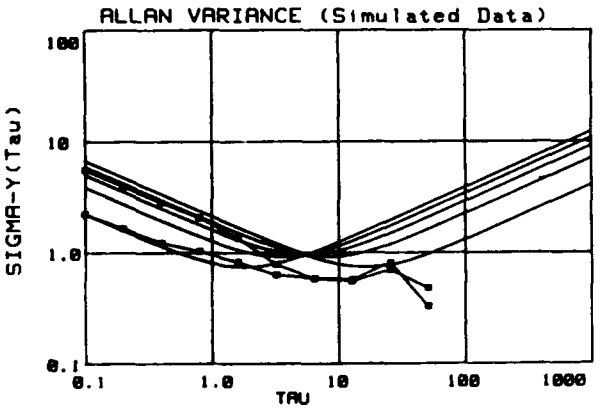


FIG. 7, THEORETICAL AND EXPERIMENTAL VARIANCES

VI. Comparisons of Time Scale Algorithms

In long-term there seems little to differentiate between the four algorithms, although there is a slight indication that the Frequency-Kalman (option 4) might be a fraction of a dB better than the others. It is clear that the various scales could depart asymptotically as $\tau^{3/2}$ from each other, but the same is true for any one of the time scales relative to some "ideal" scale. For short times ($< \tau$)the four options have about the relationships in Table 1:

TABLE 1.

Option	Short-Term Noise Level Above Frequency-Kalman
1. Time Kalman	+8 dB
2. Time-Kal (slew time)	+6.5 dB
3. Time-Kal (freq. integ)	+0 dB
4. Frequency-Kalman	0 dB

As a more important application of these Kalman algorithms, we applied them to a simulation of the NBS time scale. From other studies [Jones, et al.], we have good estimates of the white FM and random walk FM components of each clock. Table 2 lists standard deviations (σ's) for both noise types for each clock. The best clock in short-term was chosen as the reference clock, clock number 1. These values are in units of nanoseconds and correspond to a time interval of one day. To obtain values corresponding to one second, the sigmas for white FM should be divided by $(86400)^{1/2}$, and the random walk sigmas by $(86400)^{3/2}$.

TABLE 2. NBS Time Scale Parameters
(Units ns, Daily Basis)

Clock No.	White FM	Random Walk FM
1	0.5	0.55
2	2.8	0.84
3	0.6	0.83
4	9.1	3.0
5	9.9	1.7
6	9.4	1.9
7	14.3	0.86
8	11.4	2.0
9	4.7	0.55
10	2.3	0.55
11	11.4	2.1

Figure 8 represents the square root of the theoretical Allan variances corresponding to the parameters listed in Table 2. We simulated the noises for each of the eleven clocks and formed the differences between the reference clock (No. 1) and each of the other clocks corresponding to typical time scale data. This simulated data was used by the Time-Kalman and the Frequency-Kalman to define two (simulated) time scales. Since the data of each simulated clock was known, the absolute scale error for each simulated scale was calculated. (This is the value of simulation.) Figure 9 results from the calculated Allan variances for each simulated scale. Surprisingly, the turn-on transients for the time Kalman persisted for about one (simulated) month. The data for Figs. 9 and 10 come from sufficiently long averages to minimize the transients.

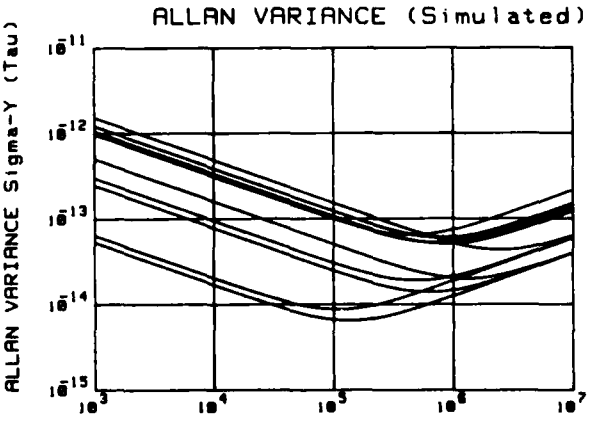


FIG. 8

THEORETICAL ALLAN VARIANCES BASED ON NBS CLOCKS

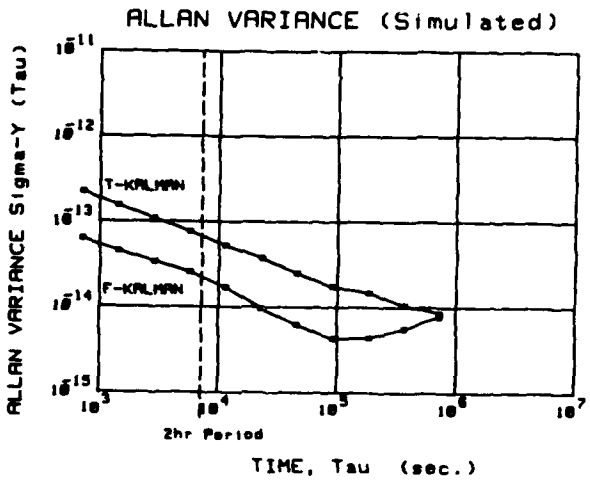


FIG. 9

OBSERVED VARIANCES FOR NBS TIME SCALE SIMULATION

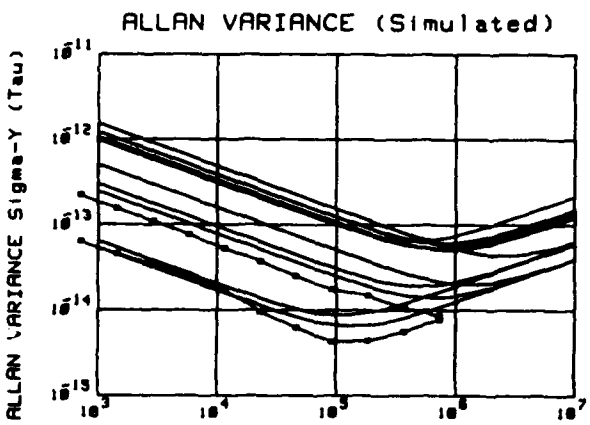


FIG. 10, NBS TIME SCALE SIMULATION

The upper curve in Fig. 9 corresponds to the typical time Kalman (option 1, above), while the lower curve corresponds to the frequency Kalman (option 4). Figure 10 superimposes Figs. 8 and 9. The integral of the frequency state of the Time-Kalman (option 3) is indistinguishable from the Frequency-Kalman (option 4).

The fundamental conclusion is that either of the Frequency-Kalman's (option 3 or 4) offers about a 10 dB improvement to the NBS time scale over the Time Kalman (option 1) for sample times shorter than a few days.

References

[1] R. H. Jones and P. V. Tryon, "Estimating Time From Atomic Clocks", Journal of Research of NBS, Vol. 88, No. 1, Jan-Feb. 1983, pp. 17-24.

[2] P. V. Tryon, and R. H. Jones, "Estimation of Parameter in Models for Cesium Beam Atomic Clocks", Journal of Research of NBS, Vol. 88, No. 1, Jan-Feb. 1983, pp. 3-16.

[3] J. A. Barnes, R. H. Jones, P. V. Tryon, and D. W. Allan, "Stochastic Models for Atomic Clocks", Proc. of 14th PTTI, Nov. 30-Dec. 2, 1982, pp. 295-306, NASA Goddard Space Flight Center, Greenbelt, MD.

[4] J. A. Barnes, "Time Scale Algorithms Using Kalman Filter--Insights From Simulation", Second International Symposium on Time Scale Algorithms, May 1982.

[5] A. Gelb, Applied Optimal Estimation, The MIT Press Cambridge, MA and London, England, Copyright 1974, The Analytic Science Corp.

[6] D. W. Allan, and J. A. Barnes, "Optimal Time and Frequency using GPS Signals", 36th Annual Frequency Control Symposium, June 2-4, 1982, pp. 378-387.

THE FRACTAL DIMENSION OF PHASE AND FREQUENCY NOISES : ANOTHER APPROACH TO OSCILLATOR CHARACTERIZATION

J.J. Gagnepain, J. Groslambert, R. Brendel

Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté - Besançon
32 avenue de l'Observatoire - 25000 Besançon - France

Summary

An evaluation is made of the possibility of applying Mandelbrot's concept of fractals to the characterization of noises and in particular of phase and frequency fluctuations. Two methods are considered. The first one consists in measuring with a gauge the apparent length of the signal : the fractal dimension of white and random walk noises are calculated with this method. The second one is based on the space filling ability of the signal. The method is applied to $1/f^{\alpha}$ noises, which are simulated on a computer, by superposition of lorentzian spectra, for integer and non integer values of the coefficient α. The dependence of the fractal dimension D as a function of the spectral density is established.

Introduction

The properties of a deterministic signal are in principle entirely known, and in particular such a signal is derivable, almost everywhere, and has a finite length. It obeys Euclide's geometry, and its dimension is just D = 1.

A stochastic signal does not follow the laws of the euclidean geometry, it cannot be described by continuous mathematical curves and its irregularities still are present even when examining with higher and higher resolution. This leads to the paradoxical conclusion : the length of this signal is not finite. The degree of irregularity can intuitively be related to its space filling ability. A white noise almost completely fills up the plane on which it is recorded. This means that its dimension rather will correspond to the dimension D = 2 of a surface. Mandelbrot[1] suggested that the degree of irregularity of a curve, or a surface, etc, can be characterized in term of a fractal dimension. For a one-dimension object the value D will be between one and two. This concept can be generalized to n-dimension objects, and in that case the fractal dimension will be such that $n < D < n+1$, i.e. the fractal dimension always is higher than the corresponding euclidean one.

The length L of a fractal signal depends on the resolution with which it is measured, i.e. on the gauge λ of the measure. Following Richardson law[2] : $L \propto \lambda^{1-D}$, where D is the Hausdorff-Mandelbrot fractal dimension. Thus measuring $L(\lambda)$ versus λ is a way to determine D.

The phase fluctuations of the output signal of an oscillator are characterized by their power spectral density $S_{\Phi}(f)$, which can be written as a series of the powers f^{α} of the Fourier frequency f. Therefore the first goal of the present work will be to determine the relation between the fractal dimension D and the exponents α of the individual spectra. This can be done by analytical calculation and/or by computer simulation.

Calculation of white and random walk noises

Following directly from Richardson law the fractal dimension D can be obtained by calculating the length L of the signal over a given period of time T as a function of a gauge λ, as shown in Fig. 1. This is done by putting straight segments λ along the curve and counting their total number N over the interval T.[3]

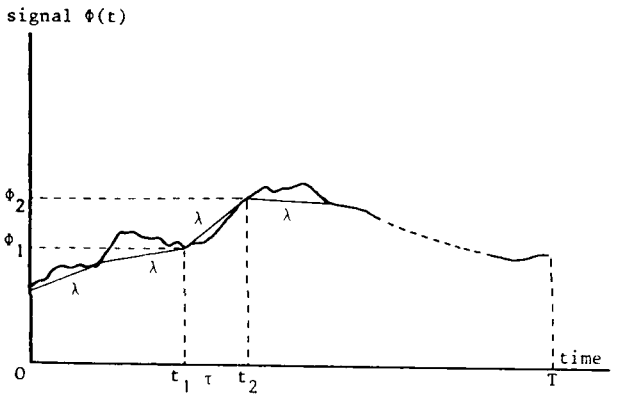


Fig. 1 : Measurement of the length $L(\lambda)$ of the signal from t = 0 to t = T by using a segment λ as a gauge

Since $L = N\lambda$ and from Richardson $L \propto \lambda^{1-D}$, thus

$$N \propto \lambda^{-D} \qquad (1)$$

The projector τ of a given segment λ, linking the points Φ_1 and Φ_2 corresponding to the values of the signal $\Phi(t)$ at $t = t_1$ and $t = t_2$, is a random variable which can be characterized by its ensemble average $\langle\tau\rangle$. The total number N of segments λ between 0 and T on the curve is

$$N = T / \langle\tau\rangle \qquad (2)$$

T will be normalized to 1.

Let $f(\tau)$ be the probability density of τ. τ is a function of the random variables Φ_1 and Φ_2

$$\tau^2 = \lambda^2 - (\Phi_2 - \Phi_1)^2 \qquad (3)$$

Therefore

$$f(\tau)\, d\tau = \iint_{\Delta D} f(\Phi_1,\Phi_2)\, d\Phi_1\, d\Phi_2 \qquad (4)$$

where $f(\Phi_1,\Phi_2)$ is the joint probability density of the variables Φ_1 and Φ_2, and ΔD is a domain in the Φ_1,Φ_2 plane such that $\tau <$ variable $\tau < \tau + d\tau$ when $(\Phi_1,\Phi_2) \in \Delta D$.

CH2186-0/85/0000-0113$1.00©1985IEEE

Considering a stationary signal $\Phi(t)$ with a gaussian distribution, the joint density is

$$f(\Phi_1,\Phi_2) = \frac{1}{2\pi\sigma^2\sqrt{1-r^2}} e^{-\frac{\Phi_1^2 - 2r\Phi_1\Phi_2 + \Phi_2^2}{2\sigma^2(1-r^2)}} \quad (5)$$

$\Phi(t)$ is caracterized by its correlation function $R(\tau)$. The standard deviation σ is given by

$$\sigma^2 = R(0) \quad (6)$$

and r corresponds to

$$r = \frac{R(\tau)}{R(0)} \quad (7)$$

Introducing new variables X_1 and X_2 given by the transformation

$$X_1 = \frac{\Phi_2 + \Phi_1}{\sqrt{2}} \qquad X_2 = \frac{\Phi_2 - \Phi_1}{\sqrt{2}} \quad (8)$$

the joint probability density takes the simpler form

$$f(X_1,X_2) = \frac{1}{2\pi\sigma^2\sqrt{1-r^2}} e^{-\frac{X_1^2}{2\sigma^2(1+r)}} e^{-\frac{X_2^2}{2\sigma^2(1-r)}} \quad (9)$$

and

$$f(\tau)\, d\tau = \iint_{\Delta D'} f(X_1,X_2)\, dX_1\, dX_2 \quad (10)$$

The $\Delta D'$ domain is represented in Fig. 2. Performing the integration (10) over $\Delta D'$ yields

$$f(\tau)\, d\tau = \frac{e^{-\frac{\lambda^2-\tau^2}{4\sigma^2(1-r)}}}{\sigma\sqrt{\pi}\sqrt{1-r}}\, \tau\, d\tau \quad (11)$$

The mean value of τ directly follows

$$\langle\tau\rangle = \int_0^\lambda \tau\, f(\tau)\, d\tau \quad (12)$$

Introducing $u^2 = \lambda^2 - \tau^2$, finally $\langle\tau\rangle$ is given by

$$\langle\tau\rangle = \int_0^\lambda \frac{\sqrt{\lambda^2-u^2}}{\sigma\sqrt{\pi}\sqrt{1-r}} e^{-\frac{u^2}{4\sigma^2(1-r)}} du \quad (13)$$

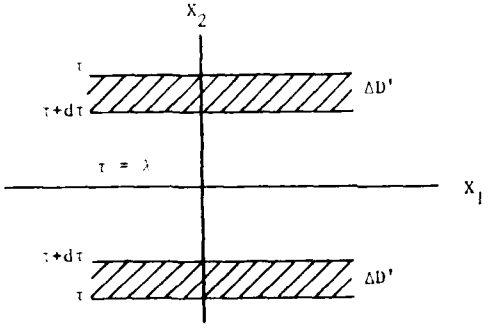


Fig. 2 : Integration domain of the random variables X_1 and X_2

White noise

White noise corresponds to independent stochastic samples Φ_1 and Φ_2 which are not correlated, and therefore to r = 0. The integral (13) cannot be exactly calculated analytically, but asymptotic values can be obtained

- for $\lambda \gg \sigma$, the integral reduces to $\int_{-\infty}^{+\infty} \sqrt{\lambda^2-u^2}\, \delta(u)\, du$, where $\delta(u)$ is the Dirac function. This gives $\langle\tau\rangle = \lambda$ and $N = \lambda^{-1}$, which corresponds to D = 1.

- for $\lambda \ll \sigma$, the integral reduces to $\int_0^\lambda \frac{\sqrt{\lambda^2-u^2}}{\sigma\sqrt{\pi}} du$, thus $\langle\tau\rangle = \frac{\sqrt{\pi}}{4\sigma}\lambda^2$ and $N = \frac{4\sigma}{\sqrt{\pi}}\lambda^{-2}$ which gives D = 2.

In the first case the gauge is larger than the noise amplitude, and therefore only the mean value of the noise is measured, leading to D = 1, as for a deterministic curve. In the second case the noise fills up the entire measurement space, and its fractal dimension goes to the topological dimension of a surface D = 2.

Random walk noise

Random walk noise is obtained by filtering white noise with a first order filter. It corresponds to a lorentzian spectrum $S(\omega) = \frac{h_0\,\omega_c}{\omega_c^2 + \omega^2}$, where ω_c is the *cut-off frequency*. The correlation function has the well known form

$$R(\tau) = \sigma^2 e^{-\tau/\tau_c} \quad (14)$$

with $\sigma^2 = h_0\omega_c/2$ et $\tau_c = 1/\omega_c$.

The solutions of equation (13) again are asymptotically calculated, with in this case $r = e^{-\tau/\tau_c}$.

- for $\lambda \gg \sigma$ the result is D = 1, as in the case of white noise anf for the same reason.

- for $\sigma > \lambda > \tau_c$ the fluctuations are measured, but the gauge λ is larger than the correlation time τ_c ; as a consequence the samples Φ_1 and Φ_2 are independent and will behave as white noise. The integral is approximated with r = 0 and yields

$$\langle\tau\rangle = \frac{\sqrt{\pi}}{4\sigma}\lambda^2 \quad \text{and} \quad D = 2 \quad (15)$$

- for $\lambda < \tau_c < \sigma$, the quantity $e^{-u^2/4\sigma^2(1-r)}$ remains close to 1 over the interval $[0,\lambda]$, and $\sqrt{1-r} \approx \sqrt{\lambda/\tau}$, with these approximations

$$\langle\tau\rangle = \frac{\lambda^{3/2}}{4\sigma\sqrt{2f_c}} \quad \text{and} \quad D = 1{,}5 \quad (16)$$

These results are summarized in Fig. 3.

Computer simulation

These results were generalized to different noises with a $1/f^\alpha$ spectrum, where the characteristic coefficient α is between 0 and 4, and does not necessarily has to be an integer.

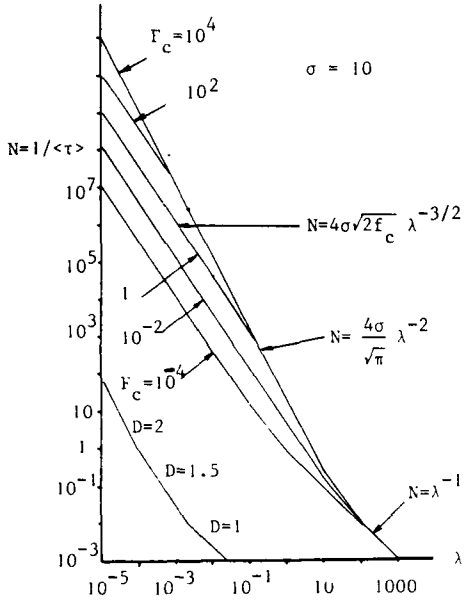


Fig. 3 : Fractal dimension of random walk noise as a function of the cut-off frequency $f_c = 1/2\pi\tau_c$ and for $\sigma = 10$

$1/f^{\alpha}$ spectrum

The $1/f^{\alpha}$ noise is generated by superposition of random walk noises with a given distribution of their cut-off frequencies, as shown by the block diagram in Fig. 4. Four elementary lorentzian spectra were used in the construction (Fig. 5), giving the final spectrum over a frequency range corresponding to 3.5 decades. The power spectral densities of the original white noise and of f^{-3} noise are presented in Fig. 6a and 6b (the calculation of the spectrum was limited to 2.5 decades by the available Fourier transform program).

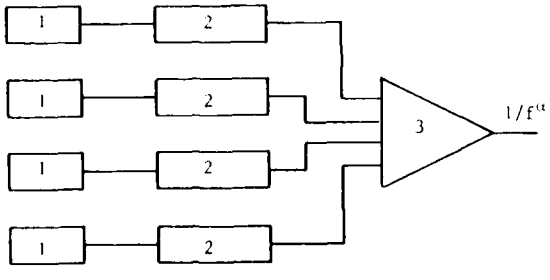


Fig. 4 : Generation of $1/f^{\alpha}$ noise.
1 - random number generators with a white spectrum ;
2 - First order digital filters with a distributed cut-off frequency ω_c ; 3 - adder

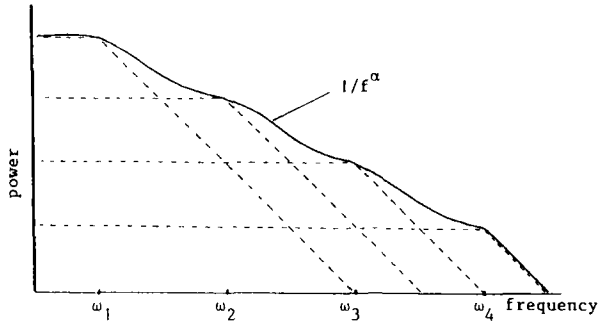


Fig. 5 : Superposition of lorentzian spectra with ω_1, ω_2, ω_3, ω_4 cut-off frequencies giving a $1/f^{\alpha}$ spectrum over the frequency range $[\omega_1, \omega_4]$

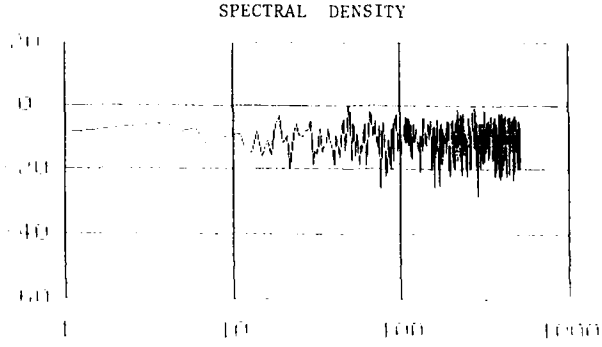


Fig. 6a : White noise spectrum

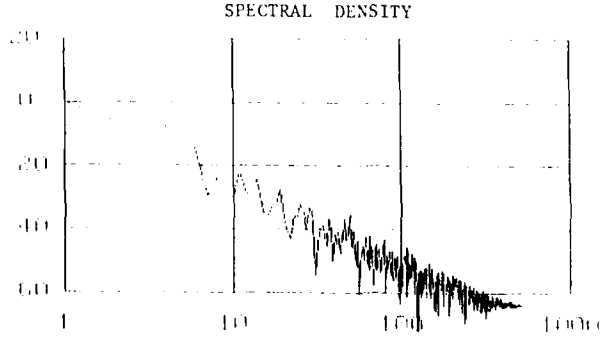


Fig. 6b : $1/f^3$ noise spectrum

Determination of the fractal dimension

The fractal dimension D can be measured by using the length method previously presented. However this method takes quite a long time. Therefore another method, called the "reticular cell counting" method (RCC) has been used.[4]

Starting from an initial square area, normalized to 1, which contains M samples of the signal (Fig. 7), the method consists in dividing the initial square into four subsquares, and then each of the four subsquares into four subsquares, and so on. After n such iterations the initial square of unit side contains 2^{2n} subsquares with side length $\lambda = 2^{-n}$. Let N be the number of squares containing at least one sample of the signal. It can be easily seen that the area S covered by the nonempty squares is $S = N\lambda^2$. Since $S \to L\lambda$ (L is the length of the curve) when the size of the squares becomes to be very small, i.e. when $\lambda \to 0$, and using $L = \lambda^{1-D}$, it appears that in this method N follows also the law given by eq. (1)

$$N \propto \lambda^{-D} \tag{17}$$

which allows one to determine D.

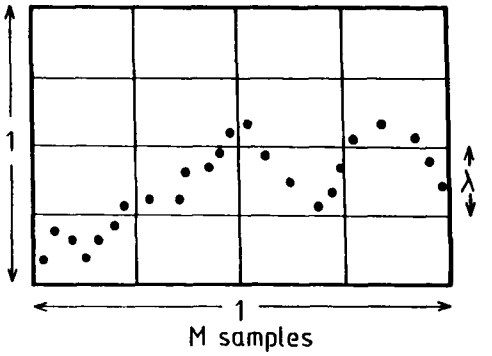


Fig. 7 : Principle of the RCC method used for measuring the fractal dimension

When λ is small enough, for instance when $\lambda < \lambda_c$, or $n > n_c$ with $(\lambda_c)^D = 1/M$ and $n_c = \text{Log } M/D \text{ Log } 2$, each sample will be in a different square (there will not be more than one sample per filled square), N will remain constant and equal to M. This gives the lower asymptot of the $N(\lambda)$ curve. The second asymptot corresponds to relation (17) and its slope gives the fractal dimension (Fig. 8). The total curve $N(\lambda)$ can be approximated by the relation (this is a phenomenological fitting)

$$N = \frac{M}{1 + (\frac{\lambda}{\lambda_c})^D} \tag{18}$$

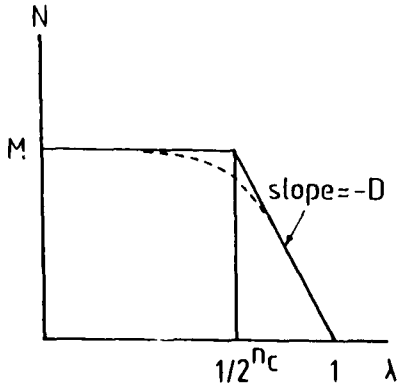


Fig. 8 : Asymptots of the N versus λ curve (in a log-log scale)

The total area covered by the nonempty squares is

$$S = N\, 2^{-2n} \tag{19}$$

2^{-2n} being the area of an elementary square after the nth iteration. Using (17) into (19) gives

$$D = 2 + \frac{\text{Log } S}{n \text{ Log } 2} \tag{20}$$

With white noise, the entire initial unit square is filled up. Therefore S = 1 and D = 2 as expected.

Using the relation $S = L\lambda$, eq. (20) can be written

$$D = 1 + \frac{\text{Log } L}{n \text{ Log } 2} \tag{21}$$

Since for a deterministic signal, the length is finite, Log L/n Log 2 → 0 for n large enough, and D → 1.

Fractal dimension and spectral density

Noises with $1/f^\alpha$ spectrum were generated, by means of the lorentzian spectrum superposition, for α = 0, 1, 2, 3, 4 corresponding to the most encountered noises, and also for the fractional values α = 0.5, 1.5, 2.5 and 3.5. The corresponding $N(\lambda)$ plots are given in Fig. 9 and 10. 8192 samples were taken in the random number sequences. The calculation of D was made by using relation (18).

In Fig. 11 is given the fractal dimension D as a function of the noise spectrum coefficient α. The relation $D(\alpha)$ is linear for $0 < \alpha < 1.5$. For large values of α, the dimension D asymptotically tends to 1.

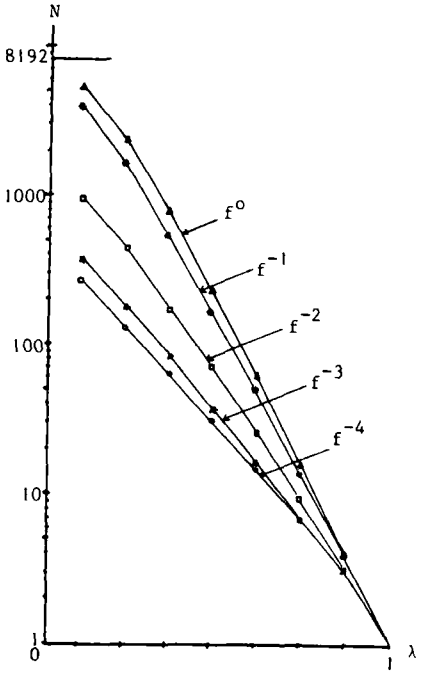


Fig. 9 : Measurement of the fractal dimension of $1/f^\alpha$ noises, for α = 1,2,3,4 by means of the RCC method

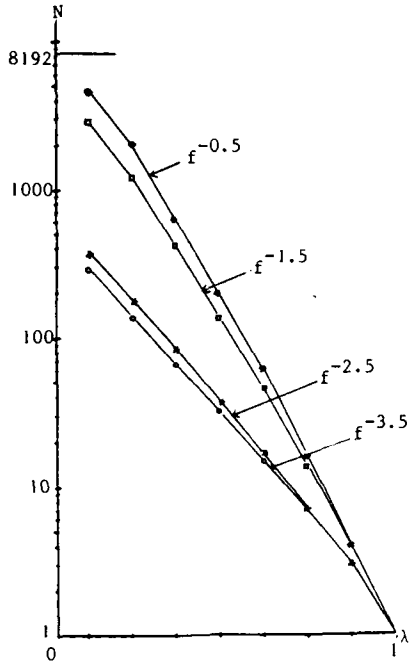


Fig. 10 : Measurement of the fractal dimension of $1/f^{\alpha}$ noises, for α = 0.5, 1.5, 2.5, 3.5 by means of the RCC method

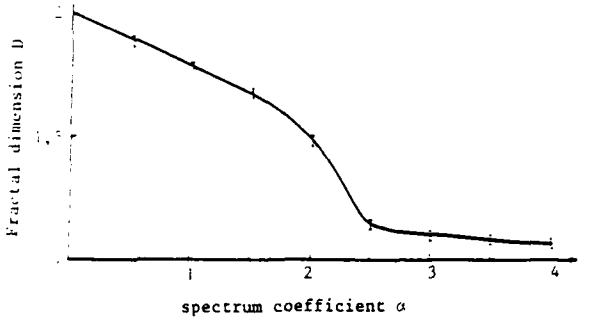


Fig. 11 : Fractal dimension D of different noises as a function of the coefficient α of their power spectral density $1/f^{\alpha}$

Random walk noise with white noise

When a signal contains noises of different types, it is important to see how it is possible to distinguish between these different contributions, if one fractal dimension can be related to each one, or rather if one single dimension is found as a combination of the individual ones.

A white noise was added to a $1/f^2$ random walk noise, as shown in Fig. 12. The corresponding $N(\lambda)$ curves are given in Fig. 13 for two levels of the white noise. This shows that the fractal dimensions of the two noises can be extracted (the 0 dB reference level corresponds to white noise and $1/f^2$ noise at the same level for f = 1 Hz).

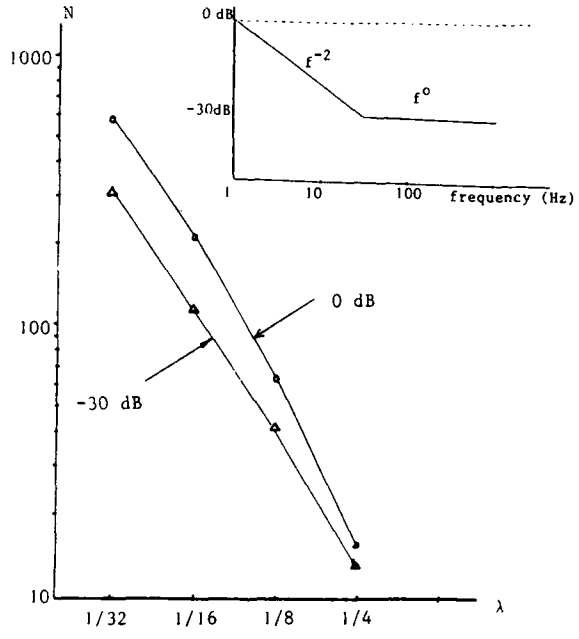


Fig. 12 : Random walk noise with white noise

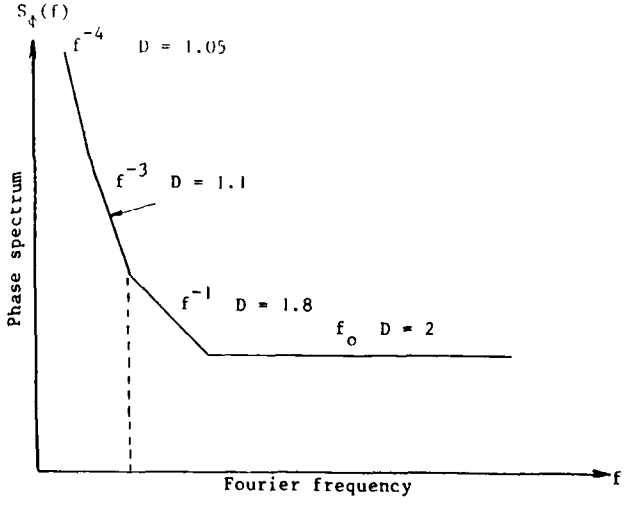


Fig. 13a : Frequency domain

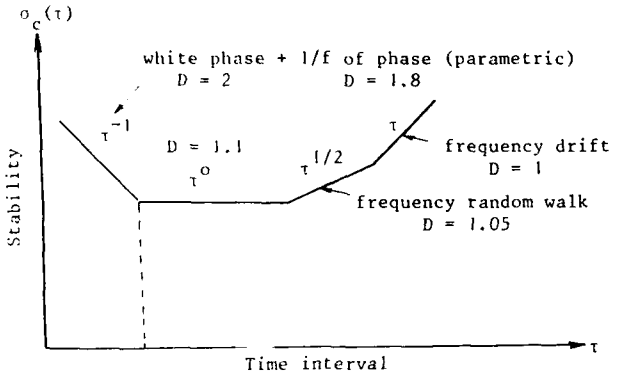


Fig. 13b : Time domain

Conclusion

The different noise contributions to the phase, or frequency, fluctuations of the output signal of an oscillator therefore can be identified by means of the fractal approach. As summarized in Fig. 13a and 13b, corresponding respectively to frequency domain and to time domain, a fractal dimension is associated to each noise. In particular, this is a possibility to distinguish between flicker phase and white phase noises, which usually cannot be separated in time domain since their Allan variance exhibits the same characteristic slope.

At the present state, the fractal approach is more qualitative than quantitative, and the efficiency of the methods for measuring the dimension D still have to be precized and their dynamical range must be improved.

These results could be generalized to a set of oscillators, using the interesting property of the fractal dimension to be just immediatly larger than the topological dimension of the space which is considered.

References

[1] B.B. Mandelbrot, The fractal geometry of nature. Freeman and Co. New-York, 1983.

[2] L.F. Richardson, General Systems Years Book, 6, 1961.

[3] B.H. Kaye, Multifractal description of a rugged fine particle profile, Particle Characterization 1, 1984.

[4] J.J. Gagnepain, C. Roques-Carmes, "Fractal approach to 2D and 3D surface roughness", 3rd Int. Conf. on Metrology and Properties of Engineering Surfaces, Teesside Polytechnic, Middlesborough, England, April 1985.

FREQUENCY STABILITY CHARACTERIZATION FROM THE FILTERED SIGNAL OF A PRECISION OSCILLATOR

by

Pierre Tremblay and Michel Têtu
Laboratoire de Recherche sur les Oscillateurs et Systèmes
Département de génie électrique
Université Laval, Québec, G1K 7P4, Canada

Abstract: This paper presents the evaluation of the frequency stability characterization of an oscillator when its signal is directly filtered instead of its phase as is usually considered in the development of the standard theory. The amplitude and the phase fluctuations of the filtered signal are expressed as functions of filter impulse response and input signals characteristics. The general expressions for the power spectral density of the amplitude and phase fluctuations of the filtered signal are then calculated. The results obtained for the phase fluctuations of the filtered signal are used to characterize the frequency stability of the oscillator. Every filter will have a particular effect on the measured phase fluctuations and we study some commonly used filters. For each filter, we give the contribution of white additive noise to the Allan variance and to the modified Allan variance (white phase noise). An experimental verification for the Allan variance in presence of additive noise shows an excellent agreement with the theoretical predictions.

Introduction

Barnes et al. [1], in their paper entitled "Characterization of Frequency Stability", have brought the major policies guiding the field of frequency stability measurement for about twenty-five years. Among other things, they wrote about time domain frequency stability measurement: "Good practice, however, dictates that the system noise bandwidth f_h should be specified with any results." Until now no means were provided to help the experimentalist performing such a measurement to evaluate this bandwidth. One was obliged to suppose a rectangular low-pass [2], [3] or a first-order low-pass [4]-[6] phase filtering in order to specify the experimental conditions and to fully characterize the oscillator under test. These two types of phase filtering are different from most of the experimental setups used to measure the frequency stability. This work presents an evaluation of the effects encountered in frequency stability characterization when a filter is directly applied to the signal of a precision oscillator. We consider the case where the oscillator signal is perturbed by internal noise and by additive noise and where the resulting signal is passed through a linear filter. The amplitude and the phase fluctuations of the filtered signal are expressed as functions of filter impulse response and input signals characteristics. The general expressions for the power spectral density of the amplitude and phase fluctuations of the filtered signal are then calculated.

A preliminary version of this work has been presented to the 1984 CPEM in Delft, The Netherlands [7].

The results obtained for the phase fluctuations of the filtered signal are used to characterize the frequency stability of the oscillator. This characterization is done either in the frequency domain by a power spectral density measurement or in the time domain by a variance measurement. A given filter will have a particular effect on the measured phase fluctuations and we study two low-pass filters; a first-order and a Nth-order Butterworth, two band-pass filters; a first-order and a Nth-order Butterworth and a second-order resonant bandpass filter. For each filter, we give the contribution of white additive noise to the Allan variance and to the modified Allan variance. However, contributions from other types of amplitude and phase fluctuations such as flicker noise or random walk could be calculated using the equations provided through this paper. These calculations will be given in ref. [8]. Experimental verification for the Allan variance is given when the oscillator signal is in presence of additive noise and when the resulting signal is filtered by a first-order and a eighth-order band-pass filter.

Filtering of a Noisy Sinusoidal Signal

Description of the Model

The model used and the calculations done in this paper will be presented extensively in reference [8]. Fig. 1 gives the block diagram representing our model. The oscillator signal is expressed by the following relation:

$$s_o(t) = A_o\,[1+\varepsilon_o(t)]\,\cos[2\pi\nu_o t+\varphi_o(t)+\theta_o], \qquad (1)$$

where A_o is the oscillator mean amplitude, $\varepsilon_o(t)$ is its relative amplitude fluctuations, ν_o is its mean frequency, $\varphi_o(t)$ is its phase fluctuations and $\theta_o(t)$ is its initial phase, which is uniformly distributed. The relative amplitude and phase fluctuations are two zero-mean wide-sense stationary random processes which are mean square continuous.

Moreover the oscillator signal is perturbed by an additive noise which is decomposed as a sum of two parts: an in phase component and an in quadrature component. The additive noise, n(t), is a zero-mean wide-sense stationary random process and is mean square continuous. This noise is then expressed as:

$$n(t) = p(t)\cos[2\pi\nu_o t+\theta_o] - q(t)\sin[2\pi\nu_o t+\theta_o], \qquad (2)$$

where the amplitudes of the two components, p(t) and q(t), are also zero-mean wide-sense stationary random processes, they are mean square continuous and uncorrelated with the oscillator relative amplitude and phase fluctuations. These two amplitudes have the same power spectral density, which is the symmetrical part

CH2186-0/85/0000-0119$1.00©1985IEEE

around the oscillator mean frequency of the power spectral density of the additive noise:

$$S_p(f) = S_q(f) = \tfrac{1}{2}\,[S_n(\nu_o+f) + S_n(\nu_o-f)]. \qquad (3)$$

These amplitudes are also correlated and their cross-power spectral densities are related to the antisymmetrical part around the oscillator mean frequency of the noise power spectral density:

$$-S_{q,p}(f) = S_{p,q}(f) = i\,\tfrac{1}{2}\,[S_n(\nu_o+f) - S_n(\nu_o-f)]. \qquad (4)$$

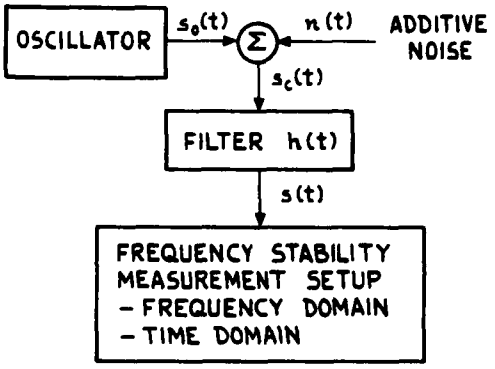


Fig. 1 - Schematic diagram of the model studied.

The composite signal resulting from the sum of the oscillator signal and the additive noise could be expressed under a form similar to the oscillator signal. At this point we must make the following hypothesis:

$$E\{\varepsilon_o^2(t)\} \ll 1\ ,\ E\{\varepsilon_o(t)\varphi_o(t)\} \ll 1 \text{ rad},$$

$$E\{\varphi_o^2(t)\} \ll 1 \text{ rad}^2,\ E\{p^2(t)\} = E\{q^2(t)\} \ll A_o^2$$

$$\text{and } E\{p(t)q(t)\} \ll A_o^2. \qquad (5)$$

The composite signal becomes:

$$s_c(t) = A_o\,[1+\varepsilon_c(t)]\cos[2\pi\nu_o t+\varphi_c(t)+\theta_o], \qquad (6)$$

where $\varepsilon_c(t)$ is its relative amplitude fluctuations and $\varphi_c(t)$ is its phase fluctuations. These processes are given by the following equation:

$$\begin{bmatrix}\varepsilon_c(t)\\ \varphi_c(t)\end{bmatrix} = \begin{bmatrix}\varepsilon_o(t)\\ \varphi_o(t)\end{bmatrix} + \frac{1}{A_o}\begin{bmatrix}p(t)\\ q(t)\end{bmatrix}, \qquad (7)$$

and have the same properties than $\varepsilon_o(t)$ and $\varphi_o(t)$. The amplitude of the additive noise in phase component contributes to the amplitude fluctuations while the amplitude of the in quadrature component contributes to the phase fluctuations.

The composite signal is passed through a linear time-invariant filter which is stable, causal and have a real continuous impulse response, h(t). The filter transfer function, H(f), is the Fourier transform of the filter impulse response.

Signal Filtering

The filtered signal is given by the convolution between the composite signal and the filter impulse response (represented by ⊛),

$$s(t) = \int_0^\infty h(\tau)\, s_c(t-\tau)\, d\tau, \qquad (8)$$

and is expressed in form similar to the oscillator signal:

$$s(t) = A\,[1+\varepsilon(t)]\cos[2\pi\nu_o t+\varphi(t)+\theta]. \qquad (9)$$

In this expression, A is the filtered signal amplitude, $\varepsilon(t)$ is its relative amplitude fluctuations, $\varphi(t)$ is its phase fluctuations and θ is its initial phase. Under the conditions described in eq. (5), the random processes $\varepsilon(t)$ and $\varphi(t)$ are found to be zero-mean if

$$A = A_o\,\alpha(\nu_o) \quad \text{and} \quad \theta = \theta_o + \Psi(\nu_o), \qquad (10)$$

where the filter transfer function has been decomposed as follow,

$$H(f) = \alpha(f)\exp[i\,\Psi(f)]. \qquad (11)$$

The relative amplitude and phase fluctuations of the filtered signal are also wide-sense stationary and mean square continuous. From eqs (8), (6), (9) and (10), these two random processes are found to be functions of the composite signal relative amplitude and phase fluctuations:

$$\begin{bmatrix}\varepsilon(t)\\ \varphi(t)\end{bmatrix} = \begin{bmatrix}h_f(t) & h_c(t)\\ -h_c(t) & h_f(t)\end{bmatrix} \circledast \begin{bmatrix}\varepsilon_c(t)\\ \varphi_c(t)\end{bmatrix}, \qquad (12)$$

where $h_f(t)$ is the direct filtering impulse response and $h_c(t)$ is the conversion filtering impulse response. They are respectively equal to

$$h_f(t) = h(t)\cos[2\pi\nu_o t+\Psi(\nu_o)] \,/\, \alpha(\nu_o) \qquad (13)$$

and

$$h_c(t) = h(t)\sin[2\pi\nu_o t+\Psi(\nu_o)] \,/\, \alpha(\nu_o). \qquad (14)$$

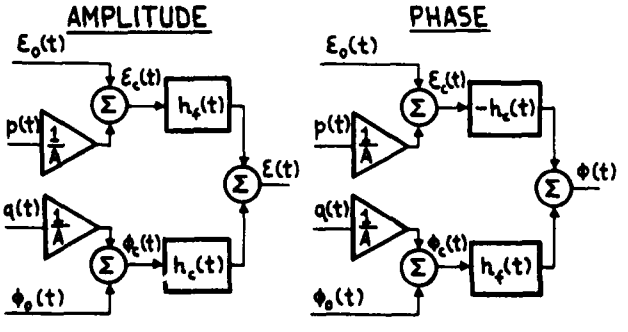


Fig. 2 - Block diagram of the signal filtering effect on the relative amplitude and phase fluctuations.

The Fig. 2 shows the signal filtering effect on the amplitude and the phase fluctuations respectively. The relative amplitude fluctuations of the filtered signal are a sum of two components. The

first is the relative amplitude fluctuations of the composite signal passed through the direct filtering and the second is the phase fluctuations of the composite signal passed through the conversion filtering. The phase fluctuations of the filtered signal are also expressed as the sum of two components. The first is the relative amplitude fluctuations of the composite signal passed through the conversion filtering and the second is the phase fluctuations of the composite signal passed through the direct filtering.

The direct filtering then relates the amplitude and phase fluctuations of the composite signal to the amplitude and phase fluctuations of the filtered signal respectively (AM-AM or PM-PM). Therefore the conversion filtering changes respectively the amplitude and phase fluctuations into phase and amplitude fluctuations (AM-PM or PM-AM).

The two random processes characterizing the filtered signal are then related to the noise added to the oscillator signal, to the oscillator signal relative amplitude and phase fluctuations, to the oscillator signal mean amplitude and mean frequency and to the filter impulse response.

The correlation function of the vector defined in eq. (12) is then calculated:

$$\begin{bmatrix} R_\varepsilon(\tau) & R_{\varepsilon,\varphi}(\tau) \\ R_{\varphi,\varepsilon}(\tau) & R_\varphi(\tau) \end{bmatrix} = \begin{bmatrix} h_f(\tau) & h_c(\tau) \\ -h_c(\tau) & h_f(\tau) \end{bmatrix} \otimes \begin{bmatrix} R_{\varepsilon c}(\tau) & R_{\varepsilon c,\varphi c}(\tau) \\ R_{\varphi c,\varepsilon c}(\tau) & R_{\varphi c}(\tau) \end{bmatrix} \otimes \begin{bmatrix} h_f(-\tau) & -h_c(-\tau) \\ h_c(-\tau) & h_f(-\tau) \end{bmatrix}, \quad (15)$$

where $R_x(\tau)$ denotes the correlation function of the process $x(t)$ and $R_{x,y}(\tau)$ stands for the cross-correlation function between the two processes $x(t)$ and $y(t)$.

By taking the Fourier transform on each side of the eq. (15), the power spectral density of the vector-valued fluctuations of the filtered signal is obtained:

$$\begin{bmatrix} S_\varepsilon(f) & S_{\varepsilon,\varphi}(f) \\ S_{\varphi,\varepsilon}(f) & S_\varphi(f) \end{bmatrix} = \begin{bmatrix} H_f(f) & H_c(f) \\ -H_c(f) & H_f(f) \end{bmatrix} \cdot \begin{bmatrix} S_{\varepsilon c}(f) & S_{\varepsilon c,\varphi c}(f) \\ S_{\varphi c,\varepsilon c}(f) & S_{\varphi c}(f) \end{bmatrix} \cdot \begin{bmatrix} H_f^*(f) & -H_c^*(f) \\ H_c^*(f) & H_f^*(f) \end{bmatrix}, \quad (16)$$

where $S_x(f)$ represents the power spectral density of the random process $x(t)$ and $S_{x,y}(f)$ denotes the cross-power spectral density between the random processes $x(t)$ and $y(t)$. In this equation, $H_f(f)$ is the direct filtering transfer function and $H_c(f)$ is the conversion filtering transfer function. These two transfer functions are respectively the Fourier transforms of eqs (13) and (14) and are given by the following relations:

$$H_f(f) = \frac{1}{2}\left[\frac{H(\nu_o+f)}{H(\nu_o)} + \frac{H^*(\nu_o-f)}{H^*(\nu_o)}\right] \quad (17)$$

and

$$H_c(f) = \frac{i}{2}\left[\frac{H(\nu_o+f)}{H(\nu_o)} - \frac{H^*(\nu_o-f)}{H^*(\nu_o)}\right]. \quad (18)$$

The direct filtering transfer function corresponds to the symmetrical part around the oscillator mean frequency of the filter transfer function, while for the conversion filtering transfer function, it is its antisymmetrical part.

Frequency Stability Characterization

Frequency Domain

The characterization of frequency stability in the frequency domain is defined as the power spectral density of the relative frequency fluctuations, $y(t)$ [1]. These fluctuations are proportional to the time derivative of the phase fluctuations. Under the condition that the second derivative of the correlation function of the phase fluctuations is continuous at $\tau = 0$, the mean square derivative of $\varphi(t)$ exists and is also zero-mean and wide-sense stationary. The power spectral density of the relative frequency fluctuations is then directly related to the power spectral density of the phase fluctuations:

$$S_y(f) = \frac{f^2}{\nu_o^2} S_\varphi(f). \quad (19)$$

The general expression for the power spectral density of the phase fluctuations is obtained from eq. (16):

$$S_\varphi(f) = |H_c(f)|^2 S_{\varepsilon c}(f) + |H_f(f)|^2 S_{\varphi c}(f) - 2\,\mathrm{Re}\{ H_f(f)\, S_{\varphi c,\varepsilon c}(f)\, H_c(f) \}. \quad (20)$$

It is shown that this power spectral density has three contributions: the converted relative amplitude fluctuations of the composite signal, its filtered phase fluctuations and a third term accounting for the cross-correlation between these two processes. Such a cross-correlation could results from previous conversion through linear filters or from non-linearities.

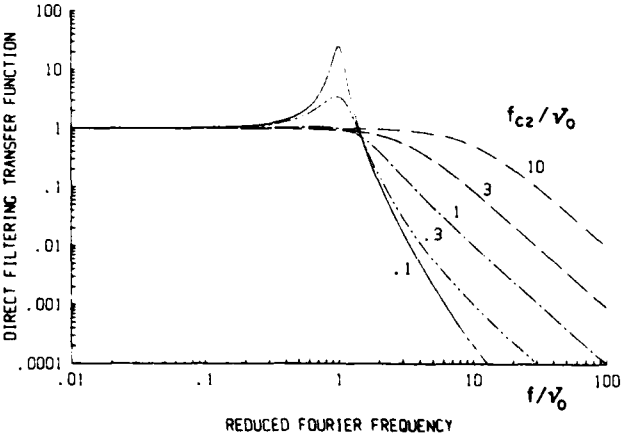


Fig. 3 - Direct filtering transfer function for a first-order low-pass filter of different cutoff frequencies.

In order to illustrate the shape of the direct filtering and the conversion filtering transfer

functions, they are respectively presented in Figs 3 and 4 for first-order low-pass filters of different cutoff frequencies [7].

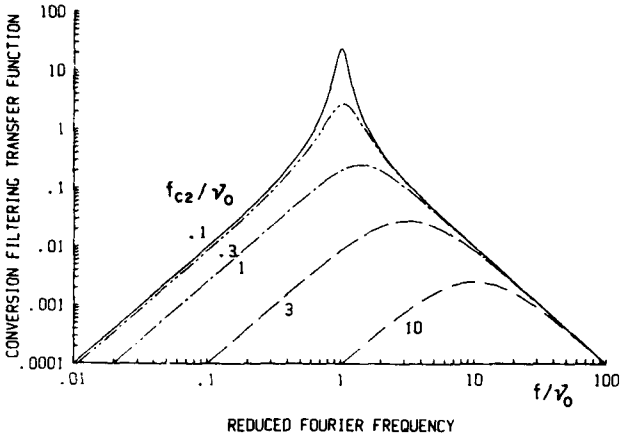


Fig. 4 - Conversion filtering transfer function for a first-order low-pass filter of different cutoff frequencies.

The direct filtering transfer function is a low-pass filter type then, the slow phase fluctuations are not affected by the filter. However, the conversion filtering transfer function is a band-pass filter type. The slow amplitude fluctuations are then always filtered out.

Time Domain

In the time domain, the characterization of frequency stability is defined as the two-sample variance of the relative frequency fluctuations averaged over a time τ [1]:

$$\sigma_y^2(2,\tau) = \frac{1}{2} E\{[\frac{1}{\tau}\int_{t+\tau}^{t+2\tau} y(\theta)\, d\theta - \frac{1}{\tau}\int_{t}^{t+\tau} y(\theta)\, d\theta]^2\}, \quad (21)$$

This variance is also often called the Allan variance [1]. Recently, Allan and Barnes have introduced another variance which they have called the modified Allan variance [9]:

$$\text{Mod } \sigma_y^2(2,\tau,n) =$$

$$\frac{1}{2} E\{[\sum_{i=1}^{n} \frac{1}{\tau}\int_{t+(i+n)\tau_o}^{t+(i+2n)\tau_o} y(\theta)\, d\theta - \frac{1}{\tau}\int_{t+i\tau_o}^{t+(i+n)\tau_o} y(\theta)\, d\theta)]^2\}, \quad (22)$$

where $\tau_o = \tau/n$. This variance has a supplementary degree of freedom, the parameter n, which is very important and must be specified each time this variance is used. It can be considered as a generalization of the Allan variance. This last one corresponds to a particular case of the modified Allan variance:

$$\text{Mod } \sigma_y^2(2,\tau,1) = \sigma_y^2(2,\tau). \quad (23)$$

In conclusion, all the results applicable to the modified Allan variance also apply to the Allan variance.

The modified Allan variance is related to the single-sided power spectral density of the relative frequency fluctuations, which is given by eqs (19) and (20), through the following equation [10]:

$$\text{Mod } \sigma_y^2(2,\tau,n) =$$

$$2\int_0^\infty S_y^s(f)\, \frac{\sin^2(\pi\tau f)}{\sin^2(\pi\tau_o f)} \cdot \frac{\sin^4(\pi\tau f)}{(\pi n\tau f)^2}\, df. \quad (24)$$

This equation is equivalent to that obtained by Lesage and Ayi [11] but it is expressed in a simpler way. With this relation, one could find the modified Allan variance for any noise source or filtering scheme.

Additive Noise: The presence of additive noise is encountered in any frequency stability measurement, then it is important to exactly know its influence. The additive noise contribution to the power spectral density of the phase fluctuations is obtained by the use of eqs (20), (18), (7), (3), (17) and (4) and gives:

$$S_\varphi(f) = \frac{1}{2A_o^2}\left[\left|\frac{H(\nu_o+f)}{H(\nu_o)}\right|^2 S_n(\nu_o+f) + \left|\frac{H(\nu_o-f)}{H(\nu_o)}\right|^2 S_n(\nu_o-f)\right]. \quad (25)$$

This power spectral density corresponds to the symmetrical part of the additive noise as if it was separately filtered. This result is then coherent with that obtained in eqs (7) and (3).

We consider the particular case where this noise is white over a sufficiently wide band with

$$S_n(f) = N_o/2. \quad (26)$$

In this case the power spectral density of the filtered signal phase fluctuations becomes:

$$S_\varphi(f) = \frac{1}{2}\left[\left|\frac{H(\nu_o+f)}{H(\nu_o)}\right|^2 + \left|\frac{H(\nu_o-f)}{H(\nu_o)}\right|^2\right] \cdot \frac{N_o}{2A_o^2}, \quad (27)$$

where the right part of the power spectral density is an equivalent transfer function for the white additive noise.

The modified Allan variance could then be calculated using eqs (24), (19) and (27) for any filtering scheme. We study some commonly used filters: a first-order low-pass, a Nth-order Butterworth low-pass, a first-order band-pass, a Nth-order Butterworth band-pass and a second-order resonant bandpass. These filters respectively have the following transfer functions:

$$|H(f)|^2 = \frac{f_{C2}^2}{f_{C2}^2 + f^2}, \quad (28)$$

$$|H(f)|^2 = \frac{f_{C2}^{2N}}{f_{C2}^{2N} + f^{2N}}, \quad (29)$$

$$|H(f)|^2 = \frac{f_{C2}^2\, f^2}{(f_{C1}^2 + f^2)\,(f_{C2}^2 + f^2)}, \quad (30)$$

$$|H(f)|^2 = \frac{f_{C2}^{2N}\, f^{2N}}{(f_{C1}^{2N} + f^{2N})\,(f_{C2}^{2N} + f^{2N})} \quad (31)$$

and

$$|H(f)|^2 = \frac{f_r^2 f^2}{Q^2 (f_r^2 - f^2)^2 + f_r^2 f^2}. \qquad (32)$$

The Nth-order Butterworth band-pass filter corresponds in fact to a Nth-order Butterworth low-pass filter followed by a Nth-order Butterworth high-pass and contains the first three filters as particular cases (with N = 1 and/or high-pass cutoff frequency equaled to zero). The results obtained are expressed as a function of an equivalent bandwidth, F_h, which depends on the type and parameters of the filter and on the oscillator mean frequency. The modified Allan variance is given by the following relation:

$$\text{Mod } \sigma_y^2(2,\tau,n) = \frac{1}{n} \cdot \frac{3}{4\pi^2\tau^2} \cdot F_h \cdot \frac{N_o}{A_o^2 \nu_o^2}, \qquad (33)$$

for long averaging times. This result compares with those obtained in refs [9], [10] and [11]. The only differences are the appearance of the equivalent bandwidth and the explicit term for the zero-frequency value of the single-sided power spectral density of the relative frequency fluctuations. This last value is found to be a factor 2 smaller than the value previously obtained through qualitative arguments. The modified Allan variance is then inversely proportional to the parameter n, to the squared value of the averaging time and to the signal-to-noise ratio. Finally, it is proportional to the equivalent bandwidth introduced in this work. This equivalent bandwidth is given in Table 1 for the particular filters previously described.

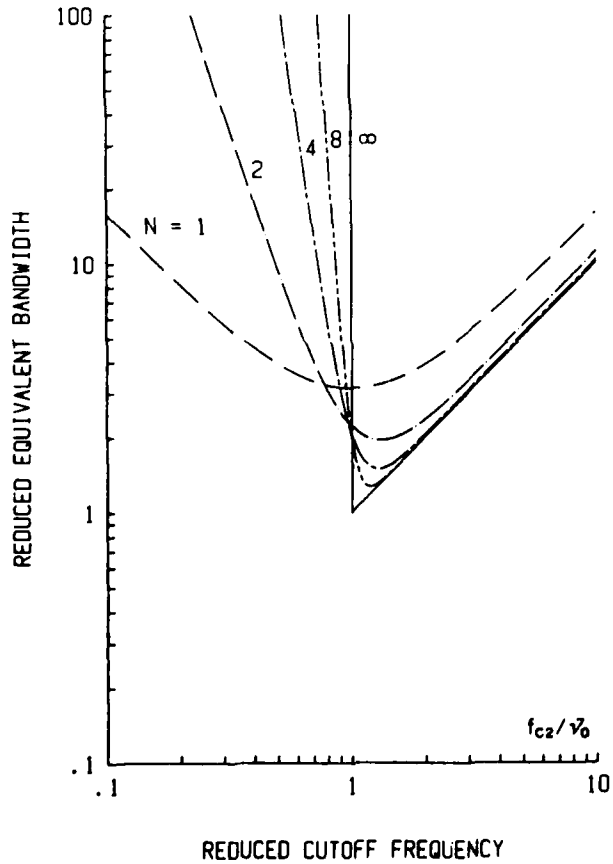


Fig. 5 - Reduced equivalent bandwidth of a Nth-order Butterworth low-pass filter as a function of the filter cutoff frequency for various orders.

The equivalent bandwidth of the Nth-order Butterworth low-pass filter is illustrated in Fig. 5 as a function of the cutoff frequency for various orders. The solid curve represents the rectangular low-pass filter which is the limit case, when N tends to infinity. At reduced cutoff frequencies much higher than 1, the equivalent bandwidth becomes directly proportional to the cutoff frequency. Particularly, for the rectangular filter, the equivalent bandwidth is exactly equal to the cutoff frequency. When the reduced cutoff frequency is smaller than 1, the equivalent bandwidth increases due to the relative attenuation of the carrier which is growing as the order of the filter increases. The reduced equivalent bandwidth is then minimized for a reduced cutoff frequency close to 1 for any order. Its value varies from π for a first-order to 1 when N tends to infinity. This optimum value is given in Table 1.

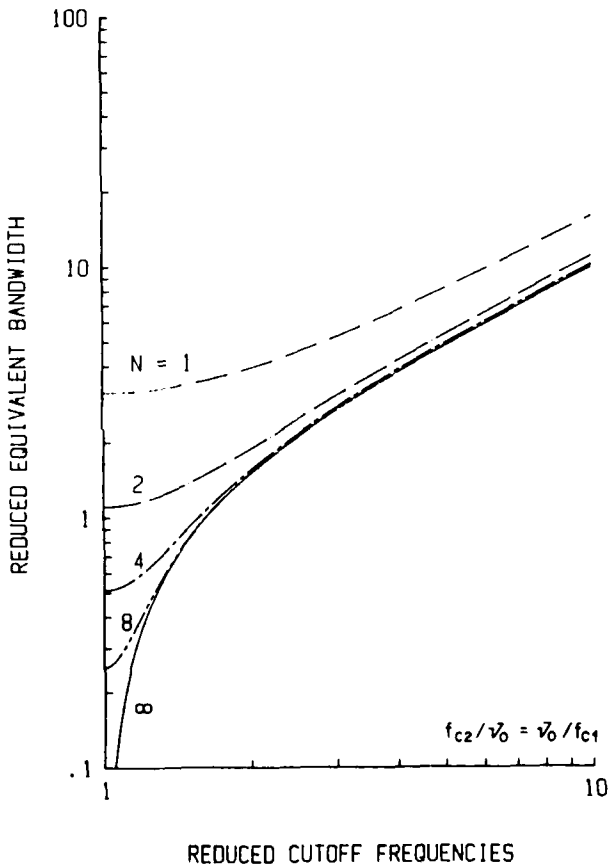


Fig. 6 - Reduced equivalent bandwidth of a geometrically centered Nth-order Butterworth band-pass filter as a function of the filter low-pass cutoff frequency for various orders.

Fig. 6 shows the equivalent bandwidth of a Nth-order Butterworth band-pass filter which is geometrically centered around the signal mean frequency. It is given as a function of the reduced low-pass cutoff frequency for various orders. The solid curve illustrates the equivalent bandwidth for the rectangular band-pass filter. Obviously, when the reduced low-pass cutoff frequency is much greater than 1, this filter has the same effect as the corresponding low-pass filter. For that particular configuration, the equivalent bandwidth is minimized when the two cutoff frequencies are the same and equal the signal mean

TABLE 1 - EQUIVALENT BANDWIDTH OF COMMONLY USED FILTERS

FILTER	EQUIVALENT BANDWIDTH (F_h)	OPTIMIZING CONDITIONS	MINIMUM F_h
first-order low-pass $2\pi f_{C2}\tau_o \gg 1$	$\dfrac{\pi\,(\nu_o^2+f_{C2}^2)}{2\,f_{C2}}$	$f_{C2} = \nu_o$	$\pi\,\nu_o$
Nth-order Butterworth low-pass $2\pi\,\sin(\pi/2N)\,f_{C2}\tau_o \gg 1$	$\dfrac{\pi\,(\nu_o^{2N}+f_{C2}^{2N})}{2N\,\sin(\pi/2N)\,f_{C2}^{2N-1}}$	$f_{C2} = (2N-1)^{1/2N}\,\nu_o$ $= \nu_o$, for $N \gg 1$	ν_o, for $N \gg 1$
second-order non-resonant band-pass $2\pi\,\min(f_{C1},f_{C2})\,\tau_o \gg 1$	$\dfrac{\pi\,(\nu_o^2+f_{C1}^2)\,(\nu_o^2+f_{C2}^2)}{2\,\nu_o^2\,(f_{C2}-f_{C1})}$	$f_{C1} = f_{C2} = \dfrac{1}{\sqrt{3}}\,\nu_o$	$\dfrac{4\pi\,\nu_o}{3\sqrt{3}}$
Nth-order Butterworth low-pass followed by high-pass $2\pi\,\sin(\pi/2N)\,\min(f_{C1},f_{C2})\,\tau_o \gg 1$	$\dfrac{\pi\,(\nu_o^{2N}+f_{C1}^{2N})\,(\nu_o^{2N}+f_{C2}^{2N})\,(f_{C2}-f_{C1})}{2N\,\sin(\pi/2N)\,\nu_o^{2N}\,(f_{C2}^{2N}-f_{C1}^{2N})}$	$f_{C1} = f_{C2} = \left[\dfrac{2N-1}{2N+1}\right]^{1/2N}\nu_o$ $= \nu_o$, for $N \gg 1$	$\dfrac{2\,\nu_o}{N}$, for $N \gg 1$
second-order resonant band-pass $\pi f_r\tau_o/Q \gg 1$	$\dfrac{\pi\,[\,Q^2\,(\nu_o^2-f_r^2)^2 + \nu_o^2 f_r^2\,]}{2\,\nu_o^2\,f_r\,Q}$	$f_r = \left[\dfrac{1}{3}\left(1-\dfrac{1}{2Q^2}\right)+\dfrac{2}{3}\left(1-\dfrac{1}{4Q^2}+\dfrac{1}{16Q^4}\right)^{1/2}\right]^{1/2}\nu_o$	$\dfrac{\pi\,\nu_o}{2Q}$, for $Q \gg 1$

Note: For white additive noise, $\text{Mod}\ \sigma_y^2(2,\tau,n) = \dfrac{1}{n}\cdot\dfrac{3}{4\pi^2\tau^2}\cdot F_h\cdot\dfrac{N_o}{\nu_o^2 A_o^2}$.

frequency. This minimum value is inversely proportional to N and tends to zero for the rectangular filter. The general optimizing conditions for a Nth-order Butterworth band-pass filter are expressed in Table 1.

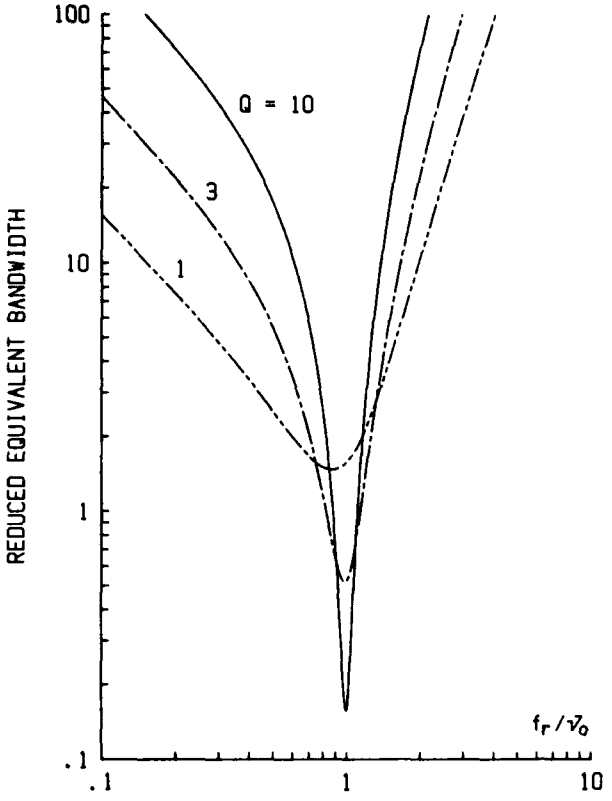


Fig. 7 - Reduced equivalent bandwidth of a selective filter with a single pair of poles as a function of the filter resonant frequency for various Q-factors.

Next, the equivalent bandwidth for a resonant bandpass filter having a single pair of poles is presented in Fig. 7 as a function of its resonant frequency for various quality factors. For each quality factor, the equivalent bandwidth is minimized at a resonant frequency which becomes closer to the signal mean frequency as the filter Q increases. This result is explicitely written in Table 1.

The various curves given in Figs 5, 6 and 7 should be used as references to evaluate the equivalent bandwidth of a particular filter scheme. For different filters than those here considered, eqs (24), (19), (27) and (33) should be used to calculate the corresponding modified Allan variance and the equivalent bandwidth. Moreover, in all time domain frequency stability measurements, one should try to approach the optimum conditions to minimize the equivalent bandwidth. It will then minimize the contribution of the additive noise which is possibly masking contributions from other more fundamental noise sources such as flicker phase noise.

Experimental Verification

An experimental verification has been made for the equivalent bandwidth of first- and eighth-order Butterworth band-pass filters which was geometrically centered around the signal mean frequency. These results are presented in Fig. 8.

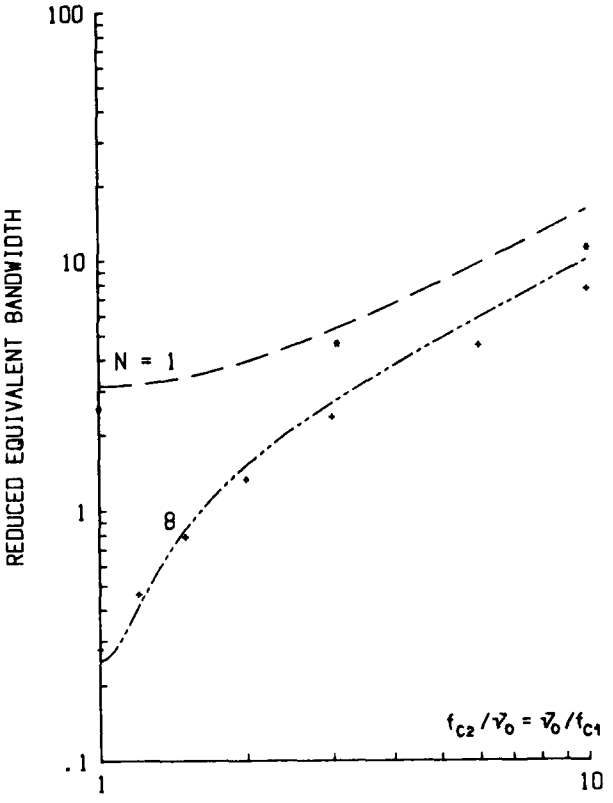


Fig. 8 - Experimental verification of the reduced equivalent bandwidth of a geometrically centered first- and eighth-order Butterworth band-pass filter.

These experimental results have been obtained by measuring the Allan variance under the various filter schemes, the additive noise level, the signal mean amplitude and its mean frequency. Next, we replaced the measured values in eq. (28) which relates the Allan variance and the equivalent bandwidth. The experimental conditions were the following: the single-sided power spectral density of the additive noise (N_0) was equal to 6 x 10^{-6} V^2/Hz over a 100 kHz band, the signal mean amplitude (A_0) was 4.0 V and its mean frequency ($_0$) was 300 Hz. The measured Allan variances have been multiplied by 3/2 in order to take into account for the effect of the dead-time between the successive frequency measurements used to calculate the Allan variance [2]. It is important to notify that the results presented here have not been normalized.

The Allan variances were measured with a 4% error and a 12% error was associated with the signal-to-noise ratio measurement which was obtained through the power spectral density of the additive noise and the signal mean amplitude. The agreement between experiment and theory is excellent since the experimental results for the equivalent bandwidth are within a factor 1.3 from the theoretical predictions illustrated by the two dashed curves.

Conclusion

One of the results arising form this work is that the phase filtering process is not readily determined from signal filtering. Next, the transfer functions for the direct filtering (AM-AM or PM-PM) and for the conversion filtering (AM-PM or PM-AM) have been derived for any filter scheme. Moreover, the additive noise contribution to the modified Allan variance is expressed in term of an equivalent bandwidth, called F_h, which is related to the filter used and to the measured signal mean frequency. By means of the tools provided in this work, the filter scheme should be chosen in order to minimize this equivalent bandwidth. Finally, a knowledge of this equivalent bandwidth is essential for a complete characterization of the frequency stability in the time domain, so it must be specified and we propose the use of Table 1 to facilitate its specification.

Acknowledgements: The authors wish to thank Mr Alain Michaud from the Laboratoire de Recherches sur les Oscillateurs et Systèmes for his numerous pertinent comments. This work was sponsored by NSERC, Ottawa, Canada and Fonds FCAC, Province of Québec, Canada.

References

[1] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leeson, T. E. McGunigal, J. A. Mullen Jr, W. L. Smith, R. L. Sydnor, R. F. C. Vessot and G. M. R. Winkler, "Characterization of Frequency Stability," IEEE Transactions on Instrumentation and Measurement, vol. IM-20, pp. 105-120, May 1971.

[2] B. E. Blair, Ed., Time and Frequency: Theory and Fundamentals. (NBS Monograph 140) Washington DC: U.S. Government Printing Office, 1974, ch. 8, pp. 151-204.

[3] K. Yoshimura, " Characterization of Frequency Stability: Uncertainty Due to the Autocorrelation of the Frequency Fluctuations," IEEE Transactions on Instrumentation and Measurement, vol. IM-27, pp. 1-7, March 1978.

[4] P. Lesage and C. Audoin, "Characterization of Frequency Stability: Uncertainty Due to the Finite Number of Measurements," IEEE Transactions on Instrumentation and Measurement, vol. IM-22, pp. 157-161, June 1973.

[5] P. Lesage and C. Audoin, "Effect of Dead-Time on the Estimation of the Two-Sample Variance," IEEE Transactions on Instrumentation and Measurement, vol. IM-28, pp. 6-10, March 1979.

[6] P. Lesage, "Characterization of Frequency Stability: Bias Due to the Juxtaposition of Time-Interval Measurements," IEEE Transactions on Instrumentation and Measurement, vol. IM-32, pp. 204-207, March 1983.

[7] P. Tremblay and M. Têtu, "Characterization of Frequency Stability: Effect of RF filtering," IEEE Transactions on Instrumentation and Measurement, vol. IM-34, pp. 151-154, June 1985.

[8] P. Tremblay, Etude de l'effet du filtrage sur la caractérisation de la stabilité de la fréquence. Ph. D. Thesis, Laval University, Québec, Canada, requirements to be fulfilled in 1985.

[9] D. W. Allan and J. A. Barnes, "A Modified "Allan Variance" with Increased Oscillator Characterization Ability," in Proceedings of the 35th Annual Symposium on Frequency Control, 1981, pp. 470-475.

[10] V. F. Kroupa, editor, Frequency Stability: Fundamentals and Measurement. New-York: IEEE Press, 1983, part III, p. 78.

[11] P. Lesage and T. Ayi, "Characterization of Frequency Stability: Analysis of the Modified Allan Variance and Properties of its Estimate," IEEE Transactions on Instrumentation and Measurement, vol. IM-33, pp. 332-336, December 1984.

THE ANALYTIC SIGNAL REPRESENTATION OF OSCILLATORS WITH APPLICATION TO FREQUENCY STABILITY ANALYSIS

Laurent-Guy Bernier and Fred E. Gardiol

Laboratoire d'Electromagnétisme et d'Acoustique
Département d'Electricité - Ecole Polytechnique Fédérale
16, Chemin de Bellerive - Lausanne 1007, SWITZERLAND

Summary

The analytic signal representation of band-pass processes is a formalism well known to radar and signal processing engineers but unfamiliar to many frequency stability specialists. In this paper it is introduced into the field of Frequency Stability Analysis and used to demonstrate several important results. An original expansion of the complex envelope of the oscillator under the high phase noise condition is given. The results are applicable to the analysis of frequency multiplication.

Introduction

Frequency stability measurements on frequency sources, as well as the many functions performed within the coherent receivers used in frequency control applications, constitute a special branch of signal processing. Among these applications the better known are Masers, other frequency standards and VLBI experiments; their common object is a particular type of narrow-band process : the stable oscillator signal. However the mathematical tools developed in the field of signal processing for the analysis of band-pass processes are seldom used by frequency stability specialists. Indeed, phase and amplitude fluctuations are generally defined axiomatically as low-pass processes and the study of the actual band-pass signal of the oscillator remains outside the conceptual framework of the problem. The present approach, on the other hand, introduces the analytic signal representation of band-pass processes as a model for the random harmonic motion of the oscillator and focuses on the relationship between the carrier, amplitude and phase fluctuations and additive noise in the actual band-pass signal of the oscillator.

Analytic Representation of Band-pass Processes

The analytic signal representation was first introduced for the analysis of signal processing in radars and in communication systems [1,2]. It is defined as follows; the analytic signal $\psi(t)$ associated to the real signal $x(t)$ is formed by adding to $x(t)$ an imaginary part equal to its Hilbert transform $\hat{x}(t)$:

$$\psi(t) \equiv x(t) + j\hat{x}(t) \tag{1}$$

where

$$\hat{x}(t) \equiv \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau}\, d\tau$$

The Hilbert transform may be considered as a quadrature filter of transfer function -j sgn(f). It follows that the double-sided power spectral density $S_{\psi\psi}(f)$ of the analytic signal vanishes for negative Fourier frequencies and is equal to twice the single-sided power spectral density of $x(t)$, noted $S^{+}_{xx}(f)$, as illustrated on figure 1.

$$S_{\psi\psi}(f) = \begin{cases} 2\, S^{+}_{xx}(f) & \forall f \\ 4\, S_{xx}(f) & \text{for } f \geq 0 \\ 0 & \text{for } f < 0 \end{cases} \tag{2}$$

Conversely, if a power spectral density vanishes for negative Fourier frequencies, the signal it represents is analytic. Hence it is easy to show under what condition the following form of signal :

$$\psi(t) \equiv A\, \gamma(t)\, \exp(j2\pi\nu_0 t) \tag{3}$$

is analytic. Considering its power spectral density,

$$S_{\psi\psi}(f) = A^2\, S_{\gamma\gamma}(f-\nu_0) \tag{4}$$

we see that it is sufficient that the low-pass modulation function $\gamma(t)$ should be band-limited to Fourier frequencies smaller than ν_0. All band-pass processes can be represented in that way and the dimensionless function $\gamma(t)$ is called the complex envelope of the phasor $\exp(j2\pi\nu_0 t)$ while A is its amplitude. The complex envelope may be expanded either into an in-phase and a quadrature components, $p(t)$ and $q(t)$, or into a real envelope $e(t)$ and phase $\phi(t)$.

CH2186-0/85/0000-0127$1.00©1985IEEE

$$\gamma(t) \equiv p(t) + jq(t) \equiv e(t)\exp(j\phi(t)) \qquad (5)$$

The Oscillator Signal under the Low-noise condition

Under the low-noise condition, the mean quadratic value of amplitude and phase fluctuations are much smaller than unity [3], hence :

$$R_{\phi\phi}(0) << 1 \text{ and } R_{\varepsilon\varepsilon}(0) << 1$$

where the mean quadratic value $< \phi^2(t) >$ is expressed in terms of the autocorrelation function $R_{\phi\phi}(\tau)$ as follows :

$$< \phi^2(t) > = R_{\phi\phi}(0) = \int_0^\infty S^+_{\phi\phi}(f)\, df \qquad (6)$$

while the mean quadratic value of the amplitude noise is defined similarly.

A first order expansion of the real envelope and phase yields the following analytic expression of the oscillator signal :

$$\psi(t) = A(1 + \varepsilon(t) + j\phi(t))\exp(j2\pi\nu_0 t) \qquad (7)$$

where $\varepsilon(t)$ represents the amplitude noise and $\phi(t)$ the phase noise. We observe, as shown on figure 2, that the amplitude noise is in-phase with respect to the carrier while the phase noise is in quadrature. Their orthogonality comes from the linear expansion of the complex envelope and is valid only under the low-noise condition.

Contribution of the Additive Noise

According to Rice's representation theorem [4,5], non-coherent band-pass processes such as the additive noise, may be represented too by a complex envelope. In this case, however, the carrier frequency ν_0 in the phasor is arbitrary and set for convenience at the oscillator nominal frequency.

$$n(t) + j\hat{n}(t) \equiv \sqrt{2}\,(c(t) + js(t))\exp(j2\pi\nu_0 t) \qquad (8)$$

Rice's theorem states that c(t) and s(t), the in-phase and quadrature components of the analytic signal expansion of the band-pass process n(t), have identical power spectral densities equal to the even part of the power spectral density of n(t) about ν_0. Comparing (7) and (8) we find that the additive noise contributions to phase and amplitude fluctuations are identical and equal to twice the inverse of the carrier or signal to noise ratio S/N in a one hertz bandwidth if a uniform power spectral density of n(t) is assumed about ν_0.

$$S^+_{\varepsilon\varepsilon}(f)_a = S^+_{\phi\phi}(f)_a = \frac{S^e_{nn}(f-\nu_0)}{A^2/2} = \frac{2}{S/N(1\text{ Hz})} \qquad (9)$$

Translated into the time domain the same result yields [3] :

$$\sigma^2(\tau) = \frac{3}{(2\pi\nu_0)^2\, S/N(B)\, \tau^2} \qquad (10)$$

Thus the so called "noise floor" of time domain systems corresponding to a τ^{-2} dependency of the Allan variance $\sigma^2(\tau)$, is in fact a statement of the signal to noise ratio in the system bandwidth B.

Effect of Linear Band-pass Filtering

The spurious amplitude to phase noise conversion produced by linear band-pass filtering and first described by Tremblay and Têtu [6] may be analysed in a similar way. Indeed the impulse response h(t) of a linear band-pass filter is itself a band-pass signal that may be represented by the following analytic signal [2] :

$$h(t) + j\hat{h}(t) \equiv H(\nu_0)\,\lambda(t)\exp(j2\pi\nu_0 t) \qquad (11)$$

where $H(\nu_0)$ is the transfer function of the filter at its nominal center frequency. Hence $\lambda(t)$ may be considered as the impulse response of a complex low-pass filter equivalent to the original band-pass filter. Expanding $\lambda(t)$ into its real and imaginary parts,

$$\lambda(t) \equiv 2(f(t) + jg(t)) \qquad (12)$$

we define a pair of real low-pass filters of impulse response f(t) and g(t). Assuming that the amplitude and phase noise processes are uncorrelated at the input of the band-pass filter, it can be shown that the amplitude and phase power spectral densities at the output are given by the following matrix equation :

$$\begin{pmatrix} S^{out}_{\varepsilon\varepsilon}(f) \\ S^{out}_{\phi\phi}(f) \end{pmatrix} = \begin{pmatrix} |F(f)|^2 & |G(f)|^2 \\ |G(f)|^2 & |F(f)|^2 \end{pmatrix} \begin{pmatrix} S^{in}_{\varepsilon\varepsilon}(f) \\ S^{in}_{\phi\phi}(f) \end{pmatrix} \qquad (13)$$

where F(f) and G(f) are the transfer functions of the real equivalent low-pass filters. It can be shown, moreover, that G(f) is a produce of the odd part of the band-pass transfer function H(f) about the center frequency ν_0. Thus if the band-pass filter is symmetrical about ν_0, G(f) vanishes and there is no AM to PM conversion. Otherwise amplitude and phase noise processes are filtered and mixed up according to (13). The equivalence between the band-pass filter H(f) and

the complex low-pass filter $\Lambda(f)$ is illustrated on figure 3.

L(f) and the Power Spectrum

The power spectrum of the oscillator may be computed easily from its analytic representation (7). Expanding the cross-correlation function of amplitude and phase fluctuations into its even and odd parts,

$$R_{\varepsilon\phi}(\tau) \equiv R^{o}_{\varepsilon\phi}(\tau) + R^{e}_{\varepsilon\phi}(\tau) \qquad (14)$$

it can be shown that the power spectrum of the oscillator is a function of the odd part only.

$$S^{+}_{xx}(f) = (A^2/2)\ \{\delta(f-\nu_0)+S_{\varepsilon\varepsilon}(f-\nu_0)+S_{\phi\phi}(f-\nu_0)+2S^{o}_{\varepsilon\phi}(f-\nu_0)\} \qquad (15)$$

This result shows that any asymmetry about ν_0 indicates a cross-correlation between amplitude and phase noise processes which indeed may be the result of an asymmetrical band-pass filtering as shown above. Equation (15) states that L(t), i.e. the power spectrum normalized by the carrier mean power [7], is a measure of the phase noise only if amplitude and additive noises are negligible and only under the low-noise condition. Otherwise the quadrature component could no longer be assimilated to the phase noise process.

The Oscillator Signal under the High Phase Noise Condition

When a clean frequency multiplication is applied on a low-noise oscillator, the mean quadratic value of the amplitude noise can be kept much smaller than unity [8]; but if the multiplication factor is high the mean quadratic value of the phase noise is bound to exceed unity. Hence under this high noise condition it is clear from (5) that the phase noise has both an in-phase and a quadrature components. Assuming that the amplitude noise is negligible, it can be shown that the autocorrelation function of the complex envelope is

$$R_{\gamma\gamma}(\tau) = \exp -[R_{\phi\phi}(0)-R_{\phi\phi}(\tau)] \qquad (16)$$

This result is well known [9,10] in communication engineering and describes the High-index FM waveform. However, we propose here an original expansion of $R_{\gamma\gamma}(\tau)$ into its in-phase and quadrature components :

$$R_{\gamma\gamma}(\tau) = \exp -(R_{\phi\phi}(0))+r_p(\tau)+r_q(\tau)$$

where : (17)

$$r_p(\tau) \equiv [\exp -(R_{\phi\phi}(0))][\cosh (R_{\phi\phi}(\tau))-1]$$

$$r_q(\tau) \equiv [\exp -(R_{\phi\phi}(0))][\sinh (R_{\phi\phi}(\tau))]$$

The first term is a constant that represents the carrier while $r_p(\tau)$ and $r_q(\tau)$ are respectively the in-phase and quadrature components of $R_{\gamma\gamma}(\tau)$. The power spectral density of the oscillator signal may be obtained by simply summing the Fourier transforms of these terms while a synchronous detection would demodulate either the in-phase or quadrature component according to the phase setting of the reference oscillator. Besides, Woodward's theorem states that both $R_{\gamma\gamma}(\tau)$, and its Fourier transform, i.e. the power spectral density, become gaussian when $R_{\phi\phi}(0)$ goes to infinity.

$$\lim R_{\gamma\gamma}(\tau) = \exp(-2\pi^2 m_2 \tau^2)$$

$$R_{\phi\phi}(0) \longrightarrow \infty$$

where (18)

$$m_2 \equiv \int_0^{\infty} S^{+}_{\phi\phi}(f)\ f^2 df$$

It is remarkable that the limit case power spectrum is a function of the second order moment m_2 of $S_{\phi\phi}(f)$ regardless of its shape [9]. Moreover, it can be shown from (17) that both the in-phase and quadrature components of the complex envelope become gaussian while the carrier vanishes exponentially when $R_{\phi\phi}(0)$ goes to infinity. Both yield equal contributions to the gaussian power spectrum.

The Case of Band-limited White Phase Noise

Our expansion of the complex envelope into its in-phase and quadrature components was verified experimentally using a pair of frequency synthesizers set at 5 MHz. One synthesizer was phase modulated by band-limited white noise and the in-phase and quadrature components were demodulated by a synchronous detection using the second synthesizer. The low-pass filtered output was fed to a spectrum analyser. Both the theoretical (solid lines) and experimental (scattered dots) power spectral densities are shown on figure 4. The theoretical spectra were obtained by computing numerically the Fourier transform of (17). The agreement is excellent except for a slight calibration error of the balanced mixer used for the synchronous detection. As expected, we observe that both spectra become independently gaussian for $R_{\phi\phi}(0)$ larger than 8. Besides it is easy to compute analytically the second order moment of a band-limited white phase noise process. Then, using (18), it can be shown that the bandwidth of the resulting gaussian power spectrum is :

$$BW = B \sqrt{2/3\, R_{\phi\phi}(0)} \qquad (19)$$

where B is the bandwidth of the band-limited white phase noise. This equation explains the experimental results published by Walls and De Marchi [8] : when $R_{\phi\phi}(0)$ is smaller than unity the signal bandwidth is B while when $R_{\phi\phi}(0)$ is larger than unity the bandwidth becomes proportional to $\sqrt{R_{\phi\phi}(0)}$, i.e. to the rms value of the phase noise fluctuations.

Conclusion

The analytic signal representation of band-pass processes is a formalism well known to radar and signal processing engineers but unfamiliar to frequency stability specialists. However the analytic signal representation provides a unified approach and a sound theoretical framework to frequency stability analysis. Moreover its application is simple and straightforward. The analysis of the AM-PM conversion produced by linear band-pass filtering leads in a simpler way to the results first published by Tremblay and Têtu. The expansion of the power spectrum into its in-phase and quadrature components is original. The theoretical results applicable to the analysis of frequency multiplication are more rigorous than those given by Walls and De Marchi, and more complete than those given by Lindsey and Chie.

References

[1] E. Bedrosian, "The Analytic Signal Representation of Modulated Waveforms", Proceedings of the I.R.E., vol. 50, October 1962, pp. 2071-2076.

[2] L.E. Franks, Signal Theory, Prentice Hall, Englewood Cliffs, 1969.

[3] J.A. Barnes, A.R. Chi, L.S. Cutler, D.J. Healy, D.B. Leeson, T.E. McGunigal, J.A. Mullen. W.L. Smith, R.L. Sydnor, R. Vessot and G.M.R. Winkler, "Characterization of Frequency Stability", IEEE Trans. on Instrum. and Meas., vol. IM-20, May 1971, pp. 105-120.

[4] J.M. Wozenkraft, I.M. Jacobs, Principles of Communication Engineering, Wiley, 1965.

[5] E.J. Baghdady, R.N. Lincoln, B.D. Nelin, "Short-term Frequency Stability : Characterization, theory and measurement", Proceedings of the IEEE, vol. 53, July 1965, pp. 704-722.

[6] P. Tremblay, M. Têtu, "Characterization of Frequency Stability : Effect of R.F. filtering", Proceedings of the CPEM 84, Delft, The Netherland, August 1984, p. 80.

[7] W.C. Lindsey, C.M. Chie, "Frequency Multiplication Effects on Oscillator Instability", IEEE Trans. on Instrum. and Meas., vol. IM-27, no.1, March 1978, pp. 26-28.

[8] F.L. Walls, A. De Marchi, "RF Spectrum of a Signal after Frequency Multiplication; Measurement and Comparison with a Simple Calculation", IEEE Trans. on Instrumentation and Measurement, vol. IM-24, no. 3, September 1975, pp. 210-217.

[9] M. Schwartz, W.R. Bennett, S. Stein, Communication Systems and Techniques, McGraw-Hill, New-York, 1966.

[10] N.M. Blachman, G.A. McAlpine, "The Spectrum of a High-Index FM Waveform : Woodward's Theorem Revisited", IEEE Trans. on Communication Technology, vol. COM-17, no. 2, April 1969, pp. 201-208.

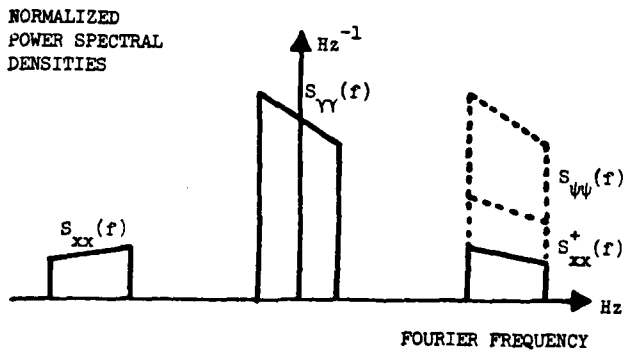


Figure 1. Comparison of the power spectral densities of $S_{XX}(f)$, $S^{+}_{XX}(f)$, $S_{YY}(f)$ and $S_{\psi\psi}(f)$. The amplitude A is normalized to unity.

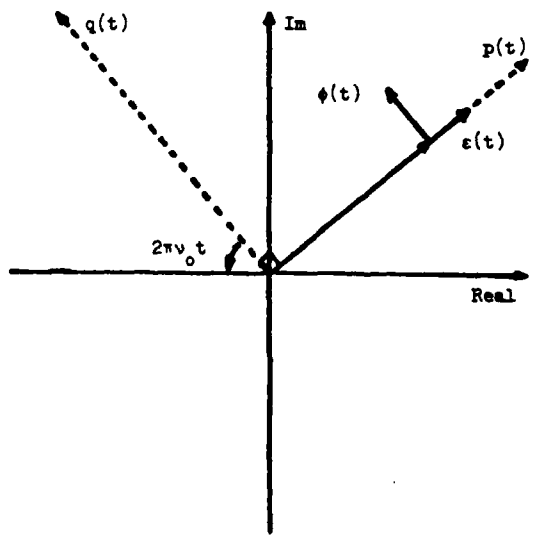


Figure 2. Representation of the signal of a low-noise oscillator in the rotating referential of the complex envelope.

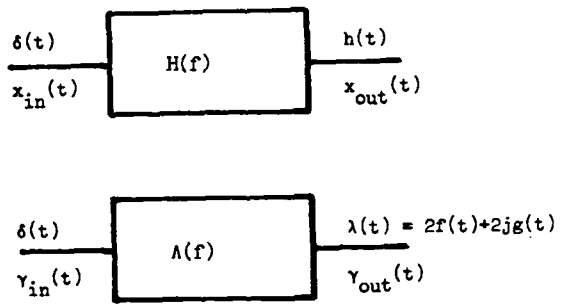
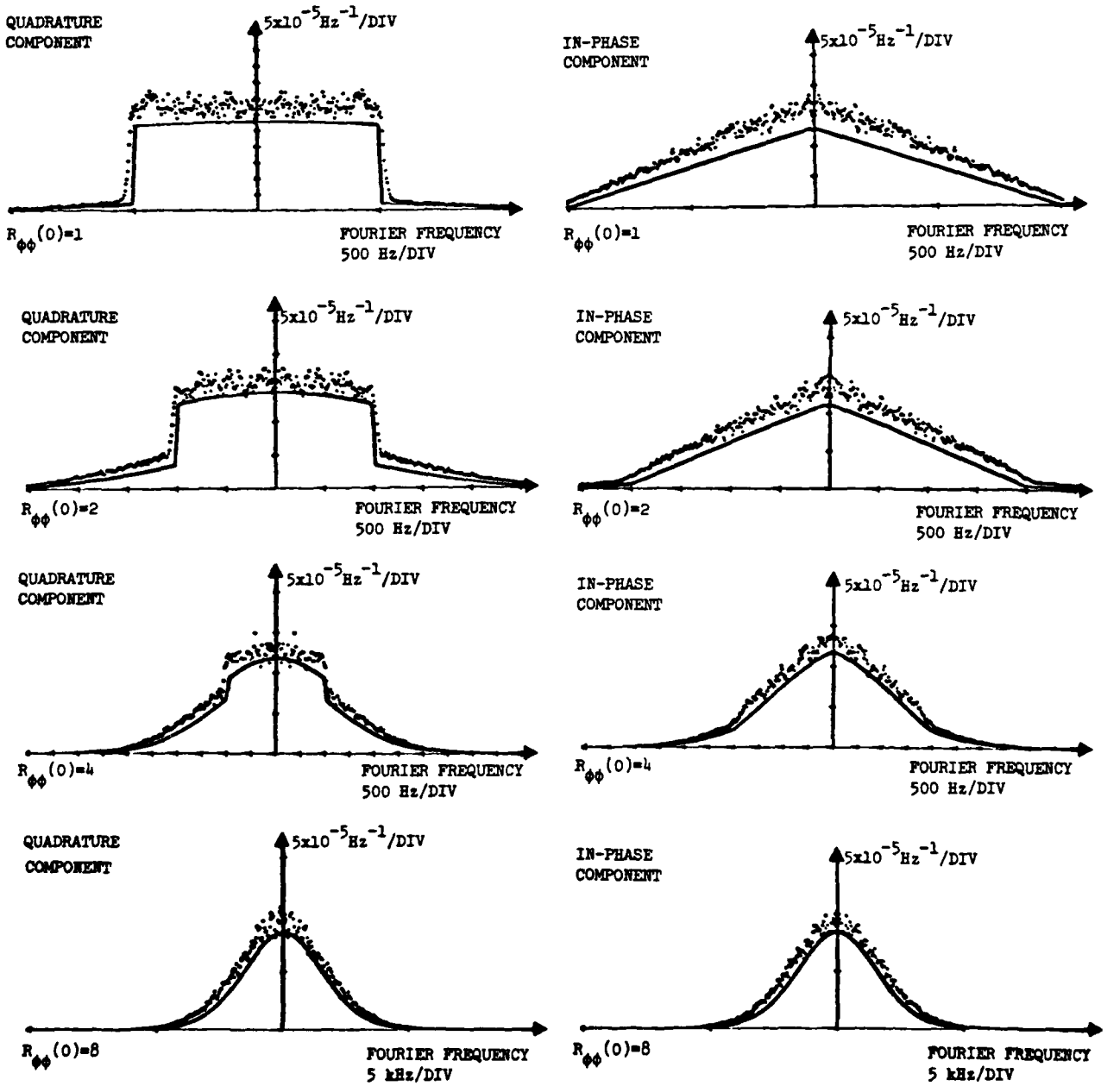


Figure 3. Equivalence between the real band-pass filter H(f) and its complex low-pass equivalent $\Lambda(f)$.

Figure 4. Power spectral densities of the in-phase and quadrature components of the complex envelope in the case of band-limited white phase noise. The quadrature component is defined as the two-sided Fourier transform of $r_q(\tau)$ while the in-phase component is defined as the two-sided Fourier transform of $r_p(\tau)$. The noise bandwidth is B.

a : $R_{\phi\phi}(0) = 1$, B = 500 Hz
b : $R_{\phi\phi}(0) = 2$, B = 1 kHz
c : $R_{\phi\phi}(0) = 4$, B = 1 kHz
d : $R_{\phi\phi}(0) = 8$, B = 1 kHz.



APPLICATION OF SPECTRUM ESTIMATION IN PHASE NOISE MEASUREMENT

Du Wulin, Xu Sanbao
Northwest Telecommunications Engineering Institute
Xi'an, PR of China

Abstract

This paper discusses a method of the phase noise measurement, based on the application of spectrum estimation. the principles of the method are described. It also presents the analyses of the variance of spectrum estimate and the selection of parameters in the measurement. An identifying experiment has been made. It is shown that this method is workable and has some advantages over other measuring methods.

Introduction

There have been many methods to measure the phase noise. Some variances which are based on the measurement in time-domain can be used to measure spectrum of phase noise $S_y(f)$, $S_\varphi(f)$. For instance, the measuring method of using the Modified Hadmardard variance to measure $S_\varphi(f)$ has been applied in some instruments. This is an indirect estimate of spectrum of phase noise, which is made by means of the parameters of random process. From the viewpoint of spectrum estimation theory, there is another direct estimate method of the phase noise, i.e, using a realization of a frequency fluctuation process, the desired phase noise spectrum can be obtained through signal processing. By using special-purpose or general-purpose microprocessors in instrument, the measurement and signal processing can be completed in the instrument in the on-line way. Compared with other measuring methods, an expectant advantage is that one may obtain the phase noise performance of a section of frequency range only through one measurement, thus reducing the time of measurement. In this paper, we have studied the principles and method of the measurement with emphasis on discussing the selection of parameters of measurement, the limitation and the adaptable area of measurement. To prove the availability of the method, an initial experiment has been made.

Principles

The fractional instantaneous frequency of signal $y(t) = \Delta\nu(t)/\nu_0$ is a random process, and can be seen as an approximate general stationary random process. through data sampling quantization values of $\bar{y} = \frac{1}{\tau}\int_0^\tau y(t)dt$ can be obtained, where τ is the gate time and the sampling period is T_s. During a limited period of measurement, the data of N $\bar{y}_k$ are a available, where $k = 1,2,---N-1$. Our purpose is to estimate the power spectrum of y $S_{y(f)}$ and $S_\varphi(f)$ with these data. $S_\varphi(f)$ is the single side power spectrum of $\varphi(t)$, and f is Fourier frequencies.

The classical spectrum estimation has been used here. From the obtained data of N, one can compute the biased estimates of the autocorrelation function of y as

$$C_y(m) = \frac{1}{N}\sum_{k=0}^{N-m-1} \bar{y}_k \bar{y}_{k+m}, \quad m = 0,---(N-1) \quad (1)$$

and

$$C_y(m) = C_y(-m).$$

the Fourier transforms of the expression are just the spectrum estimates of y, which are called a periodogram

$$I_N(\omega) = \sum_{m=-(N-1)}^{N-1} C_y(m)\exp(-j\omega m). \quad (2)$$

It can also be expressed as the Fourier transforms of limited sequence of y_k $Y(\exp(j\omega))$,

$$I_N(\omega) = \frac{1}{N}|Y(\exp(j\omega))|^2,$$

$$Y(\exp(j\omega)) = \sum_{k=0}^{N-1} y_k \exp(-j\omega k). \quad (3)$$

It is well known that $I_N(\omega)$ is not a good estimate of $S_y(\omega)$. It is not only a biased estimate (though asymptotically unbiased) but also it has too much variance. Some methods must be used to reduce the variance of estimates. We have applied the window function weighting and data averaging here. The window function weighting will result in a decrease of the resolution of spectrum estimates. Fortunately, the phase noise has ordinaryly the spectrum of power law function, and if the discrete components are not taken into account, there is not any significant fluctuation components in the area of estimate. In other words, no high resolution in the measurement of phase is necessary. By the data averaging is meant that the data of N are devided into L section, (an overlap of part of data being permissible, the number of each section being N_s). Considering the approximate statistical independence, the variance of the estimate may be decreased to 1/L by averaging the individual estimates $S_{y_i}(\omega)$. The Hanning window weighting of time-domain has been used in each section processing. Its time response is

$$w_h(n) = \frac{1}{2}\left(1-\cos\frac{2n\pi}{N_s}\right), \quad n=0,---Ns-1. \quad (4)$$

Taking into account the fact that the finite sequence of data is actually the result of the rectangular window weighting of an infinite sequence, the frequency function above can be expressed as

$$W_h(\omega) = \frac{1}{2}W_r(\omega) - \frac{1}{4}\left(W_r\left(\omega-\frac{2\pi}{N_s}\right)+W_r\left(\omega+\frac{2\pi}{N_s}\right)\right), \quad (5)$$

CH2186-0/85/0000-0132$1.00©1985IEEE

where $W_r(\omega)$ is the frequency function of a rectangular window. Therefore, in order to weight the periodogram with Hanning window, all one should do is to shift $I_N(\omega)$ in frequency-domain and to multiply it by constant coefficients, thus reducing the memory location and the number of operation.

The following is the analysis of the statistical errors in estimating the phase noise spectrum. The variance of the spectrum estimates has been proved to be

$$\text{Var}\ \hat{S}_y(k) = \frac{S_y^2(k)}{L}\left(1+2\sum_{j=1}^{L-1}\frac{L-j}{L}\,p(j)\right),$$

where

$$p(j) = \frac{j}{N_s^2 U^2}\left(\sum_{n=0}^{N_s-1} w_h(n)w_h(n+jN_d)\right)^2,$$

$$U = \frac{1}{N_s}\sum_{n=0}^{N_s-1} w_h^2(n),$$

and N_d is the displacement between the neighbouring sections.

If subsequences are not overlapped, then $p(j)=0$, for $j= 0,---L-1$, hence,

$$\text{Var}\ \hat{S}_y(k) = \frac{S_y^2(k)}{L}, \qquad L = \frac{N}{N_s}.$$

If the neighbouring subsequences overlap halfly one another, we have $p(j)= 0$, for $j= 2,---L-1$; for $N_s= 64$, $p(1)= 1/36$, and

$$\text{Var}\ S_y(k) = \frac{S_y^2(k)}{L}\left(1+2\,\frac{L-1}{L}\,\frac{1}{36}\right) \approx \frac{19}{18}\,\frac{S_y^2(k)}{L},$$

where

$$L= 2\left(\frac{N}{N_s}-1\right).$$

Compared with the non-overlap case, the overlaping processing may decrease the estimate variance.

The sensitivity of the measurement is determined by quantizing error, i.e. ± 1 error. The minimum spectrum which is limited by this error can be derived as

$$\hat{S}_y(k)_{sens} = \frac{1}{8U}\left(\frac{1}{\nu_0\tau}\right)^2 \qquad (7)$$

The necessary parameters in the measurement are: sampling period T_s, gate time τ, desired number of data N, the number of section L and the number of data in a section N_s. The decision of these parameters on one hand depends on the desired Fourier frequency range, permissible errors and the resolution of measurement, and on the other hand, it is related to the ability of the measuring equipment and its accuracy. If a low noise signal source is to be measured, the technique of error multiplication which uses frequency multiplication and mixing must be used in the same way as for an ordinary frequency measurement. To reduce the quantizing error in a counting, besides frequency measurement, the signal period may be measured via mixing signal to a lower frequency range.

Assuming the desired maxium Fourier frequency to be f_h and the minimum resolution to be F, then the number of data in a section N_s and the sampling period T_s are

$$N_s = \frac{2f_h}{F} \qquad \text{and} \qquad T_s = \frac{1}{2f_h},$$

N and L are decided by the requirement for reducing the variance. The gate time is the period during which $y(t)$ has been averaged. This process corresponds to $S_y(\omega)$ passing through a transmission system with frequency function

$$H(\omega) = S_a(\omega\tau/2),$$

acting as a low pass filter. This low pass filter can not influence the effectiveness of the measurement, because the sampling period T_s which must be greater than τ determines in fact the upper limit of the sampling frequency. To the same degree, it has the effect of pre-processing of the measured signal, filtering the frequency components above f_h and reducing the aliasing errors. It helps to reduce the quantizing error to select a larger τ, for the bits of coding output of the counter will increase with τ, and the quantizing errors will decrease. In fact, the selection of T_s is limited by τ. The results of measurement may be of significance when the quantizing error is much less than measured noise.

Experiment and Results

The arrangement and composition of the experiment are shown in Fig.1. An LC osillating source with about 100MHz frequency has been measured directly in frequency measurement. The accuracy of the gate time τ is determined by the accuracy of the clock in counter, which is much more accuracy than the stability of the measured signal source. In order to reduce the quantizing error, the gate time and sampling period are increased. The sampling period T_s which is determined on the timing device in the microprocessor is a little larger than τ. The program for the data transmission is solidified in EPROM. The obtained data has been transfered to a tape via the internal memory and then has been processed in a general-purpose microcomputer in the off-line way.

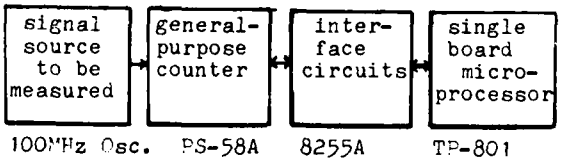


Fig.1 A measuring system

The processing of data is shown in Fig.2.

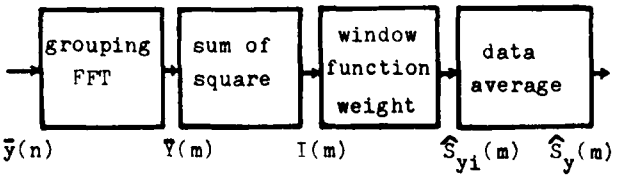


Fig. 2 Data processing

A curve of spectrum density $\hat{S}_y(f)$ which is obtained from the measured signal via measurement and data processing is shown in Fig.3.

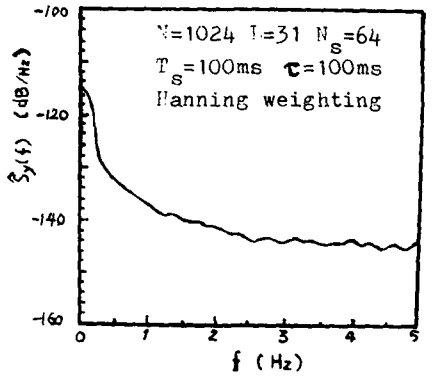


Fig.3 A set of measured results

The confidence analysis of phase noise spectrum estimate is as follows. Under the hypothesis that the frequency fluctuations obey approximately the Gaussian distribution, which is proved to be valid by our statistical test to distribution function, we have the following confidence degree

$$P\{B_1 \leq S_y(k) \leq B_2\} = 1-\alpha ,$$

in which

$$B_1 = 2I\hat{S}_y(k)/\chi^2_{\alpha/2}(2I)$$

$$B_2 = 2I\hat{S}_y(k)/\chi^2_{1-\alpha/2}(2I),$$

where α is the confidence level. when non-overlap processing is made, we have for α=0.1

$$P\{0.69\hat{S}_y(k) \leq S_y(k) \leq 1.59\hat{S}_y(k)\} = 90\%,$$

which means that the real spectrum $S_y(k)$ falls into the region of $\{0.69\hat{S}_y(k),\ 1.59\hat{S}_y(k)\}$ with the confidence degree of 0.9, i.e, the random variation of $S_y(k)$ is within $\{-1.6\ dB,\ 2.01\ dB\}$. If we make an overlap processing, this variation would be smaller. Therefore, the estimation precision of phase noise spectrum meets the engineering requirement.

Conclusion

From the analysis above and the result of the initial experiment, we can conclude that the method to measure the phase noise with spectrum estimation is workable, and it has no principle limitation. From the curve it has been seen that using the window smoothing and data averaging, the variance of the spectrum estimates may be controlled in a tolerable range. The significant advantage of this method is that one may obtain phase noise performance of a section of frequency range only in one measurement, thus reducing the time of measurement. The effect of the change in some factors on measured results will be avoided, which exists in other methods where several measurements are needed. There is also flexibility which results from using the micoprocessor. From the program for data processing here, the microprocessor built in the instrument is entirely in charge of the memory capacity and the number of operations required. The work to be done includes: using a standard measuring equipment to correct the results of the measurement by this method and to decide the accuray; planning the data proc cessing with the microprocessor built in the instrument; developing some corresponding circuits to do the measurement at the properly low frequency range and to measure the period instead of the frequency, thus the frequency f_h may be increased and the aliasing errors may be decreased.

References

1. J. A. Barnes, et al, "Characterization of Frequency Stability" IEEE Trans. Instrumentation and Measurement, Vol. IM-20, No.2, May 1971, pp. 105-120.
2. L. Peregrino, D. W. Ricci, "Phase Noise Measrrement Using A high Resolution Counter with On-line Data Processing", Proc. of the 30th Annaul Symposium on Frequency Control, USAEC, 309-316 (1976).
3. D. B. Percival, "Estimation of the Spectrum of Fractional Frequency Deviates" Proc. 32ed Ann. Symp. Frequency Control , pp. 542-548, May, 1978.
4. P. D. Welch, "The Use of the Fast Fourier Transform for the Estimation of Power Spectrum: A Method Base on Time Averaging Over Short, Modified Periodograms", IEEE on Audio and Electroacoustics, vO AU-15, No.2, pp. 70-73, June, 1967.

A COMPARISON OF FREQUENCY NOISE OF QUARTZ RESONATORS

J.J. Gagnepain

Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté - Besançon
32 avenue de l'Observatoire - 25000 Besançon - France

Abstract

The frequency fluctuations of quartz crystal resonators measured in a passive phase bridge, or in oscillators, are presented and compared, as a function of the frequencies, Q-factors, and geometries.

Introduction

The frequency range covered by quartz crystal resonators extends today from low frequency, a few MHz for thickness shear plates and miniaturized bars, up to 1 GHz for thin membrane inverted mesa type resonators. The best stabilities are obtained at frequencies around 5 MHz, but important improvements are achieved at 10 MHz. On account of their smaller size the 10 MHz resonators now have a large industrial development. In the same 1 - 10 MHz range, subminiature resonators, made by chemical etching or other processes, are also industrially developped.

Such a variety of resonators involves a large number of parameters : frequency, Q-factor, geometry, etc. The frequency stability of the resonator can be measured in an oscillator, but in that case it is not always possible to know the respective contributions of the crystal itself and of the electronics. A second method which has been used consists in measuring the fluctuations of the resonator's resonance frequency in a phase bridge, after reduction of the noise of the electronics and the driving source.[1] Most of the results which are presented were obtained using this method. In addition will be given some short term stability performances achieved in oscillators. It must be also pointed out that all the presented data, correspond to laboratory environments, and do not include the influence of vibrations, accelerations, thermal shocks, etc.

Stabilities measured in a phase bridge

The details of the measurement system can be found in reference [1]. Two identical resonators are driven at their resonance frequency by a high spectral purity source, and the phases of the two transmitted signals are compared by means of a double balanced mixer. When the two resonators are at the same frequency, and with equal Q-factors, the fluctuations of the source is rejected by 50 dB. The measure of the low frequency noise voltage at the mixer output enables to determine the frequency fluctuations of the resonators.

Typical noise spectra are shown in Fig. 1. They present mainly 1/f noise and sometimes $1/f^2$ noise. This last one can be attributed to temperature fluctuations. 1/f noise corresponds to the ultimate stability of the crystal. The 1/f noise level was normalized by measuring it at 1 Hz from the carrier. The results are given in Fig. 2 as a function of the unloaded Q-factor for a large number of resonators, from 1 MHz to 25 MHz.

A correlation between the noise level and the Q-factor can be observed. This is particularly clear when comparing the AT and BT-cut 10 MHz crystals.

The origin of 1/f noise in quartz crystal was attributed to the interaction between the acoustic wave and the thermal phonons of the lattice.[2] This was recently demonstrated by Planat,[3] who measured the noise level at low temperature, between 4K and 30K. In this temperature range the acoustic attenuation and therefore the Q-factor, have large variations. A maximum of noise was found at 20K, which corresponds to the maximum of phonon interaction as shown in Fig. 3, and this verifies the hypothesis.

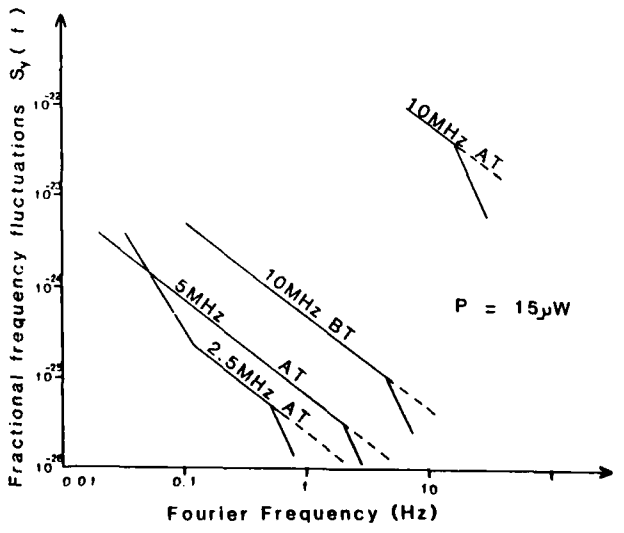


Fig. 1. Typical frequency noise spectra of 2.5, 5 and 10 MHz quartz resonators measured in a phase bridge. The solid curve shows the apparent spectrum. At Fourier frequencies above the resonator's bandwidth, the filtering effect of the resonator must be corrected, yielding the dashed line.

CH2186-0/85/0000-0135$1.00©1985IEEE

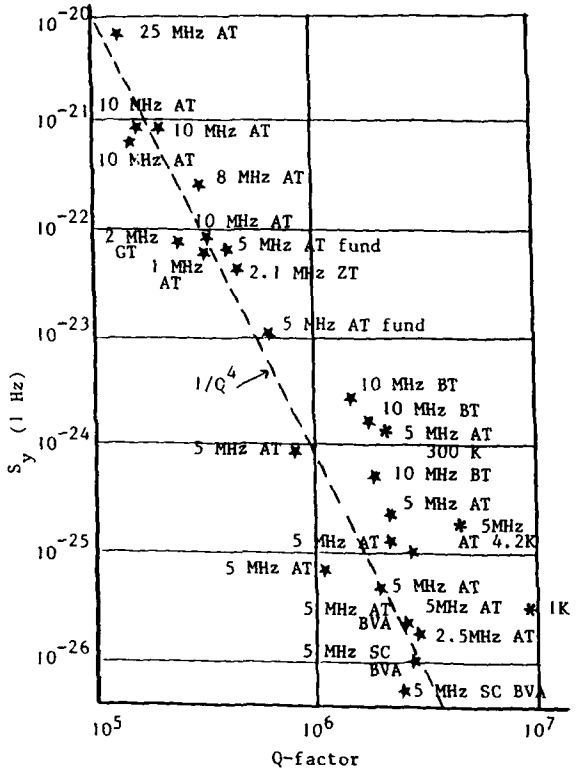


Fig. 2. 1/f frequency noise level at 1 Hz from the carrier versus Q-factor

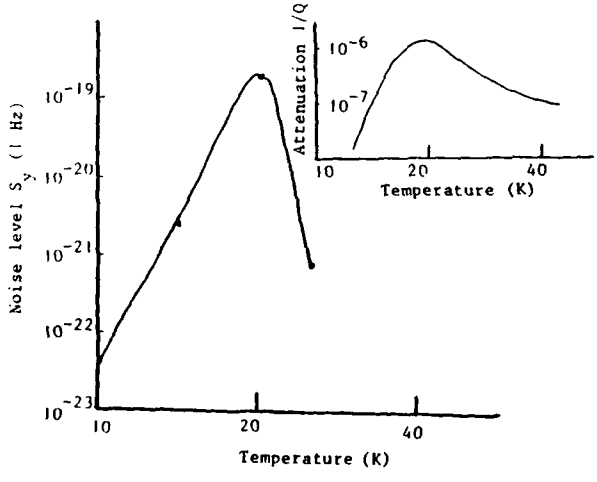


Fig. 3. 1/f noise level as a function of temperature near the maximum of phonon interaction

Stabilities measured in oscillators

Stabilities are measured by comparing, when possible, two identical oscillators, which are supposed to have the same stabilities, or by using three oscillators compared simultaneously two by two. This last method enables to characterize an oscillator by comparison with two other ones even of lower quality.

In Fig. 4 are shown the short term stabilities of five quartz oscillators. Fig. 4a corresponds to a 2.1 MHz crystal developped for watch application ; 4b to a miniaturized 5 MHz AT-cut crystal ; 4c to a 10 MHz SC-cut, 3rd overtone, QAS crystal (BVA type resonator but with adherent electrodes) ; 4d to a 5 MHz AT-cut BVA resonator ; and 4e to a 1 GHz AT-cut thin membrane resonator, made by ionic etching.

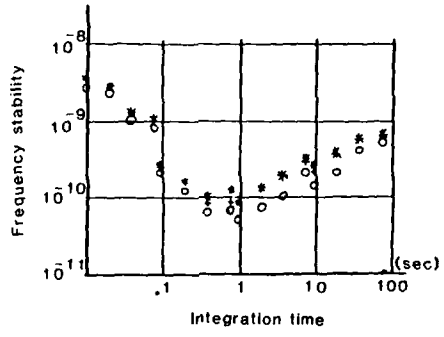


Fig. 4a. 2.1 MHz miniaturized GT-cut crystal

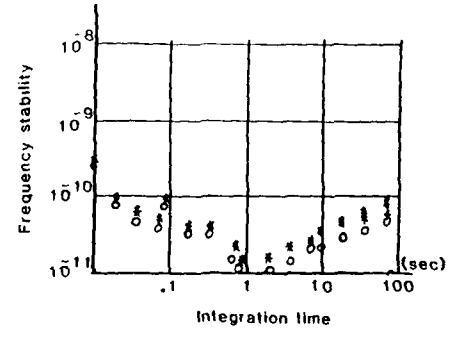


Fig. 4b. 5 MHz miniaturized AT-cut crystal

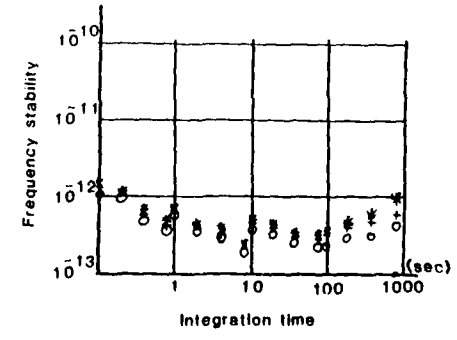


Fig. 4c. 10 MHz SC-cut QAS crystal

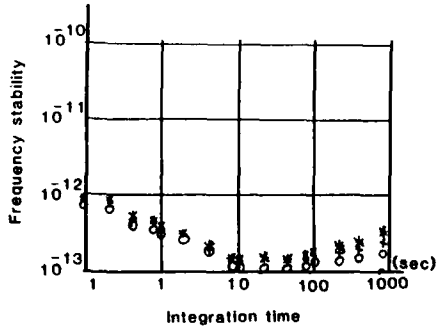


Fig. 4d. 5 MHz AT-cut BVA crystal

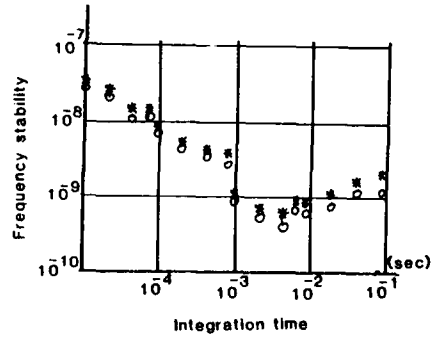


Fig. 4e. 1 GHz AT-cut crystal

These measurements indicate that the highest stabilities are still achieved at 5 MHz, and with BVA resonators. However large improvements were obtained at 10 MHz. On the other hand resonators are entering the 1 GHz range and offered now the possibility of realizing HF sources with stabilities of the order of a few 10^{-10}. Miniaturized crystals will also opened new possibilities for small size oscillators, low consumption, and low cost, when in mass production.

References

[1] F.L. Walls and A.E. Wainwright, IEEE Trans. Inst. Meas., IM-24, 1, 1975.

[2] J.J. Gagnepain, J. Uebersfeld, G. Goujon, P. Handel "Relation between 1/f noise and Q-factor in quartz resonators at room and low temperatures, first theoretical interpretation". Proceedings of the35th An. Freq. Cont. Symp., Philadelphie, 1981.

[3] M. Planat, "Temperature dependence of 1/f noise in quartz resonators in relation with acoustic attenuation", to be published.

PHASE NOISE IN CRYSTAL FILTERS

R.C. Smythe
Piezo Technology, Inc.
Orlando, Florida

Introduction

Although phase noise has been studied extensively in crystal oscillators, its occurrence in crystal filters has received scant attention. Crystal filter phase noise is of interest in at least two applications. In the first, a crystal filter is used to improve the wide-band phase noise characteristic. The second application occurs in frequency multipliers and frequency synthesizers, in which a crystal filter is used to suppress unwanted products. In both applications it is important that the low-frequency phase noise spectrum not be degraded by the filter. However, this can occur, and must be controlled.

Experimental

Phase noise in crystal filters may be measured using a differential phase noise measurement system [1, 2, 3] Figure 1 shows a set-up used in our laboratory. The reference filter is optional and should be identical to the filter under test to equalize delay in the two signal paths. All the usual considerations of phase noise instrumentation apply. In addition, the purity of the signal source is of paramount importance - first, because amplitude noise of the source is not cancelled, and second, because it may induce additional phase noise in the filter by means of (nonlinear) amplitude-to-phase conversion [4].

Figure 2 shows a typical phase noise plot for a 40 MHz AT-cut monolithic crystal filter. This two-pole, third overtone filter has a 3 dB bandwidth of approximately 6 kHz. For this measurement the signal level at the filter input is 0 dBm. The noise floor of -152 dBc (1 Hz) is set by the signal source, and is typically reached at 10 to 20 Hz offset. The noise spikes near 30 Hz are extraneous. Similar results are obtained in the 100 MHz region.

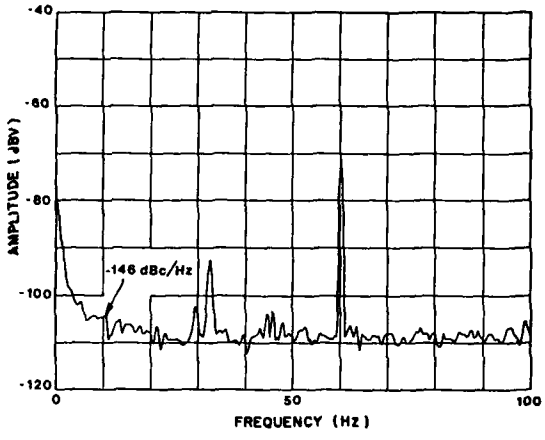


FIG. 2. TYPICAL PHASE NOISE PLOT

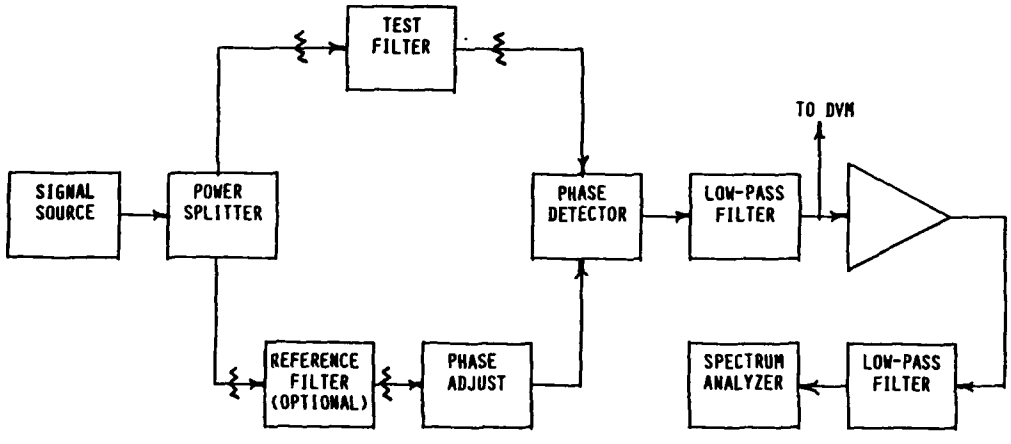


FIG. 1. FILTER PHASE NOISE MEASUREMENT

CH2186-0/85/0000-0138$1.00©1985IEEE

Discussion

As quartz resonators are a significant source of phase noise in crystal oscillators, it is not surprising that a signal passing through a crystal filter may experience phase noise degradation. In fact, the study of phase noise in crystal filters is essentially the study of noise in resonators.

Noise in resonators can be represented as random variations of resonance frequency and resistance. For both oscillator and filter applications the relation between these and phase noise depends upon the circuit in which the resonator is used. Since, in resonator measurements using a phase noise bridge, the fluctuation of phase ordinarily depends only weakly upon resonator resistance variation, the phase noise observed experimentally is attributed to frequency variation.

Frequency noise in quartz resonators has been studied experimentally [2, 3, 5, 6, 7]. At low current, it has been shown to exhibit both flicker (1/f) and random walk ($1/f^2$) behavior. Random walk of frequency has been correlated experimentally with temperature fluctuations, and can be related to the static and dynamic frequency-temperature characteristics of the resonator. The cause of 1/f frequency fluctuations is not well understood. Experimentally, a $1/Q^4$ dependence is observed [7, 8].

Measurements of resonators [7] show a large increase in frequency noise as current is increased. It would appear that at moderate current levels 1/f and thermally-induced $1/f^2$ frequency fluctuations are masked by $1/f^2$ and even higher order fluctuations associated with the higher strain levels, perhaps due to elastic nonlinearity. In a filter the magnitude of the current in each resonator may be different and in any event is frequency-dependent, but in most instances it is the moderate level effects which are of primary interest.

Our experience indicates that at moderate current levels surface-related mechanisms can contribute strongly to excess phase noise in filters and resonators. These mechanisms are controlled through careful attention to wafer preparation, plating, and cleaning.

References

[1] Howe, D.A., D.W. Allen, & J.A. Barnes, "Properties of Signal Sources & Measurement Methods," Proc. 38th Annual Symposium on Frequency Control (AFCS), pp. A1-A47; 1981.

[2] Wainwright, A.E., F.L. Walls, & W.D. McCaa, "Direct Measurement of the Inherent Frequency Stability of Quartz Crystal Resonators," Proc. 28th AFCS, pp. 177-180; 1974.

[3] Walls, F.L. & A.E. Wainwright, "Measurement of the Short Term Stability of Quartz Crystal Resonators & The Implications for Crystal Oscillator Design & Applications," IEEE Trans., v. IM-24, no. 1; 1975.

[4] Swanson, T.W., "Crystal Filter AM-PM Conversion Measurements," Proc. 32nd AFCS, pp. 250-254; 1978.

[5] Gagnepain, J.J., "Fundamental Noise Studies of Quartz Crystal Resonators," Proc. 30th AFCS, pp. 84-91; 1976.

[6] Noguchi, Y., Y. Teramachi, & T. Musha, "1/f Frequency Fluctuations of a Quartz Oscillator," Appl. Phys. Lett., v. 40, no. 10, pp. 872-873; 15 May 1982.

[7] Gagnepain, J.J., M. Olivier, & F.L. Walls, "Excess Noise in Crystal Resonators," Proc. 37th AFCS, pp. 218-225; 1983.

[8] Gagnepain, J.J., Uebersfeld, J., Goujon, G., & Handel, P., "Relation Between 1/f Noise and Q-Factor in Quartz Resonators at Room and at Low Temperatures, First Theoretical Interpretation," Proc. 35th AFCS, pp. 478-484; 1981.

DESIGNING CRYSTAL OSCILLATORS FOR IMPROVED PHASE-NOISE PERFORMANCE

A. Benjaminson

S.T. Research Corporation, Newington, Virginia 22122*

Introduction

This presentation attempts to compare the design characteristics of the basic anti-resonant crystal oscillator, with those of the Butler and DXO, both series resonant oscillators, in order to show the bases for their differing noise performances. It will also discuss the gains achieved by combined positive and negative feedback, where the latter involves the resonator, as used in the Butler and DXO, and the additional merit of voltage feedback versus current feedback, as exemplified by the DXO.

Noise Sources and Their Control

There are three types of noise generated in electronic circuits, flicker noise, thermal noise and device noise. The former is not well understood, but manifests itself through its 1/f amplitude-frequency characteristic, being very prominent at sub-audio frequencies where it phase modulates the oscillator output, but becomes inconsequential at about 100 Hz. It can also vary the phase of the oscillator feedback path, forcing the resonator frequency to shift to phase-compensate and convert the flicker noise into a $1/f^3$ effect.

Thermal, or white noise has a constant density at all frequencies. It too, phase modulates the oscillator output producing a "noise floor" that extends from the oscillator frequency to the limits of the circuit passband. Thermal noise can also disturb phase within the oscillator loop in the same manner described for flicker noise, and results in $1/f^2$ noise.

We are fortunate to a degree, that the phase modulation of the feedback path within the oscillator is limited by the narrow bandwidth of the resonator, resulting in the rapid decrease of $1/f^3$ and $1/f^2$ noise beyond the half-bandwidth values on either side of the operating frequency. Since bandwidth is equal to the quotient of frequency divided by resonator Q, the maintenance of high crystal Q is vitally important in this regard.

Thermal noise which is a function of operating temperature only, can be lumped with noises produced by irregular movement of charge in transistors and diodes, in that device noise density changes very gradually with frequency while white noise, as the adjective implies, is uniform over the entire electromagnetic spectrum. Device noise is rated in terms of noise figure, for transistors and diodes, and depends on device current, and on source to input resistance ratios.

Aside from selecting circuit configurations and devices to provide a minimum noise figure, the only way to beat thermal noise is to overpower it, literally. Thermal noise, measured in power per Hertz of bandwidth, equals -174 dBm at room temperature. Since we are concerned only with its phase noise component, which averages one-half of the total, the phase noise power amounts to -177 dBm, or 3 dB less. Thus an otherwise perfect oscillator, generating 1 milliwatt (0 dBm) will exhibit a noise floor of -177 dBc. Allowing for the higher noise/signal ratio at the sustaining amplifier's input, the device noise, and the less than optimum impedance matching, one is lucky to achieve -170 dBc, at frequencies far enough away from the carrier (operating) frequency to be beyond $1/f^3$, $1/f^2$, and 1/f noise generating sources, with a 0 dBm output level.

Limitations of Anti-Resonant Oscillator Circuits

A type of circuit generally called the Pierce, or Colpitts oscillator as shown in Figure 1, comprises a tuned circuit in which the inductance is produced by the resonator operating above series resonance and tuned by two capacitors in series. A transistor is connected as shown, to amplify the voltage across C_1 sufficiently, to provide the power absorbed by the three resistances and sustain the process of oscillation.

The resistances represent the input loss of the transistor, the loss in the resonator, and the power consumed by the external load. The emitter resistor produces broadband negative feedback, is generally included for impedance and gain control, and helps to reduce 1/f modulation.

The power output of anti-resonant oscillators is limited by crystal dissipation, and the need to maintain crystal Q. The Q of the crystal is reduced by a factor of r_x/R_T where $R_T = r_i + r_x + r_L$, limiting the amount of output power to 10-30% of that dissipated in the crystal. Assuming a crystal dissipation of .5 mW (a moderately high value), output power may be about .1 mW or -10 dBm. Since the highest noise to signal ratio occurs at the input to the transistor, let us assume an amplifier gain of typically, 10 dB, a noise figure of 4 dB and a total output level, at the collector of -2 dBm. This will produce a phase noise L (f) = -177 + 10 + 4 + 2 = 161 dBc.

Finally, the output level of -10 dBm must be presented to a buffer amplifier whose own sideband noise is typically, L (f) = -177 + 10 + 3 = -164 dBc, assuming the input level of -10 dBm, and a N.F. of 3 dB. This will reduce the overall value of L (f) to slightly above -160 dBc, a fairly common value for this type of oscillator under the assumed conditions.

The limitations of anti-resonant oscillators then stem from reduced output power being limited to a fraction of the allowable crystal dissipation, by the need to maintain a high operating Q for the crystal.

High Q is required to hold down the bandwidth of the circuit and reduce the $1/f^3$, and $1/f^2$ noise. Depending on the use for which the oscillator is designed, some compromise may possibly be considered.

* Currently affiliated with Systematics General Corp.

CH2186-0/85/0000-0140$1.00©1985IEEE

Benefits of Negative Feedback Through the Resonator

Series mode oscillators such as the Butler and DXO, have several advantages over the Pierce-Colpitts group. They not only provide more power output, but can also increase the effective phase-slope of the resonator when negative feedback is applied through the resonator. The higher output power can improve the signal-to-noise-ratio appreciably above that of anti-resonant circuits and requires less gain in the buffer amplifier, while the enhanced phase-slope reduces the levels of $1/f^3$ and $1/f^2$ noise.

The Butler oscillator however, is very difficult to design because of the compromises necessary to achieve a well-balanced circuit. These stem from the need to proportion the circuit impedances to the series resistance of the resonator. As shown in Figure 2, the crystal is in series with the input impedance of the transistor, and is driven by the tuned circuit acting as a transformer for the load impedance. The resultant conflict is similar to that previously described for the Pierce-Colpitts group. In this oscillator the input impedance must be low, and the equivalent series load resistance must be high, to avoid a severe reduction in the operating Q of the resonator. A heavy load is advantageous in this case since the high output level is beneficial as mentioned previously, however the load can't be reduced without adverse effect on circuit Q.

While low input impedance can be obtained by using sufficient dc transistor current, since $R_{in} = .026/I_C$, gain control is inhibited by the high fixed series resistance of the crystal. Gain control of the Butler by varying transistor current is very difficult to achieve. The use of an attenuator, or voltage divider, as will be described for the DXO, is worth considering for the Butler.

Differential Crystal Oscillator[1]

The theory of operation of the DXO was described in the referenced paper including the phase-slope multiplication effect, whose benefits have been demonstrated in the laboratory, and plotted in Figure 3.

The method used to achieve negative feedback in the DXO, is fundamentally different from that used in the Butler oscillators. The latter, obtain negative feedback by connecting the crystal in the positive feedback loop, in what can be called current feedback, since the transistor current passes through the crystal and the transformed load resistance. This is an improvement on the Pierce-Colpitts group, since the crystal is now outside the tuned circuit, lowering the crystal current to equal the transistor current rather than 5 to 10 times higher, and improving the signal-to-noise-ratio in doing so.

The DXO however, separates crystal excitation completely from the transistor by driving the crystal from the tuned circuit through a current-limiting resistor, to form a voltage-divider at the transistor input and produce voltage feedback. The resonator is thereby isolated from both load and transistor currents being dependent only on the tank voltage and the current-limiting resistor. The tank voltage in turn, is a function of the supply voltage (and ALC, if used), while power output is determined by the ac current in the transistors, and the load impedance. This situation provides a basic oscillator design with considerable freedom of choice, since crystal current, transistor current, and the amount of negative feedback are virtually independent variables.

The differential amplifier can provide the high gain needed to benefit from the flicker noise reduction produced by the M_Q factor ($A_o \times \beta_n$). An unbypassed emitter-to-emitter resistor is not usually required since the high open-loop gain is reduced by negative feedback to the reciprocal of β_n.

It is interesting to note that both bases are driven in phase, with the positive input slightly higher than the negative input, while the collector currents are out-of-phase due to this small differential. This conditions raises the input impedance of the positive side by approximately $\beta_p/(\beta_p - \beta_n)$, while the negative input has a negative input impedance, $\beta_n/(\beta_n - \beta_p)$.

Limiting action (in the absence of ALC), creates less distortion than in other oscillators since the higher ac current andout-of-phase base and collector voltages on the positive feedback side, start the limiting action by reducting the effect of β_p. Since β_p is required to be only slightly larger than β_n, ($\beta_p - \beta_n = 1/A_o$) a small reduction has a marked effect on the signal level.

Transistors can be selected and biased to provide low noise figures and the high output levels needed to maintain a low phase-noise. Transistors such as the MRF-904, which has a N.F. of less than 2 dB at collector currents to 10 mAdc, have been used to provide output power levels of more than 10 mW. Such levels can be obtained with crystal currents below 1 mA in high impedance crystals, such as 5th overtone SC units, or by the use of a crystal impedance transformer, with lower impedance fundamental, or 3rd overtone crystals.

The DXO favors the use of high impedance crystals which are normally difficult to handle in other circuits, and should make them more popular where their reported lower aging rates are required.

A typical plot of single-sideband phase noise is shown in Figure 4. The low noise levels at 10 to 100 Hertz offset from the carrier can be seen. The noise level above about 500 Hertz offset has reached the noise floor of the measurement system used.

Conclusion

The DXO has been shown to be capable of very low-noise operation over a wide range of operating conditions, especially with high impedance resonators.

Additional work is planned to examine its application to a variety of vhf resonators, both SAW and bulk-wave types, and also to Lateral Field Resonators.

Acknowledgment

This work has been supported by the Frequency Control and Timing Branch of the USA ERADCOM in Fort Monmouth, New Jersey.

Reference

[1]A. Benjaminson, "Balanced Feedback Oscillators", Proc. 38th Annual Symposium on Frequency Control, pp. 327-333, 1984.

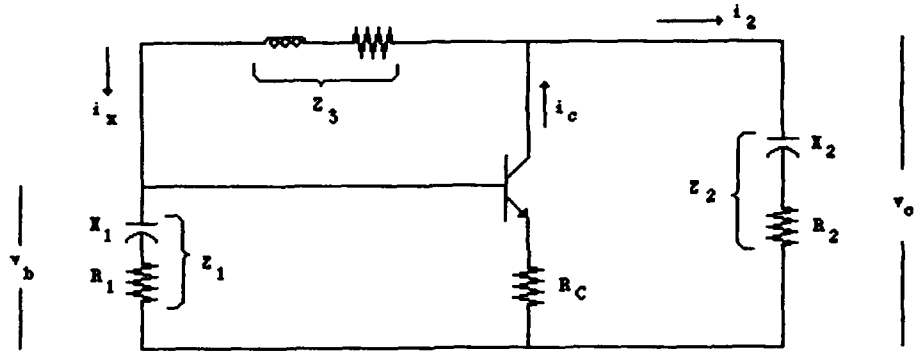
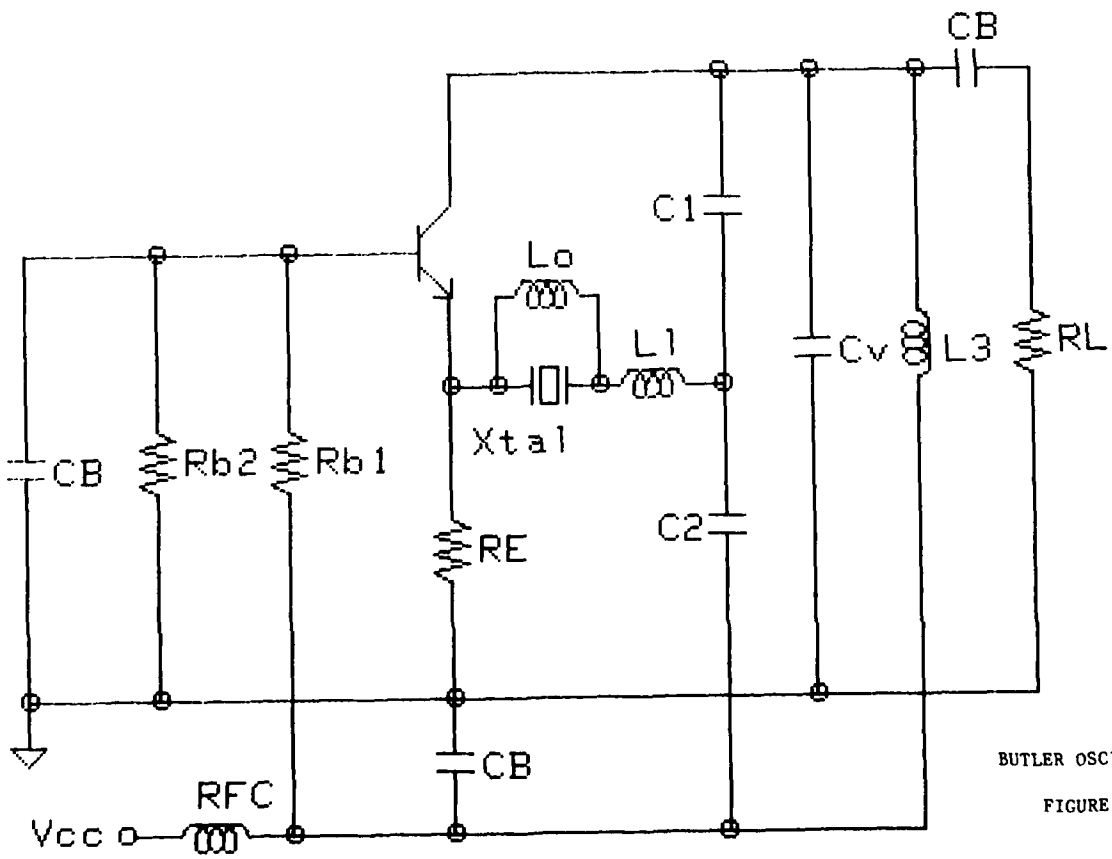
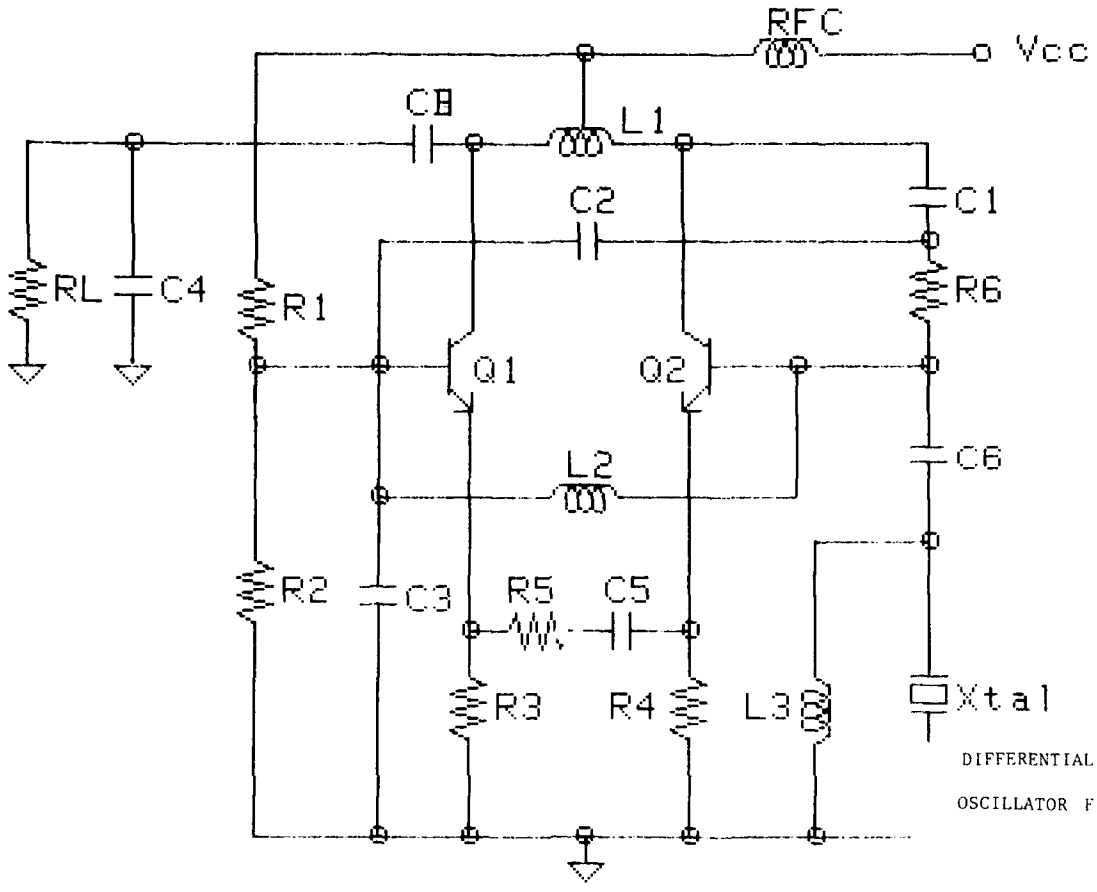


FIGURE 1. BASIC OSCILLATOR CIRCUIT



BUTLER OSCILLATOR

FIGURE 2



DIFFERENTIAL CRYSTAL OSCILLATOR FIGURE 3

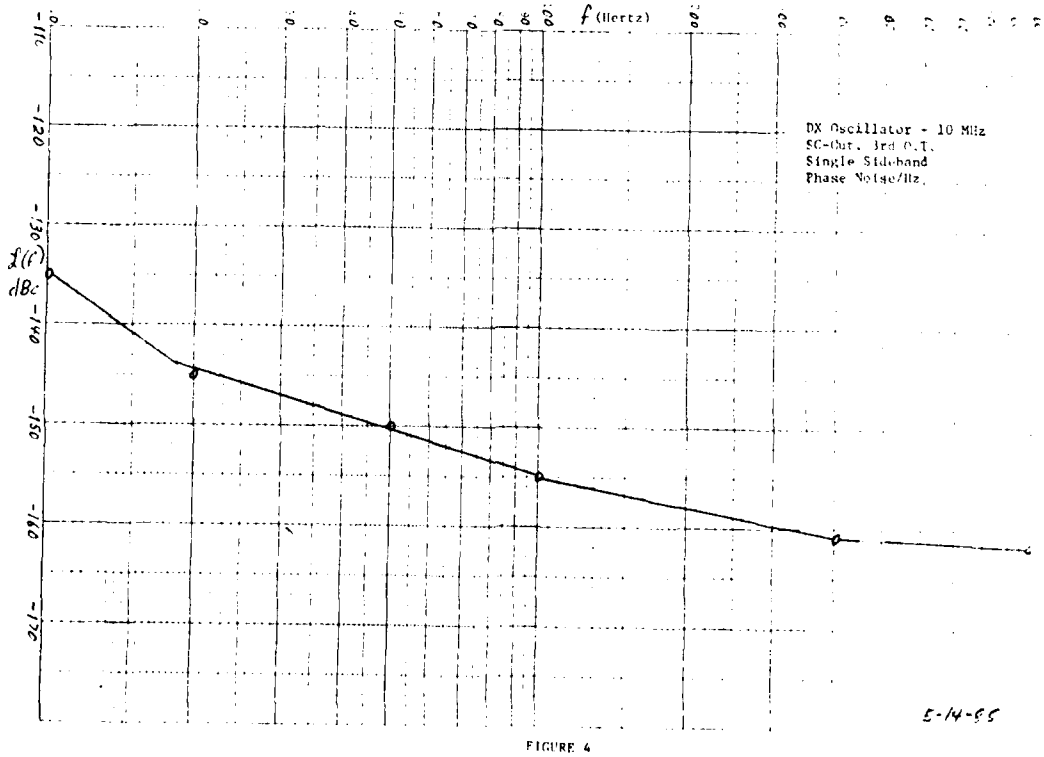


FIGURE 4

DEVELOPMENT OF A GPS TIME-TRANSFER RECEIVER AND TIME COMPARISON RESULTS

Michito IMAE, Makoto URATSUKA, Chihiro MIKI,
Takao MORIKAWA, Kosuke AKATSUKA and Kazuyuki YOSHIMURA

Radio Research Laboratory,
Ministry of Posts and Telecommunications

4-2-1, Nukui-kitamachi Koganei, Tokyo 184, JAPAN

Summary

The Radio Research Laboratory (RRL) has made the international time comparison since August 1984 using two GPS time-transfer receivers, one of which was developed by RRL and the other by a Japanese electric manufacturer, independently.

The performances of the receiver developed by RRL are ; time fluctuation of 6 to 20 ns at carrier-to-noise-density ratio $(C/No)_B$ of +40 to +55 dBHz and frequency stability of $2 \times 10^{-8}/\tau$. The over-all receiver delay was measured by using a GPS signal simulator with accuracy of 15 ns. The receiver does not use the ionosphere compensation data from the GPS satellites, but uses its own model based on total electron contents data measured at RRL.

Precision of 10 to 20 ns (rms value) is obtained in time comparisons between RRL and USNO or other organizations under the common view schedules. The international time comparison results are sent to BIH together with data of the atomic clocks and the primary frequency standard of RRL to contribute to TAI.

Introduction

The Radio Research Laboratory has been long making the international time comparison through the northwest Pacific network of Loran-C, but it is not directly connected with the north Atlantic ones, and has a precision too low for use of time transfers to contribute to TAI by sending our clock data.

The Global Positioning System (GPS) (1) gave possibility of direct time-transfers between RRL and US/European organizations. We started to develop a home-made receiver in 1983 and successfully completed it in 1984. We also have a GPS time-transfer receiver independently developed by a Japanese electric manufacturer.

By using the two receivers, RRL has been making the international time comparison on a routine base with precision of 10 to 20 ns since August 1984. The receiving schedule is based on the common-view schedule (2) which is calculated by NBS. We have been sending time comparison data through GPS to BIH, together with data of commercial Cs clocks and the RRL Cs beam primary frequency standard (3).

This paper presents the constitution and performances of the receiver developed by RRL, and time comparison results using the two receivers.

Receiver hardware

The block diagram of the GPS time-transfer receiver developed by RRL (GTR-1) is shown in figure 1. The 1.575 GHz signal from a GPS satellite is received by a one-turn herical antenna, which has a beam width of about 100 degrees and is automaticaly controlled in its direction by a computer. The received signal is filtered, amplified by a low noise amplifier, down-converted to a 75 MHz IF signal and transmitted to the receiver main unit. The IF signal, which is bi-phase modulated by the C/A code at clock rate of 1.023 MHz and by 50 bps navigation data (4-5), is despreaded by a correlation circuit which is constructed by a delay-lock loop (DLL) (6), and then demodulated by a Costas loop (7) to get the 50 bps data.

The pseudorange is measured with the internal time interval counter by counting the time interval between 1 kpps signal from a C/A code generator and 1 pps signal from the standard clock. As the measured pseudorange has an ambiguity of 1 ms, we eliminate it by the computer software.

The received 1.5 GHz signal from GPS has about +/- 4.8 kHz doppler frequency shift. We absorb it by controlling the VCO of 65.42 MHz with the ephemeris data sent from the satellites. The computer controls the whole system and makes data acquisition and processing.

Figure 2 shows the one-turn herical antenna (right side of the boom), a higher gain antenna (left side of the boom) which is going to be tested, an antenna controller and the front end of the receiver. Figure 3 shows the receiver main unit with the computer.

Receiver software

Function of software

The receiver is controlled by the mini-computer which has a Z80A CPU with 4 MHz clock. It has a 14 inche character display, two drives of 5 1/4 inche size of mini-floppy disk and a line printer. The computer software is written by Assembler and compiler-type Fortran, and its size is about 40 kbytes including data area. The functions of the software are ;

1) control of the hardware such as C/A code setting and VCO frequency setting,
2) pseudorange measurement, collecting and decoding of the 50 bps navigation data,
3) data processing and computation for time comparison results,
4) management of the receiving schedule.

We can get the time comparison results of (UTC(RRL) - GPS Time) every 5 sec. The data are printed out to a printer and written on a mini-floppy disk after each observation.

Ionosphere effect compensation

In the case of one-way time-transfer such as the GPS time-transfer, compensation of the ionosphere effect on the propagation time of the signal is very important. The ionosphere effect depends on the frequency of the signal and the total electron contents (TEC) along the path, and is approximately written by the

CH2186-0/85/0000-0145$1.00©1985IEEE

following equation(8).

$$t_{ion}=40.5 \times Nt/f^2/c \quad (ns)$$

where Nt is the TEC along the path in m^{-2}, f is the frequency of the signal in Hz and c is the speed of the light in m.

The navigation data sent from the GPS satellites include the ionosphere compensation terms. However we do not use them, but use our own model of the TEC for the compensation. Our model is based on the TEC data measured at RRL by receiving a VHF signal trasmitted from Engineering Test Satellite Type-II (ETS-II), the first geostationary satellite of Japan launched in 1977.

We calculated monthly averages of the TEC data measured every day for 1979 to 1984. The magnitude of the TEC yearly changes according to solar activities, but the daily pattern averaged over a month shows almost the same for each year. Therefore, for the compensation of the ionosphere effect, we use the daily TEC-pattern averaged for each month with its magnitude changes due to solar activities. We are examining the difference between our model and that of GPS, but at least it seems that our model gives better compensation for night time than GPS's model.

Performances of the receiver

Stability of DLL

Figure 4 shows rms fluctuations of the measured pseudorange versus carrier-to-noise-density ratio (C/No) of the input signal to the receiver. In our case C/No of 40 to 50 dBHz can be obtained at a normal receiving condition with the one-turn herical antenna, which corresponds to fluctuation of 10 to 20 ns. Figure 5 shows the stability (square root of the Allan variance) of the measured pseudorange at constant C/No (50 dBHz). It denotes a characteristic of 2×10^{-8}/tau.

Receiver delay measurement

For accurate time-transfers, estimation of the receiver delay is one of the most important factors. We measured the receiver delay by using a system shown in figure 6. The signal from the simulator of the GPS signal is inserted from just after the antenna and going through the receiver to be detected and measured by a time-interval counter with respect to the reference signal from the simulator.

The measured result is shown in Table 1. It shows the accuracy of about 15 ns, which is limited mainly by the precision of reading out on the oscilloscope used, and can be improved.

We also made a measurement of the temperature coefficent of the delay of the receiver main unit to get +3 ns/°C. As the receiver main unit is placed in the room where the temperature is controlled within 2 degrees, the effect seems to be negligible. But, for a more accurate time-transfer, we need a temperature compensation.

Time comparison results

International time comparison

Figure 7 shows results of the international time comparisons between UTC(RRL) and UTC(USNO) via GPS and Loran-C during the term of December 1984 to April 1985. For the GPS time comparison we use GPS SV#6 and SV#9 under the common view schedule. It shows the frequency difference of about 1×10^{-13} between UTC(RRL) and UTC(USNO). The peak variation is within about 50-60 ns and the rms value less than 20 ns for each satellite. In the figure, the lack of data is mostly due to the failures of our receiver. The time comparison result via Loran-C shows very low precision due to the indirect link of the Loran-C networks.

Figure 8 shows the frequency stabilities (square-root of Allan variance) of the time transfer results in figure 7. The stability of the time-transfer via GPS is several times better than the Loran-C.

Domestic time comparisons

Figure 9 shows results of the time comparisons between RRL and Tokyo Astronomical Observatory (TAO) by GPS and Loran-C. The GPS time-transfer result, obtained under the common view schedule, includes all GPS satellite available at present, six satellite (SV#6, 8, 9, 11, 12 and 13), but there is plotted one point a day for each satellite. The peak variation is within 100 ns and the rms variation about 30 ns. In this case the stability of the GPS time comparison results seems for averaging time of less than a few days 2 or 3 times better than the Loran-C, but for longer averaging time not so better because of the direct short-range connection of the Loran-C between the both organizations.

We also have a GPS time-transfer receiver independently developed by a Japanese manufacturer. Figure 10 shows time comparison results acquired in experiments using these two receivers, conducted between RRL Headquarters and International Latitude Observatory of Mizusawa (ILOM) or Kashima Space Center of RRL under common view schedules. In both case the variations are within 100 ns.

With ILOM we are making a time transfer experiment through a TV link of the Japanese broadcasting satellite BS-2. The comparison between the two links , via GPS and via BS-2, is under examination.

At Kashima we have a Very Long Baseline Interferometer (VLBI) station to make the joint experiments with NASA for measurement of the crustal plate movement and with USNO for time-transfer, and both started last year. The GPS time-transfer experiment between RRL Headquaters and Kashima was made to examine the connection of the three sites -RRL Headquarters, RRL Kashima station and USNO- by GPS during the VLBI experiments.

Time-transfer links of RRL

Figure 11 shows our overall system of time-transfer using space links(9). The GPS receivers are used mainly for time transfers between RRL and US/European organizations. The result of the GPS time transfer will be calibrated once a month by that of VLBI time transfer whose accuracy is expected to be less than 10 ns and precision around sub-nanosecond. The accuracy may be improved to 1 ns level in the near future.

We are going to start time-transfer experiments using Geostationary Meteorological Satellite (GMS-3) with CSIRO in Australia in 1985 and Shanghai Observatory in China in 1986. GPS time transfers will be also done in parallel with those via GMS-3.

We are also going to establish an accurate time-transfer system by using newly designed spread spectrum equipments for time-transfer and an automobile earth station therough Japanese communication satellite CS-2. It uses a two-way time-transfer technique. Accuracy of ns level has been obtained in a preliminaly experiment,

which is almost as good as that in the CS-1 experiment (10). We are planning to use the CS-2 time-transfer link to calibrate the other ones, such as those via GPS, GMS-3, BS-2 and VLBI, in domestic experiments.

Conclusion

Reported were the hardware and software of the GPS time-transfer receiver developed by RRL and time comparison results using it. The receiver shows enough performances for the present international time-transfer. By using it with the other one, we have been able to make direct time-transfers between RRL and US/European organizations, and make contributions of our atomic clocks and RRL primary frequency standard to TAI. For a more precise and accurate time-transfer via GPS, we are planning to improve our receiver by :

1) using a higher gain antenna,
2) improving of the DLL loop, and
3) compensating the ionosphere effect on real time.

Acknowledgements

The authors wish to thank the staffs of TAO and ILOM for their cooperations of the GPS time-transfer experiments.

References

(1) R. J. Milliken and C. J. Zoller, "Principle of Operation of NAVSTAR and System Characteristics", Journal of Navigation, 25, 2, 1978.
(2) D. W. Allan and M. A. Weiss, "Accurate Time and Frequency Transfer During Common-View of a GPS Satellite", Proc 34th Ann. Freq. Control Symposium, 1980.
(3) K. Nakagiri et al, "Accuracy Evaluation of the RRL Primary Cesium Beam Frequency Standard", Prc. 38th Ann. Freq. Control Symposium, 1984.
(4) A. J. Van Dierendoch, "The GPS Navigation Message", Journal of Navigation, 25, 2, 1978.
(5) C. R. Payne Jr., "Navstar Global Positioning System: 1982", Proc. 3rd International Geodetic Symposium in Satellite Doppler Positioning, 1982.
(6) R. C. Dixon, "Spread Spectrum System", John Wiley and Sons Inc., 1976.
(7) J. J. Spilker Jr., "Digital Communications by Satellite", Prentice Hall, 1977.
(8) K. Davies, "Ionospheric Radio Propagation", NBS Monograph, 80, 1965.
(9) K. Yoshimura et al, "Research Activities on Time and Frequency Transfer using Space Links", special issue of Proc. IEEE, Nov., 1985 (to be published).
(10) M. Imae et al, "Time-Comparison Experiments with a Small K-band Antenna and SSRA Equipments via Domestic Geostationary Satellite", IEEE trans. on IM, IM-32, 1, 1983.

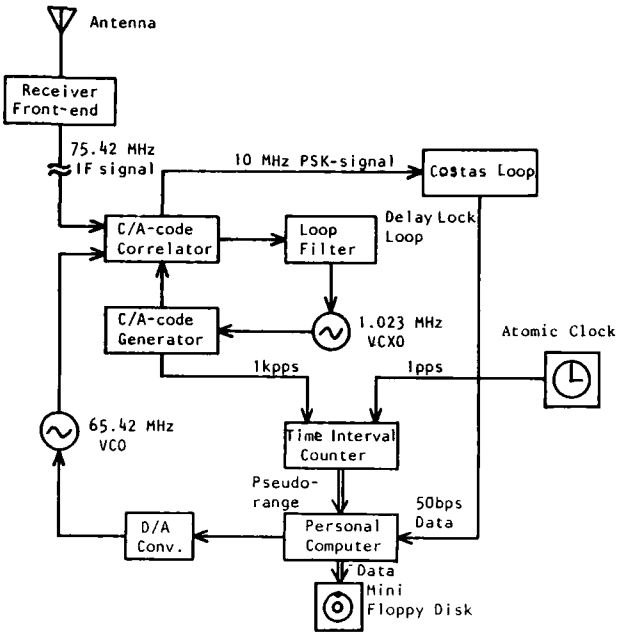


Figure 1. Block diagram of GPS time-transfer receiver (GTR-1) developed by RRL.

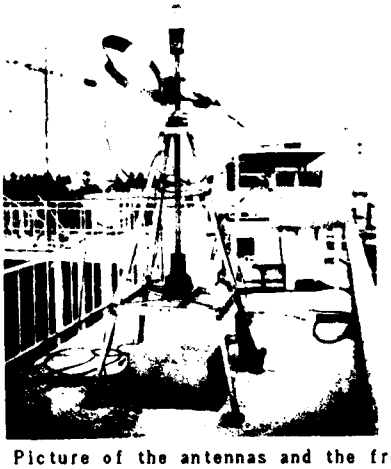


Figure 2. Picture of the antennas and the front end of the receiver.

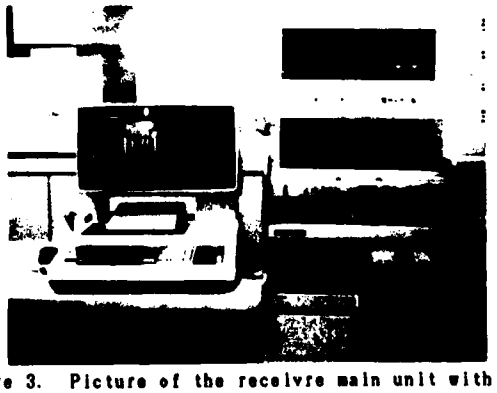


Figure 3. Picture of the receivre main unit with the computer.

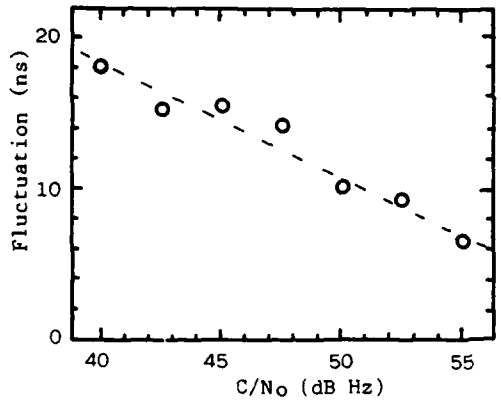


Figure 4. RMS fluctuations of the measured pseudorange versus C/No of the input signal to the receiver.

Table 1. Measured result of the receiver delay.

	measured value
Modulation delay in the simulator ; Tm	953 ± 10 ns
Cable delay in the measurement system ; Tc	352 ± 5 ns
Round trip time (the simulator to the receiver) ; Tt	1956 ± 10 ns
Receiver delay ; Tr=Tt-(Tm+Tc)	651 ± 15 ns

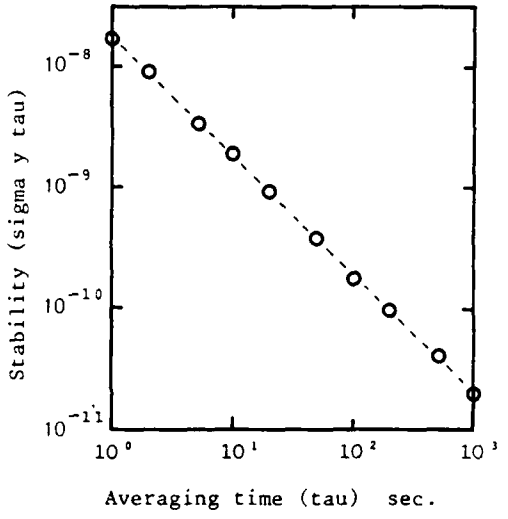


Figure 5. Stability of the measured pseudorange at constant C/No (50 dBHz).

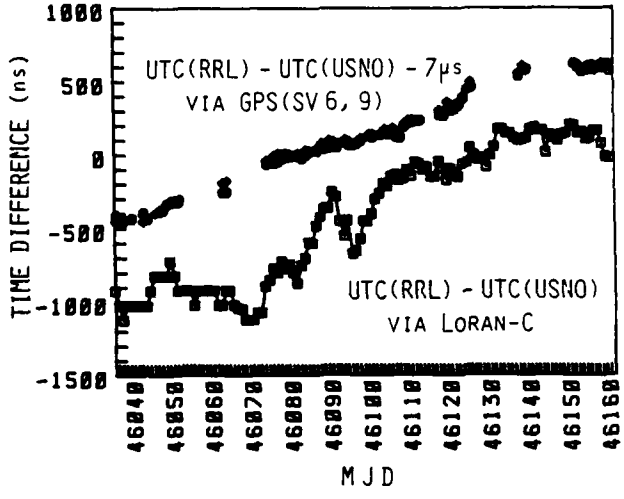


Figure 7. Time comparison results between UTC(RRL) and UTC(USNO) via GPS and Loran-C.

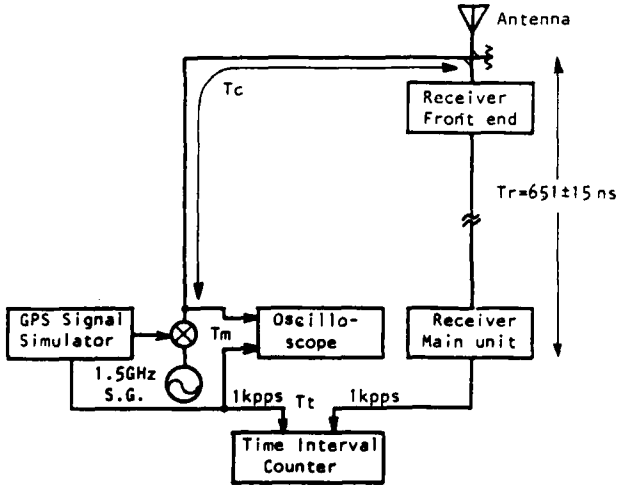


Figure 6. Measurement system of the receiver delay.

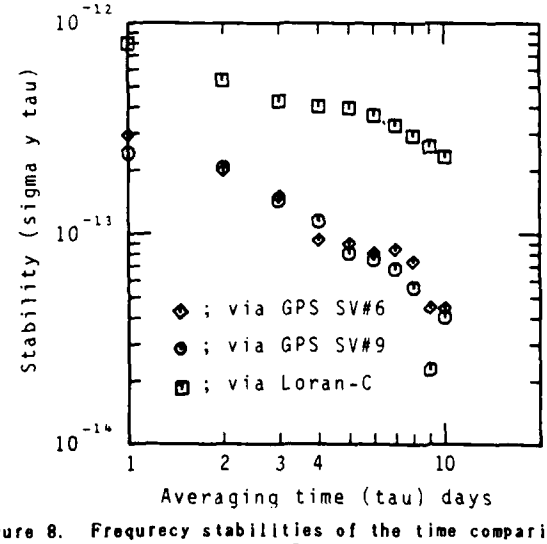


Figure 8. Frequrecy stabilities of the time comparison results in figure 7.

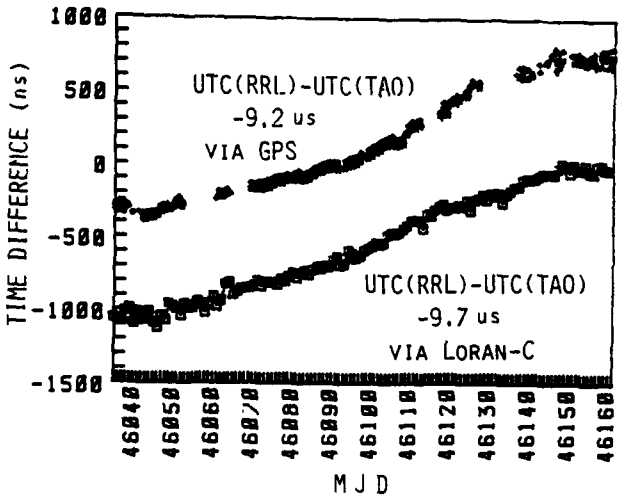


Figure 9. Time comparison results between UTC(RRL) and UTC(TAO) via GPS and Loran-C (for each result, frequency offset of 2×10^{-13} is eliminated).

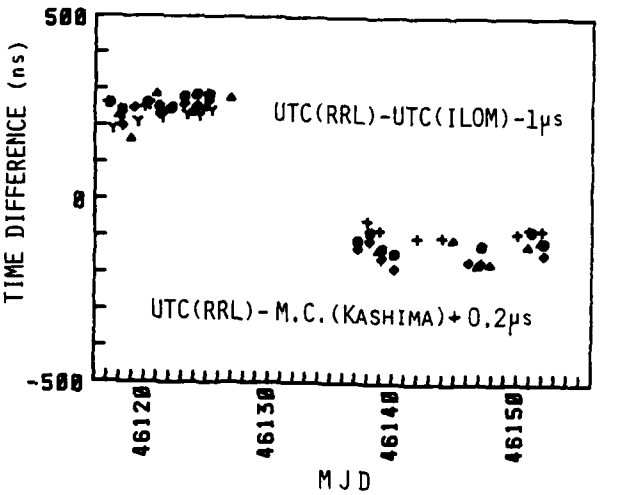


Figure 10. Domestic time comparison results using two GPS receivers.

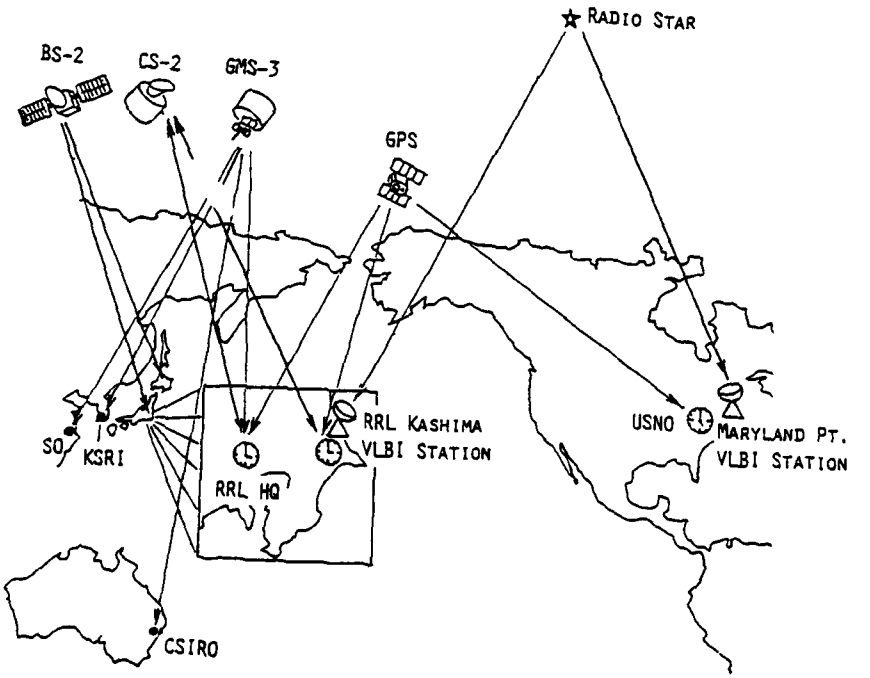


Figure 11. Overall time-transfer systems of RRL using space links.

COMMERCIAL GPS RECEIVER FOR TIME AND FREQUENCY EQUIPMENT APPLICATIONS

R.L. Lewis
G.F. Knoernschild
N.B. Hemesath

Collins Government Avionics Division
Rockwell International Corporation
400 Collins Rd. NE
Cedar Rapids, IA 52498

Abstract

The NAVSTAR Global Positioning System (GPS) currently being readied for production by the Department of Defense will provide significant improvements in the accuracy and cost-of-ownership features of future electronics systems used for navigation, positioning, and time and frequency reference applications. The Collins Avionics Group of Rockwell International has developed a commercial version of a modular sensor unit called the NAVCORE I™ GPS receiver which derives position, velocity and time data from the satellite signals. Digital data outputs are updated at the rate of one complete solution per second making the sensor function applicable to a wide range of dynamic and static product applications. This paper describes the features and capabilities of the timing version of the NAVCORE I™ receiver and describes the interfaces required for its use as a sensor in time and frequency product applications.

Introduction

With the recent contract award by the Department of Defense for the first production GPS user equipment, the production phase of the third element of the NAVSTAR GPS program has begun implementation. The first two elements, satellite vehicles and ground command stations, have been underway for over a year, with the first production satellites scheduled for launch starting in October of next year. Based on the current launch schedules worldwide satellite coverage will be available for 2D navigation by late 1987, with full 3D navigation coverage by mid-1988.

While the primary focus of the GPS program is on its revolutionary improvements in accuracy and cost-of-ownership for navigation and positioning systems, it offers the same kind of benefits for time and frequency reference product applications. Furthermore, these benefits can be realized several years sooner, with precision time signals already available for 16-20 hours per day from the development satellites, and full worldwide time coverage expected to be available by the beginning of 1987.

Commercial Applications

Collins Government Avionics has been actively involved in the development of military GPS user equipment since 1974. During that time as we have participated in the transition from advanced technology concepts to application hardware, we have also participated in the promotion of the system benefits for worldwide commercial applications. Two years ago, when we became convinced that the C/A code signals would ultimately be made available for commercial uses without restrictions or usage charges, we initiated product plans to apply the applicable parts of our hardware and software development work to non-military applications. This paper describes the first commercial product resulting from that effort, with emphasis on its possible time and frequency reference system applications.

Time & Frequency System Applications

Theoretically, GPS time receivers would be able to continuously deliver precision timing signals within the average of the eighteen atomic clocks on board the satellites, if satellite positions were always precisely known and if there were no ionospheric refraction to degrade signal transmission. For each user application the architectural complexities required to reduce the effect of these error sources to yield the desired accuracy must be weighed against the mechanization cost of the function.

In Collins developmental programs during the past two years, timing signal accuracies of less than 10 nsec have been demonstrated using both the NBS-designed common view receiver and a modified military P-code GPS receiver equipment which provides both position and time outputs. Timing signal accuracies of less than 100 nsec are now being consistently demonstrated using stand-alone C/A code receivers.

NAVCORE I™ Product Features

Although Collins Government Avionics is not directly involved in the development of time and frequency reference products, time information is a standard output of the navigation solution from the recently introduced NAVCORE I™ commercial navigation sensors.

The NAVCORE I™ receiver is a single-channel, sequential-tracking receiver that computes position, velocity and time solutions from the C/A code using the LI frequency. It is designed to perform the navigation sensor function for vehicles with speeds up to 600 knots and accelerations of 1 g. Through software modifications to force the velocity terms to zero and optimize the Kalman filter parameters, a customized version of the system software has been implemented to specifically meet the requirements of a basic GPS time sensor for use as a building block in a wide range of time and frequency reference products.

Operating Features

A single channel sequential design is used with the tracking circuitry dwelling upon each of four satellites for one fourth second, thus making measurements for a complete solution update once per second. After the receiver antenna position has once been established through an internal four-satellite position solution, integrity of the time transfer function is maintained with single satellite signal reception. Tracking is interrupted only to acquire satellites (as needed) and to collect ephemeris data periodically. During these intervals tracking is halted for periods of six seconds, which is a standard data subframe interval, with velocity terms held constant to span the data collection intervals. For time applications, this update interval has negligible effect on system performance since the Kalman filter models the frequency error of the receiver's internal frequency standard.

Operating software is designed for minimal demands on the user. For example, under normal operation the receiver provides a 1-pulse-per-second (pps) signal which is automatically slaved to UTC and a digital data message which identifies time at the pulse. No operator inputs are needed under normal startup conditions because the receiver has "keep alive" memory and a low-power coarse time reference source. At power-down, the last computed

CH2186-0/85/0000-0150$1.00©1985IEEE

position and almanac parameters for all satellites are retained in random access memory. When power is returned, the stored data and approximate knowledge of time from the internal reference allow satellite acquisition to begin automatically. The NAVCORE I™ receiver implementation uses a TCXO rather than an ovenized frequency standard, so no warm-up time is needed, and the acquisition process begins immediately with the application of power. The first solution is typically obtained in two and one half minutes.

The operating software also has a "cold start" mode which is used when critical memory has been lost through battery replacement or other equipment repair. In this mode the operator is requested to designate a satellite for acquisition and to enter the receiver's position within the nearest degree of latitude and longitude. Using this data the receiver is able to bootstrap itself into normal operation within approximately 20 minutes, since almanac data for all satellites must be collected before the normal acquisition procedure can begin.

System Architecture

The NAVCORE I™ receiver architecture has been specifically devised to achieve low production implementation cost. Special effort was made to utilize off-the-shelf commercial components and low-cost technologies wherever possible without compromising operational reliability. For example, along with the selection of a TCXO, several stages of conversion were employed to distribute system gain at a number of frequencies. This approach has reduced the risks of regeneration and yielded significant producibility benefits. Figure 1 shows the detailed frequency plan which is implemented.

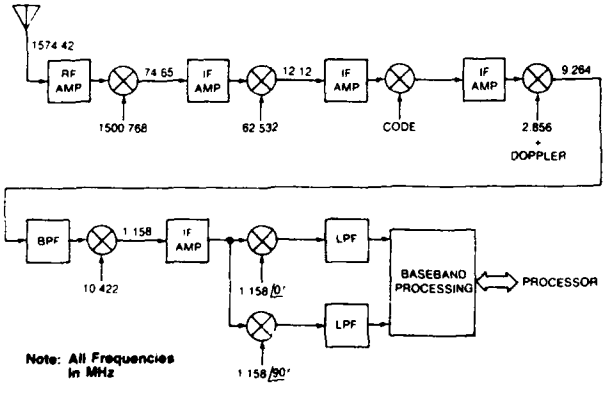


Figure 1. NAVCORE I™ Frequency Plan

This architecture was selected to meet the anticipated cost and performance requirements of a core GPS sensor which could be easily integrated into a widely varied range of navigation, positioning, and time products. Functional partitioning and physical packaging alternates have been implemented to enhance this building block concept.

Packaging

Functional partitioning of the NAVCORE I™ receiver mechanization is shown in figure 2, with the dotted lines showing the individual hardware modules. The same five hardware modules are used in both the navigation and time transfer versions of the receive, with alternate EPROM's used to implement the different operating software for each version.

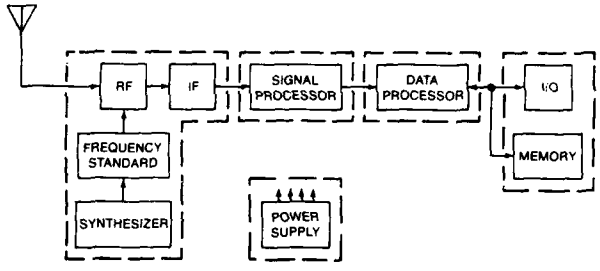


Figure 2. NAVCORE I™ Functional Partitioning

Two alternate physical packaging designs have been implemented to provide installation options for the user. Photos of the two configurations are shown in figure 3. The single unit standard package version has a volume of less than 270 cubic inches and weighs less than 8 pounds. The half-height low profile package is designed to provide more installation alternatives, and can be mounted in a 19-inch rack unit only 3-1/2 inches high. Either configuration can be mounted horizontally or vertically with no cooling air required.

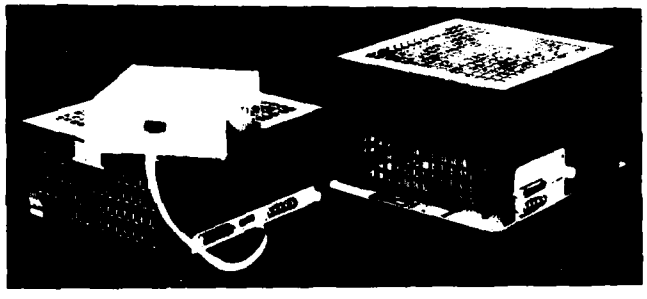


Figure 3. NAVCORE I™ Hardware Configurations

Performance Specifications

The physical and performance specifications for the NAVCORE I™ receiver are summarized in figures 4 and 5. Both laboratory and field testing are proving these specifications to be

SIZE	19.5 x 18.7 x 12.1 CM
WEIGHT	3 KG
POWER	25 WATTS
TEMPERATURE (AMBIENT)	–20 °C TO +55 °C
ALTITUDE	–300 TO 17,000 METERS
SHOCK (OPERATING)	6G
VIBRATION	DO-160A, CURVE B
HUMIDITY	95% NONCONDENSING

Figure 4. Physical and Environmental Parameters

POSITION ERROR	50 METERS SPHERICAL (SEP)
VELOCITY ERROR	0.5 METERS/SEC (1 SIGMA) EACH AXIS
TIME ERROR	100 NANOSECONDS (1 SIGMA)
TIME TO FIRST POSITION SOLUTION	LESS THAN 5 MINUTES (WARM START) LESS THAN 20 MINUTES (COLD START)
MAXIMUM VELOCITY	600 KNOTS
MAXIMUM ACCELERATION	1G

Figure 5. Performance Specifications

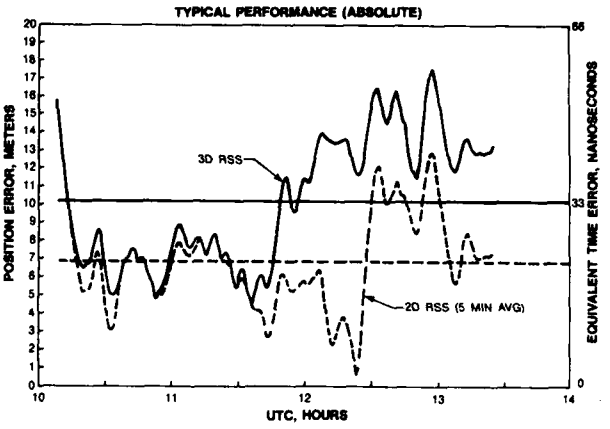


Figure 6. Measured Position Error

conservative. Most of the test results to date have concentrated on measurement of position accuracies rather than time accuracies because of the relative ease and speed with which comparisons can be made to known GPS position standards. Position measurement tests typically indicate that the spherical error (SEP) is in the range of 20-25 meters and that the velocity error is 0.1 to 0.2 meters per second. Using the established conversion factor of 0.3 meters/nanosecond, this corresponds to time errors in the range of 60-75 nanoseconds. Figure 6 shows some typical position error and equivalent time error plots recently taken over an interval of approximately two hours from one of the pre-production receivers.

User Interface Options

Since its inception the NAVCORE I™ receiver has been designed to be a low-cost black-box GPS sensor. In addition to the 1-pps analog timing pulse output port, the set also has two RS-232 input/output ports. One is bidirectional for control and display functions. In addition to giving the user complete command of the set, this port outputs much health and status information, such as identities of the satellites being tracked and GDOP of the constellation, which are of value to the user's application. The other port is unidirectional and outputs a variety of processed digital information such as current position, time and time correction solutions. Figure 7 shows how the NAVCORE I™ sensor fits within a generic block diagram for time and frequency product applications.

In concept, this approach provides a high degree of flexibility for equipment manufacturers in the design of products for the time and frequency reference system market. In practice, this approach is providing equipment manufacturers with a proven GPS sensor for a new generation of low-cost, high-performance products.

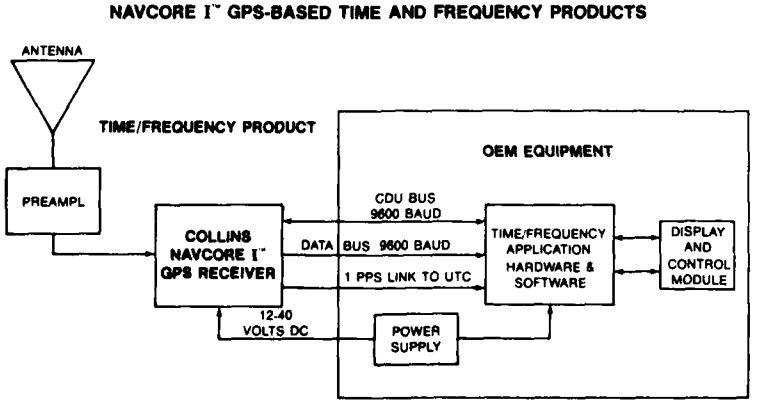


Figure 7. Time and Frequency Application

39th Annual Frequency Control Symposium - 1985

THE STATE-OF-THE-ART MEDIUM TERMINAL (SAMT) TIME AND FREQUENCY DISTRIBUTION SYSTEM

A. Vulcan and M. B. Bloch

Frequency Electronics, Inc.
Mitchel Field, New York 11553

Abstract

The SAMT Satellite Ground Station Distribution System uses redundant Cesium Frequency Standards in conjunction with automatic signal monitoring and switchover equipment to provide 5 MHz, 1 MHz, 1 PPS, and TOD to converter assemblies and other users at the ground station site. The unique features of the SAMT Distribution System are extremely low vibrational and operationally induced phase and amplitude perturbations and output isolation exceeding 100 dB.

Automatic switchover is accomplished by a method which ensures that the phase change at the system timing outputs is less than 0.001 degrees when switching from the on-line to standby atomic standard. Glitch-free performance is obtained by utilizing a dynamic phase matching system which incorporates A to D and D to A converters, and a micrologic memory which automatically slews the output phase of the standby Atomic Standard to agree with that of the on-line unit prior to switchover. A disciplined oscillator provides the 5 MHZ, 1 MHz and 1 PPS system outputs with extremely low phase jitter and a noise floor of -170 dBc/Hz.

Various sine wave and pulse output signals are fed to the system users by high isolation amplifier modules. Cascode circuitry and careful layout ensures that the isolation between the outputs exceeds 105 dB. This isolation permits changing terminations or removing output cables with less than 1 millidegree equivalent phase crosstalk.

All of the modular subassemblies used in the SAMT Distribution System are encapsulated to reduce vibration induced phase modulation. Drawers and modules can be removed and replaced while maintaining normal operation of the station. The modules are monitored for faults, and alarm signals are fed to a subsystem controller which ascertains the type of fault and feeds it to a central computer.

Introduction

The State-of-the-Art Medium Terminal (SAMT) Satellite Time and Frequency Generation and Distribution System was developed to supply basic timing signals for the ground station complex. Special circuitry and packaging techniques are utilized to ensure that operation of the station is not affected by normal user intervention. This requires that non-reciprocal, active isolators be used at all outputs to ensure that changing load impedances does not introduce phase or amplitude perturbations on other outputs which feed high order frequency multipliers and phase locked synthesizers. The atomic frequency standards feed a 5 MHz ultra low-noise, disciplined, quartz oscillator which suppresses phase glitches due to standard switchover. A computer controlled fault detection and module bypass capability ensures that no single point failures can occur. All of the modular subassemblies are carefully packaged to minimize vibration-induced phase noise. The total system is packaged in a standard seven foot high rack.

System Description

Figure 1 is the Functional Block Diagram of the SAMT System. Dual atomic frequency standards A1 and A2 provide outputs at 5 MHz, 1 MHz, 1 PPS and TOD. These standards receive redundant DC operating power from independent power supply drawers A6 and A7. Each atomic standard also has a fault output to flag the failure of any of the timing outputs or loss of internal phase lock. Each of the four signal outputs from the atomic standards feeds an Electronic Switch Module (ESM) which detects signal integrity and selects the appropriate standard for use as the system reference. The electronic switch modules use cascaded pin diode switches, having a total isolation in excess of 120 dB and insertion loss of less than 2 dB. This high degree of isolation is necessary to prevent beating between the sources. The 5 MHz and 1 MHz are selected by the diode switches, while the 1 PPS and TOD outputs are selected with low distortion JFET's. Selection of the appropriate standard for system reference input is accomplished either automatically or manually. In the automatic mode, the computer establishes priority and upon failure of the on-line standard, the operating standby is automatically selected by the ESM's. In the manual mode, which is used for maintenance purposes, selection of either standard is controlled by the operator. At no time can both standards be connectd to the system input.

The 5 MHz and 1 MHz from the selected atomic standard is fed via the ESM to the DTF bypass module A3 which, under normal operating conditons, feeds the selected 5 MHz signal to the Disciplined Time Frequency Standard (DTF) 5 MHz synchronization input. At the same time, the bypass module terminates the 1 MHz input and routes the 5 MHz and 1 MHz DTF outputs to the distribution amplifiers. Hence, the DTF is normally locked to the selected atomic standard and the 5 MHz and 1 MHz system signals are derived from the low noise DTF outputs. Special circuitry within the DTF effectively reduces output phase changes to less than 1 millidegree when the input 5 MHz locking signal is amplitude or phase perturbed.

The 1 PPS signal from the on-line ESM feeds a 1 PPS amplifier module which provides eight buffered 1 PPS outputs. The outputs are derived from low-noise, high-speed, complementary drivers which have less than 150 picoseconds of phase jitter. The 1 PPS amplifier accepts an external 1 PPS input, upon computer command, to provide the system output.

In the event of a DTF failure, the bypass module routes the on-line ESM outputs directly to the distribution amplifiers, bypassing the DTF. In this mode of operation, the system outputs retain the long term stability and accuracy of the atomic standard but the phase noise is degraded by 5 to 10 dB and glitch-free performance is compromised since the narrow-band, phase-locked flywheel in the 5 MHz and 1 MHz path is circumvented. Operation of the bypass module is determined by computer software which selects the bypass mode upon receipt of a fault signal from the DTF. A fault signal is generated if the integrity of the phase-locked loop, output signal, or oscillator oven temperature is compromised. The parameters are a sensitive indication of the performance of the DTF.

The 1 MHz and 5 MHz Bypass Module outputs drive appropriate multiple output, high isolation, distribution amplifiers. These amplifiers incorporate removable modules for each channel with 105 dB of isolation.

CH2186-0/85/0000-0153$1.00©1985IEEE

Residual phase noise is at least 8 dB below that of the input signal so that the amplifiers are essentially transparent to the input signal during normal operation. The 5 MHz and 1 MHz distribution amplifiers have provisions for software selection of external inputs, as does the 1 PPS amplifier. Each amplifier module has an internal voltage regulator to effectively eliminate the effect of ripple and noise on the B+ line, as well as reduce conducted emissions from the module. All of the modules used in the SAMT system incorporate this feature. In addition, each of the dual DC bus inputs is fused so that if a short circuit occurs within the module, the fuse opens and the overall bus remains active. Thus, power supply redundancy is maintained.

System power is supplied from redundant Power Supply and Battery Backup Units. Each of these rack panel assemblies provides +28 Vdc to the respective drawers. Internal 24-volt 23 ampere-hour batteries provide eight hours of system backup in the event of AC mains failure. Since the battery voltage and the AC power supply voltages are diode or-gated, and the outputs from each power supply drawer are diode or-gated within the modules, changeover to the standby power supply or battery is accomplished without introducing a B+ voltage interruption. The small, 3 to 4 volt bus change that does occur is effectively eliminated by the module regulators, and no timing error is observed.

Performance Characteristics

Table 1 itemizes the specifications for the SAMT Time and Frequency Distribution System. It is seen that a total of 20 sine wave outputs, eight 1 PPS timing pulses, and two Time-of-Day serial data streams are provided. The sine wave signals are derived from the DTF in the normal operating mode or the on-line Cesium Standard in the bypass mode. The output signal stability is determined by the atomic standard or DTF at sampling times more or less than 100 seconds respectively. An SC-cut resonator and low noise Pierce oscillator circuit provides the phase noise performance shown. The isolation between sine wave outputs is determined by the cascode amplifier modules. Special packaging techniques ensure that each output channel is physically isolated with RFI feedthru filters incorporated in the alarm and power interfaces. The high isolation ensures that phase perturbations are held below 0.0016 degrees when changing loads on the outputs or removing adjacent amplifier modules. In the SAMT system, phase changes at the outputs are multiplied by a factor of approximately 2000. This corresponds to a 3.2 degree change in phase at the X-Band output frequency.

Linear AC power supplies in conjunction with sealed lead-acid batteries, provide the DC operating power of +24 Vdc to the various assemblies. Low-noise linear supplies are utilized to minimize noise and ripple on the B+ lines whose effects would be multiplied just as phase noise or discrete spurious signals.

Subassembly Description

Cesium Standard

The Cesium Standard used for the SAMT application incorporates a hybridized, high stability design developed specifically to have low vibration induced phase noise. A functional Block Diagram of the Cesium Standard is shown in Figure 2. The main chassis consists of the following major assemblies.

Table 1

SAMT Performance Characteristics

OUTPUTS	
5 MHz at +13 dBm, 12 outputs	
1 MHz at +13 dBm, 8 outputs	
1 PPS, TTL, 8 outputs	
TOD, TTL, 2 outputs	
STABILITY	
$+7 \times 10^{-12}$ accuracy	
$\pm 5 \times 10^{-12}$ for 5 years	
$\pm 2 \times 10^{-12}$ from -10°C to +50°C	
150 PSEC jitter on 1 PPS	
PHASE NOISE	
10 Hz, -130 dBc/Hz	
100 Hz, -145 dBc/Hz	
1,000 Hz, -160 dBc/Hz	
10,000 Hz, -173 dBc/Hz	
ISOLATION	
-105 dB between all 5 MHz outputs	
-110 dB between all 1 MHz outputs	
0.0016° maximum phase perturbation with changing loads or standard switchover	
0.01 dB maximum amplitude perturbation	
POWER REQUIREMENTS	
70 watts, DC operation	
100 watts, AC operation	
8 hours of battery capacity	

CESIUM BEAM RESONATOR
LOOP ELECTRONICS
MODULATOR/MULTIPLIER
SYNTHESIZER
5.0 MHZ OCVCXO
1 MHZ GENERATOR
CLOCK
POWER SUPPLY

The CBR, Assembly A1, is a passive device, utilizing the hyperfine energy transition of the cesium atom which (in an ambient controlled magnetic field and a perturbation frequency at 9.192+ MHz) presents a DC signal intensity which is a function of perturbation frequency. If the perturbation frequency is modulated, an AC error signal is generated which indicates the center of the cesium resonance signal when the AC signal is zero. Hence, the CBR acts as a very narrow band discriminator and is used as the frequency determining element in a closed servo loop.

The output of the CBR detector is a low-level 83.33 Hz sine wave with an intensity of approximately 10^{-13} amps. This signal feeds the loop electronics, where it is amplified by a low-noise FET amplifier with 40 dB of gain. The FET output is further amplified by a bandpass amplifier whose output feeds a synchronous detector. The other input to the synchronous detector is an 83 Hz sine wave derived from the synthesizer. A fixed 2 millisecond delay compensates for the CBR time delay so that the inputs to the synchronous detector have the proper phase relationship. The phase detector output, which is zero

volts at lock, is filtered and fed to the 5 MHz OCVCXO modulation input. The OCVCXO frequency is pulled in a direction which tends to null the phase difference between the CBR output and the OCVCXO synthesized output. Hence, frequency locking is obtained.

The crystal oscillator assembly uses a double oven modified Pierce circuit with an SC-cut, 5th overtone quartz crystal to supply an ultra-high stability 5 MHz signal. The use of the SC-cut crystal ensures low phase noise performance in the SAMT operating environment.

Selected cesium resonators are used for the SAMT application to minimize frequency stability of the output signal. Figure 3 illustrates σ vs. τ over a sampling range of 10^3 to 10^7 seconds.

A frequency divider, amplifier, and filter assembly generates a low distortion 1 MHz output. The serial time-of-day and 1 PPS outputs are produced from 5 MHz in the A9 clock module. Operating voltage for the various assemblies is generated from the dual +28 Vdc inputs in DC to DC converter A11.

Disciplined Time - Frequency Standard

The DTF circuit shown in Figure 4 provides output frequencies of 5 MHz, 1 MHz, and 1 PPS. It is phase-locked to a 5 MHz input and time synchronized to an external 1 PPS input. During normal operation, holding register 1 is enabled and register 2 is disabled. The VCXO frequency is phase-locked to the external 5 MHz input and the control voltage at the input to the electronic phase shifter remains constant. The memory storage circuit contains a voltage comparator which receives inputs from the D/A converter and the phase detector output, causing the up-down counter to count up or down resulting in a binary number. This number is stored in the non-destructive register and converted to a DC voltage which feeds the VCXO via the loop filter. This locks the oscillator with a resolution of 2.5×10^{-12}. Upon a perturbation of the 5 MHz reference input (due to Atomic Standard switchover or an operator induced phase step), holding register 1 is disabled and register 2 enabled, holding the VCXO input voltage and frequency constant. The electronic phase shifter is slewed until the output of register 2 agrees with register 1. This condition occurs when the 5 MHz input phase is shifted to agree with the value prior to interruption. At this time normal phase locking is restored and register 1 enabled. Extremely fast fault detection circuitry is required to initiate the relocking sequence before an error propagates to the VCXO input. The full sequence takes less than 10 seconds, during which time the system outputs are derived from the memorized DTF output.

Distribution Amplifier

Figure 5 is a block diagram of the twelve channel, 5 MHz Distribution Amplifier, configured for eight outputs. A double-throw, double-pole input switch selects the 5 MHz system input or an external source for distribution. The selected input feeds a 1x12 passive power splitter whose outputs drive the 12 amplifier modules. Each module receives redundant fused DC power inputs which are combined by gating diodes. The resulting +24 Vdc voltage is regulated to +18 Vdc by a three terminal, linear, low-noise, regulator. The regulator output feeds the amplifier circuit. The fuses are incorporated so that if a short circuit occurs within the module, the load will be removed from the power bus and the rest of the system will operate normally. Each module has a bilevel fault output which changes state when the output signal drops below a predetermined level. An analog monitor output whose voltage is proportional to the RF level, is also provided for level monitoring purposes.

Figure 6 is the low-noise cascode amplifier schematic diagram. The two-transistor circuit is temperature compensated to operate over a broad temperature range with a gain change of less than 0.2 dB from -20 to +65°C. Emitter degeneration in the input stage provides DC and AC stability and sets the gain of the module. A sample of the amplifier output is peak detected and fed to a comparator whose inverting input is biased with a reference voltage which determines the point at which the alarm is actuated. Typically, a decrease of 2 dB or greater in the output level is considered a fault condition. The comparator output illuminates a front panel fault indicator and feeds an inverting circuit which generates a TTL compatible alarm output. The amplifier is packaged in a shielded assembly, measuring 5 inches by 3 inches by 0.4 inches.

Power Supply Description

Figure 7 is a functional block diagram of the power supply drawer. Two redundant drawers are used in the SAMT system. A linear AC to DC power supply converts the line voltage to +28 Vdc. This power supply provides operating voltage for the system and simultaneously charges the batteries during normal operation. Five independently fused outputs feed the various drawers that comprise this system. Critical parameters are monitored and presented as visual displays and remote outputs. Fault signals are generated if any of the following conditions occur: Blown Fuse, Overheating, Low Output Voltage, or AC/DC Power Supply Fault.

Two 12 volt, 23-ampere hour, sealed lead-acid batteries are connected in series to provide the +24 Vdc output in the battery mode of operation. The batteries are diode or-gated with the power supply output to provide an uninterruptible voltage. Battery charging current, discharging current, and voltage are monitored on a front panel meter.

Since each module in the SAMT system has an internal voltage-regulator and, in the AC mode of operation, load regulation is provided by the DC power supply in the power supply drawer, dual regulation is incorporated. Consequently, ripple and injected noise are reduced to levels below -120 dBc. In addition, bus voltage variations caused by replacement of modules or system surges are not propagated to on-line active circuitry.

Mechanical Design

Figure 8 shows the overall SAMT system. A single 72 inch high rack encloses the 7 drawers. Starting at the top, these drawers are the two Cesium Standard/ESM Assemblies, DTF/Bypass/1 PPS Amplifier Assembly, 5 MHz Distribution Amplifier, 1 MHz Distribution Amplifier, and two Power Supply/Battery Backup Units.

Summary

A State-of-the-Art Medium Terminal Time and Frequency Generation and Distribution system is described which was developed to provide the basic reference signals for a high performance satellite ground terminal. The design emphasizes low phase noise, low spurious signals, negligible vibration induced sidebands, and high reliability. An eight hour battery backup capability is incorporated.

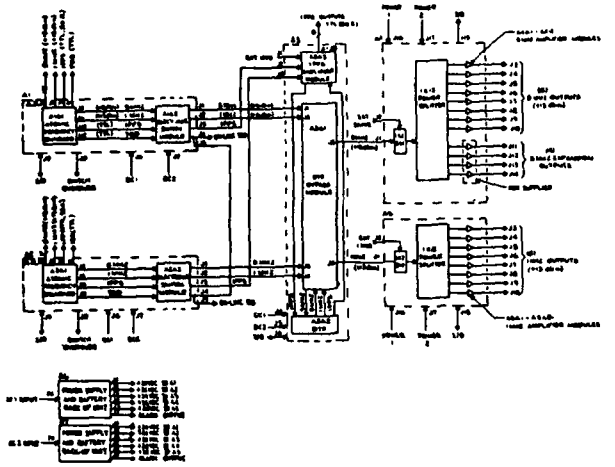


Figure 1. Block Diagram, Signal Generator Group (SGG) Model FE-5115A

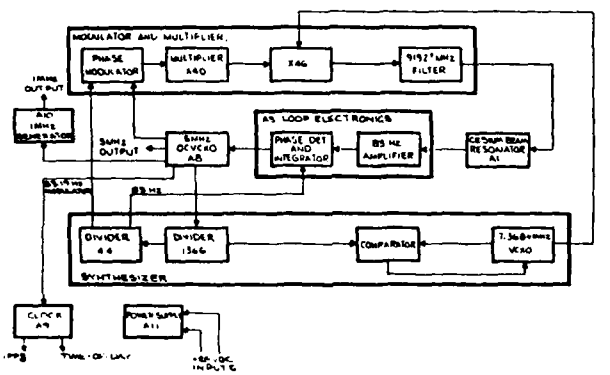


Figure 2. Block Diagram, Cesium Standard

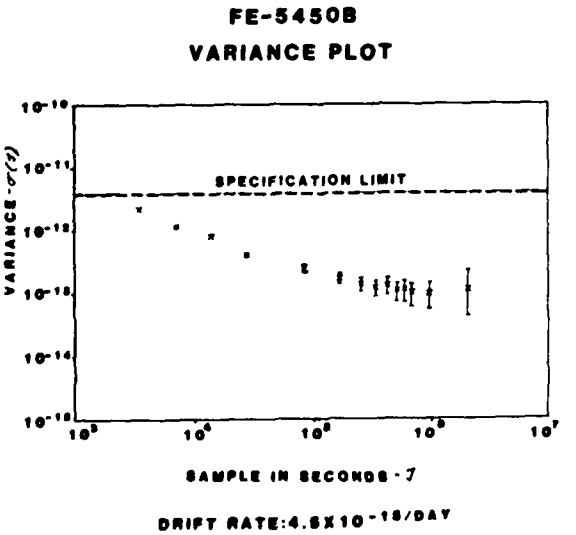


Figure 3. Cesium Standard Stability

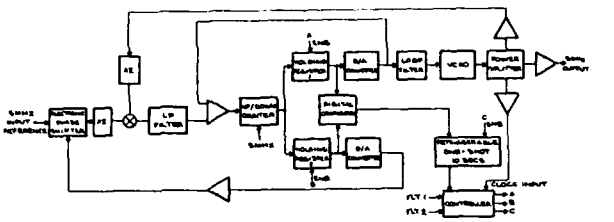


Figure 4. Block Diagram, Disciplined Time Frequency Standard

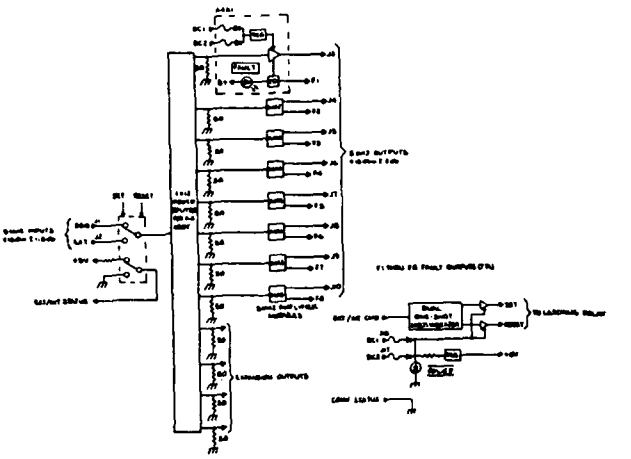


Figure 5. Block Diagram, 5 MHz Amplifier

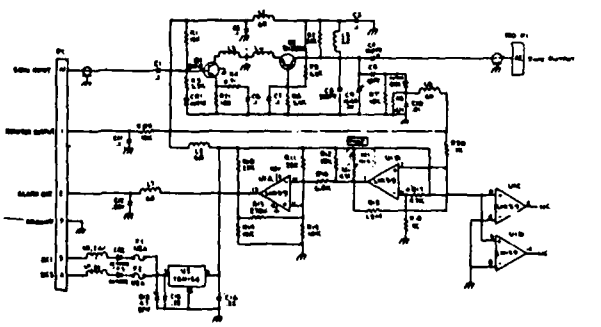


Figure 6. Schematic Diagram, Amplifier Module, 5 MHz

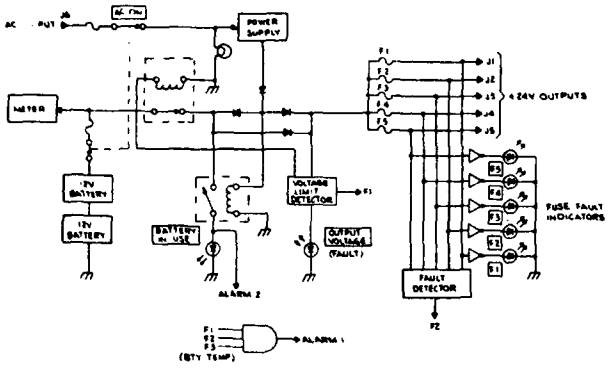


Figure 7. Block Diagram, Power Supply/Battery Backup Unit

Figure 8. State-of-the-Art Medium Terminal (SAMT) Frequency/Time Standard, FE-5115A

A 9.2 GHz SUPERCONDUCTING CAVITY STABILIZED OSCILLATOR

Bokuji KOMIYAMA

Radio Research Laboratory
Koganei Tokyo 184, Japan

Abstract

The experimental results on the frequency stabilization of a Gunn diode oscillator by the 9.2 GHz TM010 niobium superconducting cavity (unloaded Q = 3×10^{8}) are described, where the cavity is used as a frequency determining element in the frequency discriminator circuitry. The frequency stabilization technique was reviewed. It was shown that the frequency pulling by a square law detector tuning circuit in the discriminator considerably degrades the frequency stability of a stabilized oscillator (SCSO) and also shown that the pulling can be sufficiently reduced by proper design of the discriminator. The frequency stability of the SCSO was measured by comparing it with the auto-tuned hydrogen maser. The frequency stability of $\sigma_y(\tau)$ reached 1.4×10^{-14} at τ = 500 s, where the linear frequency drift was removed. The measured S/N of the discriminator shows that even with the comparatively low Q cavity, short-term frequency stability of $1.1\times10^{-14}/\sqrt{\tau}$ will be achieved in the SCSO. The frequency fluctuation due to mechanical deformation of the cavity caused by the tilt and vibration of the dewar, however, seems to be serious. The measured accelelation sensitivity of the SCSO was 6.5×10^{-8}/g for the vibration frequency less than 80 Hz.

Introduction

A superconducting cavity stabilized oscillator (SCSO) has received considerable attention [1]-[3] due to its excellent short-term frequency stability, which results pimarily from the ultra-high Q and the small temperature coefficient of superconducting cavities and also from the simplification of temperature regulation at cryogenic temperature.

The SCSO with short-term frequency stability of $\sigma_y(\tau) = 6\times10^{-16}$ has been reported by Stein etal [4], where the X-band Gunn diode oscillator is stabilized by the superconducting cavity with loaded Q of 10^{10}.

For frequency stabilization, superconducting cavity resonators fabricated from bulk niobium are preferred [5]. While superconductor-sapphire cavities are studied for frequency stabilization recently [6], [7], measured properties of them have not been obtained yet. Several methods have been developed to manufacture niobium cavities [8]-[10]. The investigation on superconducting microwave cavities, however, has been performed mainely in the field of high energy physics in order to utilize them as linear accelerators, where major interests in superconducting cavities are how to realize high acceleration fields in cavities without degradation of Q.

Therefore, works on superconducting cavities for precision frequency stabilization are few and detailed data are published only for the cavity fabricated by the method developed at Stanford University [5], where the cavity is assembled by electron-beam welding and then alternately fired in ultra-high vacuum and chemically polished [8]. In addition, high quality niobium cavities, whose Q's are as high as 10^{10}, seem to be available still only in limited laboratories.

We have studied on a 9.2 GHz SCSO in order to apply it to evaluation of the Cs primary frequency standard of Radio Research Laboratory (RRL) and to a starting oscillator of frequency synthesis in the infra-red [11]. This paper presents experimental results on the frequency stabilization of a Gunn diode oscillator by the 9.2 GHz TM010 superconducting cavity. The cavity was machined from a solid rod of niobium and was processed by electropolishing, anodizing and degassing under an ultra high vacuum condition [10].

The frequency of the Gunn diode oscillator was locked to the resonant frequency of the cavity by the technique of a cavity frequency discriminator [12]. The frequency stabilization technique was reviewed. It was shown that the frequency pulling caused by a square law detector tuning circuit in the discriminator degrades considarably the frequency stability of an SCSO and also shown that the pulling can be sufficiently reduced by a proper design of the discriminator.

The performance of the SCSO was measured by comparing it with the auto-tuned hydrogen maser. The results show that even with the comparatively low Q (3×10^{8}) cavity, an excellent short-term frequency stability will be achieved in the SCSO. The frequency fluctuation due to mechanical deformation of the cavity caused by the tilt and vibration of the dewar, however, seems to be serious.

9.2 GHz superconducting cavity resonators

The cross-sectional view of the fabricated TM010 9.2 GHz cavity is shown in Fig. 1. The cavity of spherical shape was composed of two half cells which were machined from a solid rod of niobium. The two halves were electron-beam welded at the center. The cavity was designed by the group of National Laboratory for High Energy Physics (KEK). Surface preparation of the cavity was done by electropolishing, anodizing and by degassing under an ultra-high vacuum condition [10]. After the surface processing, a copper tubulation and a coaxial probe with a ceramic window were attached to the cavity. Then the cavity was sealed off after evacuating. In our case, the geometry of the spherical shape aimed at improvement of the flow of electrolyte inside cavities during electropolishing. Its effect for increasing quality factor, however, was not observed.

The cavity was mounted in a vacuum can which was immersed within a helium dewar of 130 mm diameter. The basic design of the cryogenic apparatus was similar to [13].

CH2186-0/85/0000-0159$1.00©1985IEEE

Frequency stabilization technique

A blockdiagram of the SCSO is shown in Fig. 2 where the superconducting cavity is used as a frequency determining element in a frequency discriminator circuitry. The detail of this frequency stabilization technique is described in ref. [12]. The divided power output of the Gunn diode oscillator employed by the frequency discriminator is phase modulated at a frequency of f_m which is much larger than 3 dB bandwidth of the loded Q of the cavity. The reflected power from the cavity is detected by a square law detector and after amplified, this signal is product-detected with the modulation oscillator of the proper phase. Finally, the output voltage Vx of the discriminator is

$$V_x = 4bR_v\eta AP_dQ_U(\omega - \omega_0)/\omega_0$$
$$= 4bR_v\eta AU(\omega - \omega_0)$$

where b and Rv are the current sensitivity and the video resistance of the square law detector respectively, η : the ratio of the first sideband to the carrier amplitude, A : the combined gain of the amplifier and mixer, Pd : the dissipated power of the cavity, $\omega_0=2\pi f_0$: resonant frequency of the cavity, $U=P_dQ_U/\omega_0$: the stored energy, and Q_L and Q_U are the loaded and unloaded Q's of the cavity respectively.

By using Laplace notation, the frequency fluctuation of the SCSO $\Delta\omega(s)$ is

$$\Delta\omega(s) = \frac{G(s)}{1+G(s)}\Delta\omega_r(s) + \frac{1}{1+G(s)}\Delta\omega_G(s) + \frac{G(s)}{1+G(s)}\frac{e_n(s)}{K_d} \quad (1)$$

where $\Delta\omega_r(s)$ is the fluctuation of the resonant frequency of the cavity, $\Delta\omega_G(s)$: the frequency fluctuation of the free-running Gunn diode oscillator, $e_n(s)$: output noise voltage of the discriminator, $K_d=4bR_v\eta AU$: the sensitivity of the discriminator and G(s) is the open loop gain. The third term of the right side in eq. (1) is the frequency fluctuation caused by S/N of the discriminator, and when $\Delta\omega_G(s)$ is fully suppressed, the short-term frequency stability of the SCSO is limited by the S/N.

Frequency pulling caused by the square law detector tuning circuit

Square law detector tuning circuits in the frequency discriminator can cause frequency pulling. Fig. 3 shows an equivalent representation of the detector circuit. Here, R_0 is the source resistance, Rj the junction resistance of the diode and R, L and C are the equivalent series resistance, inductance and capacitance of the detector circuit respectively.

The reflected signal Vr from the cavity is expressed [12] as

$$V_r=V_i\{|\Gamma| e^{j(\omega t+\psi)} - \eta(e^{j(\omega+\omega_m)t} - e^{j(\omega-\omega_m)t})\}$$

where $|\Gamma| e^{j\psi}$ is the reflection coefficient of the cavity. By taking only terms at the modulation frequency, the detected signal Vd by the square law detector is

$$V_d \propto |\Gamma| \sin(\psi + \phi - \frac{\phi_1 + \phi_2}{2}) \times \sin(\omega_m t + \frac{\phi_1 - \phi_2}{2}) \quad (2)$$

and

$$\phi = -\tan^{-1}\{2Q_d(\omega - \omega_d)/\omega_d\}$$
$$\phi_1 = -\tan^{-1}\{2Q_d(\omega + \omega_m - \omega_d)/\omega_d\} \quad (3)$$
$$\phi_2 = -\tan^{-1}\{2Q_d(\omega - \omega_m - \omega_d)/\omega_d\}$$

where $\omega_d=2\pi f_d=1/\sqrt{LC}$ is the resonant frequency of the tuning circuit and $Q_d=\omega_d L/(R_0+R+Rj)$ is the loaded Q of the tuning circuit. Eq. (2) was derived assuming

$$\{2Q_d(\omega - \omega_d)/\omega_d\}^2 << 1$$
$$\{2Q_d(\omega + \omega_m - \omega_d)/\omega_d\}^2 << 1 \quad (4)$$
$$\{2Q_d(\omega - \omega_m - \omega_d)/\omega_d\}^2 << 1$$

For the frequency locking, Vd = 0. Then,

$$|\Gamma| \sin(\psi + \phi - \frac{\phi_1 + \phi_2}{2}) = 0 \quad (5)$$

By using $\tan^{-1}x=x-(x^3/3)$ for $x<<1$, we have from eq. (3), (4)

$$\phi - \frac{\phi_1 + \phi_2}{2} = -(Q_d/\omega_d)^3\omega_m^2(\omega - \omega_d) \quad (6)$$

We get the following at frequency close to the resonant frequency of the cavity

$$|\Gamma| \sin\psi = -\frac{2\beta}{\beta + 1}\delta$$
$$|\Gamma| \cos\psi = \frac{\beta - 1}{\beta + 1} \quad (7)$$

where $\delta=2Q_L(\omega - \omega_0)/\omega_0$ and β is the coupling coefficient of the cavity. By substituting eq.(6), and (7) into eq.(5), we obtain

$$(f - f_0)/f_0 = -P_L(f_d - f_0)/f_0 \quad (8)$$

where $|\phi-(\phi_1+\phi_2)/2|<<1$ and $|P_L|<<1$ are assumed. P_L is the pulling factor and given as

$$P_L = \{2(\beta - 1)/\beta\}\{(Q_d)^3/Q_L\}(f_m/f_0)^2 \quad (9)$$

f_d may vary with both temperature and drive level of the diode.

With $\beta=1.3$, $Q_L=10^8$, $Q_d=500$, $f_d=f_0=9.3$ GHz and $f_m=1$ MHz, typical values in our case, we obtain $P_L = 6.7\times10^{-9}$. Then, the fractional frequency fluctuation of f_d of 2×10^{-5} will result in a fractional frequency fluctuation of 1.3×10^{-13} in the SCSO. This example shows that reduction of the pulling factor is very important in an SCSO with a comparatively low Q superconducting cavity or in a low temperature crystal oscillator [14]. Besides by increasing Q_L, P_L can be reduced by lowering f_m or Qd.

Operating condition and design of the feedback loop

The cavity operating temperature is 1.6 K. Q_U and Q_L of the cavity at this temperature were 1.3×10^8 and 3×10^8 respectively and the resonant frequency was 9.289 GHz. Fig. 4 shows the measured temperature

coefficient of the cavity. The variation in the effective electrical length of the microwave resonator with temperature is the dominant effect on the resonant frequency shift above 1 K and the penetration depth of the rf magnetic fields, which changes the electrical length, varies exponentially with temperature [5]. In our case logarithm of the fractional frequency plotted in Fig. 4, however, dose not change linearly with inverse of temperature. The temperature coefficient is 1.5×10^{-9}/K at 1.6 K. The temperature of the cavity is regulated within 10 uK [15].

The modulation frequency is selected as 1 MHz conventionally [12]. η is set equal to 0.2. e_n in eq.(1) results primarily from noise voltage of the detector diode. We use a biased Schottky barrier diode as the square law detector in order to obtain better S/N of the discriminator. The detector is connected to the cavity through the circulator in order to reduce the unwanted reflected power from the diode to the cavity. The measured Qd was 150 which was lowered on purpose to reduce P_L. With the parameters mentioned above, P_L is calculated as 1.4×10^{-10}, which is sufficiently small. The measured value of bR_v was 1.8×10^{3} A^{-1}, which is not the highest one available because of intentional degradation of Qd.

The signal to noise ratio of the discriminator is proportional to the stored energy in the cavity. Fig. 5 shows the frequency shift of the SCSO vs input power of the cavity. Input power was increased from -36 dBm to -18 dBm and decreased to -36 dBm again. The curves show hysiterisis. What mechanism is responsible for the hysiterisis is under investigation. The results indicate that lower operating power is preferable in order to reduce frequency fluctuation due to fractional power variation. The cavity is operated at -30 dBm, that is, at a stored energy of 4.9×10^{-9} J, which is less by about two order of magnitude than in the case of ref.[12]. The calculated power spectrum of fractional frequency Sy(f) from the measured S/N of the discriminator was, however,

$$S_y(f) = 2.5 \times 10^{-28} / \text{Hz}$$

, which results in $\sigma_y(\tau)=1.1\times10^{-14}/\sqrt{\tau}$. Because we have not had a reference signal stable enough to evaluate the short-term frequency stability of the SCSO so far, we can not determine the optimum stored energy level in the cavity yet. The coefficient of fractional frequency shift over fractional power variation is 1.2×10^{-10} at the operating power level.

The output of the detector is amplified so as to be $A=10^3$ with the amplifier which has notch filters of both 2nd and 3rd harmonics of the modulation frequency. Temperature of the Gunn diode oscillator is regulated within 1°C. The measured frequency stability of the free running Gunn diode oscillator is expressed approximately as $\sigma_y(\tau)^2=(6\times10^{-9})^2+(8\times10^{-9}\tau)^2$ and the maximum frequency shift due to room temperature is less than 500 kHz. The tuning rate and output power of the Gunn diode oscillator are 900 kHz/volt and 20 mW respectively

We designed the feedback loop so that the fractional frequency fluctuation of the SCSO due to the Gunn diode oscillator will be suppressed to less than $\sigma_y=1\times10^{-15}$ for $\tau\gtrsim1$ s. The gain cross-over frequency is 158 kHz and the loop gain increases 6 dB/octave from the gain cross-over frequency down to 38 Hz and 12 dB/octave below 3 Hz. The cavity resonance adds a roll-off of -6 dB/octave for frequencies higher than the cavity loaded bandwidth [12]. The loop filter comprises two stage proportional and integral action network, one of which time constants is selected equal to f_b^{-1} in order to cancel a first order lag transfer function caused by the cavity resonance. $f_b=f_0/2Q_L$ is half of the cavity bandwidth. Theoretically, the loop will eventually track out any change of frequency of the Gunn diode oscillator. The maximum frequency offset of the Gunn diode oscillator over which the loop will hold lock is, however, limited by the dynamic range of the bias circuit in our case and is a few mega herz, which is sufficient for the Gunn diode oscillator used.

The loop characteristics are confirmed by measuring the spectral density of frequency fluctuation, Sy(f, Gunn), of the free running Gunn diode oscillator and Sy(f, system), of the SCSO respectively, because from eq.(1), the following relation is there between both spectra.

$$S_y(f, \text{system}) = \left|1 + G(s)\right|^{-2} S_y(f, \text{Gunn})$$

The results are shown in Fig. 6. Sy(f, system) was obtained by measuring the spectral density of the output voltage Vx of the discriminator in Fig. 2 and by neglecting the roll-off added by the cavity resonance. Fig. 6 shows that the predicted loop gain is achieved and frequency fluctuation of the Gunn diode oscillator is sufficiently suppressed.

Frequency stability measurement of SCSO

The frequency of the SCSO was measured by comparing it with the the 9.24 GHz signal frequency-multiplied from 5 MHz output of one of the auto-tuned hydrogen masers of *RRL*. The measured fractional frequency stability is shown in Fig. 7 where the linear frequency drift is removed for $\tau \gtrsim$ 50 s and the stability of the SCSO is assumed to be the same as that of the hydrogen maser. The stability reaches 1.4×10^{-14} at τ=500 s. The relative frequency stability of the two identical auto-tuned hydrogen masers measured at 5 MHz is also shown in Fig. 7, one of which was used as the reference oscillator for measuring the stability of the SCSO. The measurement bandwidth is 10 Hz for both cases and for the measurement of the SCSO, the lowpass filter with 12 dB/octave is used. For $\tau \lesssim$ 200 s, better results are obtained for the stability between the SCSO and the hydrogen maser than for that of the two hydrogen masers. The dotted line shows the calculated frequency stability $\sigma_y(\tau)=1.1\times10^{-14}/\sqrt{\tau}$ from the measured S/N of the discriminator, as given before, at the operating condition. These results indicate that a better short-term frequency stability than measured one will be achieved in the actual SCSO. At 500 s, the influence of the flicker floor seems to appear.

A typical example of the frequency drift ofthe SCSO is shown in Fig. 8. The record of room temperature is also given. The frequency was measured at averaging time of 14 s. After replenishing the outer dewar with liquid nytrogen, the large frequency variation is observed. The direction of frequency variation changes near the middle of the record, while room temperature increases monotonously with time. The fractional frequency variation is within 3×10^{-12} for more than 11 h. The drift is probably caused by mechanical deformation of the cavity due to the tilt of the dewar, which is changing with the level of liquid nytrogen.

The phase noise power spectrum $S_\phi(f)$ (rad^2/Hz) of the SCSO containes many discrete components. These components show a strong correlation with corresponding discrete ones of the vibration spectrum $S_{gr}(f)$ (g^2/Hz) of the dewar which was measured at the top flange of it for vertical direction. Fig. 9 shows the ratio of $S_\phi(f)$ and $S_{gr}(f)$ for corresponding discrete components between both spectra. The data were taken without taking any vibration isolation to the system. The plots go down with 20 dB/decade slope for vibration frequency less than 80 Hz. From the results, the acceleration sensitivity of the SCSO is calculated as $6.5\times10^{-8}/g$.

Fig. 10 shows the measured $S_\phi(f)$ of the SCSO after taking vibration isolation between the dewar and the floor and also between the vacuum rotary pump and the floor. The frequency multiplied signal from the 5 MHz crystal oscillator was used as a reference one. There is still a large discrete component which corresponds to the rotational speed of the pump (24.2 Hz) and this value is larger than the one predicted from the measured $S_{gr}(f)$ by an order of magnitude. We assume that this is because of insufficient vibration isolation for horizontal direction between the dewar and the pump through evacuating tubes. The contribution of this component to the measured $\sigma_y(\tau)$ of the SCSO in Fig. 7 is, however, negligible small because the lowpass filter used in that measurement attenuates it by about 15 dB.

The results of Fig. 8, 9 and 10 show that reduction of mechanical deformation of the cavity is extremely important in the SCSO.

Discussion

Besides frequency fluctuation sources of the SCSO mentioned above, several frequency noise sources are shown in ref. [13]. We, here, briefly discuss frequency fluctuation caused by dispersion of the waveguide and residual AM components of the phase modulator among those sources.

(1) Dispersion of the waveguide: The upper and lower sidebands and the carrier all have different phase velosities. By similar calculation to the frequency pulling by the detector tuning circuit, the corresponding fractional frequency offset $\Delta f/f_0$ is given as

$$\frac{\Delta f}{f_0} = -\frac{1}{8Q_L}\cdot\frac{\beta-1}{\beta}\cdot\frac{x^2y^2}{(x^2-y^2)^{3/2}}\cdot(f_m/f_0)^2\cdot L$$

$$x = 2f_0/c$$

$$y = \gamma/a$$

where c : the speed of light, a : width of rectangular waveguide, L : length of waveguide. For a typical example of our case, where L = 2 m and WRJ-10 is used,

$$\Delta f/f_0 = 1.4\times10^{-15}$$

The fractional frequency fluctuation will be much small compared to the offset.

(2) Residual AM components of phase modulator: It is pointed out that reduction of AM components of the phase modulator contributed much to the improvement of the frequency stability [4]. In our case, the power in AM sidebands of the modulator was about -38 dB below the phase modulation sidebands at the operating condition and we use the modulator without such servo control as described in ref. [12]. By a simple test of applying hot air to the modulator with a blower , we could not find the fractional frequency variation larger than 3×10^{-13}.

We are now trying to reduce P_L, instead of by lowering Qd, by lowering f_m and consequently the gain cross-over frequency by an order of magnitude than current system, which will result in the improvement of the S/N of the discriminator by about 6 dB. For better long-term frequency stability, lower operating temperature of the cavity and better vibration isolation of the dewar will be necessary.

We have operated the cavity for more than one year, and between experiments it has been stored in the vacuum can at room temperature. No degradation of Q has been observed yet.

Conclusion

The experimental results on the frequency stabilization of a Gunn diode oscillator by the 9.2 GHz TM010 niobium superconducting cavity are presented. The frequency stabilization technique was reviewed. It was shown that the frequency pulling by a square law detector tuning circuit considerably degrades the frequency stability of the SCSO and also shown that the pulling can be sufficiently reduced by proper design of the discriminator. The performance of the SCSO measured by comparing it with the auto-tuned hydrogen maser showed that even by comparatively low Q cavity, excellent short-term frequency stability will be achieved in the SCSO. The frequency fluctuation due to mechanical deformation of the cavity caused by the tilt and vibration of the dewar, however, seems to be serious and works on a cavity mounting system, which is immune against the tilt and vibration of the dewar, will be one of the most important subjects in SCSOs.

Acknowledgement

The author would like to thank T.Furuya and T.Nakazato of KEK for their invaluable advice and discussions in fabricating cavities. He also thanks K.Yoshimura for his useful discussion and encouragement and Y.Yasuda for his continuous encouragement to perform this work.

References

[1] S.R.Stein,"Space applications of superconductivity:resonators for high stability oscillators and other applications", Cryogenics, pp.363-371, July 1980

[2] D.M.Strayer, G.J.Dick and E.Tward,"Superconductor-sapphaire cavity for an all-cryogenic SCSO",IEEE Trans. Magn., Mag-17, pp.512-515, May 1983

[3] A.G.Mann and D.G.Blair,"Ultra-low phase noise superconducting-cavity stabilized microwave oscillator with application to gravitational wave detection",J. Phys. D:Appl Phys.,16, pp.105-113 1983

[4] S.R.Stein and J.P.Turneaure,"Superconducting-cavity stabilized oscillators with improved frequency stability",Proc. IEEE, 63, pp.1249-1250, Aug. 1975

[5] S.R.Stein,"Application of superconductivity to precision oscillators" ,Proc. 29th Ann. Symp. on Frequency Control, pp.321-327, 1975

[6] D.G.Blair and S.K.Jones,"High Q sapphaire loaded superconducting cavities and application to ultra stable clocks",IEEE Trans. Magn., Mag-21, pp.142-145 March 1985
[7] G.J.Dick and D.M.Strayer,"Development of the superconducting cavity maser as a stable frequency source",Proc. 38th Annu. Symp. on Frequency Control ,pp.435-446, 1984
[8] J.P.Turneaure and N.T.Viet,"Superconducting Nb TM010 mode electron-beam welded cavities",Appl. Phys. Lett., 16, pp.333-335, May 1970
[9] H.Diepers and H.Martens,"Superconducting niobium TE011 mode cavities with high critical magnetic field and high Q value",Phys. Lett., 38A, 5, pp.337-338, Feb. 1972
[10] T.Furuya, T.Hosoyama, T.Kato, Y.Kojima and O.Konno,"C-band superconducting cavity research at KEK", in Proc. 1979 Linear Accelerator Conference, Montauk, p.194, 1979
[11] B.Komiyama and Y.Yasuda,"Superconducting cavity stabilized oscillators by self-injection locking", Electronics and Communication in Japan, 63-B, 4, pp.50-57, 1980
[12] S.R.Stein and J.P.Turneaure,"The development of the superconducting cavity stabilized oscillator", Proc. 27th Annu. Symp. on Frequency Control, pp.414-420, 1973
[13] S.R.Stein,"The superconducting-cavity stabilized oscillator and an experiment to detect time variation of the fundamental constants", Ph. D. Dissertation, Stanford University 1974
[14] B.Komiyama,"Quartz crystal oscillator at cryogenic temperature", Proc. 35th Annu. Symp. on Frequency Control, pp.335-339, 1981
[15] B.Komiyama,"New design approach to precision thermostats at liquid-helium temperature", to be published in Rev. Sci. Instrum.

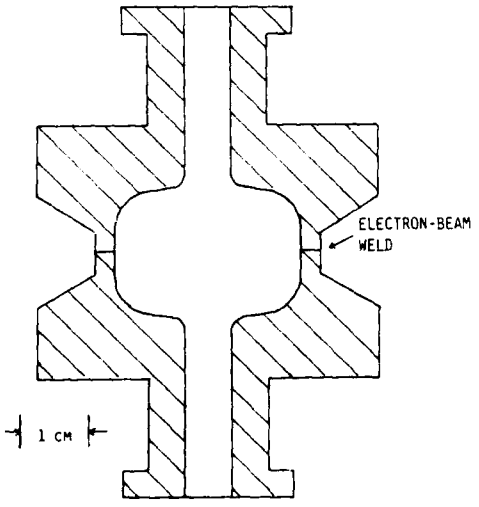


Fig. 1. Cross-sectional view of TM_{010} mode spherical cavity.

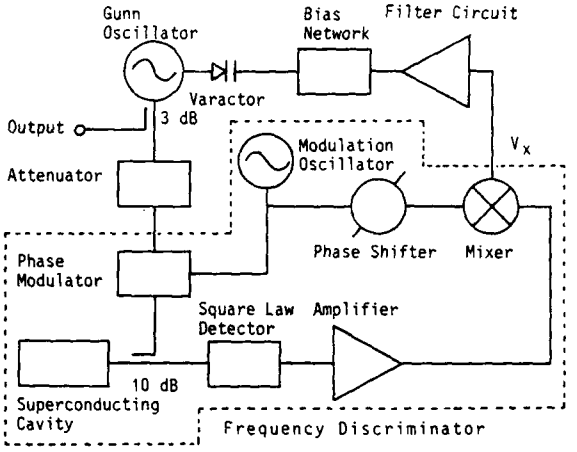


Fig. 2. Block diagram of the SCSO. The output voltage of the frequency discriminator is denoted as V_x.

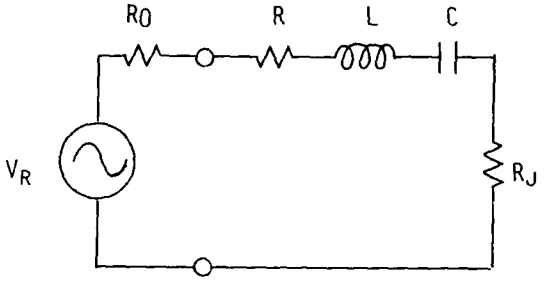


Fig. 3. Square law detector equivalent circuit. V_r is the reflected signal from the cavity.

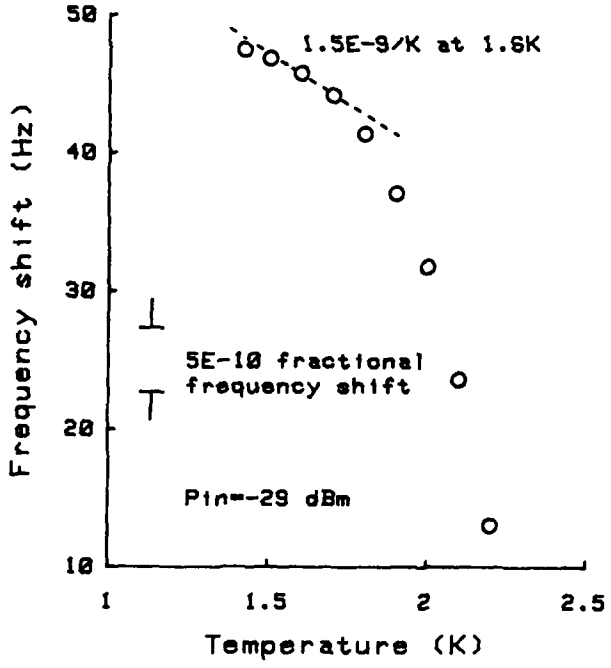


Fig. 4. Frequency variation versus operating temperatuare of the cavity. Incident power into the cavity is -29 dBm. The zero on the vertical axis is chosen arbitarily.

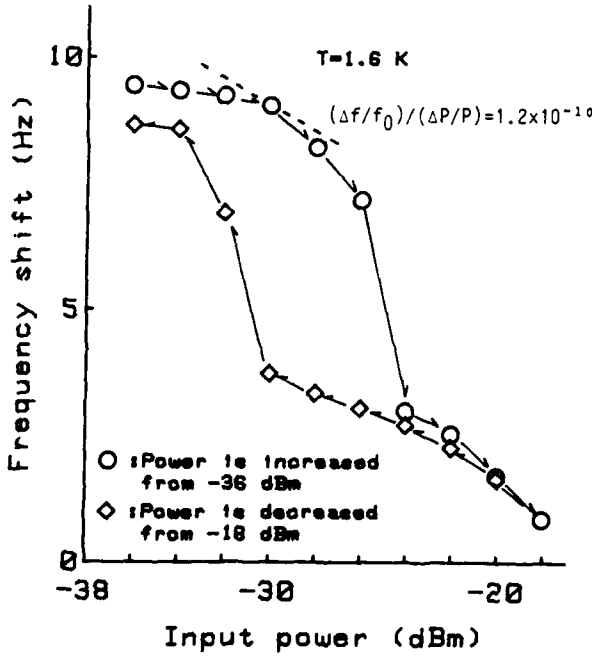


Fig. 5. Frequency shift of the SCSO versus input power of the cavity. Cavity operating temperature is 1.6 K. The zero on the vertical axis is chosen arbitarily.

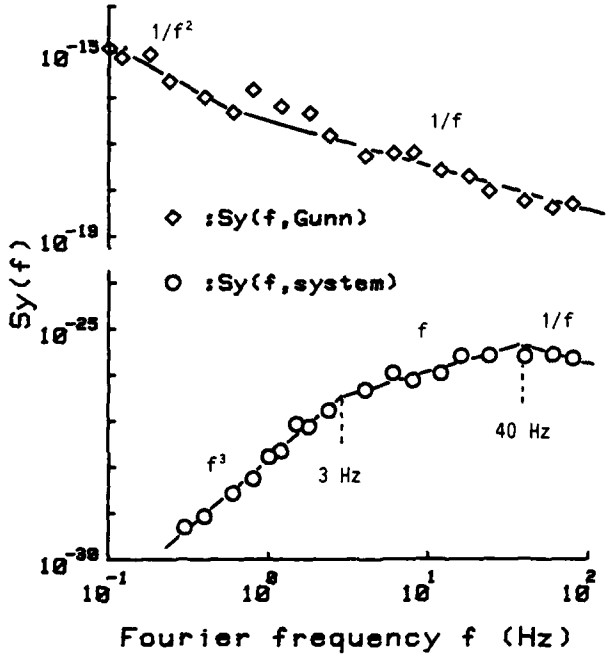


Fig. 6. Spectral density of fractional frequency fluctuation. $S_y(f, Gunn)$ is for the free running Gunn diode oscillator and $S_y(f, system)$ for the SCSO.

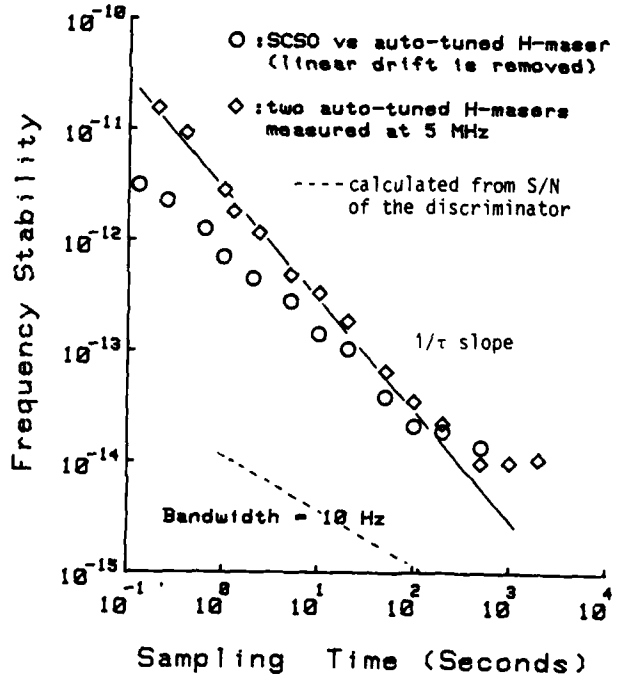


Fig. 7. Measured fractional frequency stability. Circle shows $\sigma_y(\tau)$ of the SCSO, where linear frequency drift is removed for $\tau \geq 50$ s and number M of independent samples is ; M=100 for $\tau \leq 5$ s, M=50 for 10 s $\leq \tau \leq$ 50 s and M=25 for $\tau \geq 100$ s. Square shows $\sigma_y(\tau)$ of the two identical auto-tuned hydrogen masers, where M=100 for $\tau \geq 50$ s, M=50 for 100 s $\leq \tau \leq$ 500 s and M=25 for $\tau \geq$1000 s. Dotted line is frequency stability of the SCSO calculated from the S/N of the discriminator.

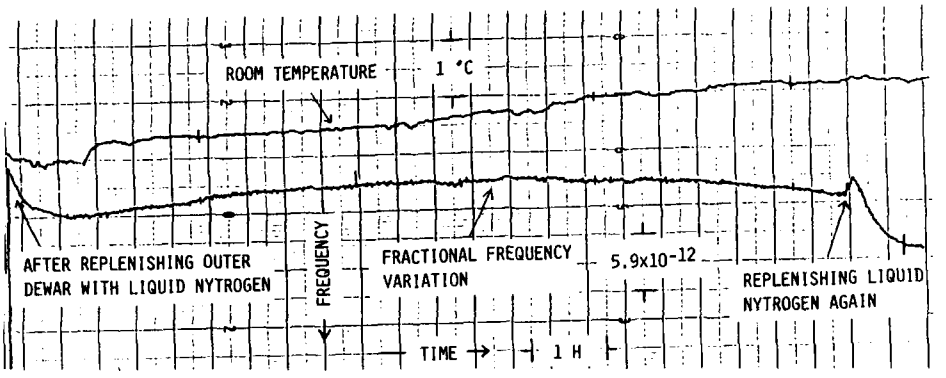


Fig. 8. Examples of frequency drift of the SCSO. The record of room temperature is also shown.

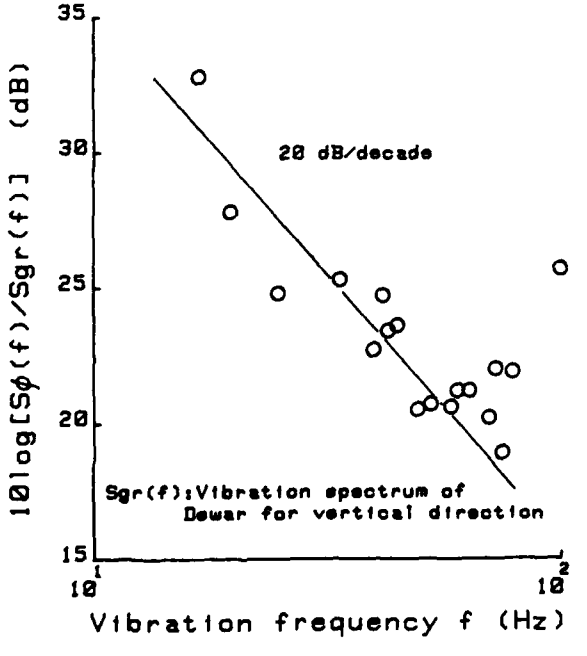


Fig. 9. The amplitude ratio of the discrete components between the phase noise power spectrum $S_\phi(f)$ of the SCSO and the vibration spectrum $S_{gr}(f)$ of the dewar for vertical direction.

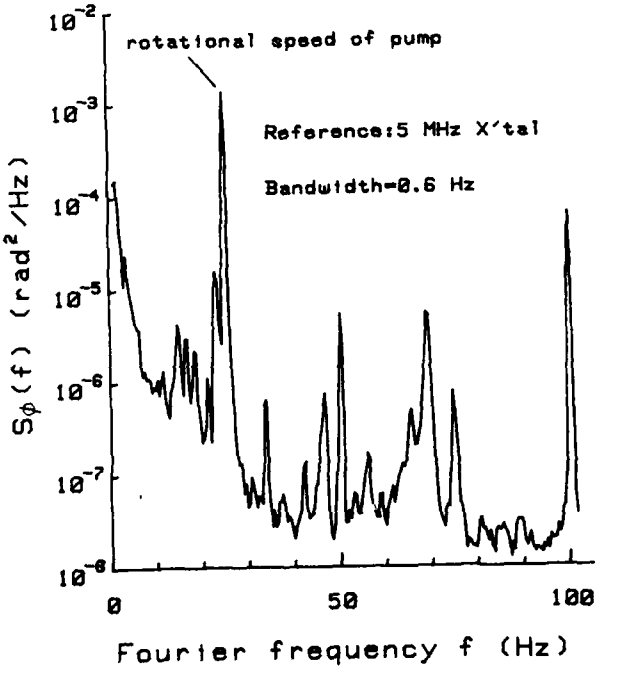


Fig. 10. Phase noise power spectrum of the SCSO. The frequency-multiplied signal from the 5 MHz crystal oscillator is used as a reference one. The fourier frequency of the highest peak corresponds to rotational speed of the pump.

REDUCING TCXO ERROR AFTER AGING ADJUSTMENT

R.L. Clark

McCoy Electronics Company
Mt. Holly Springs, PA 17065

Abstract

Temperature compensated oscillators require periodic adjustment to correct for aging drift. For internally compensated oscillators, where the oscillator is varactor tuned over temperature, the aging adjustment is performed by varying a reactive element in the oscillation loop. However, the varactor represents a different portion of the total load reactance after aging adjustment thereby altering the original frequency sensitivity to varactor bias changes. This altered sensitivity disturbs the original frequency versus temperature characteristics and produces a trim skew error. To fully exploit the benefits of digital compensation, and to satisfy trim skew error specifications for certain new military programs such as SINCGARS, a method of reducing trim skew error for TCXO's is required.

An alternative approach to performing the aging adjustment would be to change the DC bias voltage of the varactor. If the oscillator frequency versus voltage curve is totally linear, the circuit sensitivity to varactor bias voltage (ppm/V) will remain constant and no additional error in the frequency vs. temperature characteristics will be introduced. Voltage linearity requirements for both series and parallel resonance oscillators are discussed along with an error analysis of TCXO employing a parallel resonant crystal. Comparative data is presented for both capacitor and varactor trimmed oscillators. Both the data and the error analysis suggest that varactor trimming may be a viable technique for reducing trim skew error in TCXO's.

Introduction

Temperature compensated oscillators require periodic adjustment to compensate for aging drift. For internally compensated oscillators, where the frequency is varactor tuned over temperature, the capacitive trim sensitivity is altered after aging adjustment and therefore the original temperature characteristics may be disturbed. This disturbance is termed trim skew error.

Gruen et al[1] introduced the effect and Galla and McVey[2] formulated approximations for trim skew error for several circuit configurations utilizing a capacitor trimmer. However, little progress has actually occurred towards reducing the effect of trim skew error.

For low tolerance oscillators, trim skew error is not important, but as tolerances become tighter, it can represent a significant portion of the total specified frequency tolerance.[2] Figure 7 shows typical trim skew error for a TCXO.

This paper examines a method of reducing trim skew which utilizes the varactor bias voltage as the frequency trim mechanism. If the frequency vs. varactor voltage curve is totally linear, and varactor voltage is used to temperature compensate and adjust for aging drift, trim skew will reduce to zero. A critical parameter then becomes the amount of non-linearity in the frequency vs. voltage curve. An approximation for trim skew error is derived for a parallel resonant crystal employing a series varactor. The resulting error is compared to the error which would result from using capacitive trimming as the aging adjustment mechanism. It is shown that the use of voltage trimming can significantly reduce trim skew error. To quantify trim skew, it is necessary to examine the underlying causes of error.

Source of Error

Trim skew error can easily be illustrated by examining the common relationships for frequency pull (ppm) and trim sensitivity (ppm/pf).

$$\frac{\Delta f}{f} = \frac{C_1 \times 10^6}{2(C_o + C_L)} \quad \text{(ppm)} \qquad (1)$$

$$\frac{\partial \left(\frac{\Delta f}{f}\right)}{\partial (C_L)} = \frac{-C_1 \times 10^6}{2(C_o + C_L)^2} \quad \text{(ppm/pF)} \qquad (2)$$

As C_L is varied in (1) to correct for aging drift, the capactive trim sensitivity given by (2) is altered. Unless the total range of the temperature compensating reactance is changed in proportion to the new sensitivity, the TCXO will not be properly compensated after aging adjustment.

For oscillators employing an inductor or capacitor as the aging adjustment mechanism, the elimination of trim skew requires that the slope of the frequency vs. varactor voltage curve be independent of incremental changes in the frequency adjust trimmer. Capacitive trim sensitivity is the slope of the frequency vs. voltage curve. One method of making the slope of the frequency vs. voltage curve independent of the aging trimmer would be to design a network which has a constant capacitive trim sensitivity. However, constant trim sensitivity implies a linear relationship between load capacitance and frequency. Achieving this linearity is a difficult requirement, even over small ranges of load capacitance. By increasing the capacitive sensitivity, the total load capacitance variation required to compensate the oscillator over temperature may be reduced. The trim sensitivity will be more constant over a smaller range of load capacitance, but the oscillator may become too sensitive to component aging and supply voltage variations.

If the capacitive sensitivity cannot be made constant, it is necessary to weight the temperature compensating voltage such that the load reactance change becomes a function of both temperature and aging adjustment. This weighting requires that the temperature compensating reactance range be increased or decreased in proportion to the changes of trim sensitivity brought by aging adjustment.

Voltage Weighting

One method of weighting the temperature compensating reactance is to use varactor bias voltage for temperature compensation and aging adjustment. Aging adjustment can be implemented in such a manner by letting

$$V_b = V_t - V_a \qquad (3)$$

CH2186-0/85/0000-0166$1.00©1985IEEE

where

V_b= varactor bias voltage

V_t= temperature compensating voltage

V_a= aging trim voltage

Figure 1 shows a typical realization of a circuit which will perform this function. Physically, the voltage control port of any voltage controlled oscillator can be used to adjust for aging drift.

If the frequency vs. voltage curve of such a circuit is totally linear, no trim skew error will result. Essentially, the temperature compensating voltage is weighted by the non-linear varactor voltage to capacitance transfer function such that the range of capacitance variation over temperaure becomes a function of the aging trim voltage. As the capacitive trim sensitivity is increased (or decreased) by aging adjustment, the varactor reactance change over temperature is decreased (or increased). In this context, capacitive trim sensitivity becomes meaningless in determining trim skew, and voltage sensitivity becomes the critical parameter.

Linearity Requirements

Of course, if the frequency vs. voltage curve is totally linear, no error will be introduced by aging adjustment. Because it is virtually impossible to produce an oscillator which has total linearity, we now seek an expression for trim skew error when total linearity is not present.

To determine exact trim skew error, one would subtract the frequency at the minimum and maximum voltages required to temperature compensate the oscillator before and after aging adjust. The worst error and greatest non-linearity typically occur at the lowest and highest varactor bias voltage. For AT crystals, the lowest and highest voltages will typically occurs at the crystal turning points. One would expect the greatest trim skew error to occur around the crystal turning points. Error can be determined at these voltages (or turning points) by simply substituting the effective load capacitance into the function relating frequency to load capacitance. Error would then be found from

$$E = f(C_L) - f(\bar{C}_L) + a \tag{4}$$

where a represents the maximum anticpated aging, and C_L and $\bar{C}_L$ are the effective load capacitances, before and after aging adjustment, respectively. However, as the greatest error can occur either at the low voltage, or the high voltage, it is necessary to evaluate expression 3 at four different load capacities, or the corresponding voltages. These voltages would correspond to the lowest and highest voltages which would be encountered if aging were either positive or negative. For design purposes, an approximation is desired.

Assuming no discontinuities, good linearity, and monotonacity in the frequency versus voltage characteristics, trim skew error can be approximated by

$$E = \Delta V_t \ (S1 - S2) \tag{5}$$

where S1 and S2 represent the linearized slopes of the frequency versus voltage curve, evaluated before and after aging correction. This is illustrated in Figure 3. By letting P represent the pull required to temperature compensate the oscillator, the voltage change required to pull the oscillator can be expressed as

$$\Delta V(T) = P/S1, \tag{6}$$

Expression (4) can now be rewritten as

$$E = P\left[1 - \frac{S2}{S1}\right] \tag{7}$$

Although equation (5) is useful if one has obtained an accurate voltage versus frequency curve, it does not aid in the selection of circuit components. Therefore, we seek a relationship between circuit values, maximum aging, crystal pull and trim skew error. It is assumed in the derivation of this expression that the crystal Co is not resonated out, and that the circuit is in the configuration of Figure 2. Similar expressions can be derived for other configurations such as series mode operation.

For the circuit of figures 2 and 3.

$$\frac{1}{C_L} = \frac{1}{C(V)} + \frac{1}{C_T} \tag{8}$$

and

$$\frac{1}{C_T} = \frac{1}{C_1} = \frac{1}{C_2} \tag{9}$$

Varactor capacitance can be modelled by

$$C(V_b) = \frac{K}{(\phi + V_b)^{\alpha}} \tag{10}$$

Where K and α are diode constants, ϕ is the contact potential, and V_b is the varactor bias voltage.[4] The slope of the frequency versus bias voltage curve can be obtained from

$$\frac{\partial\left(\frac{\Delta f}{f}\right)}{\partial V_b} \tag{11}$$

$$= \frac{\partial\left(\frac{\Delta f}{f}\right)}{\partial C_1} \cdot \frac{\partial C_1}{\partial C(v)} \cdot \frac{\partial C(v)}{\partial V_b} \tag{12}$$

$$= \frac{-C_1 \cdot K \cdot C_L \cdot 10^6}{2(C_o + C_L)^2 \cdot C(v) \cdot (\phi + V_b)^{\alpha}} \quad \text{(ppm/V)} \tag{13}$$

For constant voltage sensitivity, the product of the partial derivatives given by expression 12 must be made constant. It is interesting to note that the derivatives relating frequency to load capacitance and load capacitance to varactor capacitance in (12) will vary during aging adjustment, regardless of the trimming mechanism used. By utilizing voltage trimming, one introduces the ability to vary the derivative relating varactor capacitance to varactor bias voltage. This effect allows the product to be nearly constant, although the individual terms are non-linear, and may change significantly at different values of V_b.

Trim skew error may now be approximated by

$$E = P\left[1 - \frac{\partial\left(\frac{\Delta f}{f}\right)}{\partial V_b}\Bigg|_{V_b = V_1 + \Delta V_a} \div \frac{\partial\left(\frac{\Delta f}{f}\right)}{\partial V_b}\Bigg|_{V_b = V_1}\right] \tag{14}$$

where V_1 is the nominal compensation voltage and V_a is the voltage required for aging adjustment. Expression 14 can be written in terms of circuit values as

$$= P \left[1 - \frac{(C_O + C_L)^2 \cdot \overline{C_L}^2 \cdot (V_1 + \Delta V_A + \phi)^{\alpha-1}}{(C_O + \overline{C_L})^2 \cdot C_L^2 \cdot (V_1 + \phi)^{\alpha-1}} \right] \quad (15)$$

Obviously, a number of trade-offs exist in the selection of TCXO circuit values and it is rarely possible to optimize for all parameters simultaneously. For low trim skew error, the key factor is voltage linearity. The more traditional techniques of achieving linearity are to resonate Co from the circuit or to add linearization networks.[3] However, the former technique may be impractical at low frequencies, while the latter requires additional circuit components and time-consuming component selection. We now examine an alternative technique of linearization whereby the diode parameters are variables. If an inflection could be found in the frequency vs. voltage curve, the slope would be totally linear around this point and would reduce to zero. An inflection would occur, of course, when

$$\frac{\partial^2\left(\frac{\Delta f}{f}\right)}{\partial V_b^2} = 0$$

For $C_T >> C(v)$ this inflection occurs when

$$V_b = \frac{K}{C_O} \cdot \left[\frac{\alpha - 1}{\alpha + 1}\right]^{1/\alpha} - \phi \quad (16)$$

Equation 15 was programmed into a digital computer and circuit values were manipulated while subject to the relationships between C(V), C_L and C_t developed earlier. The inflection voltage indicated by (16) was used as a starting value. Crystal parameters were chosen as those typical for a 3.2 MHz fundamental AT resonator. The crystal parameters and aging were as follows:

C_1 = .012 pF

C_O = 3.2 pF

p = 40 ppm

a = 3 ppm

It was found that trim skew could be reduced to zero and that the optimal V_1 for zero error could be varied with other circuit values. The presence of an inflection is of significant value in obtaining good linearity. A general algorithm to optimize error subject to practical circuit constraints was not developed, but simple value manipulation yielded zero error when

α = 1.1

V = 3.25 volts

K = 300 pF

C_t = 160 pF

All other circuit parameters can be derived from these values. The resulting voltage sensitivity is 21 ppm/V at a load capacitance of 47 pF.

Because 16 is an approximation based on incremental values around a point of inflection, one must examine the exact frequency change at the voltage endpoints when determining trim skew error to a high degree of accuracy.

Calculating the voltage range required to compensate the above crystal indicates that the voltage must vary from

V_L = 2.344 V

to

V_H = 4.155 V

The required voltage change to correct +/-3ppm of aging drift is .1356 V. Substituting the resulting load capacitances into (3) yields an error of 2.2×10^{-8}.

For camparison purposes, the error resulting from using a trimmer capacitor was calculated. The resulting error for a frequency change of +/-3ppm was .135 ppm. This is a significant increase in error when contrasted to voltage trimming. In addition, the values chosen are also nearly optimal for capacitive trimming.[2]

The reduction in trim skew error can be graphically illustrated by plotting the derivative of the frequency versus voltage curve. Figures 4 and 5 show the derivatives of the frequency vs. voltage curves (voltage sensitivities) for capacitive and voltage trimming. Note that the turnover in the derivative curve will coincide with the inflection in the frequency vs. voltage curve. Clearly, the voltage sensitivity remains much more constant when voltage is used as the aging trim mechanism.

Experimental Results

As the most critical factor in determining low trim skew for this method is the frequency linearity, our first task was to ensure linearity.

Although 16 indicated that an inflection does not occur for $\alpha < 1$, we have experimentally been able to find an inflection for α as low as 0.7. It is the author's opinion that poor fit of the varactor model used is responsible for this error. Figure 6 shows frequency versus voltage for a typical oscillator used in the experiment. Centering the varactor bias around the inflection would, of course, minimize trim skew error. Graphically determining the trim skew error from this curve indicates a value of approximately .23 ppm. The trim skew error was also graphically computed for the case of capacitive trimming by varying the frequency with a load trimmer and using the new slope of the frequency vs. voltage curve in 5. This error was approximately .55 ppm. Figures 7 and 8 show frequency vs. temperature for voltage trimming and capacitive trimming. The oscillators were offset by +/-3ppm and results normalized such that the curves intersected at +30°C. An obvious reduction in error has occured.

Summary

Provided that good voltage linearity is present, trim skew error may be substantially reduced by using varactor bias voltage to adjust for aging drift. The technique is really no more difficult than using the control port of voltage controlled oscillator as the aging adjustment mechanism.

Additionally, if the proper diode characterisitics are present, good linearity may be achieved without resonating C_O from the circuit.

References

1. Gruen, H.E., Dominguey, C.D., Hardt, I.E., Kidman, R.D., "Frequency Temperature Compensation Techniques for Quartz Crystal Oscillators," Quarterly Report No. 9, Contract PA 36-039 AMC-022 82 (E), Oct., 1964.

2. Galla, W.D., McVey, E.S., "TCXO Error Due to Aging Adjustment," Proc. 34th Annual Symposium on Frequency Control, 1980.

3. Frerking, M.E., Crystal Oscillator Design and Temperature Compensation, pg. 122-125, Van Nostrand Reinhold, New York, 1978.

4. Gray, P.R., Meyer, R.G., Analysis and Design of Analog Intergrated Circuits, pg. 2-9, Wilex, New York, 1984.

5. Sherman, J.A., "Trim Sensitivity, A Useful Characterization of a Resonator," pg. 181-185, Proc. 33rd Annual Symposium on Frequency Control, 1979.

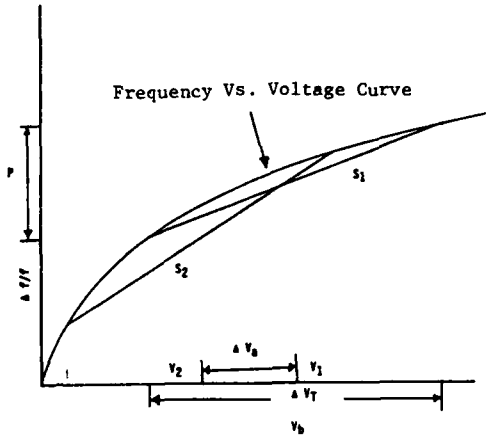


Figure 3. Linearity Requirements And Siope Characterization

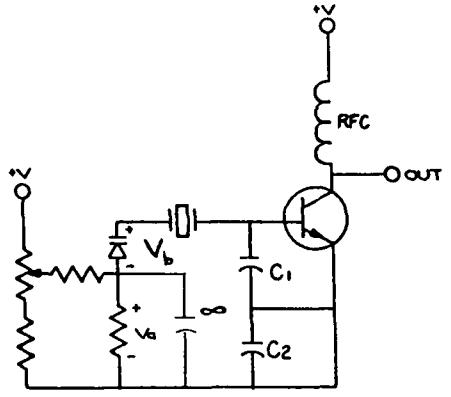


Figure 1. Typical Oscillator Utilizing Voltage Trimming

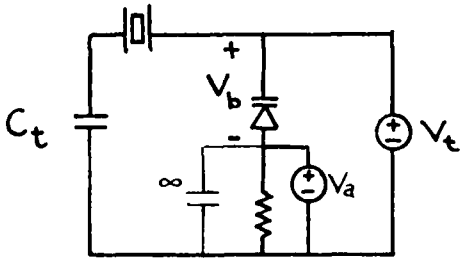


Figure 2. Crystal And Associated Components For Circuit Of Figure 2

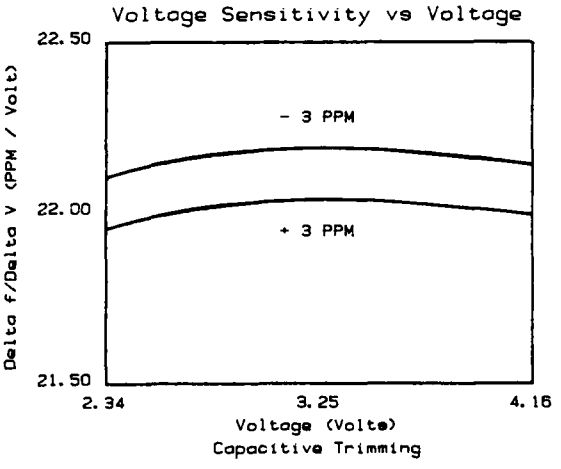


Figure 4. Derivative of Frequency Vs. Voltage (Voltage Sensitivity) As Frequency Is Pulled +/-3 PPM Using A Capacitor

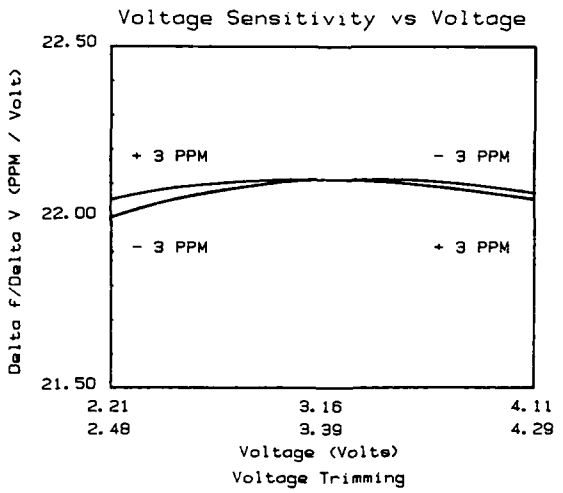


Figure 5. Derivative Of Frequency Vs. Voltage Curve (Voltage Sensitivity) As Frequency Is Pulled +/-3 PPM Using Voltage Trimming. The Higher Frequency Corresponds To The Higher Voltage Range.

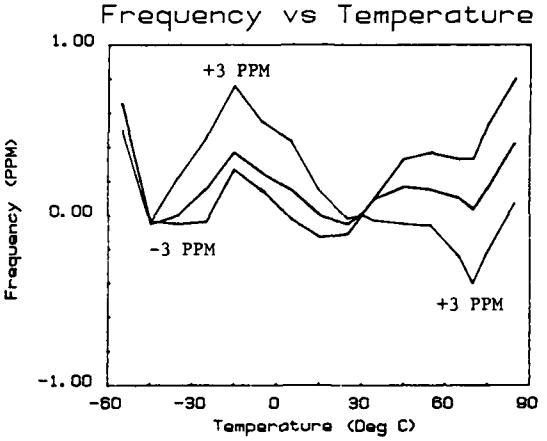


Figure 7. Relative Frequency Deviation As Frequency is Pulled +/-3 PPM Using A Capacitor

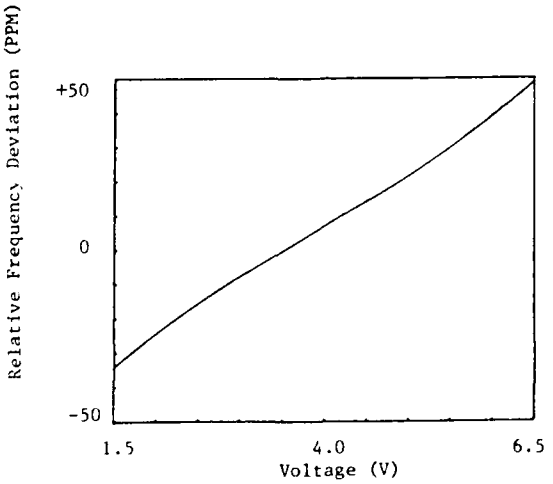


Figure 6. Typical Frequency Vs. Voltage Curve For Test Oscillator

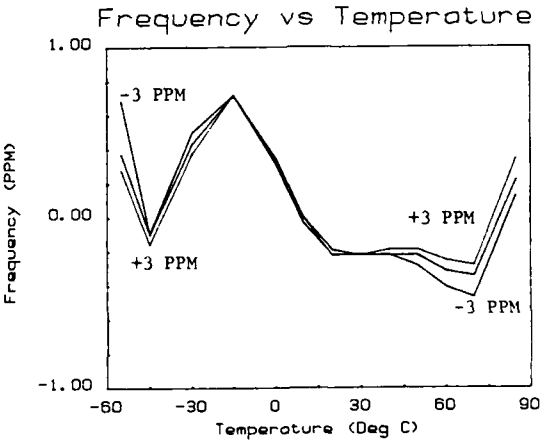


Figure 8. Relative Frequency Deviation As Frequency Is Pulled +/-3 PPM Using Voltage Trimming

RESULTS OF CONTINUED DEVELOPMENT OF THE DIFFERENTIAL CRYSTAL OSCILLATOR

Albert Benjaminson

S.T. Research Corporation, Newington, VA *

Summary

Since the first paper on Balanced Feedback Oscillators was presented at the 38th Annual Frequency Control Symposium,[1] the development of this new oscillator, now called the Differential Crystal Oscillator (DXO), has continued. We can now report on additional improvements achieved partly through the use of a computer-aided design program developed at S.T. Research.

Sideband phase noise levels, below those reported previously for 10 MHz low-noise oscillators, have been measured. These have been obtained through the DXO's ability to generate a high ratio of output power to crystal power dissipation, and to do this at transistor operating current levels selected for the lowest noise figure independent of crystal current. These, combined with a novel automatic level control circuit (ALC) that does not reduce transistor current, has made these results possible.

The DXO is a series-resonant oscillator that uses negative feedback through the resonator to stabilize the sustaining amplifier within the resonator's bandwidth, and to reduce its gain to nearly unity beyond that bandwidth. Within the resonator's bandwidth this has the effect of multiplying the phase slope of the resonator, thereby increasing the short-term stability with respect to circuit and component variations.

Positive feedback is applied independently of the negative feedback while automatic level control is achieved by controlling the positive feedback thru a low-noise attenuator. In this way we have been able to reduce phase noise sidebands to levels of -140 dBc at 10 Hz, and -160 dBc at 100 Hz offset from an SC-cut 10 MHz 3rd O.T. DX Oscillator.

The Differential Crystal Oscillator

Continued design, analysis and testing of the DXO since last year's introduction[1] has clarified its advantages and defined its limitations. This paper will consider four aspects of its design; its broadband noise performance, its influence on 1/f noise, the types of ALC it can use, and its operation with SC and overtone crystals.

Broadband Noise

Broadband noise comprises thermal noise and excess noise, such as the shot-noise and partition noise produced in semi-conductors, exclusive of 1/f noise. The Noise Figure quoted for transistors expresses the noise generated in excess of thermal noise under specified operating conditions.

The total noise level produced by the resistive component of an impedance is -174 dBm at room temperature. On average, one-half of this power will be evidenced as amplitude modulation of the oscillator's output power, and one-half as phase modulation. Since we are concerned only with the latter, its level will be -177 dBm plus the excess noise generated by the active devices.

It is at this point that we face the physical limits of oscillator design. An examination of Figure 1 which illustrates the basic DXO schematic shows an output voltage of 2v rms at each collector across collector loads of 1000 ohms each. This produces 6 dBm into each load at the operating frequency. A portion of the power going into the negative feedback branch is dissipated in the crystal (approximately 0.4 mW). This load is not present beyond the resonator bandwidth as the crystal impedance rises much above the series resistor value, and reduces the amplifier gain to unity. This makes it difficult to judge the noise level at sideband frequencies beyond the resonator bandwidth since no N.F. information is available for unity gain amplifiers. It can be stated that the S/N ratio for sideband frequencies beyond the 1/f region will be

$$S/N = +6 - 177 + N_O + N_A \text{ dBc}$$

where N_O and N_A are the noise figures for the oscillator and following amplifier, respectively.

Transistor noise figures are specified either with source resistances at the value for lowest noise figure, or at 50 ohms. The dc operating current is also specified, and a reasonably high gain circuit is implied, or specified, for the N.F. test. While the source resistance for the sustaining amplifier's input can be determined by calculating the load impedance transformed by β_p to the base, the gain for the sustaining amplifier outside the resonator bandwidth is only unity and places the N.F.'s values in doubt.

Phase-noise tests have been run on a measuring system patterned after N.B.S. design.[2] At sideband frequencies above about 200 Hz, the oscillator noise merges with the noise floor of the measuring system.

The lowest noise floor measured to date is greater than -165 dBc, in a one Hertz band, using either 2N4134's, or MRF-904's operating with collector currents between 3 and 5 mADC. A typical phase-noise plot is shown in Figure 4.

Additional testing will be done to try for higher output power levels without a parallel increase in noise figures. The DXO design has an edge here in that its two transistors can produce 3 dB more power than a single transistor, and its balanced configuration greatly reduces second harmonic distortion.

* Currently affiliated with Systematics General Corp.

CH2186-0/85/0000-0171$1.00©1985IEEE

Since the crystal current is a function of the ac collector voltage and not the collector current, higher power output is best achieved with high currents and low load impedances. It also skirts the problem, when using high frequency transistors, of limited BVCEO values.

1/f Noise

One of the unique features of the DXO is the use of dual feedback; negative feedback via the resonator, and positive feedback for sustaining the oscillations. As analyzed in the original paper, the use of dual feedback can increase the frequency/phase slope of the resonator by a factor designated M_Q. This is due to the bridge-like configuration of the differential connection. For example, when a four-arm bridge containing a resonant circuit is adjusted for balance, the phase-slope approaches infinity and any frequency shift in the driving voltage or tuned circuit, induces a sharp change in phase angle. Since the output of a perfectly balanced bridge is zero, a bridge-type oscillator must operate off-balance to the degree required to sustain oscillation. Hence in the DXO the unbalance, expressed as the difference between the feedback arms, is limited to

$$\beta_p - \beta_n = 1/A_o$$

where β_p is the positive feedback factor, β_n, the negative feedback factor, and A_o is the open-loop gain of the differential amplifier.

This limits the multiplying factor M_Q to the value $M_Q = A_o\beta_n$ as derived in the referenced paper.

A value of M_Q equal to unity still provides a small improvement in the circuit's resistance to changes due to 1/f noise. Values of M_Q up to ten or more can be realized, and produce reductions in phase noise of 10 dB or more in the region from 0 Hz to 100 Hz over conventional oscillators.

This factor is related to the negative feedback, provided through the crystal via the β_n path. As described in the literature, the factor $(1 + A_o\beta_n)$ is responsible for both the gain stabilization of the sustaining amplifier at resonance, and for its resistance to changes in loop gain and phase angle. Thus the application of negative feedback via the crystal in the DXO produces a very stable, resonator-tuned sustaining amplifier, with the added merit of separating crystal current from transistor current.

SC and Overtone Resonator Operation

The version of the DXO described in this paper uses a balanced single-tuned circuit to achieve gain and produce waveform symmetry through the summation of the differential amplifier collector currents. The bandwidth of the tuned circuit is controlled by the value of loaded Q employed, where the 3 dB bandwidth is equal to f_o/Q_L. A low Q_L is generally desirable to reduce frequency shifts due to reactance changes, and a value of 20 is a good starting value. When SC-cut crystals are used, the Q may need to be raised, depending on the overtone ratio and the relative activity of the B- and C-modes. No additional mode rejection traps are needed with the DXO, eliminating one of the great nuisances involved in the design of parallel-resonant mode oscillators for overtone and SC-cut crystals.

The negative feedback divider uses the voltage across the crystal at series-resonance to control the amplifier gain. This favors the use of high impedance crystals since, for a given value of β_n, the crystal-feed resistor increases with crystal impedance and reduces dissipation in both the tuned circuit and the crystal. The lowest sideband noise has been achieved with fifth-overtone SC-cut crystals, and with third overtone crystals operating through an auto-transformer with impedance ratios of 4 to 9.

The auto transformer is wound on a small ferrite core with enough turns to achieve resonance with circuit and crystal capacitances. The transformer also replaces the shunt inductance often used in series-mode oscillators to resonate with the crystal's shunt capacitance.

Automatic Level Control

Since the sustaining amplifier's open-loop gain, A_o, must be maintained at a high level to realize the benefits indicated from negative feedback, it creates a problem in the application of ALC. Level control is essential to achieve the maximum performance possible through linear operation of an oscillator. In most oscillator circuits, the output of a buffer amplifier is rectified to provide a dc control voltage for ALC. The control voltage is then applied to the base of the sustaining amplifier, or the current control transistor. In both methods, the result is a reduction in sustaining amplifier gain, bringing the open-loop gain to unity where stable oscillations can develop at the set amplitude.

This approach would not be satisfactory with the DXO unless the value of $\beta_p - \beta_n$ is preset to a small number whose reciprocal is slightly below the maximum available gain, to satisfy the relationship, $\beta_p - \beta_n = 1/A_o$. If β_p is controlled instead, A_o can remain at the high design value required for a large M_Q factor and only a small variation in β_p will be required to achieve stable operation. This is the approach taken in the designs shown in Figure 2 and Figure 3.

The ALC control voltage is obtained in both circuits by direct rectification of the DXO output voltage, through the use of a voltage-doubler circuit, and is applied to the β_p control element.

A voltage divider is used in both the circuits shown, a varicap arrangement in Figure 2, and a JFET as a variable resistor in Figure 3. The two-varicap circuit shown in the first figure provides for cancellation of the frequency-pulling effect of a single varicap, since the divider constitutes a capacitive load across the tuned circuit and thus affects the resonant frequency of the tank. Since the varicaps account for only a few percent of the total tank reactance and are controlled through a long time constant R-C filter, they do not influence the phase-noise spectrum.

When A_o is high, β_p is almost equal to β_n, (slightly larger by $1/A_o$) resulting in β_p tracking in changes in β_n. The latter, of course, changes with crystal impedance, primarily from load capacitance adjustments, or from changes in temperature, resulting in little variation of the control voltage under normal operating conditions.

Another ALC circuit is shown in Figure 3, in which a resistance divider is substituted for the capacitance divider of the previous positive feedback control circuit. The control element in this scheme is a JFET, used as a variable resistor, again under the control of the rectified oscillator output voltage. JFET's such as the 2N4416 perform well in this circuit.

Since the available output voltage from the DXO is very large, little or no voltage gain is required in the buffer amplifier. Isolation from load variations is essential however and an emitter follower,[3] preferably with a Darlington-connected pair of transistors is usually adequate. If a mixer, or other 50 ohm element, is to be driven from the oscillator, a small broadband R.F. transformer can be used for impedance conversion, such as the models manufactured by MiniCircuit Labs.

An alternative output arrangement, useful with the popular 50 ohm hybrid line amplifiers, is to wind a few turns of wire around the DXO tank circuit, with a 25-50 ohm resistor in series to match the amplifier. Gains of from 10 dB to 30 dB are available with these amplifiers, and some are rated for reverse isolation values of 20-30 dB, and noise figures as low as 3 dB to 4 dB.

Conclusion

The DXO has been shown to be a useful oscillator design, suitable for high impedance overtone-mode resonators, and particularly for the SC-cut 3rd and 5th overtone units for frequencies from 1 MHz to at least 50 MHz. The relatively large available output power, and low phase-noise characteristics should favor it as a driver for multiplier chains, mixers, and transmitter exciters

Acknowledgements

This work was supported by the Frequency Control and Timing Branch, ET&DL of the U.S. Army Electronic Research and Development Command under Contract DAAK20-84-C-0386.

The author also acknowledges the help of my colleague, Paul Toste, in developing and testing the computer programming involved.

References

[1] A. Benjaminson, "Balanced Feedback Oscillators", Proc. 38th Annual Symposium on Frequency Control, pp. 327-333, 1984.

[2] F.L. Walls, "Description of Normal Operation and Circuit Diagram for the New NBS Phase Noise Detector", Time and Frequency Division, National Bureau of Standards, Boulder, Colorado.

[3] F.L. Walls, S.R. Stern, J.E. Gray and D.J. Glaze, "Design Considerations in State-of-the-Art Signal Processing and Phase Noise Measurement Systems", Proc. 30th Annual Symposium on Frequency Control, pp. 269-274, 1976.

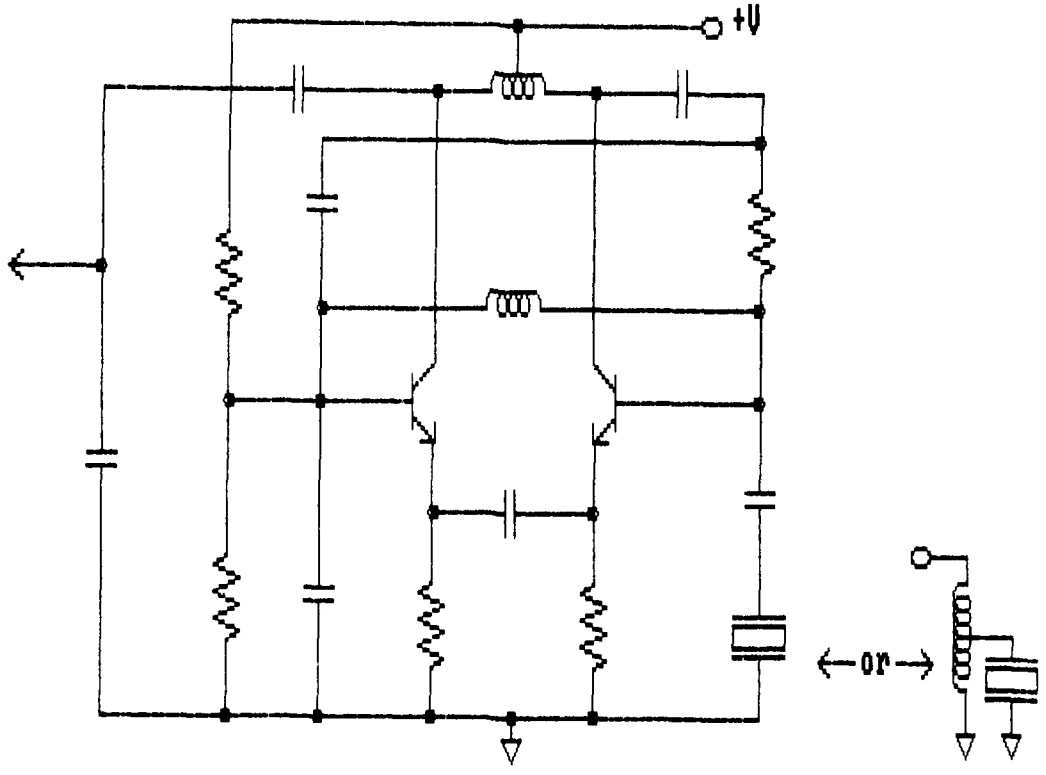


Figure 1 Basic Differential Oscillator Circuit

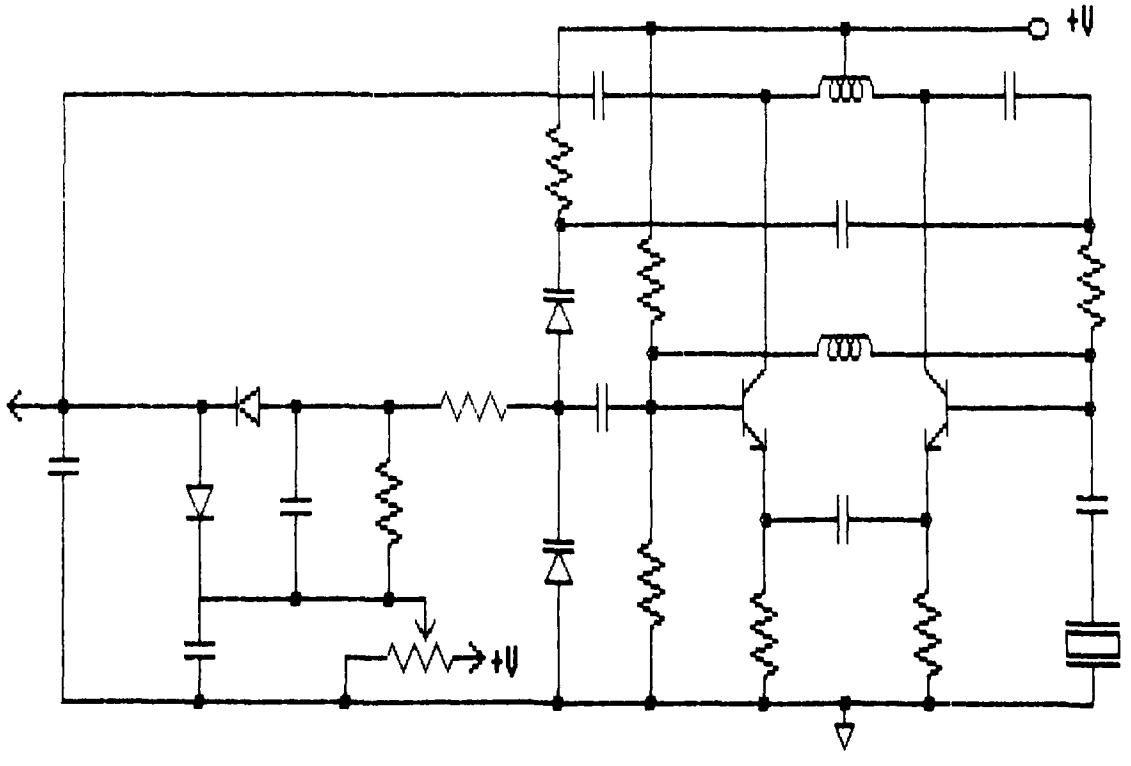


Figure 2 DXO with Varicap ALC

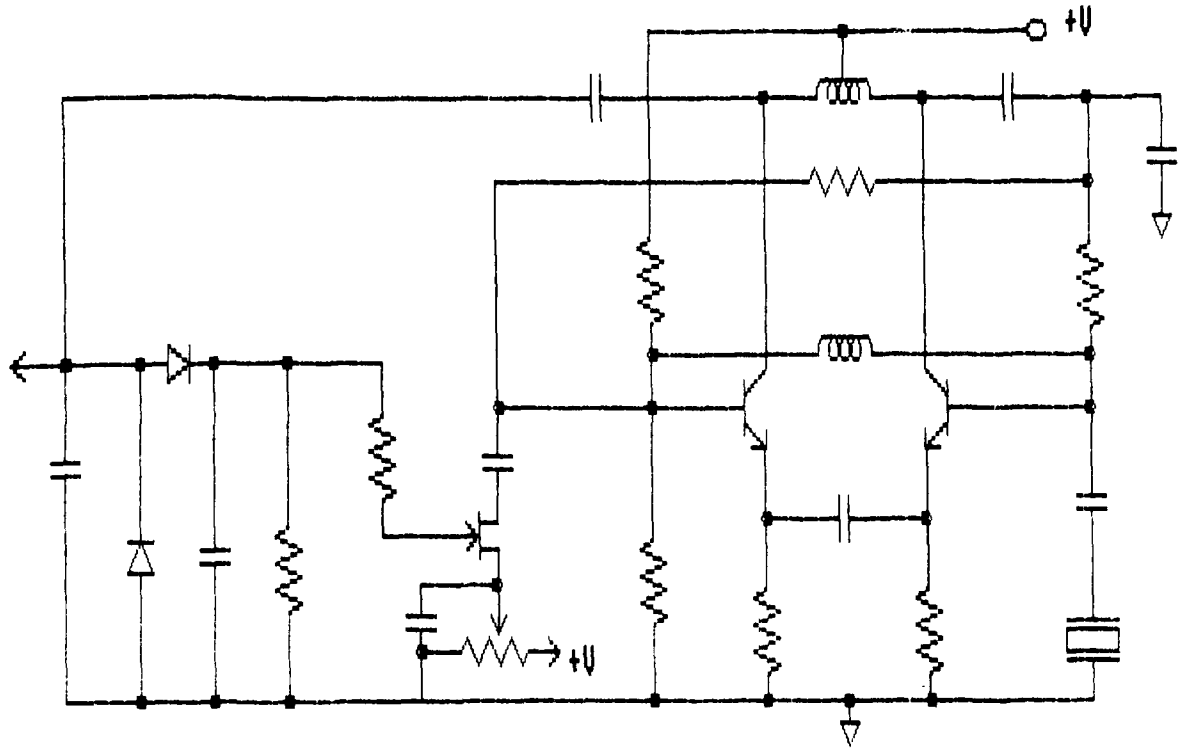
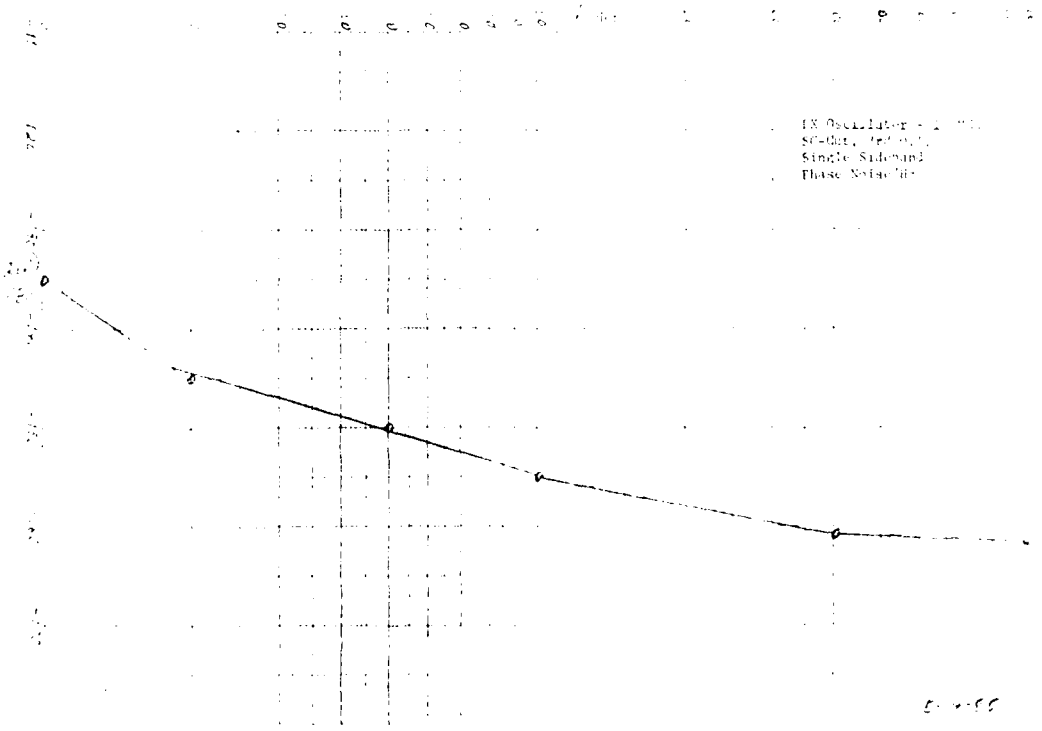


Figure 3 DXO with JFET ALC



COMPUTER ANALYSIS OF COLPITTS CRYSTAL OSCILLATOR

Takehiko ADACHI, Motoyoshi HIROSE, Yasuo TSUZUKI

Yokohama National University
Tokiwadai, Hodogaya-ku
Yokohama, Japan 240

SUMMARY

In this paper, a novel analyzing method of crystal oscillators is described. In an ordinary oscillator, such as colpitts oscillator, an oscillation can be achieved by connecting a crystal resonator to an active circuit having a capacitive reactance and a negative resistance . In the present method, the negative resistance and equivalent capacitance of the circuit are first calculated by use of a circuit simulation program based on nonlinear models of active devices. Then the oscillation frequency and its signal level are obtained for a specified crystal resonator. In this work calculations have been made on a colpitts oscillator widely used in the frequency range up to 30 MHz. The important design parameters of the oscillator, such as the level and the frequency dependences of both the equivalent capacitance and the negative resistance, the oscillation frequency and level and the oscillation frequency change due to the power supply voltage variation, have been calculated with satisfactory agreements with measurements. It is concluded from these results that this analyzing method is useful for the analysis and design of crystal oscillators.

1. INTRODUCTION

In most of the work dealing with the analysis of crystal oscillators, the nonlinearity of active devices has not been taken into consideration. In these works, the oscillation frequency and negative resistance have been calculated by use of the linearized equivalent circuits of active devices. This is because of the difficulties of solving the nonlinear equations in general and of the lack of any good models for the active devices.[1][2][3] Although a variety of circuit simulation programs based on the nonlinear models of active devices have in recent years been developed and used widely at universities and factories, their applications have been restricted mainly to the simulation of digital circuits and operational amplifiers.[4][5] The reason is as follows: the transient analysis routine has usually been used to calculate the steady state response of the nonlinear circuits. In the analysis of the crystal oscillator containing a very high quality factor resonance circuit, however, a large amount of calculations is necessary to reach the steady state solution. Therefore, use of these circuit simulation programs for this type of analysis has been considered to be impractical.

In an ordinary oscillator such as a colpitts oscillator, however, an oscillation can be achieved by connecting a crystal resonator to an active circuit having a capacitive reactance and a negative resistance. Due to the nonlinearity of the active device, both the magnitude of the equivalent capacitance and that of the negative resistance change with the oscillation signal level. Accordingly the steady state oscillation is attained at the point where the magnitude of the negative resistance becomes equal to that of the equivalent resonator resistance. The oscillation frequency is determined from the condition where the magnitude of the capacitive reactance of the active circuit at the oscillation signal level is equal to the inductive reactance of the resonator. Therefore, if the negative resistances and the equivalent capacitances of the active circuit are calculated at various signal levels in advance, the oscillation frequency and its signal level when a specified crystal resonator is connected can easily be determined. It should be noted that the calculation of the equivalent capacitances and negative resistances can easily be made by the simulation program, if the active circuit contains no resonance circuits.

The principle of calculating the equivalent capacitance and the negative resistance is as follows: First, the voltage response of the oscillator circuit driven by a current source connected in place of a crystal resonator has been calculated by use of the transient analysis routine. Second, from the steady state voltage response, its fundamental frequency component with the same frequency as the driving current has been calculated by FFT. Finally, from the amplitudes of this component and the driving current, and their phase difference, the equivalent capacitance and the negative resistance can be calculated.

In order to evaluate the validity of the present analyzing method, calculations have been made on a colpitts oscillator in common use in the frequency range up to 30 MHz . SPICE2 is used as a circuit simulation program. The important design parameters of the oscillator, such as the level and frequency dependencies of the equivalent capacitance and the negative resistance, the frequency and level of oscillation, and the frequency change due to the power supply voltage variation, have been calculated and compared with the measured results.

CH2186-0/85/0000-0176$1.00©1985IEEE

2. PRINCIPLE OF ANALYSIS

In this section, the principle of the analyzing method is described. Fig.1 shows the circuit configuration of a colpitts oscillator. The equivalent one port impedance of the active circuit seen from the terminal 2-1' is expressed by a negative resistance -R and a capacitance CL connected in series. The capacitance Cs in Fig.1 is a variable capacitor used to adjust the total capacitance of the circuit seen from the terminal 1-1' so that the oscillation frequency becomes the desired value. In general, the equivalent capacitance and the negative resistance are functions of frequency and the level of drive current. On the other hand, the impedance of the crystal resonator, which is expressed by a series circuit of an equivalent resistance Re and an inductive reactance Xe, has almost no signal level dependency. When a specified resonator is connected to the active circuit, an oscillation is obtained at the frequency and the current level where the following conditions are satisfied;

$$Re = |-R| \tag{1}$$

$$Xe = 1 / (2\pi f \cdot CL) \tag{2}$$

Therefore, if the frequency and level characteristics of the equivalent capacitance and the negative resistance have been calculated beforehand, the oscillation frequency and its signal level can be calculated for a specified crystal resonator.

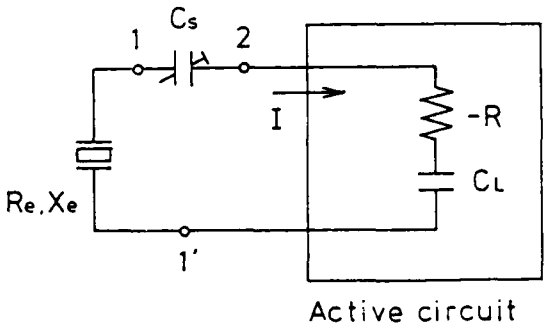


Fig.1 Configuration of colpitts crystal oscillator

3. METHOD OF CALCULATION

In this section, detailed calculation procedures of the equivalent capacitance and the negative resistance using SPICE2 circuit simulator are described.[6]

(1) At first the operator inputs the calculation parameters to a computer and runs the SPICE2 program. These parameters are the circuit structure, the values of the circuit elements, the model parameters of transistors, the frequency and the magnitude of a driving current source, and other control parameters specific to SPICE2 operation. Then, SPICE2 makes the following calculations (2) through (4).

(2) The DC voltages and currents are obtained by the DC analysis routine for all of nodes and branches respectively to find the operating point of transistors.

(3) The voltage response at the terminal 1-1' is calculated by the transient analysis routine when the active circuit is driven by the current source I*u-1(t)*sin(2 ft). Where the U-1(t) is the Dirac's unit function. In this calculation, the DC voltages obtained in procedure (2) are used as the initial conditions.

(4) The fundamental frequency component is calculated from the steady state voltage responses by FFT. The magnitude V of this component and the phase difference p between this component and the driving current are then obtained.

(5) The equivalent capacitance CL and the negative resistance -R are calculated by substituting V, p, I into the following equation.

$$CL = I/\{2\pi f \cdot V \cdot \sin(p)\} \tag{3}$$

$$-R = (V/I)\cos(p) \tag{4}$$

Using these negative resistance and equivalent capacitance, the oscillation frequency and current level of the entire oscillation circuit can be calculated in the following manner. The level characteristics of the negative resistance and the equivalent capacitance are calculated at a frequency close to the resonance frequency of the resonator under consideration. Fig.2 shows an example of such characteristics at a specified power supply voltage Vccl. In this figure, the current value I1, where the magnitudes of the circuit negative resistance and the equivalent resonator resistance become equal, shows the driving current of the resonator. The magnitude of the equivalent capacitance CL1 at this current level I1 becomes the equivalent capacitance at the steady state oscillation. Substituting this equivalent capacitance, the series resonance frequency fs, the quality factor Q, the parallel capacitance Co, and the capacitance ratio r, into the following equation, yields the desired oscillation frequency.

$$f = fs' \{ A - \sqrt{A - (1 + 1/r')} \}^{1/2} \qquad (5)$$

where $A = 1 + 1/(2r') - 1/(2Q')$

$$fs' = fs \{ 1 + (a/r) \}^{1/2}$$

$$r' = r \{1 - a/r\}/(1 - a)$$

$$Q' = Q (1 + a/r)^{1/2}$$

$$a = Co/(Co + CL)$$

The frequency change due to the power supply voltage variation is also one of the important characteristics of an oscillator, and can be calculated in the following way. The level characteristics of the negative resistance and the equivalent capacitance are calculated at a power supply voltage Vcc2 (Vcc2>Vcc1), as shown in Fig.2 by the dotted line. In this figure, the value of current I2 at A2 gives the oscillation current of the resonator, and the value of the equivalent capacitance CL2 at B2 yields its value at the steady state oscillation for the power supply voltage Vcc2. Substituting the values of CL1 and CL2 into the equation(5), the oscillation frequencies f1 and f2 are obtained for the two supply voltage Vcc1 and Vcc2, respectively. The fractional frequency change is calculated by substituting f1, f2, Vcc1, and Vcc2 into the following equation;

$$df/(f \cdot dVcc) = (f2 - f1)/(f1(Vcc2 - Vcc1)) \qquad (6)$$

4. ESTIMATION OF ANALYSIS CONDITION

Calculations have been made on a sample oscillator which is designed to operate in the frequency range between 10 and 30 MHz. Fig.3 shows the circuit configuration of this oscillator. Cs was made infinity during the calculation described below. A modified Gummel-Poon model built in SPICE2 was used as a transistor model.[7] The equivalent circuit of this model is shown in Fig.4. The calculations were made by using the model parameters measured on the used transistor. (Appendix 1)

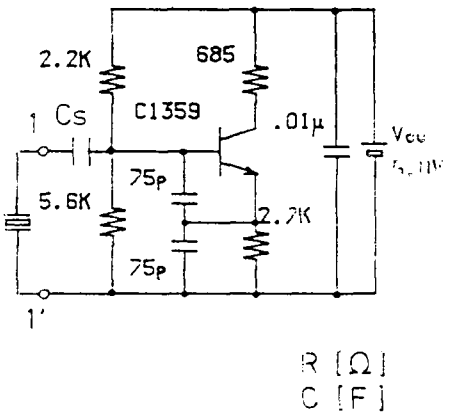


Fig. 3 Crystal oscillator under test

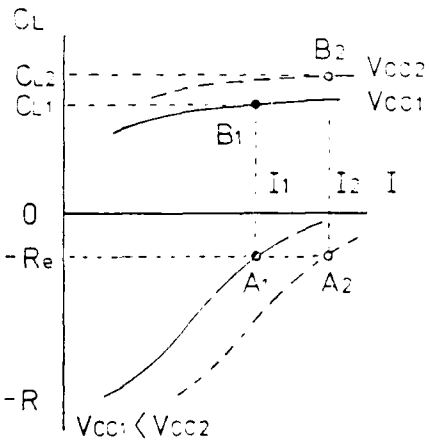


Fig. 2 Change of equivalent capacitance and negative resistance caused by change of current level and supply voltage

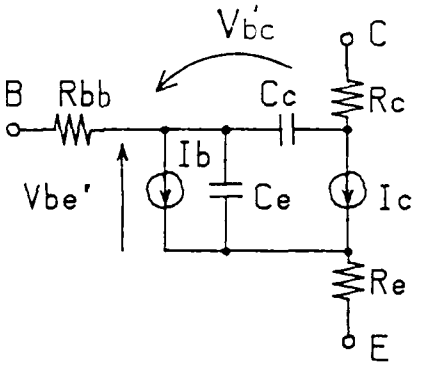
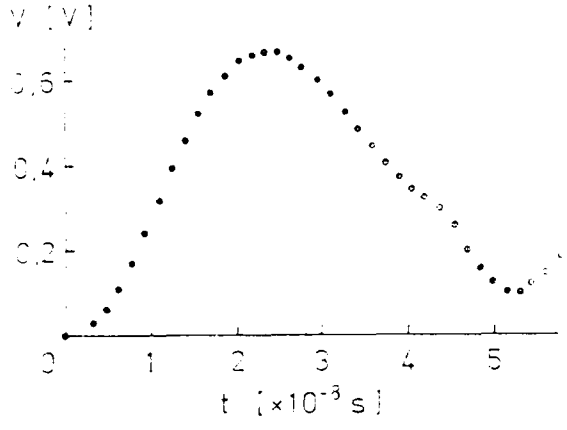


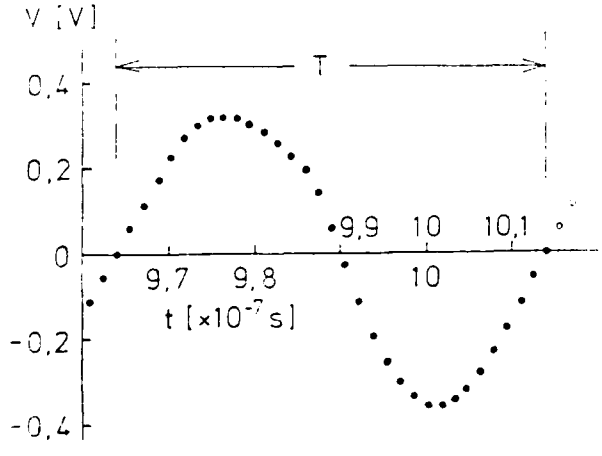
Fig. 4 Modified Gummel-Poon mode

In this section, the calculation conditions of the transient analysis to obtain sufficient accuracies for the values of the equivalent capacitances and the negative resistances are studied . Fig.5(a) and (b) show waveforms of the voltage response at the terminal 1-1' of Fig.1 when the magnitude and frequency of the driving current are 2.1mA and 20 MHz, respectively. Fig.5(a) shows the waveform during its first cycle and (b), that during the 20th cycle. Although, a considerable transient asymmetry is present around the horizontal axis in Fig.5(a), a good symmetry is obtained in Fig.5(b) with almost no transient effects. In this case the period of the response was found to coincide with that of the driving current. Fig.6 shows the convergence characteristics of the relative differences dT/To, dCL/CL, and dR/R against the number of period. It is evident that the calculation over 20 periods is sufficient to reach the steady state solution.

The relation between the number of calculation points within one period and the relative errors have also been studied. It was shown that 32 points in one period are sufficient to obtain the accuracies of better than 0.1 percent for both the equivalent capacitances and the negative resistances. Thus, in the following calculations, we will take 32 calculation points in one period and perform calculation over 20 periods for the transient analysis. HITAC M-160 II computer is used for the calculation. The CPU time necessary for the calculation of the equivalent capacitance and the negative resistance for one value of the current source was found to be about three minutes.



(a) Voltage response wave form in first period



(b) Voltage response wave form in 20th period

Fig.5 Voltage response calculated by transient analysis of SPICE2(G1)

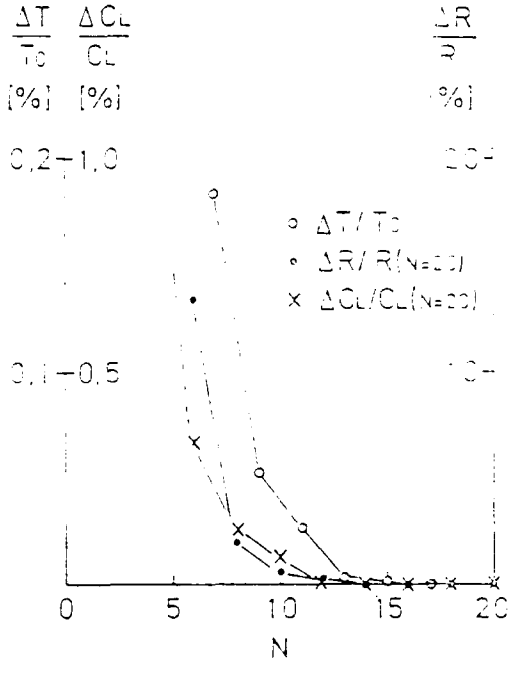


Fig.6 Convergences of voltage response time period, equivalent capacitance, and negative resistance with respect to time period number

5.LEVEL CHARACTERISTICS OF NEGATIVE RESISTANCES AND EQUIVALENT CAPACITANCES

Fig.7 shows the current level characteristics of the equivalent capacitance and the negative resistance of the oscillator circuit shown in Fig.3 at 20 MHz and 30 MHz. The solid lines show the calculated results and the symbols show the corresponding measured results. Measurements have been made by using the bridge method developed previously by the authors[8]. Good agreements have been obtained between the calculations and the measurements.

6.FREQUENCY CHARACTERISTICS OF EQUIVALENT CAPACITANCE AND SIGNAL LEVEL

Fig.8(a) and (b) show the frequency characteristics of the equivalent capacitance and the signal current level which yield a constant 16 ohm negative resistance. The solid lines show the calculated results and the circles, measured results. Satisfactory agreements have been obtained between the calculated and measured characteristics. In Fig.8(b), the frequencies at which the current becomes zero give upper and lower frequency limits within which oscillations can build up with the resonator equivalent resistance of 16 ohms. The Dotted line in Fig.8(a) shows values of the equivalent capacitance at the small signal current levels. These values correspond to those obtained by a linearized analysis. By comparison of the solid and dotted lines in Fig.8(a), it is evident that the linearized analysis gives considerably different values for the equivalent capacitances at actual operating signal levels. At the frequency where the solid and dotted lines intersect, the equivalent capacitance remains almost constant against the change of the the equivalent resistance of the resonator. Therefore, this frequency is the most suitable frequency for the operation.

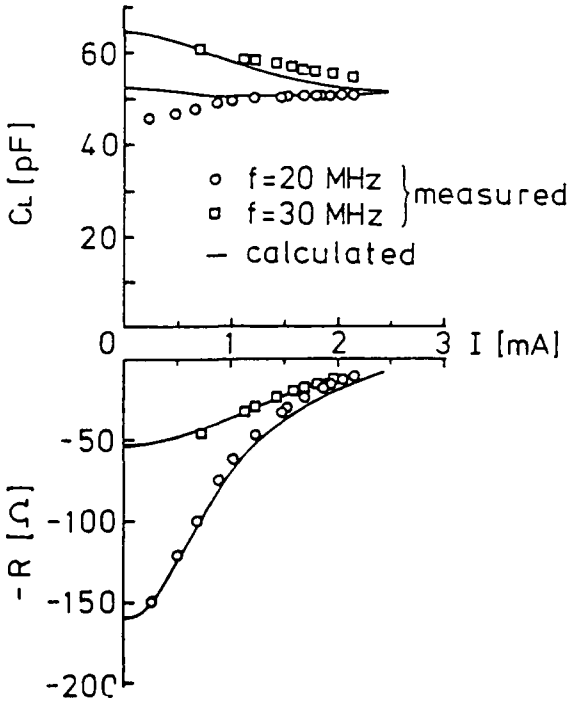
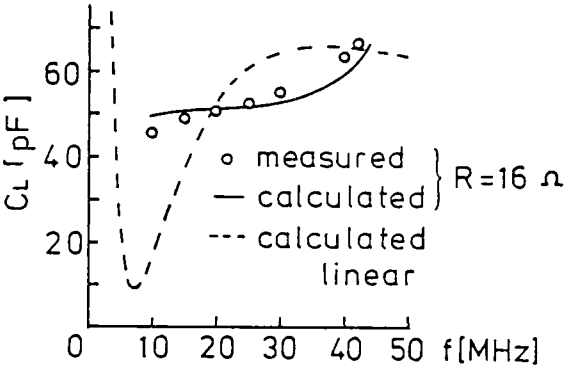
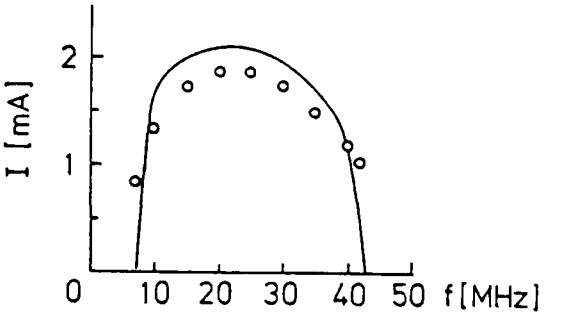


Fig. 7 Current level characteristics of equivalent capacitance and negative resistance



(a) Frequency characteristics of equivalent capacitance



(b) Frequency characteristics of crystal driving current

Fig. 8 Frequency characteristics of equivalent capacitance and crystal driving current for R=16Ω

7.FREQUENCY CHANGES DUE TO DC SUPPLY VOLTAGE VARIATION

Fig.9 shows an example of the operating frequency change due to the power supply voltage variation calculated at 20 MHz, the optimum frequency for the operation of this oscillator. The equivalent parameters of the resonator used for this calculation are listed in Table 1.[9] The solid line shows the calculated results and the circles, the measured results. A good agreement between calculation and measurement is obtained for the frequency change against the emitter resistance. From this result, it can be said that the calculation of the frequency stability can be made by the present method. Calculations of such properties for different circuit parameters will be useful for the design of oscillation circuits.

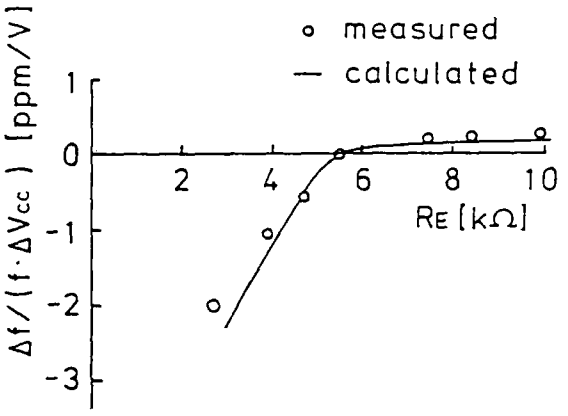


Fig.9 Relation between frequency variation caused by supply-voltage variation and emitter resistance

Table 1 Equivalent Circuit Parameter of Crystal Resonator

25.0[°C] 0.1[mA]

fs[Hz]	R1[Ω]	r	Q	Co[pF]
19,998,544	16.3	255	5.46×10^4	2.28

8.CONCLUSION

An analyzing method of colpitts type oscillators has been proposed. Calculations have been made on a sample oscillator widely used in the frequency range from 10 to 30 MHz. It has been shown from comparison between calculation and measurement that the equivalent capacitance, negative resistance and oscillation current can be calculated within 10 percent errors. This means that the oscillation frequency can be calculated within a few ppm error for a typical resonator in this frequency range. It has also been shown that the frequency change due to the power supply voltage variation can be calculated within an accuracy of the order of 0.1 ppm. The results have also shown that the level and frequency characteristics of the equivalent capacitance, the negative resistance, and the current of oscillation can easily be obtained. There is an optimum operating frequency at which the oscillation frequency change becomes almost independent of the equivalent resonator resistance. Moreover, the dependencies of the operating frequency changes due to power supply voltage variations on the values of the emitter resistance have been shown. It is concluded from these results that this method is useful for the analysis and design of crystal oscillators.

REFERENCES

[1] R. Nardin and J. Ho; "Computer Design and Analysis for High Precision Oscillators", Proc. of 28th AFCS, pp.237-242, May, 1974.

[2] S. Kodama and Y. Sato; "An analysis of Unwanted Frequency Oscillation in Crystal Controlled Oscillator", Proc. of 31st AFCS, pp.417-424, May 1977.

[3] Marvin E. Frekinq; "Crystal Oscillator Design and Temperature Compensation", Van Nostrand Reinhold Company, 1978

[4] D.J. Blattner;"Choosing the Right Programs for Computer-Aided Design", Electronics, Vol. 49, NO. 9, pp.102-105, April 1976

[5] G.D. Hachtel and A. Sangiovanni-Vincentelli; "A Survey of Third-Generation Simulation Techniques", Proc. of the IEEE Vol.69, NO.10, pp.1264-1280, Oct. 1981.

[6] L.W. Nagel; "SPICE2: A Computer Program to Simulate Semiconductor Circuits", ERL Memo ERL-520, University of California, Berkeley, May 1975.

[7] H. Daseking;"An Advanced Circuits-Oriented Integrated ·Circuit Bipolar Transistor model", 20th Midwest Symposium of Circuit and Systems Technical Digest, Part 2, pp.599-604, 1974.

[8] M. Toki and Y. Tsuzuki; "Measuring Method of Equivalent Series Capacitance and Negative Resistance of Crystal-Oscillator Circuits", Electronics Letters, Vol. 19, NO. 23, pp. 959-960, Nov.1983.

[9] Y. Tsuzuki, M. Toki, T. Adachi and H. Yanagi ; "Automatic Measurement of Parameters of VHF Quartz Crystal Resonators", Poc. of 33rd AFCS, pp.201-205, May 1977.

[10]T. Adachi, M. Hirose, and S. Moriyama; "Measuement of Transistor Model Parameters for Computer Aided Design of Electronic Circuits", IECE Trans., Vol.J67-c, NO.5, May 1984

Appendix 1 Transistor model

In Fig.4, Ib and Ic express the base and collector current, respectively. Ce and Cc express the capacitance of the base-emitter junction and the base-collector junction, respectively. In the case of using a transistor in the positive-bias active region, Ib, Ic, Ce, and Cc are expressed as follows:

$$Ic=\frac{Is}{Qb}\exp\left(\frac{V'be}{Nf\cdot Vt}\right)$$

$$Ib=\frac{Is}{Bf}\left[\exp\left(\frac{V'be}{Nf\cdot Vt}\right)-1\right]+Ise\left[\exp\left(\frac{V'be}{Ne\cdot Vt}\right)-1\right]$$

where

$$V'be=Vbe-Ib\cdot Rbb-(Ib+Ic)Re$$

$$V'bc=Vbc-Ib\cdot Rbb-Ic\cdot Rc$$

$$Qb=\frac{1}{2}\left(1+\frac{V'c}{Vaf}\right)\left[1+\sqrt{1+4\frac{Is}{Ik}\left[\exp\left(\frac{V'be}{Nf\cdot Vt}\right)-1\right]}\right]$$

$$Rbb=\frac{3(Rb-Rbm)(\tan Z-Z)}{Z(\tan Z)^2}+Rbm$$

$$Z=\frac{\pi}{24}\sqrt{\frac{Irb}{Ib}\left(\frac{144\cdot Ib}{\pi^2\cdot Irb}-1\right)}$$

$$Ce=\frac{d}{d\,V'be}\left(\frac{Tff}{Qb}If\right)+\begin{cases}Cje\left(1-\dfrac{V'be}{Pe}\right)^{-Me} & (V'be<fc\cdot Pe)\\ Cje(1-fc)^{-Me}\left[1+\dfrac{Me(V'be-fc\cdot Pe)}{Pe(1-fc)}\right] & (V'be\geq fc\cdot Pe)\end{cases}$$

where

$$Tff=Tf\left[1-Xtf\left(\frac{Itf}{If+Itf}\right)^2\right]$$

$$If=Is\left[\exp\left(\frac{V'be}{Nf\cdot Vt}\right)-1\right]$$

The model parameters of the used transistor were measured by the method developed previously by the authors. Table A1 shows the measured model parameters.

Table A1 Model Parameters of Transistor (2SC1359)

Parameter	Measured Value	Parameter	Measured Value
Is (A)	7.888×10^{-16}	Cje (F)	2.234×10^{-12}
Nf	1.005	Vje (V)	0.6617
Bf	64.87	Mje	0.2844
Ise (A)	3.006×10^{-13}	Tf (sec)	7.563×10^{-10}
Ne	1.663	Xtf	-0.5467
Rb (Ω)	39.64	Itf (A)	4.020×10^{-3}
Rbm (Ω)	-34.11	Fc	0.9667
Irb (A)	6.587×10^{-4}	Cjc (F)	1.916×10^{-12}
Ikf (A)	0.1790	Vjc (V)	0.4889
Vaf (V)	223.9	Mjc	0.1916
Re (Ω)	0.0	Tr (sec)	7.917×10^{-6}
Rc (Ω)	0.0	Nr	1.096

RECENT DEVELOPMENTS IN SYNCHRONIZATION AND TRACKING WITH SYNCHRONOUS OSCILLATORS

T. Flamouropoulos *
V. Uzunoglu #
M. H. White *

* Lehigh University
Fairchild Laboratory #161
Bethlehem, Pennsylvania 18015

Fairchild Communications & Electronics Company
Germantown, Maryland 20874

Abstract

The Synchronous Oscillator (SO) is a network that exhibits many characteristics useful in communications and signal processing systems. A primary feature of the SO is synchronization and tracking. The performance of SOs as carrier and clock recovery systems in QPSK modems is superior to PLL recovery systems. The multifaceted nature of the SO is responsible for this improved performance. Initial results have been obtained in modeling the nonlinear behavior of this network. The model predicts the amplitude and frequency of free-running oscillations as well as the tracking range of the SO.

Introduction

The fact that nonlinear oscillators can be synchronized with an injected signal is well known. Research on such oscillators however is not widely reported in communications or other relevant literature. The literature that is available refers mostly to the injection locked class of oscillators. The works of van der Pol [2] and R. Adler [4] are early studies that analyzed these oscillators. The Synchronous Oscillator (SO) presented here represents a departure from the established classification of injection locked oscillators. The behavior of the SO does not conform to that predicted for injection locked variety. V. Uzunoglu discusses the differences between the SO and other oscillators in [1]. The purpose of this paper is to briefly describe the main features of the SO, present data of SO performance in communication systems and present theoretical results derived from an analysis of the SO.

The Synchronous Oscillator

The Synchronous Oscillator (SO) is a synchronization and tracking network. The network is a modified colpitts oscillator as shown in Figure 1. The input signal is injected into the SO via the bottom transistor Q_2. The output is taken at the collector of Q_1 which is the active element of the oscillator proper. The capabilities of the SO are many and diverse, that range from synchronization and tracking, to frequency division, bandpass filtering and amplitude to phase conversion.

The primary attribute of the SO is synchronization and tracking. Figure 2 illustrates the behavior of the SO as recorded by an HP 3577A Network Analyzer. The analyzer injects a single frequency signal of known amplitude and frequency into the SO. The analyzer simultaneously monitors the output of the SO at the injected frequency through a narrow band filter. The frequency of the injected signal is swept from a lower to an upper limit while the amplitude is held constant, e.g. -40 dBV for Figure 2. The abscissa of the plot in Figure 2 represents the injected frequency and the ordinate represents both gain and phase of the network. The flat region of the gain curve is the tracking range where synchronization occurs. In this frequency range the phase is approximately linear.

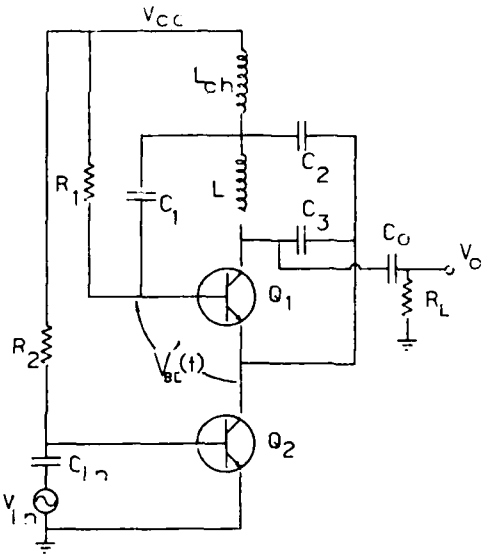


Figure-1: The Synchronous Oscillator Circuit.

The amplitude of the injected signal plays an important role with regard to the tracking range of the network. Variation of the injected amplitude causes a change in the width of the tracking range. A stronger signal produces a wide tracking range while a weaker signal leads to a narrrow tracking range. Figure 3 shows a set of transfer characteristics for injected amplitudes of -10 dBm, -20 dBm, and -30 dBm. The increase in gain with decreasing signal is due to the constant output amplitude of the oscillator.

CH2186-0/85/0000-0183$1.00©1985IEEE

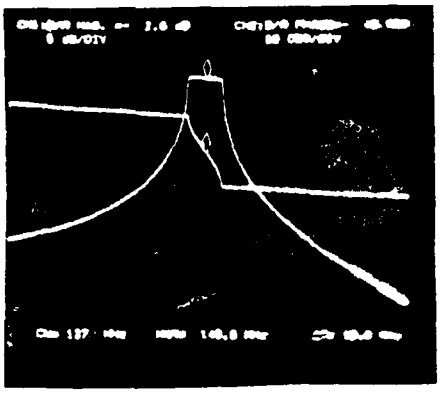


Figure-2: The transfer function of a Synchronous Oscillator.

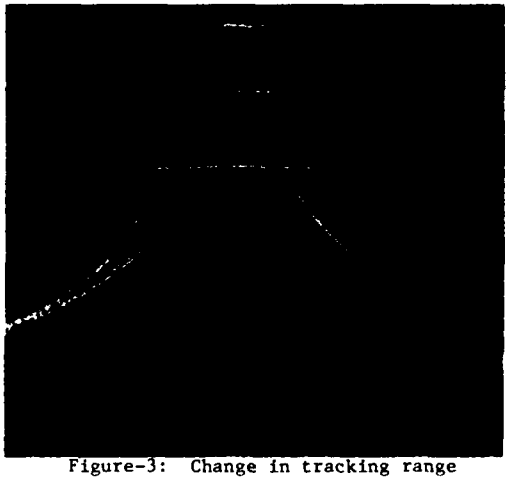
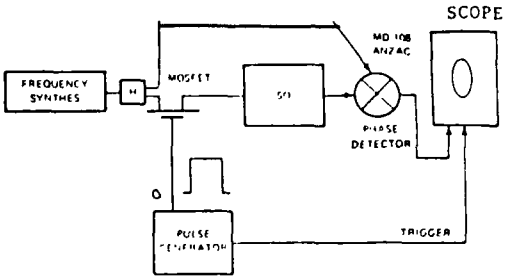


Figure-3: Change in tracking range due to different input magnitudes.



Measurement setup for phase acquisition in a Synchronous Oscillator.

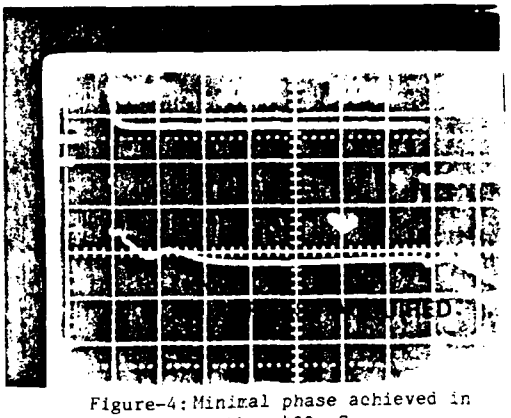


Figure-4: Minimal phase achieved in less than 400 nS.

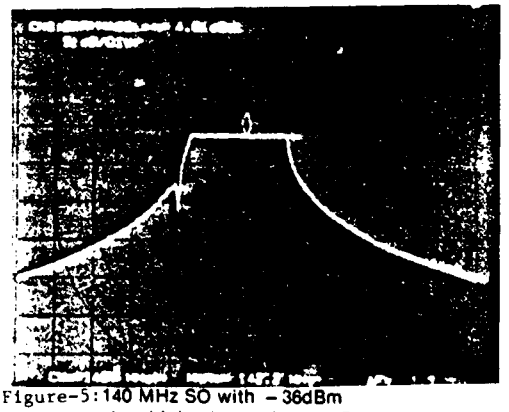


Figure-5: 140 MHz SO with – 36dBm signal injection and – 11 dB carrier-to-noise ratio

The curves in Figures 2 and 3 are steady state measurements. Investigation of transient phenomena in the SO requires description of the process of signal acquisition. This process is best described by the parameter known as acquisition time. Acquisition time is the interval between the moment the injected signal is introduced to the input of the SO and the moment at which steady state synchronization is achieved. An illustration of the acquisition process is presented in Figure 4. A carrier signal is injected into the SO and at the same time compared to the output of the SO through a phase detector. The carrier frequency is offset by 100 KHz from the natural frequency of the SO. At first there is a large phase difference between the carrier and the SO output, however within 400 ns the phase difference becomes minimal and constant, indicating steady state synchronization. When the carrier frequency is equal to the natural frequency of the SO, then the acquisition time can become as low as 100 ns.

An important consideration in synchronization is the quality of the injected signal. The SO can track signals that are literally buried in noise. Figure 5 shows a gain phase plot of an SO injected with a signal that has a -11 dB S/N ratio. The transfer characteristic for noisy input is almost identical to that obtained with a noiseless signal as shown in Figure 6. Conventional synchronization networks utilizing PLLs require S/N ratios that are greater than zero. It is interesting that while the SO rejects noise within the tracking range, it is able to accept and synchronize to frequencies that are outside the tracking range. Such frequencies are super and sub harmonics as well as certain ratios of integers of the natural frequency of oscillation. Figure 7 shows a scope photo of the SO waveform synchronized to an input that is 3/2 the natural frequency of oscillation.

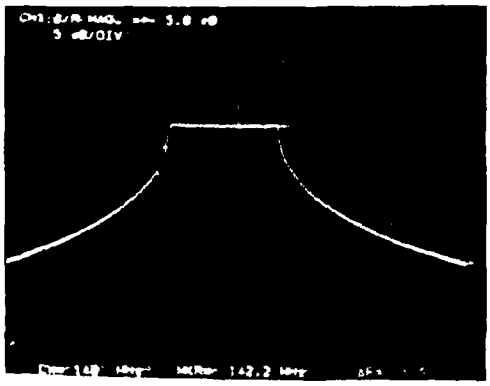


Figure-6: 140 HMz SO with −36dBm signal injection under low noise conditions

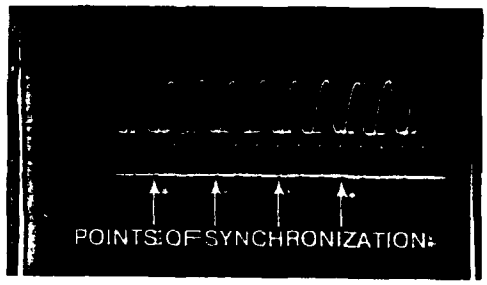
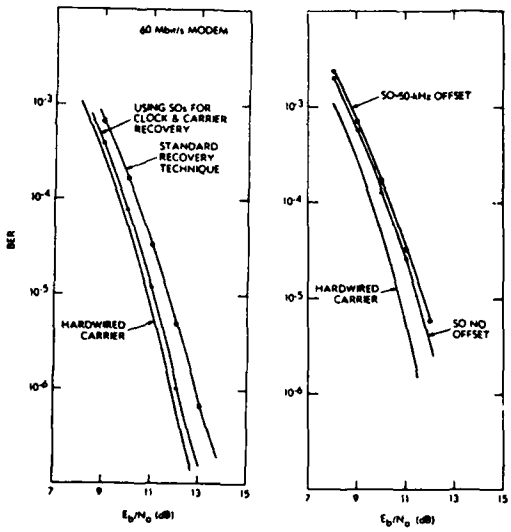
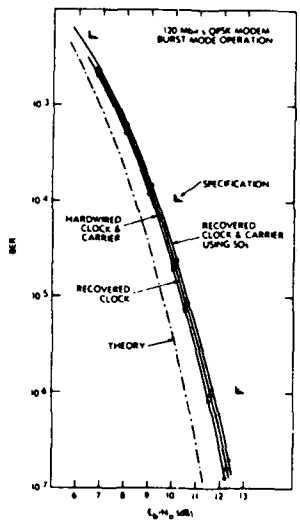


Figure-7: Synchronization in a divide by 3/2 process



(a) zero offset (b) ±50KHz offset



BER Performance for 60 Mbit/S Modem

BER Performance for 120 Mbit/S Modem

$\overline{E_b/N_o(dB) = C/N}$ (dB) - 10 log(BW/BR), where C = carrier power, N = noise power, BW is the bandwidth of the IF filter and BR the bit rate.

Figure-8: BER performance of QPSK modem containing Synchronous Oscillators.

These characteristics are instrumental in improving the performance of the SO when used as a synchronization network in QPSK modems. A bottom line indicator of the performance of SOs in a modem is the measurement of Bit Error Rate (BER). The measurement is made using a pseudo random bit sequence. Figure 8 shows measurements of Bit Error Rate as a function of carrier to noise ratio for 60 Mbit/s and 120 Mbit/s QPSK modems. The measurements indicate that modems containing SOs perform within .2 dB of modems that have a hardwired carrier, as opposed to recovering the carrier from the signal. The SO modems also consistently perform better than the required standard and considerably better than modems using conventional recovery techniques, i.e. PLLs.

Theory

Nonlinear oscillators are widely used in the communications field and other disciplines. The full potential of nonlinear oscillators however remains unknown and unrealized. This situation stems mainly from the inability to adequately model the nonlinear behavior of oscillators. Van der pol made the earliest attempt to analyze a nonlinear oscillator [2]. Kurokawa has proposed another methodology recently [3]. An analysis of the synchronization process in injection locked oscillators was done by R. Adler [4].

Most of the existing theoretical work is inadequate in describing SO transfer characterist ics and synchronization behavior. The biggest obstacle in all nonlinear analyses of oscillators is the characterization of the "negative resistance" that maintains stable oscillations. In the case of the SO this negative resistance translates to the large signal transconductance of the active element Q_1. An analysis done at Lehigh University [5] reveals that the large signal transconductance of Q_1 defined as the fundamental harmonic of the collector current over the base emitter voltage is

$$G_{m1} = \frac{2\,g_{m1}}{\frac{q}{kT}|V_{BE}|}\,\frac{I_1\left(\frac{q}{kT}|V_{BE}|\right)}{I_0\left(\frac{q}{kT}|V_{BE}|\right)} \qquad [1]$$

where g_{m1} is the small signal transconductance of Q_1 and $|V_{BE}|$ is the amplitude of the voltage across the base emitter junction as illustrated in Figure 9. The symbol I denotes a modified Bessel function. Only the fundamental harmonic of the collector current is considered because all otner harmonics generated at Q_1 are assumed to be attenuated by the high-Q feedback circuit.

A plot of the ratio G_{m1}/g_{m1} on a log-log scale is presented in Figure 10. As expected, the transconductance decreases as $1/|V_{BE}|$ for increasing oscillations. Stable oscillation is achieved when the G_{m1} decreases to a level that is just sufficient to counterbalance the resistive losses of the oscillator. A piecewise linear approximation of this curve is introduced for purposes of further analysis. G_{m1} is approximated as

$$G_{m1} = g_{m1} \qquad \text{When} \quad |V_{BE}| < \frac{2kT}{q}$$

and [2]

$$G_{m1} = 2\,\frac{kT\,g_{m1}}{q|V_{BE}|} \qquad \text{When} \quad |V_{BE}| > \frac{2kT}{q}$$

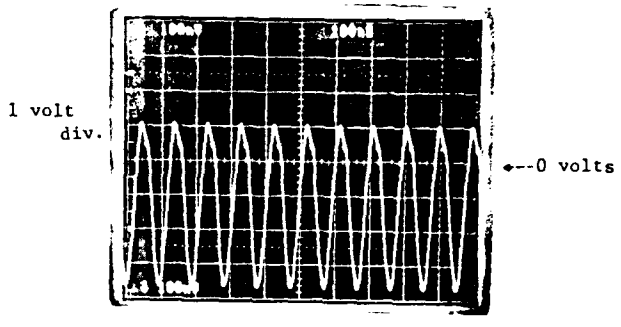


Figure-9: Voltage across the base-emitter junction. Most of the waveform is below zero volts.

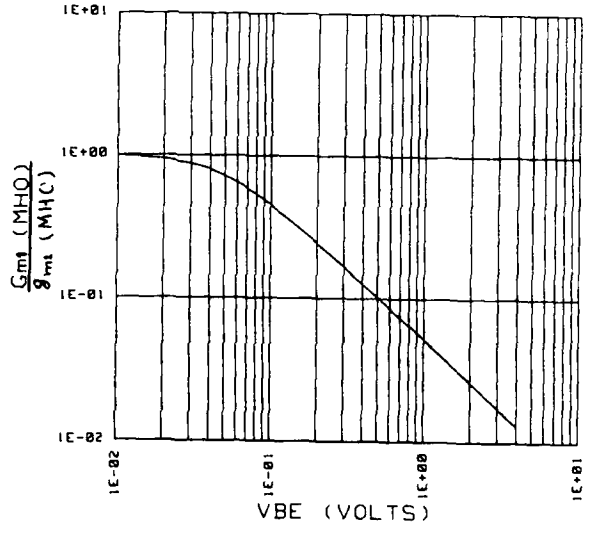


Figure-10: Ratio of transconductances as a function of base-emitter voltage

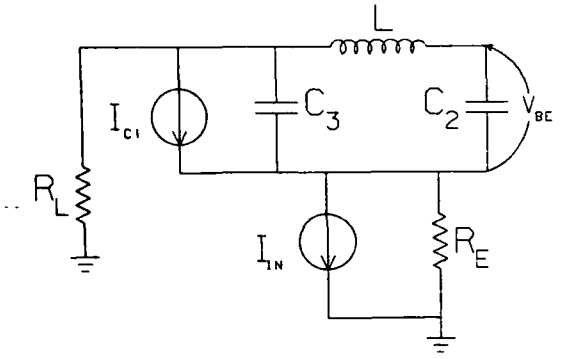


Figure-11: The ac equivalent circuit of the Synchronous Oscillator.

The expression for G_{m1} can now be used in an ac analysis. The ac equivalent circuit of the SO is presented in Figure 11. Capacitor C_1 has been replaced by a short circuit, because the capacitor is large with respect to the natural frequency of the SO. The transistors are replaced by ac current sources. The resistor R_E is the output conductance of Q_2 and I_{IN} is the injected signal. The transconductance relation mentioned earlier refers to the ratio I_{C1} over V_{BE}. A nodal analysis of the equivalent circuit leads to a transfer function of V_{BE} in terms of the injected current. V_{BE} is chosen for analytical convenience. It can be shown that the output voltage of the SO across the load R_L is approximately equal in amplitude and 180 degrees out of phase with V_{BE}. The transfer function is expressed in terms of the magnitude of V_{BE} and I_{IN}.

$$\left[\omega^2(\omega_o^2-\omega^2)^2(R_T C_3)^2+\left[(\omega_1^2-\omega^2)+R_T\omega_1^2 G_{m1}\right]^2\right]|V_{BE}|^2 = \left[R_E\omega_1^2\,|I_{IN}|\right]^2 \qquad [3]$$

Where $\omega_o^2=\frac{C_3+C_2}{LC_3C_2}$, $\omega_1^2=\frac{1}{LC_3}$ and $R_T=R_E+R_L$

When $|V_{BE}| < 2kT/q$ equation [3] is a linear small signal transfer function. When $|V_{BE}| > 2kT/q$ the transfer function becomes nonlinear. If the square term in this equation is distributed and G_{m1} is replaced by its expression, equation [2], then the transfer function takes the form

$$A(\omega)|V_{BE}|^2 + B(\omega)|V_{BE}| + C(\omega) = 0 \qquad [4]$$

where

$$A(\omega) = \omega^2(\omega_o^2-\omega^2)^2(R_T C_3)^2 + (\omega_1^2-\omega^2)^2$$

$$B(\omega) = 4\omega_1^2(\omega_1^2-\omega^2)R_T\bar{I}_{C1}$$

$$C(\omega) = 4\omega_1^4 R_T^2\bar{I}_{C1}^2 - \left[R_E\omega_1^2\,|I_{IN}|\right]^2$$

where the injected current has been moved across the equal sign and incorporated in the coefficient C(w) and $\bar{I}_{C1}$ is the dc current flowing through the collector of Q_1.

The solution for $|V_{BE}|$ is simply the quadratic formula.

$$|V_{BE}| = \frac{-B(\omega) \pm \sqrt{B(\omega)^2 - 4A(\omega)C(\omega)}}{2A(\omega)} \qquad [5]$$

A restriction is placed on this solution by the discriminant in the square root. The discriminant must always be non-negative if the solution is to remain a real quantity. For frequencies where the discriminant becomes negative, the solution does not exist. When the injected current is zero the solution predicts a free-running oscillation frequency and amplitude for the SO. When the injected current is nonzero there is a range of frequencies for which the discriminant is non-negative and a solution does exist. This is the tracking range of the SO.

TRACKING BANDWIDTH

$$\Delta\omega = \frac{\sqrt{2}\,g_{m2}|V_{IN}|}{[C_2+C_3]|V_{BE}|} \qquad [6]$$

Notice that the tracking range is proportional to the injected amplitude as observed in the experimental measurements of the transfer characteristic of the SO in Figure 3.

Conclusion

The SO is a network that displays a variety of peculiar and desirable characteristics. As a synchronization network, it has an output voltage that is independent of input frequency and amplitude within the tracking range. The SO synchronizes to an injected input in a time interval demonstrably shorter than that of a PLL used in a similar application. The acquisition of a noisy signal is performed almost as effortlessly as the acquisition of a noiseless signal. The SO is capable of signal processing functions such as frequency division, amplitude to phase conversion with its linear phase response and narrowband filtering with its high skirt selectivity. Such abilities enable the SO to perform consistently better than the allowed specifications or competing networks when used as a carrier or clock recovery system in QPSK modems. Analytical efforts at Lehigh University aimed at modeling the unusual behavior of the SO have yielded expressions that describe important parameters such as large signal transconductance, tracking range and free-running operation.

References

[1] Uzunoglu, V.
"Synchronization and Tracking in Synchronous Oscillators".
Proceedings of the 37th Annual Symposium on Frequency Control, pages 91-97. U.S Army ERADCOM and IEEE, 1-3 June, 1983.

[2] van der Pol, B.
"Forced Oscillations in a Circuit with Non-linear Resistance".
Phil. Mag. S.7 3(13):65-80, January, 1927.

[3] Kurokawa, K.
"Some Basic Characteristics of Broadband Negative Resistance Oscillator Circuits".
BSJT :1937-1955, July, 1969.

[4] Adler, R.
"A Study of Locking Phenomena in Oscillators".
Proc. IRE 34:351-357, June, 1946.

[5] Flamouropoulos, T.
"A Study of Nonlinear Transconductance in Synchronous Oscillators".
M.S. Thesis, Lehigh University, January, 1985.

A NOVEL SYNTHESISER FOR MINIATURE SSB RADIO EQUIPMENT

C.K. Richardson

Plessey Electronic Systems Research
Roke Manor Laboratories,
Romsey, SO5 OZN, U.K.

ABSTRACT

There is an increasing requirement for small portable HF SSB radio equipment.

A major problem encountered when designing such equipment is the lack of a simple inexpensive frequency synthesiser design which combines high resolution, low noise, low power requirements and small size. Two popular techniques normally used in modern SSB radio equipment include the Fractional N and Multi Loop systems. Both of these are difficult to miniaturise without loss of performance.

The limitations of the above synthesiser systems are discussed in this paper. An alternative technique which meets the main requirements is also covered.

This system employs a simple divide by N loop with a wide channel spacing. Interpolation between the channels is achieved by controlling the loop reference oscillator frequency. The interpolation can be modified by time sharing between two discrete frequencies produced by the oscillator, or an analogue proportional control can be used. Both methods are discussed. Details of an LSI chip designed for synthesiser systems using this technique is also given.

INTRODUCTION

This paper gives a brief description of a novel high performance synthesiser system suitable for use in SSB radio equipment. The main features of the design are high switching speeds between programmed frequencies, small channel spacing and low power consumption.

Frequency synthesisers employed in small VHF portable radio equipment almost invariably use a divide by N frequency/phase lock loop. This type of synthesiser configuration is uncomplicated and inexpensive and can be directly programmed by BCD decade switches or other convenient means.

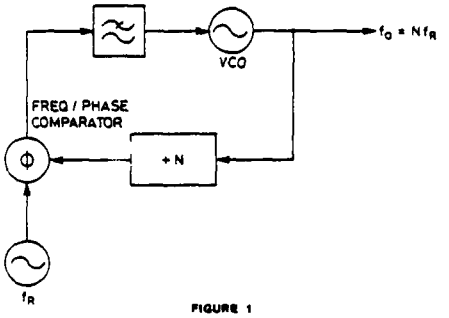


FIGURE 1

A SIMPLE DIVIDE BY N SYNTHESISER

A block diagram of a conventional divide by N synthesiser is shown in Figure 1. The output signal from a voltage controlled oscillator is divided down by a programmable divider. The programmable divider is designed to divide by any whole number up to a maximum determined by the complexity of the device. The output signal from the divide by N block is compared to the frequency of a fixed reference source fR by a frequency/phase comparator. The signal from this device controls the VCO so that the output signal from the divider is phase locked to the reference oscillator waveform. It will be appreciated that the output frequency of the VCO is N times the reference frequency fR. If fR is the channel spacing frequency, then the desired channel can be selected by the setting of N.

The main disadvantage of this simple system for SSB applications, is that the bandwidth of the loop, and therefore the speed at which the VCO can be controlled, is proportional to the comparison frequency fR. A low pass filter in the control line to the VCO is required in order to prevent modulation of the VCO by the pulsed output signal from the comparator. This modulation results in unwanted sidebands on the output signal. In order to reduce the unwanted sidebands to an acceptable level, the loop bandwidth normally has to be designed to be between 1/100 to 1/1000th of the reference frequency. The bandwidth depends on the type of frequency/phase comparator used as well as the required level of unwanted sidebands. Therefore, if a small channel spacing is required, the VCO may take an unacceptably long time to settle on the correct frequency after changing channel. Also, changes in VCO frequency due to power supply variation may not be corrected by a very slow loop.

CH2186-0/85/0000-0189$1.00©1985IEEE

METHODS OF IMPROVING PERFORMANCE

Two of the more popular methods of overcoming the channel spacing versus switching speed limitations will now be briefly described.

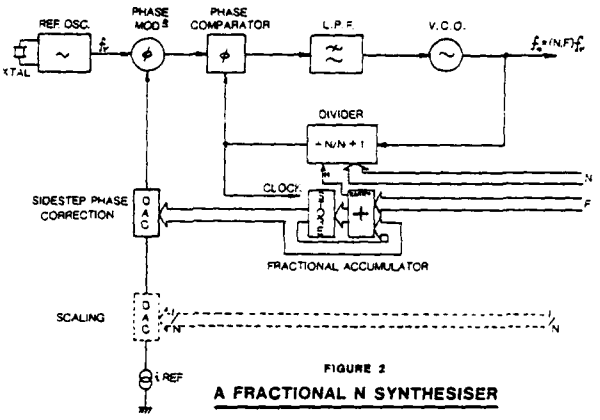


FIGURE 2
A FRACTIONAL N SYNTHESISER

The Fractional N Synthesiser (Fig. 2)

This method consists of side-stepping the basic count of the variable divider from N to N+1 for one reference cycle period at such a rate that the averaged division has a value N (1-F) + (N+1)F = N+F, where F is the effective fraction of the time spent at the N+1 value. In this way, a division ratio having a fractional component is attainable - hence the term "fractional-N" [1]. The resulting average output frequency is (N+F).fr.

The effect of sidestepping the division ratio in this way is to generate a ramp waveform of a period of $\frac{1}{F.fr}$ at the output of the frequency-phase comparator. This signal may then propagate to the VCO via the loop filter and cause unwanted frequency modulation.

One method used to reduce this unwanted signal is to cancel it with a separately generated compensation signal added in anti-phase. A correction ramp waveform can be generated by accumulating the fractional part using an adder and store and feeding this digital signal to a DAC. The output signal from the DAC then consists of the required analogue compensation waveform. This type of synthesiser can give high performance but is complex and requires high speed components, particularly in the frequency/phase comparator. It is not intrinsically suitable for incorporation in low power equipment.

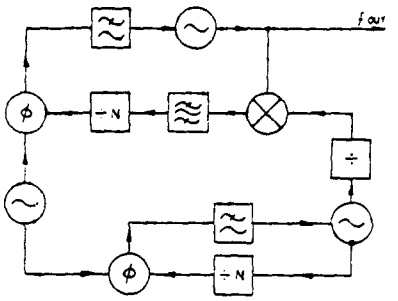


FIGURE 3
MULTI-LOOP SYNTHESISER

Multi Loop Synthesiser (Fig. 3)

This method utilises at least two divide by N synthesiser loops [2]. Both have relatively wide channel spacing. One synthesiser output signal is divided down in frequency and then mixed with the output signal of the second synthesiser to produce an offset frequency. It follows that the channel spacing of the offset synthesiser is also reduced by the divide ratio and interpolates between the channels of the undivided synthesiser loop. Switching speeds are not altered by the action of the divider, therefore small channel spacing consistent with high switching speeds can be achieved by use of this method.

The main disadvantage of the multiple loop solution is the production of unwanted signals. The use of multiple VCOs operating in the same frequency range within a small space results in the generation of unwanted IPs. In order to achieve a satisfactory level of unwanted signals, VCO and divide circuits need to be screened and isolated from one another. Multiple loop synthesisers are not recommended for small HF radio equipment.

The Controlled Reference Synthesiser

The proposed controlled reference synthesiser is shown in Figure 4.

The synthesiser uses two divide by N counters N1 and N2. The N1 counter is connected in the simple synthesiser loop described before, and divides by using the normal technique of counting down from a programmed value to zero. When zero is reached, a pulse is sent to the frequency/phase comparator and the counter simultaneously resets to the programmed value. The cycle is then repeated. It will be appreciated therefore, that N1 divides the VCO frequency by the programmed count. N2 is connected so that its counter logic is pre-set to the programmed count when N1 has momentarily reached zero. N2 then counts down by one for each cycle of the VCO frequency to zero, and latches in the zero state. N2 is later reset to its programmed value when N1 reaches its zero state again.

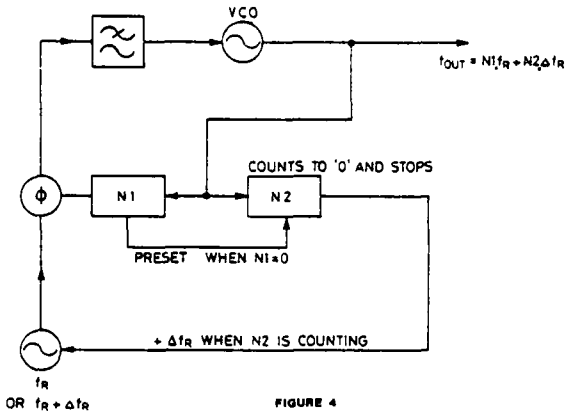


FIGURE 4

A CONTROLLED REFERENCE SYNTHESISER

N2 is arranged to give a '0' output when it is stopped at zero and a '1' output voltage when it is counting. Therefore, the reference frequency is fR when N2 is stopped and fR + ΔfR when it is counting. During a single cycle of the reference signal, the frequency of the reference oscillator is increased by ΔfR for a period proportional to the setting of N2 and inversely proportional to the setting of N1.

It can be shown that the relationship between the output frequency of the synthesiser and the setting switches is as follows:-

$$fo = fR.N1 + N2.\Delta fR$$

The channel spacing is equal to ΔfR and is therefore independent of fR.

It will be noted that the simple system shown in Figure 4 will only operate correctly if the N2 count does not exceed the N1 count. This limitation on the resolution can be overcome by connecting a third counter N3 in parallel with N2. Additional logic can then be used to increase the reference oscillator frequency by a further increment Δ2fR when N3 is counting.

The output frequency then becomes:-

$$fo = N1.fR + N2.\Delta fR + N3.\Delta 2fR.$$

By connecting additional counters in this way, each having a count equal to or less than N1, any resolution can be achieved.

EXAMPLE

A synthesiser is required to cover 100 - 200MHz in 1Hz steps.

If N2<N, fR could be 10kHz.

$$\therefore fo = 10^4.N1 + 1.N2$$

i.e. for fo = 123.456789MHz

N1 Setting = 12345 (x 10kHz)

N2 Setting = 6789 (Hz)

Several variations of the synthesiser system are possible. The system can be made more flexible if the pulsed output signal from the N2 counter is integrated so that it then becomes a steady D.C. signal with a voltage proportional to the setting of N2 and inversely proportional to the setting N1. This signal can then be used to control proportionally the frequency of the reference signal instead of switching it between two frequencies.

The same formula can be used for this alternative arrangement -

$$fo = fR.N1 + \Delta MAX\ fR.N2$$

ΔMAX fR is the maximum increase in reference frequency when N2 is continuously counting, (N1 = N2)

A method of realising a proportionally controlled linear and stable reference signal is shown in Figure 5.

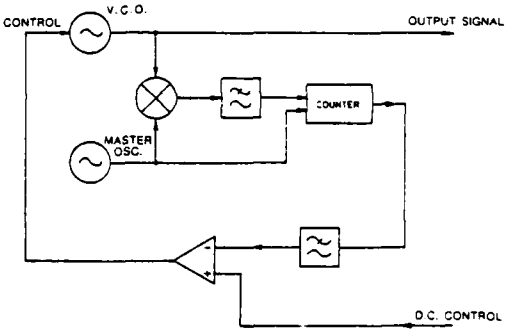


FIGURE 5

CONTROLLED REFERENCE OSCILLATOR

OPERATION OF THE CONTROLLED REFERENCE OSCILLATOR

The VCO output signal is mixed with a signal from a fixed frequency oscillator on a similar frequency. The difference frequency from the mixer is fed to a counter device via a low pass filter. The counter is reset to zero by a transition of the difference frequency waveform. The counter counts a pre-determined number of cycles of the fixed frequency oscillator output signal and stops until reset by another input transition. The output signal from the counter is a high level logic signal when the counter is counting and a low level when it is stopped. This signal, after filtering by a low pass filter, consists of a DC level which is proportional to the difference frequency between the two oscillators.

The DC signal is compared with a DC reference signal in a comparator and is then fed back to control the variable oscillator frequency. By this means, the variable oscillator frequency is automatically adjusted until the two comparator inputs are at the same level.

By use of this technique, the difference frequency between the master oscillator and the slave oscillator can be accurately controlled by a DC input signal.

THE ADVANTAGES OF THE CONTROLLED REFERENCE SYNTHESISER

The main advantage of the controlled reference synthesiser over other systems is that the spectral purity of the output signal is similar to that of a simple single loop synthesiser designed for the wider of the two channel spacings. The pulsed control signal from the N2 loop is at the same frequency as, and phase locked to, the reference frequency. The reference frequency is controlled by the mark space ratio only. Since perturbation of the reference phase is at the reference frequency rate, extra spurious signals passed to the VCO by the loop filter, are minimised.

A dual modulus high speed divider can be used to divide the VCO frequency prior to the N stages, as is normal in a single loop synthesiser. The remainder of the synthesiser circuits can therefore operate at low frequencies enabling low power consumption to be achieved.

The controlled reference synthesiser can be considerably less complex than similar designs of the same performance.

DISADVANTAGES

The N2 interpolation loop is in effect a frequency lock loop. There is therefore a small error in the synthesiser output frequency which is dependent of circuit conditions and may vary with temperature, power supply voltage etc.

By use of carefully designed circuits, this error has been reduced, in practice, to 5Hz at 50MHz over a temperature range of -20 to +70^{0}C.

AN LSI CONTROLLED REFERENCE SYNTHESISER

A C-MOS LSI 10 decade programmable divider I.C. has been designed and manufactured for use in a controlled reference synthesiser system. A block diagram of the I.C. circuit is shown in Figure 6. Five 2 decade dividers are programmed by a 41 bit serial to parallel shift register. The BCD format, programme inputs to the dividers are fed from a 40 bit latch driven by the shift register. The 40 bit serial programme data is shifted down the register by the data clock. The data is then stored in the latch and the register is reset to the all zero state, ready for a new data entry.

A block diagram of the complete synthesiser using the I.C. is shown in Figure 7.

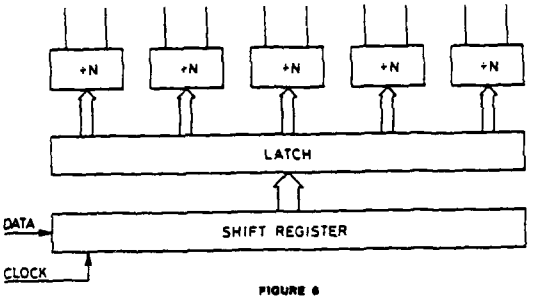


FIGURE 6
L.S.I. DIVIDER SYSTEM

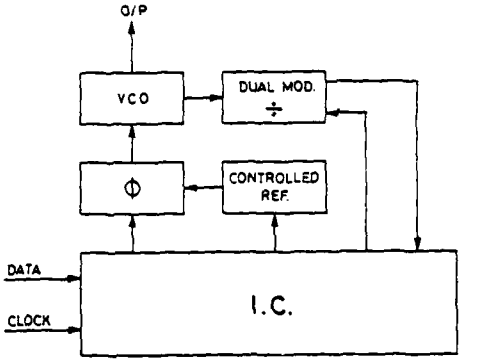


FIGURE 7
HIGH RESOLUTION SYNTHESISER

REFERENCES

Digital PLL Frequency Synthesisers Theory and Design, Englewood Cliffs, NJ: Prentice - Hall, 1983.

[1] pp. 124-141

[2] pp. 309-351

HYBRID MINIATURE OVEN QUARTZ CRYSTAL OSCILLATOR

John Ho

Frequency Electronics, Inc.
Mitchel Field, New York 11553

Summary

A new generation of small, fast warm-up precision ovenized crystal oscillator has been developed and is available for military and commercial use. The overall size is only 4.2 in^3 and weighs 3.8 oz. (Figure 1). The unit will warm-up from -54°C to within $1PP10^7$ in less than 5 minutes. The short term stability is measured at better than $3PP10^{12}$ per 10 seconds and long term stability is less than $2.5PP10^{10}$/day over 40 days.

Introduction

Two different models are presented in this paper with test results and comparisons. The models are the FE-2211 and the FE-2185. The major difference is that the FE-2211 has higher stability than the model FE-2185 (Figures 2 and 3).

Oscillator Package Design

The major innovation is that both models are designed with a new concept of heating the crystal by coupling and controlling the heat to the crystal blank directly through its crystal pin and the support elements (Figure 4). Therefore, a ceramic substrate with thick film heater elements printed next to the crystal pins for fast heat transfer are designed. the control sensor is located and coupled between the crystal pins and heater substrate for accurate temperature control, so that thermal overshoot to the crystal blank is minimized.

To improve the oscillator frequency stability with temperature, it is also desirable to have a Colpitts oscillator hybrid and oven control hybrid in a stable thermal environment. Figure 5 shows a block diagram and Figures 6, 7 and 8 show the thermal package and arrangements to reduce the mass of the assembly and the interconnection between circuits. The substrate packge is approximately 0.6" x 0.6" x 1.4". This package contains a Colpitts hybrid, a dual heater controller hybrid, heat sensor, crystal and heater elements. The output amplifier hybrid is located outside the thermal package. Figure 9 shows a summary of the specifications.

Hybrid Packages

Colpitts Oscillator Hybrid

This circuit is packaged in a 0.5" x 0.5" x 0.2" sealed 16-pin flatpack, capable of operating a supply voltage between 5 to 10 volts, and is designed to both AT cut or SC cut crystals. When the circuit is operating with an SC cut crystal, the B mode frequency will be automatically suppressed to assure the C mode desired frequency is operating. The hybrid circuit also provides VCO capabilities in the $8PP10^7$ range.

RF Amplifier Hybrid

This amplifier circuit is also packaged in the same 0.5" x 0.5" x 0.2" 16-pin flatpack as the Colpitts oscillator. This circuit can be used in every oscillator in the frequency of 5 to 15 MHz. The following is a summary of specifications:

A. Frequency Range: 5 MHz to 15 MHz
B. Power Supply Voltage Range: 10 to 20 Volts
C. Operating Current: 2 mA to 10 mA, settable by external pin arrangements
D. RF Power Input: 0 dBm $\pm$ 1.5 dB
E. RF Power Output: Settable 3 to 13 dBm
F. Maximum Ouptut Harmonics: 40 dB
G. VSWR for Input and Output: 1.2:1
H. Input and Output Isolation: 50 dB minimum

Oven Control Hybrid

The oven control circuit is packaged in a sealed 0.5" x 0.38" x 0.133" 12-pin flatpack. It is a linear DC control system, containing a temperature sensing bridge network with a built-in reference voltage, a high gain comparator and two heater driving circuits. The control system is capable of controlling two heater elements; one for normal heat control, the other for warm-up booster heater, ranging from 1 to 20 watts.

Test Results

1. Test results of frequency vs. ambient temperature, for the Model FE-2211 (Figure 10) show the double oven system is considerably better than the single oven control system by a factor of six to one.

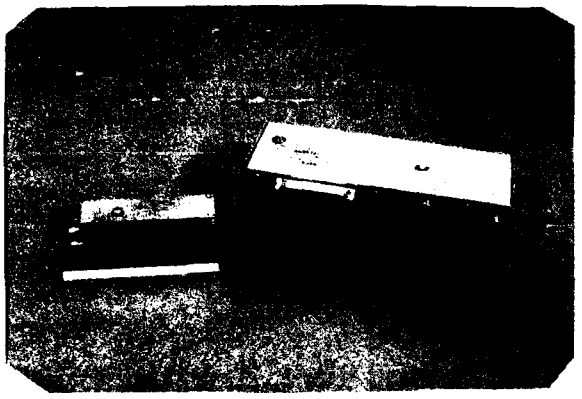
2. Test results of frequency vs. ambient temperature for the Model FE-2185 are shown in Figure 11.

3. The frequency aging data of two typical units are shown in Figures 12 and 13.

4. The frequency short term stability of both models is shown in Figure 14. Typically, between 1 to 10 seconds the RMS fractional frequency deviation is $3PP10^{12}$.

5. Warm-up data for is shown in Figures 15 and 16.

CH2186-0/85/0000-0193$1.00©1985IEEE



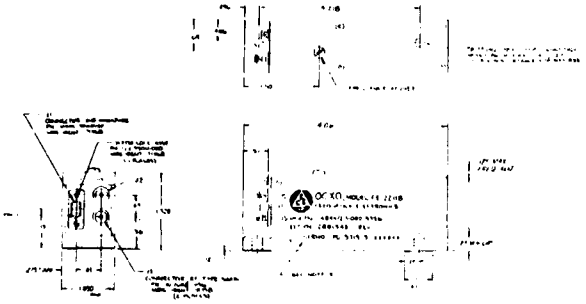


Figure 2. Oven Controlled Crystal Oscillator, Model FE-2211B, Outline

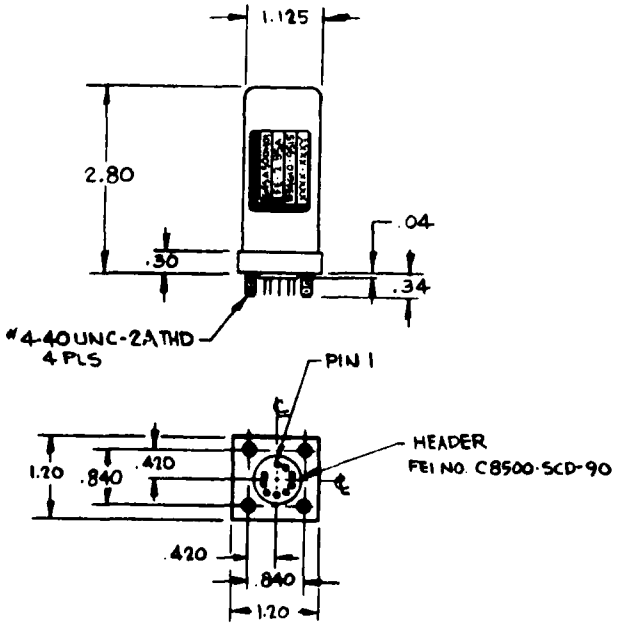


Figure 3. Oven Controlled Crystal Oscillator, Model FE-2185A, Outline

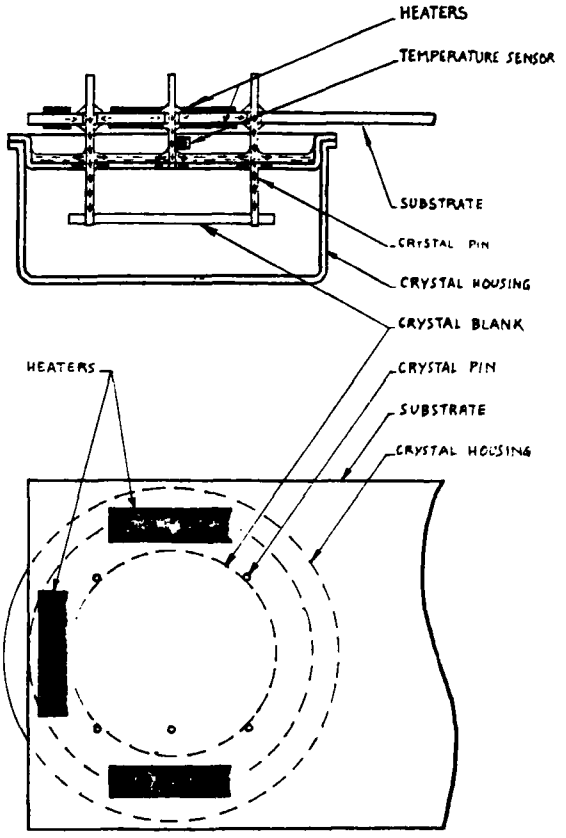


Figure 4. Crystal, Open View

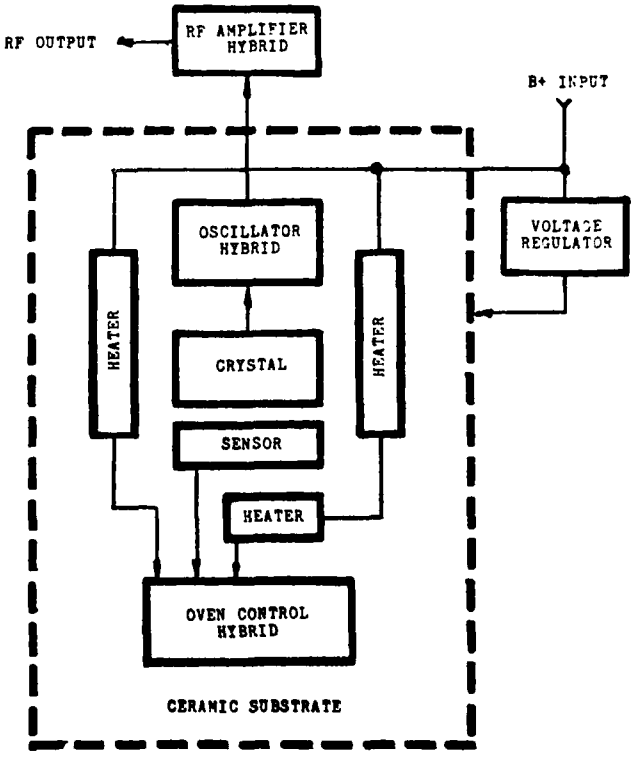


Figure 5. Block Diagram

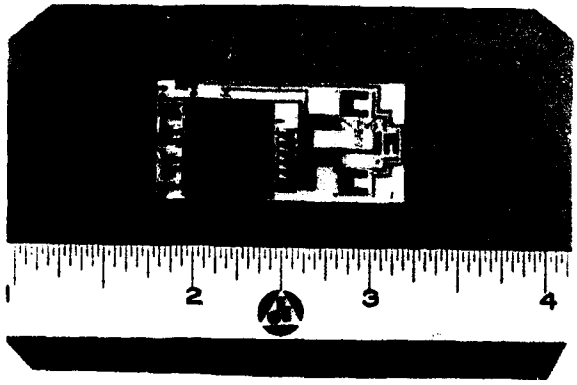


Figure 6. Thermal Package Arrangement

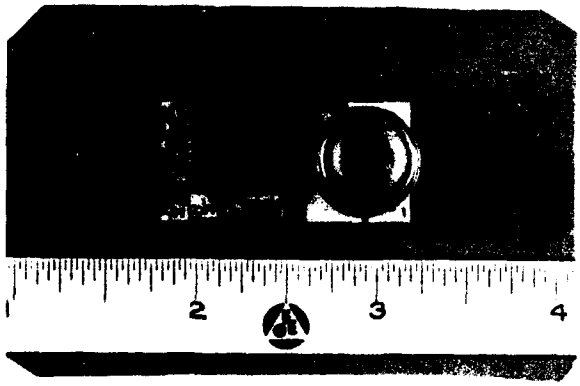


Figure 7. Thermal Package Arrangement

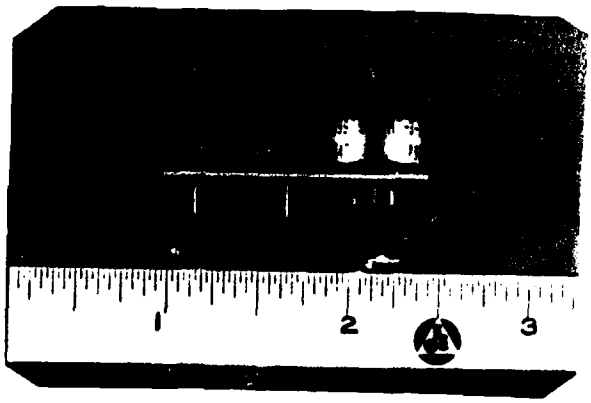


Figure 8. Thermal Package Arrangement

ITEM	MODEL FE-2211	MODEL FE-2185
FREQUENCY	10 MHz	10 MHz
TYPE OF CRYSTAL	3RD OVERTONE, SC CUT TO-8, 3 POINT MOUNT	3RD OVERTONE, SC CUT TO-8, 3 POINT MOUNT
WARM-UP TIME @ 25°C	2 MIN. 1 $PP10^{8}$ 5 MIN. 1 $PP10^{9}$	2 MIN. 8 $PP10^{8}$ 4 MIN. 1 $PP10^{8}$ 5.5 MIN. 5 $PP10^{9}$
OVEN CONTROL SYSTEM	DOUBLE OVEN CONTROL	SINGLE OVEN CONTROL
SHORT TERM STABILITY	3 $PP10^{12}$/SEC	3 $PP10^{12}$/SEC
LONG TERM AGING	1 DAY, 2 $PP10^{10}$/DAY	1 DAY, 8 $PP10^{10}$/DAY
TEMP STABILITY -10°C TO +60°C	4 $PP10^{9}$	2.5 $PP10^{8}$
PEAK POWER	15 WATTS	8 WATTS
OPERATING POWER	1.9 WATTS	0.95 WATTS
"g" SENSITIVITY IN THE BEST AXES IN THE WORST AXES	<1 $PP10^{10}/g$ <4 $PP10^{10}/g$	<1 $PP10^{10}/g$ <4 $PP10^{10}/g$
WEIGHT	3.85 OZ.	4.5 OZ.
VOLUME	4.2 IN^3	3.5 IN^3

Figure 9. Summary Specification

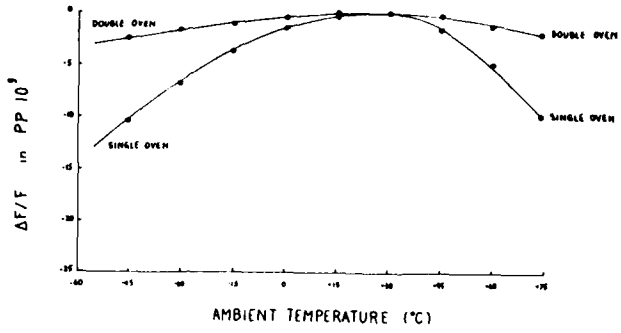


Figure 10. Temperature Stability, FE-2211A

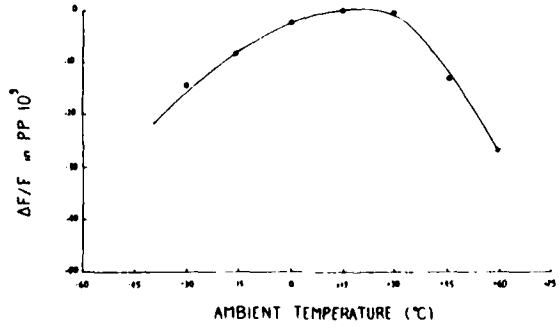


Figure 11. Temperature Stability, FE-2185A

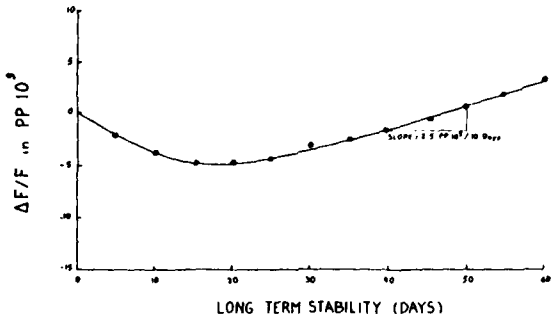


Figure 12. Aging Rate, FE-2211A

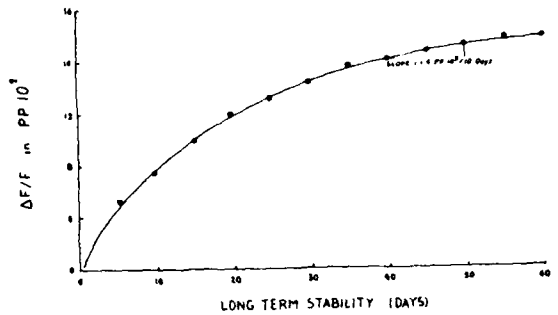


Figure 13. Aging Rate, FE-2185A

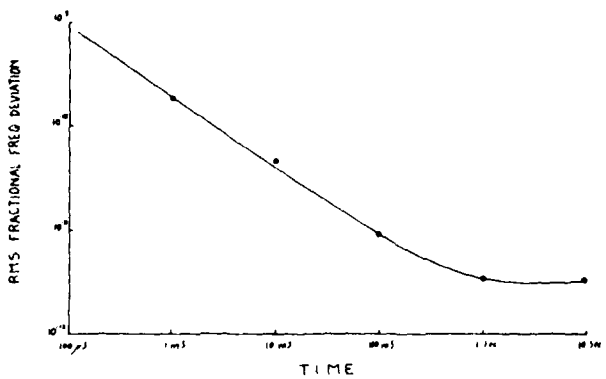


Figure 14. Short Term Stability, FE-2185A and FE-2211A

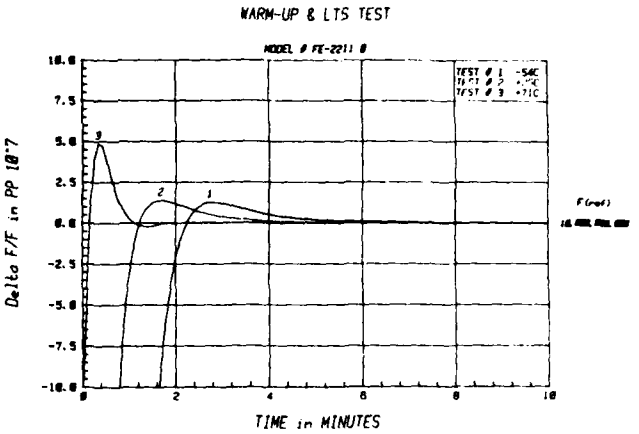


Figure 15. Warm-up Data, FE-2211B

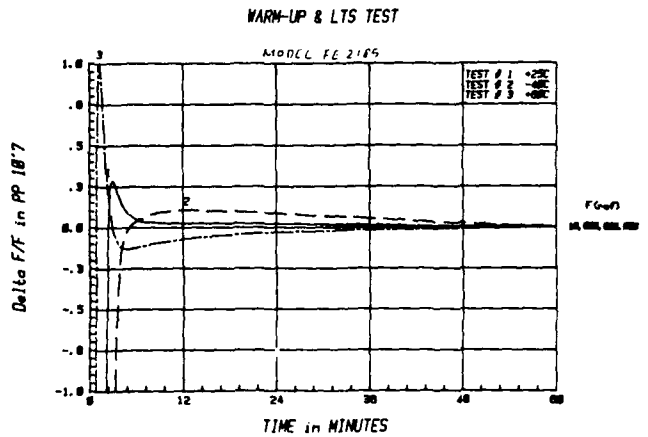


Figure 16. Warm-up Data, FE-2185A

LOW NOISE, VHF CRYSTAL-CONTROLLED OSCILLATOR UTILIZING DUAL, SC-CUT RESONATORS*

M.M. Driscoll
Westinghouse Defense and Electronics Center
Baltimore, Maryland 21203

Summary

This paper describes the design and performance of a novel dual-crystal-controlled VHF oscillator circuit exhibiting output-signal phase-noise sideband spectra substantially lower than those obtainable using conventional single-resonator oscillators. The use of multiple (two) crystal resonators effects a 5 dB reduction in phase-noise sideband level for carrier offset frequencies f ≤ 1 kHz. The oscillator circuit is symmetrical and includes two low-selectivity tank circuits (to prevent oscillation at crystal B mode resonances) so that each resonator is operated in an identical manner.

The SC-cut crystal resonators are operated at a high drive level (5 to 8 mW), and individual oscillator output signals are extracted through the resonators, which are each series-connected to identical common-base buffer amplifiers. At modulation rates in excess of the resonator loaded bandwidth, additional oscillator-signal phase-noise suppression occurs as a result of the resonator narrow-band transmission responses. Also, because the load (buffer amplifier) additive noise spectra are uncorrelated, signal frequency summation at the load circuit outputs effects a 3 dB reduction in output-signal phase-noise sideband level at high (f ≥ 10 kHz) modulation rates. Phase-noise sideband levels of −134 dB/Hz at f = 100 Hz, −164 dB/Hz at f = 1 kHz, and −181 dB/Hz at f ≥ 10 kHz have been demonstrated using 80 MHz prototype oscillators. Subsequent measurement of individual oscillator-resonator short-term frequency stability show that oscillator output-signal phase-noise sideband levels below f = 1 kHz are a result of resonator instability rather than sustaining-stage transistor phase noise.

In addition, the use of identically operated resonators is shown to provide a means for reducing device output-signal frequency instability due to mechanical stress by using resonators exhibiting equal and opposite acceleration-frequency sensitivities.

Introduction

Figure 1 shows a functional schematic diagram of a low-noise, crystal-controlled oscillator recently developed at Westinghouse. The novel features of the circuit that play an important role in reducing the level of output-signal phase noise are (1) use of an SC-cut crystal resonator capable of linear operation at high (5 mW) levels of crystal dissipation, and (2) connection of the oscillator crystal resonator in series with the load (common base amplifier) circuit so that the frequency selectivity of the resonator is utilized to band-limit the oscillator output signal spectra.

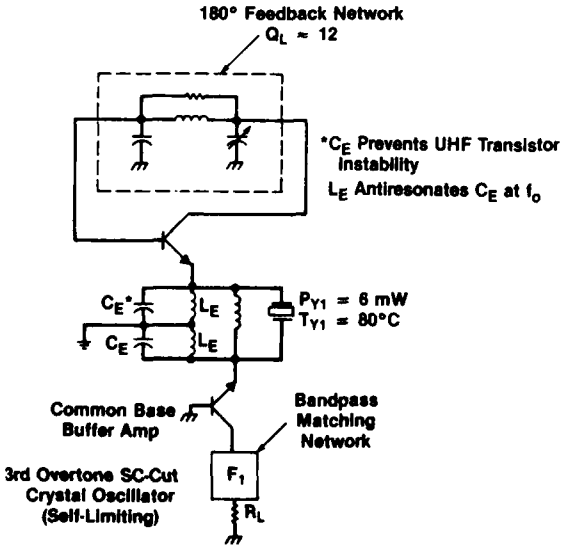


Figure 1. SC-Cut Crystal Oscillator Functional Schematic Diagram

The spectral density of the fractional phase fluctuations of the output signal of the circuit of figure 1 (over a carrier offset-frequency range 10 MHz ≥ f ≥ 10 Hz) can be expressed as:

$$S_{\delta\phi}(f) = 2\left[\left(\frac{K1}{f} + K2\right)\left(\frac{1}{d\phi/df}\right)^2\left(\frac{1}{f}\right)^2 + \frac{K3}{f} + K4\right] \quad (1)$$

Where K1 = Oscillator (open loop) flicker noise constant
K2 = Oscillator (open loop) white noise constant
f_o = Carrier frequency
$d\phi/df$ = Effective (loaded) oscillator resonator phase slope $= 2Q/f_o$ and
Q = Effective (loaded) oscillator resonator Q
K3 = Buffer amplifier flicker noise constant
K4 = Buffer amplifier white noise constant = FKT

The important assumption that has been made with regard to equation 1 is that short-term frequency instability in the quartz resonator itself does not have a predominant effect and may be neglected.

Figure 2 shows in general form the resulting oscillator short-term stability. $\mathcal{L}(f) = 10 \log [S_{\delta\phi}(f)/2]$ and is defined as the ratio of the power in one phase-noise sideband, on a per-hertz bandwidth spectral density basis, to the total signal power, at Fourier frequency f from the carrier. As shown in the figure, there is a conversion of oscillator (open loop) white phase noise to white frequency noise (20 dB/decade) and flicker of phase noise to flicker of frequency noise (30 dB/decade) at frequencies less than the oscillator resonator half-bandwidth. This is the result, in the closed loop oscillator, of the conversion of (open loop) oscillator phase uncertainty to (closed loop) frequency uncertainty due to the requirement in the oscillator of maintenance of $2n\pi$ radians closed-loop phase shift. The degree of phase-to-frequency conversion is determined by the oscillator resonator phase slope or group delay.[1]

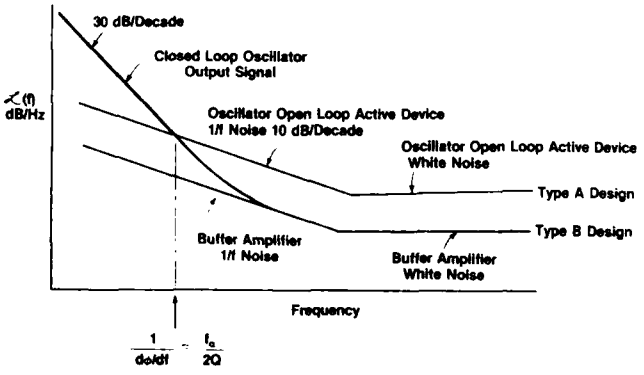


Figure 2. Typical Phase-Noise Sideband Spectra for VHF Crystal Oscillator

*This work was sponsored by the Electronic Systems Division of the Air Force Systems Command

CH2186-0/85/0000-0197$1.00©1985IEEE

At frequencies in excess of the resonator half bandwidth, the output-signal spectral behavior can be influenced by details of the oscillator circuit design. For the figure 1 type circuit (denoted by the circuit B curve in figure 2), the phase noise spectrum is further attenuated by virtue of the narrow-band selectivity of the resonator impedance characteristic.

Figure 3 shows a comparison of measured output-signal phase-noise sideband spectra for the circuit of figure 1 and an earlier developed production circuit using an AT-cut crystal resonator. The data in figure 3 was obtained by phase-locking two identical oscillator circuits with actual measurements made after times-2 frequency multiplication.

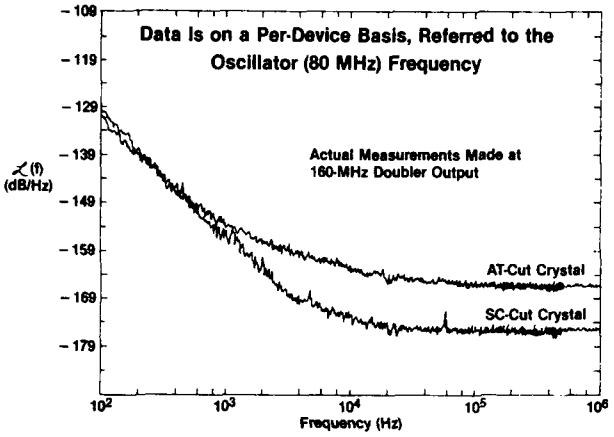


Figure 3. Short-Term Stability for AT-Cut and SC-Cut Crystal Oscillators. Data Is on a Per-Device Basis, Referred to the Oscillator (80 MHz) Frequency.

As shown in the figure, the circuit of figure 1 provides a means for achieving substantial reduction in output-signal phase-noise level for carrier offset frequencies in the range $f \geq 1$ kHz. For $f \leq 1$ kHz, however, no improvement is obtained. The lack of spectral improvement was originally attributed to the assumption that the oscillator circuit transistor flicker noise level (K1 term in equation 1) was substantially unchanged in the new circuit, as was the value of resonator loaded Q. It occurred to the author that one method for obtaining lower oscillator-signal phase-noise levels, then, was to consider increasing the oscillator closed-loop signal group delay through the use of multiple (two) crystal resonators.

Dual Crystal Controlled Oscillator

The concept of using more than one crystal resonator in a single oscillator may not be novel. However, it appears that no actual multiple crystal oscillator circuit is in current use, possibly due to the problems associated with required tracking between resonators. As will be explained, development of the SC-cut crystal makes such a circuit design practical. Additionally, the concept of extracting multiple correlated oscillator output signals through each of separate load circuits represents an important novel aspect of the subject circuitry.

Figure 4 shows a simplified schematic diagram for the subject oscillator circuit. The circuit is similar to that of figure 1, except that two cascaded transistor stages are used in the oscillator sustaining circuit, each connected to separate crystal resonators feeding separate load circuits. The result is that the effective oscillator closed-loop signal transmission phase slope, $d\phi/df$ (or group delay, $d\phi/d\omega$), is 2½ times larger than for the figure 1 circuit. The reason the group delay is more than simply twice that for the figure 1 circuit can be explained as follows.

Operation

Static measurement of the figure 1 circuit $d\phi/df = 2Q/f_o$ indicates that the effective resonator Q is typically 70 percent of the unloaded resonator Q. The degradation in effective Q is largely a result of additional resonator loading from the resistive portion of the sustaining

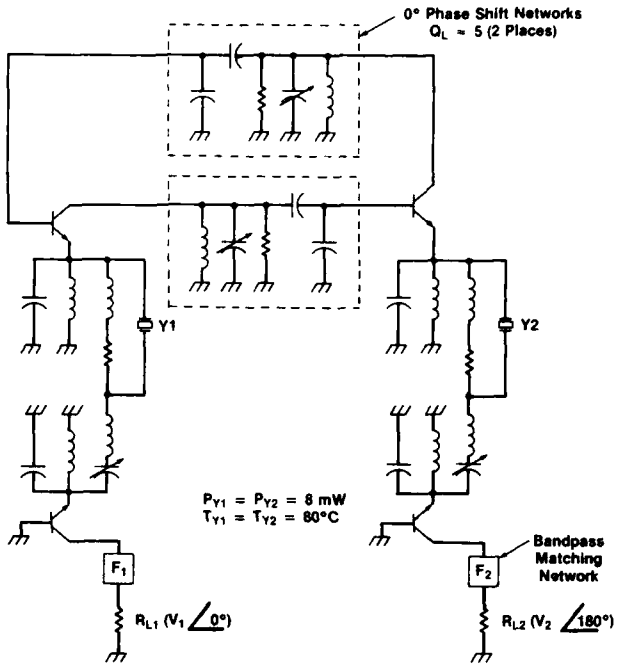


Figure 4. Dual SC-Cut Crystal Oscillator Functional Schematic Diagram

stage and buffer amplifier impedances facing the crystal resonator. Because the oscillator circuit of figure 1 is self-limiting, steady-state operation requires a reduction in small-signal (closed loop) excess gain. This reduction is accomplished (with increasing signal level) when Q1 in figure 1 current-limits or "turns off" during a portion of the instantaneous RF signal waveform. As a result of partial conduction in Q1, the resistive portion of the Q1 emitter-to-ground impedance "seen" by the crystal unit increases.

It is the fundamental Fourier component of this impedance, together with the input impedance of the common base buffer amplifier, that effects the resonator Q degradation. For the figure 1 circuit, small-signal excess gain is set typically at 3 to 5 dB. In other words, Q1 must cut off for a sufficient portion of the RF waveform cycle to reduce the Q1 signal gain by 4 dB. In the figure 4 circuit, current limiting is identically employed in each transistor stage, but reduced limiting (2 dB) is required in each individual stage. Static measurement of the figure 4 circuit frequency-phase sensitivity indicates that individual resonator Q degradation is about 10 percent ($Q_L \approx 0.9\ Q_U$).

Implementation

The phase shift network in figure 4 can be implemented in any number of ways. Both can be configured as identical pi networks, as shown in the figure. Selection of network reactive and resistive component values allows the desired network voltage gain and selectivity (bandwidth) to be obtained. The cascaded selectivity of the two phase-shift networks must be sufficient to prevent circuit oscillation not only at unwanted resonator overtone modes of vibration but, for the case of the SC-cut resonator, at the so-called B mode (which occurs at a frequency approximately 10 percent higher than for the desired C mode). Selection of 20 to 25 percent bandwidth in each network suffices. It should be noted that 180 degree phase shift networks could be employed in place of the 0 degree networks. If this is done, oscillator outputs are nominally in phase.

It is necessary, in the circuit of figure 4, for the series-resonant frequency of each of the two resonators to either remain constant or to drift identically over time and temperature. The frequency change exhibited by a temperature controlled SC-cut resonator [housed in a simple, readily available TO-8 controller such as the WJ-334334 manufactured by Watkins Johnson allowing rather crude ($\Delta T = \pm 10°C$)

temperature control can be held to less than $\pm 2 \times 10^{-7}$. In addition, aging rates for these resonators are vendor-guaranteed at less than 5×10^{-10}/day. When one compares the estimated long-term 1-PPM non-tracking frequency change between resonators (due to temperature change and aging) to the 10-PPM estimated loaded resonator bandwidth, the result is a shift in the oscillator operating frequency from a condition in which each resonator impedance is pure resistive (series resonance) to a condition in which each resonator impedance has a small imaginary component (8 degree impedance phase angle).

For frequency uncertainties of about 10 percent of the resonator bandwidth, the resultant operating frequency remains substantially centered in the resonator series-resonant transmission response, and oscillator performance remains unaffected.

The additional advantage of the use of the new (figure 4) circuit involves the extraction of oscillator signal current through separate load circuits. At low modulation rates the oscillator output signals are correlated. At modulation rates in excess of the resonator bandwidths, the signal spectra, including the additive noise of the load (buffer amplifier) circuits, is uncorrelated by virture of the high out-of-band impedance exhibited by the resonators. Thus, amplitude summation (using a properly phased power combiner) or frequency summation (using a balanced mixer) of the individual oscillator output signals can be used to effect an additional 3 dB reduction in net output-signal phase-noise levels at high modulation rates, as shown in figure 5.

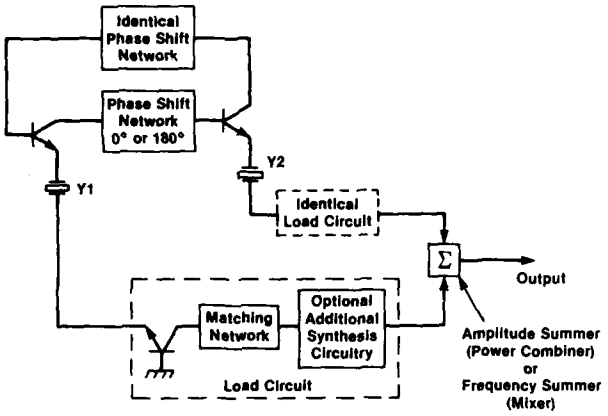


Figure 5. Dual Crystal Oscillator Output-Signal Summation for Reduction of (Uncorrelated) Load-Circuit Additive Noise

Figure 6 shows a detailed schematic diagram for the two prototype oscillators that were constructed. Sustaining-stage tuning procedure is quite straightforward. Oscillation is initiated by lightly coupling an 80-MHz signal generator output into the oscillator circuit and tuning C1 through C4 for a maximum signal at either RF output port. C1, C2 are then tuned for maximum level at RF output 1, and C3, C4 are tuned for maximum level at RF output 2.

To measure oscillator-signal phase noise, one of the prototypes was modified for voltage tuning as shown in figure 7. A tuning sensitivity of 35 Hz/volt was obtained. The relative phase change between the two oscillator output signals was less than ± 0.5 degree over a 3 to 10 volt tuning voltage range, indicating excellent tracking between the Y1 and Y2 tuning circuits.

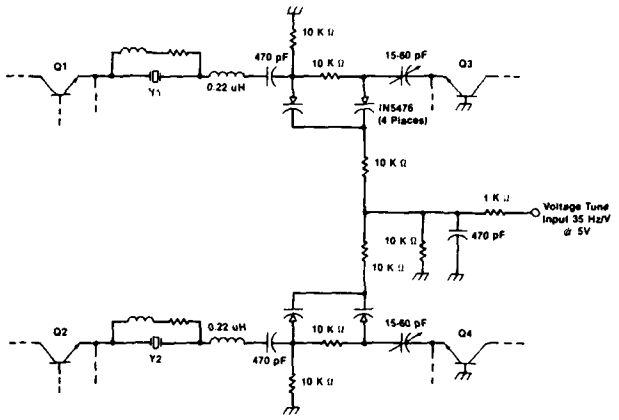


Figure 7. Oscillator Modification for Voltage Tuning

Figure 8 shows the phase noise measurement test setup using a Hewlett Packard 3047A Automated Phase Noise Measurement System. Figure 9 shows the measurement results. Figure 10 shows a comparison of phase noise sideband spectra, with the effects of instrumentation noise removed, for single AT-cut, single SC-cut, and dual SC-cut Westinghouse oscillator designs. As shown, use of the dual-crystal circuit has resulted in substantial (4 dB) additional improvement in obtainable output-signal phase-noise spectral level over the entire measurement band.

Resonator Short-Term Stability

In connection with recently observed unit-to-unit variation in the output-signal phase-noise sideband level of production VHF AT-cut crystal-controlled oscillators, measurements of actual crystal resonator short-term frequency stability have been conducted at Westinghouse us-

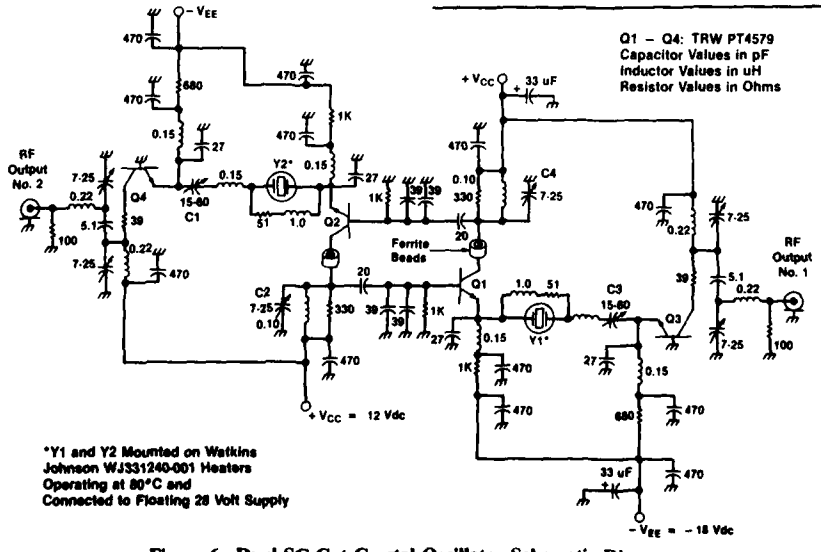


Figure 6. Dual SC-Cut Crystal Oscillator Schematic Diagram

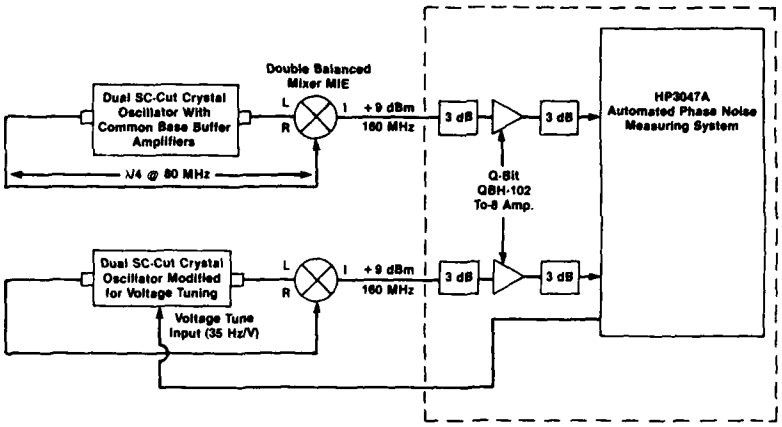


Figure 8. Dual SC-Cut Crystal Oscillator Phase-Noise Measurement Test Setup

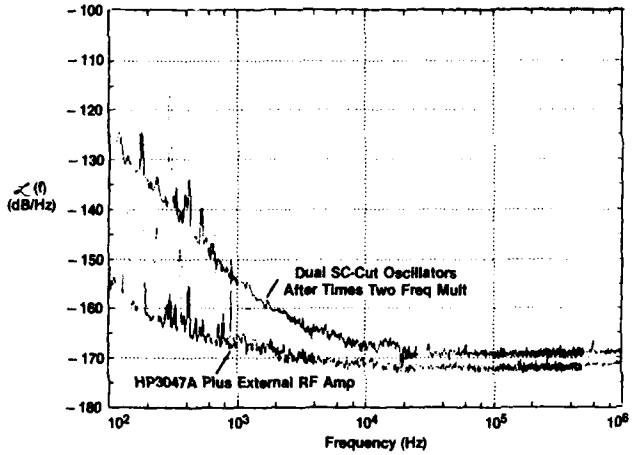


Figure 9. Comparison of Dual SC-Cut Crystal Oscillators and HP3047A Instrumentation Phase Noise Spectra

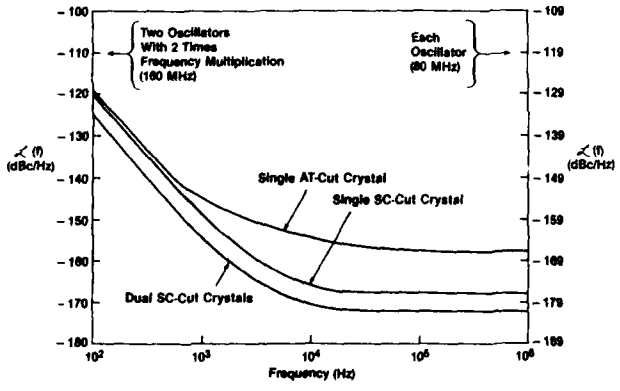


Figure 10. Comparison of Single and Dual Crystal Oscillator Phase Noise With Effects of Instrumentation Noise Removed

ing techniques similar to those described in work reported earlier.[2,3] A separate paper dealing with the details of this work is being prepared for possible presentation at the 1985 IEEE Ultrasonics Symposium.

As part of these experiments, the third overtone SC-cut crystals were removed from one of the prototype dual crystal oscillators. (Unfortunately, the second prototype oscillator had been delivered to another facility, and the resonators used in the second prototype could not be recovered.)

Subsequent measurement of SC-cut crystal short-term frequency stability was made using the HP3047A by driving each of the SC-cut resonator under test (6 mW drive) and a second, previously measured, "low noise" AT-cut resonator (1 mW drive) from a common low-noise crystal oscillator. The test setup is similar to that used by Elliot and Bray.[3]

Figure 11 shows the measurement results, after appropriate data conversion, as fractional frequency spectral densities.[4] Also plotted in the figure is the predicted effect, on oscillator output signal spectrum, of the use of these two resonators in the dual-crystal oscillator circuit. The prediction is based on the fact that, in the oscillator circuit, the output signal frequency is the mean of the resonant frequencies of each of the two crystal-plus-tuning-reactance networks (assuming equal resonator loaded Q).

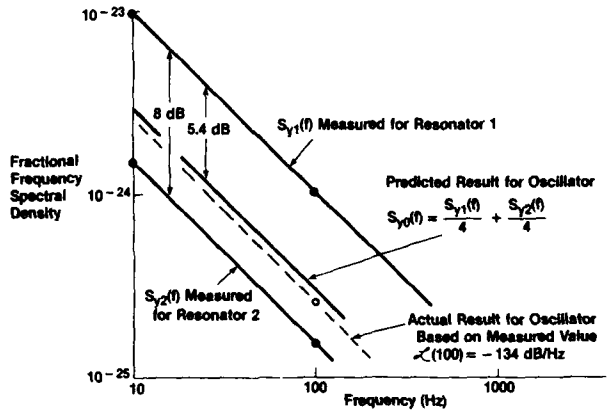


Figure 11. Comparison of Measured Short-Term Frequency Stability of Oscillator Signal and Individual Oscillator Resonators

Further, a frequency change in one of the resonators (assuming stationary frequency in the other) produces half that change in the oscillator output signal frequency, and visa versa. Thus:

$$S_{yo}(f) = \frac{S_{y1}(f)}{4} + \frac{S_{y2}(f)}{4} \qquad (2)$$

where $S_{y1}(f)$ and $S_{y2}(f)$ represent the spectral density of the fractional frequency fluctuations of the individual resonators, and $S_{yo}(f)$ represents the corresponding effect on the oscillator output signal.

Figure 11 also shows the figure 10 measured oscillator output-signal phase-noise spectrum converted to fractional frequency spectral density

format. The results show that the current limitation in oscillator output-signal short-term frequency stability at modulation rates below 1 kHz is due to instability in one of the oscillator crystals (Y1), and they indicate that further improvement is possible using selected "low noise" resonators.

In addition, because of the nature of the circuit, it appears possible to obtain reduction in output-signal frequency sensitivity to circuit mechanical/acoustic stress through the use of resonators exhibiting equal and opposite force-frequency sensitivities.[5,6]

Conclusions

A novel low-noise crystal-controlled oscillator configuration has been developed. The use of two identically operated, SC-cut crystal resonators, with output signal (current) extraction through the resonators themselves, has been shown to offer significant reduction in oscillator output-signal phase-noise sideband spectra.

Measurement of individual oscillator resonator short-term frequency stability indicates additional reduction in oscillator output-signal phase-noise levels can be obtained at modulation rates below 1 kHz by screening resonators for lowest $S_y(f)$ performance.

An additional feature of the dual-resonator oscillator configuration is the potential improvement obtainable in output-signal frequency change due to mechanical/acoustic circuit stress through the use of matched resonators exhibiting equal and opposite force-frequency sensitivities.

References

1. D.B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proc. of IEEE,* Vol. 54, No. 2, February 1966, pp. 329-330.
2. F.L. Walls and A.E. Wainwright, "Measurement of the Short-Term Stability of Quartz Resonators and the Implications for Crystal Oscillator Design and Applications," *IEEE Trans. on Inst. and Measurement,* Vol. IM-24, No. 1, March 1975, pp. 15-20.
3. S. Elliot and R.C. Bray, "Direct Phase Noise Measurements of SAW Resonators," *Proc. 1985 IEEE Ultrasonics Symposium,* pp. 180-185.
4. D. Halford, J.H. Shoaf, and A.S. Risley, "Spectral Density Analysis: Frequency Domain Specification and Measurement of Signal Stability," *Proc. 27th Annual Symposium on Frequency Control,* USAEC, Fort Monmouth, N.J., June 1973, pp. 421-430.
5. A. Ballato, "Resonators Compensated for Acceleration Fields," *Proc. 33rd Annual Symposium on Frequency Control,* USAEC, Fort Monmouth, N.J., June 1979, pp. 322-336.
6. R.L. Filler, "The Effect of Vibration on Frequency Standards and Clocks," *Proc. 35th Annual Symposium on Frequency Control,* ERADCOM, Fort Monmouth, N.J., June 1981.

A SATELLITE OSCILLATOR FOR VERY PRECISE ORBITOGRAPHY : THE DORIS PROGRAM

By A. DEBAISIEUX*, JP. AUBRY*, E. GERARD* and M. BRUNET**

* CEPE - Compagnie d'Electronique et de Piézo-Electricité
95100 - ARGENTEUIL - FRANCE

** CNES - Centre National d'Etudes Spatiales
31055 - TOULOUSE - FRANCE

- ABSTRACT -

The increasing capabilities and expected operational time of modern satellites require high performances oscillators manufactured with approved technologies. For radar altimetry of the POSEIDON program, precise positionning of the satellite is also required.

The main topic of the DORIS equipment is to allow very precise orbitography, with a goal of 10 cm through radial velocity measurement (acc. 0.3 mm/s). The exact positionning is computed from Doppler measurements (dual frequency receiver) between the local oscillator and ground references (global doppler tracking network).

The drift of the local oscillator must be lower than 4.10^{-13}/ mn. in the satellite. This oscillator must present a short term stability of 3.10^{-13} for σ = 10s, a frequency temperature stability lower than 1.10^{-10} between 0 to 40°C (with slope up to 2°C / 15 mn). Total ageing of this oscillator must be lower than 1.10^{-7} in 10 years.

CEPE has manufactured various prototypes of such an oscillator, by using BVA and QAS design resonators which show the best short term stability.

The actual design of the complete oscillator, including thermal and mechanical frame is shown.

Typical results are shown with 10 MHz BVA SC 5th OT, QAS SC 5th OT and QAS 3rd OT resonators.

I - INTRODUCTION

In the past few years, satellite based communication, localization, positionning became an area of great interest, either for scientific, commercial or military applications.

The US satellite TOPEX, to be launched by ARIANE in 1989, will allow a first joint oceanographic mission between NASA and CNES <1>.

The main objective of the POSEIDON project is the study of circulation of oceans. This project is based on the development of two specific equipments, an highly precise radar altimeter and a very precise orbit determination system (DORIS). Both of them will flight on TOPEX 1989 and on SPOT 3 (1991).

The DORIS instrument will be previously tested on SPOT 2, to be launched in 1987.

The goal on accuracy of waveheight measurement requires an orbit determination error less than 10cm.

The oscillator used in the DORIS receiver must then allow low noise doppler measurements.

This paper describes the oscillator specifications and the results obtained on various prototype oscillators using BVA or QAS resonators. Flying model status regarding SPOT 2 is also presented.

II - DORIS RECEIVER

The DORIS system (Doppler Orbitography and Radio positionning Integrated from Space) consists in uplink doppler measurements on signals emitted by about 50 earth beacons. Each beacon has a coverage area about 20°.

The accuracy in orbit determination is required lower than 10 cm.

The doppler measurements will then be processed through perturbation models (gravity, atmospheric drag...) to lead to precise orbit calculation. For example, a specific gravity potential model will be developed by CNES and iteration during the first months of operation will allow the adjustment of the parameters of this model.

The uplink doppler measurements between ground beacon and orbit receiver were choosen because of their simplicity. Measurements are performed and dated on board and sent to earth via telemetry.

Velocity measurements errors are mostly due to short term instabilities of oscillators (0.2 mm/s --> 5.10^{-13}).

Other noise sources add a 0.1 mm/s inaccuracy.

In addition, the frequency drift of the local oscillator during the time the satellite passes over a given beacon has a major impact on the global processing.

Dual frequency operation allows to reduce most of the atmospheric propagation effects (Ionospheric

CH2186-0/85/0000-0202$1.00©1985IEEE

error).

Meteorological parameters detected around the beacon are sent to the satellite. A correction model allows the reduction of tropospheric errors.

The DORIS instrument design is shown on figure 1. Doppler measurements are performed on two frequencies (401.25 MHz and 2 036.25 MHz). All the beacons are continuously transmitting at the same time on the same frequency. The system knows a priori the doppler frequency of the signal it has to track. This knowledge avoids (or limits) false data acquisition and allows the selection of a given beacon when two or more are simultaneously received.

More details on the DORIS system and receivers can be founded in <1>.

Beacon and local oscillator are critical items in this system.

Basic requirements of the local oscillator come out from :

- short term stability (Doppler measurement : 0.3 mm/s accuracy requires $\sigma_y(\tau) < 5.10^{-13}$ for τ = 10 s).

- mid term stability (frequency drift during the pass above one beacon must be lower than $1.5.10^{-13}$/mn).

III - DORIS OSCILLATOR SPECIFICATIONS

1 - Oscillator specification

The uplink dual doppler receiver of the DORIS instrument requires very high stability oscillator. One just has seen that the doppler measurement accuracy requires (in both beacons and receivers) a short term stability better than 5.10^{-13} (for τ = 10s) and mid term stability better than $1.5.10^{-13}$/ mn (during 15 minutes).

Additionnal requirements come out from the nominal conditions of operation in the satellite.

These conditions are :

- temperature range : 0 - 40°C
- maximum temperature drift rate : 2°C/15mn
- maximum DC power : 3 W (0 - 40°C under vacuum).

Furthermore, this oscillator must be integrated in the receiver on earth, then, because of the small bandwidth of phase locked loop, the frequency variation between atmospheric pressure and vacuum must be very low ($< 1.10^{-8}$ is specified).

Table I gives the summary of the main specified parameters and performances of the DORIS oscillator.

2) Qualification test sequence

Along with the performance specification, the qualification / verification test sequence is also specified.

The measurements which have to be performed on MQ and MV oscillators are as described on table III.

These tests have been asked for providing a real insurance that the up to date performances which are required (short term and mid term stabilities) could (or would be) preserved under the nominal conditions of the satellite (noisy DC supply, turn on under vacuum, temperature drifts under vacuum, etc....).

IV - CRYSTAL RESONATORS

1) Resonator design

From the oscillator specification it comes out that the resonator must show simultaneously the following performances.

- potentially low ageing ($< 1.10^{-10}$/ day)
- high radius of curvature around turn over temperature (SC cut)
- low short term stability
  ($\sigma_y(\tau) < 5.10^{-13}$ at τ = 10 s)
- low pressure sensitivity

Furthermore, those crystal resonators must be vibration resistant and the technology will have to be approved by the Space Agency.

From the above criteria, the QAS or BVA design resonator is the only one which can meet the DORIS specification < 3, 4 >.

The table II gives summary of the typical performances which can be achieved with BVA and QAS design resonators. The most sensitive parameters regarding the DORIS mission have been considered (G sensitivity, warm-up time capability ... are not included).

The three kinds of resonators :

- BVA 10 MHz SC cut 5th overtone
- QAS 10 MHz SC cut 5th overtone
- QAS 10 MHz SC cut 3rd overtone,

are then competitors to be utilized in DORIS oscillators

2) Quality insurance

To be approved for use on DORIS SPOT 2, the oscillator has to make use of components, technology, materials and processes in agreement with the space approved rules. The resonator then has to go accross a flying approval procedure.

The QAS or BVA design being a new technology, it would have to be submitted to qualification level B 1 of the SCC 3501 generic specification to reach the project quality level.

In fact, in QAS design, most of the manufacturing processes are the same that the approved one in classical resonators. Disregarding ultrasonic

machining, QAS resonator is quite a classical component.

To be in time with the launch schedule for SPOT 2, it was agreed, for QAS technology, to ask for a B 2 level of VOQ, strengthened by mechanical tests on additional VOQ resonators (shocks, vibrations).

The BVA approval for space would have to go through a more complete qualification procedure because of the major differences between standard and BVA designs.

V - DORIS OSCILLATOR DESIGN

1) Basic design

The mechanical design of a Doris oscillator is depicted in figure 2 <5>.

The oscillator is on a PCB which also receives the resonator. This PCB is inserted into a metallic box (40 . 40 . 40 mm^3) in such a way to keep a good thermal contact between the resonator and its surrounding.

The heating power transistors and the oven control circuitry (hybrid package) are welded on the sides of the metallic cube.

This assembly is fixed between two PCBs which receive the DC voltage regulators, isolators and output amplifiers. For thermal point of view this structure is as symmetrical as possible.

The oven control is PID to avoid any thermal dynamic behaviour or overshoot during high rate temperature drifts.

2) Doris prototypes

Various prototypes of Doris oscillators have been manufactured and tested. They utilize either QAS SC cut 3rd overtone (HC 40 or T 21 11), or QAS SC cut 5th overtone (T 35 16) or BVA SC cut 5th overtone resonators as described in IV.

Various thermal and mechanical configurations have been tested and optimized by light iterative changes of the basic design.

The main parameters to take into account to make the final choice of oscillator design are :

- frequency stability vs temperature range
- short term stability
- mean slope (per minute on 15 mn measurements)
- ageing

The expected quadratic mean slope value is 2.10^{-13}/ mn, calculated from 20 successive slopes. Each slope is calculated from 90 measurements of 10 s time duration, i.e. 15 mn of observation (total measurement time = 5 hours).

Taking into account that maximum thermal drift can be 2° / 15 mn between 0 and 40°C, one must achieve a frequency temperature stability around 1.10^{-10} between 0 and 40°C.

Table IV summarizes the results which have been achieved with three kinds of prototypes.

The QAS 3rd overtone resonator is the most powerfull regarding frequency-temperature drift, but BVA is the most powerfull regarding ageing.

Figures 3 to 6 give the typical results of these oscillators regarding the most important DORIS requirements.

- short term stability : figures 3a to 3c give an histogram calculated from $\simeq$ 60 calculations of $\sigma_y(\tau)$ for τ = 10 s and n = 30 samples each time (i.e. : 30 x 60 $\simeq$ 1800 samples) (QAS 3rd OT, QAS 5th OT, BVA 5th OT).

- frequency temperature drift : figures 4a to 4c give f(T) between - 10 and 50°C for each type under normal pressure and under vacuum.

- mid term stability : figures 5a to 5c give frequency drift calculations , successive mean slopes of frequency drift during 15 minutes (τ = 10 s) and the standard deviation of frequency around mean slope (p) for each point.

- spectral phase noise densities are shown on figures 6a and 6b.

To summarize f(T) behaviour under normal pressure and vacuum, at stable temperature or under temperature drift ..., the "thermal vacuum test" described on figure 7, has been applied on QAS 5th overtone OUS oscillators.

Additional results can be obtained from this test:

- retrace
  - normal pressure 0° C : $< 6.10^{-10}$
  - vacuum 0° C : $< 6.10^{-10}$
- DC power consumption during warm up
  - normal pressure 0° C : < 8 W
  - vacuum 0° C : < 8 W
- DC power consumption after warm up
  - normal pressure 0° C : < 6 W
  - vacuum 0° C : < 3 W

In conclusion, simultaneously quite all of the parameters of table I.

3) Technological stresses

In all the space projects one has to mixed the performance requirements of the mission and the stresses applied by the quality level insurance required by the program.

That is to say that one has to utilize components, processes, material which have to be qualified or approved.

We have already pointed out this point in paragraph IV, regarding quartz crystal resonators.

This selection was also applied on the choice of transistors, thermistors, varactors and also for hybrid packages and related technology (design, testind, welding ...). Thermal insulation materials (Delrin ...) have also to be selected within a qualified material list.

VI - FLYING MODELS STATUS

The choosen resonator was a QAS 5th overtone. A shortened procedure for qualification was agreed with CNES (B 2 + level).

As far as it was possible, all of the technological stresses (if they did not reduce performances) were taken into account in the material list, components list and procedure list.

Three high reliability oscillators are under construction :

- M I (Identification Model) : it has to be exactly representative of flying model design. Components must be of the same type (but not space qualified) than in high reliability models.

- M Q (Qualification Model) : it must have exactly the same design, components (high reliability) and use the same procedures that the flying model will have. Qualification test sequence of the complete oscillator will be performed on it.

- M V (Flying Model) : it must be identical to M Q , but testing levels are lower than in M Q.

The identification model is already constructed. It is under testing procedure. It meets the specification on all the points of table I.

M Q and M V are ready for final assembly.

Figure 8 gives picture of M I DORIS oscillator.

C O N C L U S I O N

Nowadays, the specifications of equipments used for satellite based communication, localization, positionning, time dissemination ... call for very sophisticated oscillators.

The radiopositionning instrument, DORIS, joined to the radar altimeter will be able to provide very accurate waveheight measurements.

The DORIS (Doppler Orbitography and Radiopositionning Integrated from Space) system consists in uplink doppler measurements on signals transmitted from about 50 earth beacons.

The target accuracy on waveheight and on satellite orbitography requires an Ultra Stable Oscillator. This one must show low short term stability less than 5.10^{-13} (σ = 10 s), low mid term stability (frequency drift during one pass) $p < 2.10^{-13}$ / mn and low ageing (1.10^{-10} / day) under the nominal conditions of the mission (temperature range 0 - 40°C, maximum temperature slope 2°C / 15 mn, low power consumption).

BVA and QAS design resonators have been used in various prototypes in the DORIS project.

These prototypes allow to build up manufacturing and control processes used for the flying model oscillators.

The complete POSEIDON equipment will be launched on TOPEX and on SPOT 3, DORIS will be tested on the SPOT 2 flight.

The DORIS oscillator developped by CEPE meets the specifications applying for SPOT and TOPEX mission. It would be usable in many other space projects. Its dimensions are very small (102 . 74 . 94 mm).

References

<1> JP. AGUTTES and P. RAISONVILLE :
The French payload on board TOPEX

<2> M. BRUNET :
Problèmes rencontrés dans les applications spatiales des oscillateurs
Séminaire sur les Etalons de Fréquence, BESANCON 1985 p. H 1.

<3> JP. AUBRY and A. DEBAISIEUX :
Further results on 5 MHz and 10 MHz resonators with BVA and QAS designs.
Proc. of 38th AFCS 1984.

<4> JP. AUBRY :
Résonateurs de hautes performances pour applications aérospatiales.
Séminaire sur les Etalons de Fréquence, BESANCON 1985 p. M 1.

<5> C. BEAUVY, G. MAROTEL and P. RENOULT. :
High performances from a new design of crystal oscillator.
Proc. of 15th PTTI 1983.

Frequency	10 MHz	
. Operating temperature range	0 - 40° C	maximum temperature drift slope : 2° C / 15 mn
. Short term stability $\sigma_y(\tau)$	$< 5.\ 10^{-11}$	τ = 10 s 8 h after turn on
. Medium term stability Mq = $\sqrt{\frac{\sum x^2}{n}}$	$< 2.\ 10^{-13}$/mn	8 h after turn on under stable conditions p = least square fit amoung 90 measurements of τ = 10s time duration n = 20 samples
. Long term stability $\frac{\Delta f}{f}$	$< 1.\ 10^{-10}$/day	after one month continuous operation
	$< 2.\ 10^{-7}$/10 years	
. Pressure sensitivity $\frac{\Delta f}{f}$	$< 5.\ 10^{-8}$	between 760 mmHg and 10^{-5} T
DC power consumption	< 6 W	normal pressure 0 - 40° C
	< 3 W	under vacuum 0 - 40° C
Phase noise spectral density		
- at fm = 10 Hz	$\mathcal{L}(f)$: 120 dBc / Hz	
- at fm = 100 Hz	$\mathcal{L}(f)$: 130 dBc / Hz	
- at fm = 1 KHz	$\mathcal{L}(f)$: 145 dBc / Hz	
- at fm = 10 KHz	$\mathcal{L}(f)$: 155 dBc / Hz	
. Vibration sensitivity $\frac{\Delta f}{f}$	$< 5.\ 10^{-9}$	after nominal verification/ qualification vibration levels
. DC variation sensitivity $\frac{\Delta f}{f}$	$< 1.\ 10^{-11}$	20 V ± 1 %
. Load charge sensitivity $\frac{\Delta f}{f}$	$< 1.\ 10^{-10}$	50 Ω TOS < 2

TABLE I : SPECIFICATION OF DORIS OSCILLATOR

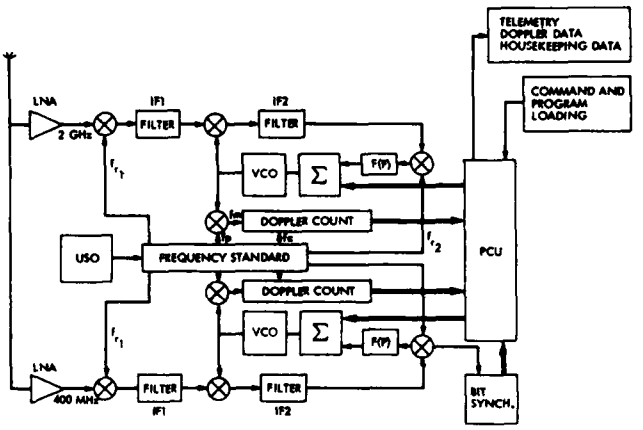


Figure 1

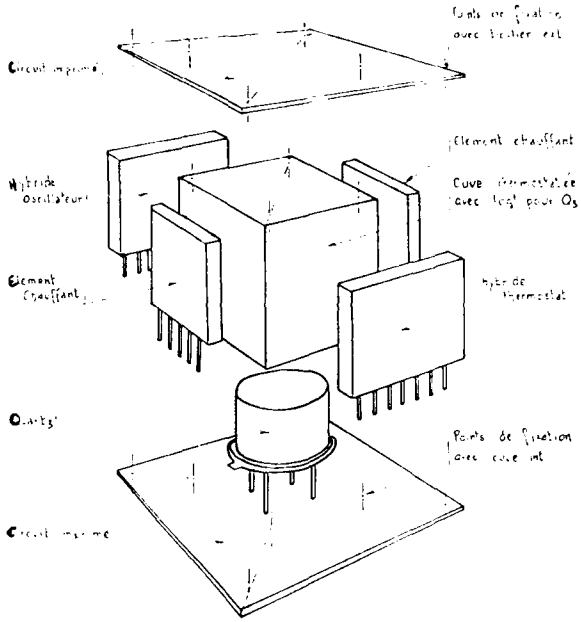
Block diagram of the DORIS receiver

type	frequency overtone cut	$\sigma_y(\tau)$ τ = 10 s	ageing	Pressure sensitivity 760 mmHg - 10^{-5} T		lower mechanical resonant frequencies (1)
				min	max	
B V A T 35 16	10 MHz SC 5th OT	$< 5.\ 10^{-13}$	$< 2.\ 10^{-11}$ / day	$1.\ 10^{-9}$	$1.\ 10^{-8}$	2800 Hz
Q A S T 35 16	10 MHz SC 5th OT	$< 5.\ 10^{-13}$	$< 1.\ 10^{-10}$ / day	$5.\ 10^{-9}$ $5.\ 10^{-9}$	$2.\ 10^{-7}$ $5.\ 10^{-8}$	4600 Hz 3500 Hz
Q A S T 21 11 (HC 40)	10 MHz SC 3rd OT	$< 5.\ 10^{-13}$	$< 1.\ 10^{-10}$ / day	$1.5\ 10^{-9}$	$6.\ 10^{-9}$	2200 Hz

TABLE II : STATE OF THE ART CRYSTAL RESONATORS FOR DORIS OSCILLATOR.

Measurements / Tests	Initial measurements	Thermal cycling and dim pressure	Thermal cycling under vacuum	Mechanical Tests	Final measurements	Reflow and aging
	E 1	E 2	E 3	E 4	E 5	E 6
T 1 Frequency	X	X	X	X	X	X
T 2 Output level	X			X	X	
T 3 Consumption	X	X	X	X	X	X
T 4 Short term stability	X	X	X		X	
T Mid term stability	X	X	X		X	
T Long term stability						X
T Voltage						X
T Vibration Sensitivity	X				X	
T Harmonics - N.A. Harmonics	X				X	
T Temperature - Temperature		X	X			
T Pressure Sensitivity			X			
T Load Sensitivity	X				X	
T Acceleration Sensitivity	X				X	
T Acceleration				X		
T Magnetic Susceptibility				X		
T Miniaturization						

TABLE III : QUALIFICATION TEST SEQUENCE OF DORIS OSCILLATOR



STRUCTURES D'ASSEMBLAGE DU PILOTE

Fig. n° 2

		$\frac{\delta f}{f}$ (T) (T = -40°C)	$\sigma_y(\tau)$ τ = 10 s	med term stability T= stable	/24h /1mn	aging /day	sensitivity to : pressure	load	(V)² voltage	$\mathcal{L}$ (f) 10 Hz; 100 Hz; 1 KHz; 10 KHz (dBc/ Hz)
Q1, Q2, Q3 2 A.T. 5th OT 3rd OT	atm.	3.10^{-10}	$\leq 3.10^{-13}$	5.10^{-11}/mn					5.10^{-11}	-117; -142; -155; -155
	vacuum	1.10^{-10}	$\leq 3.10^{-13}$	5.10^{-11}/mn		*				
Q1, Q2, Q3 2 A.T. 5th OT	atm.	4.10^{-10}	$\leq 5.10^{-13}$	5.10^{-11}/mn					5.10^{-11}	-117; -142; -155; -155
	vacuum	4.10^{-10}	$\leq 5.10^{-13}$	5.10^{-11}/mn		*				
Q1, Q2, Q3 B.T. 5th OT	atm.	1.10^{-10}	$\leq 5.10^{-13}$						5.10^{-11}	-137; -142; -155; -155

* After 50 hours under vacuum only

TABLE IV : PERFORMANCES OF DORIS PROTOTYPES

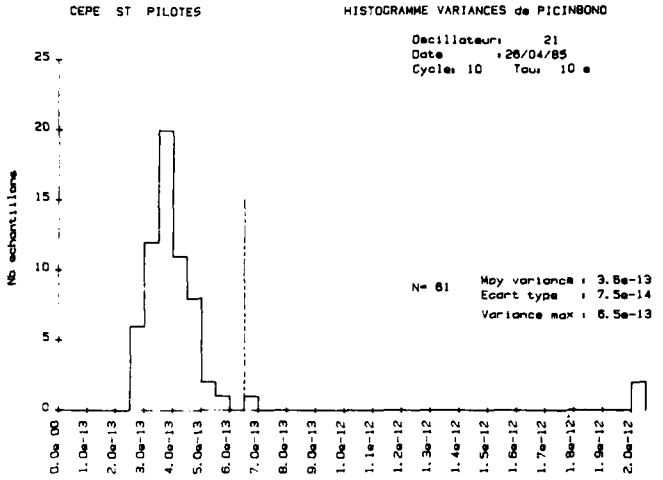


Figure 3a

Short term stability QAS 3rd OT

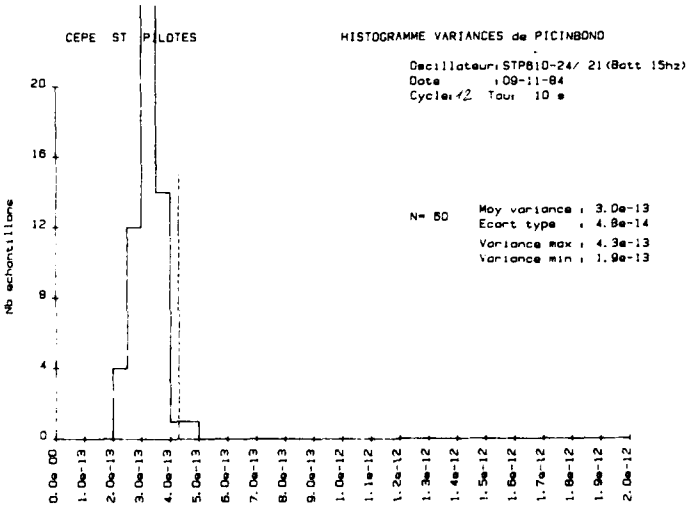


Figure 3b

Short term stability QAS 5th OT

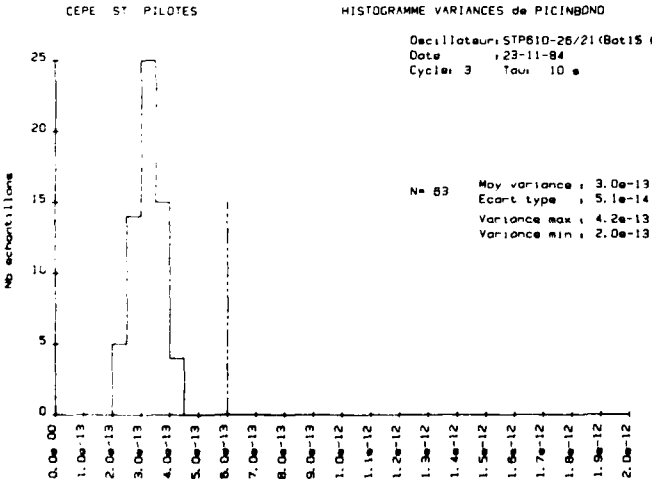


Figure 3c

Short term stability BVA 5th OT

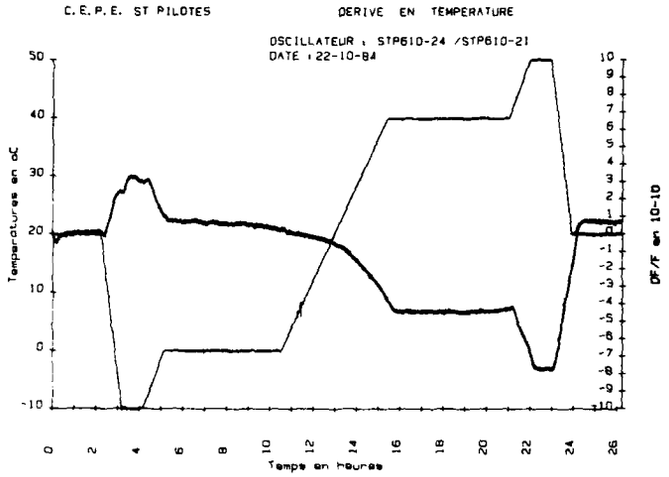


Figure 4a

Frequency t emperature drift QAS 5th OT

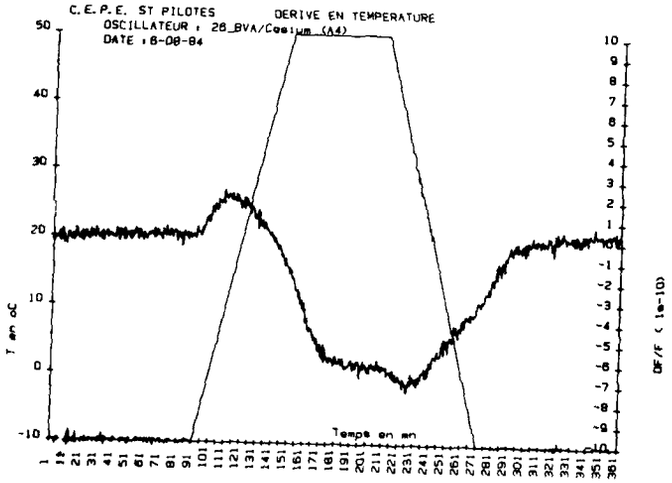


Figure 4 b

Frequency temperature drift BVA 5th OT

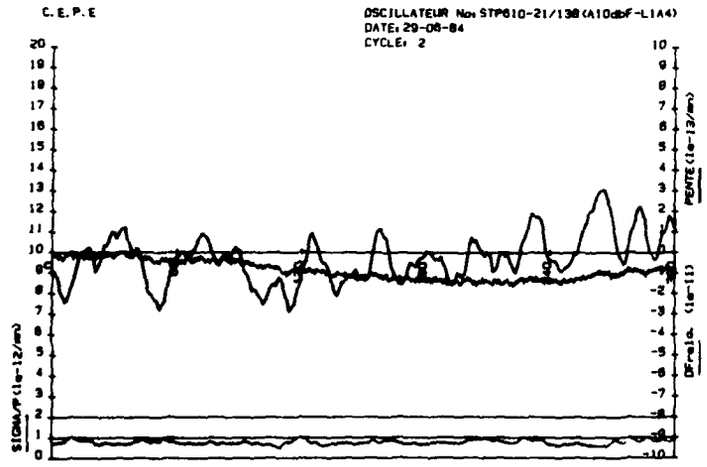


Figure 5a

Med term stability QAS 3rd OT

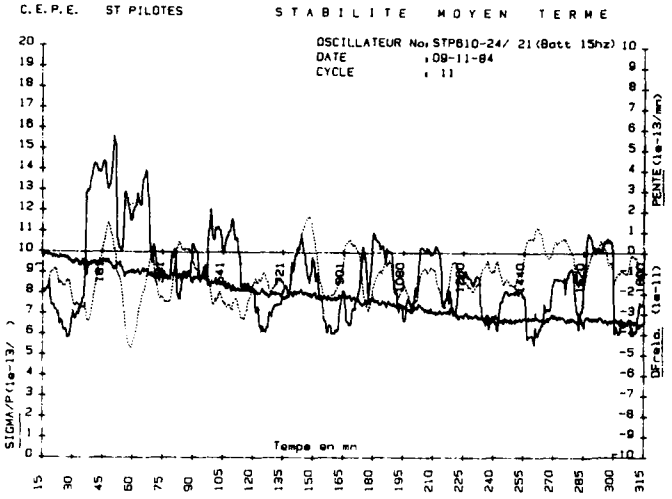


Figure 5b
Med term stability QAS 5th OT

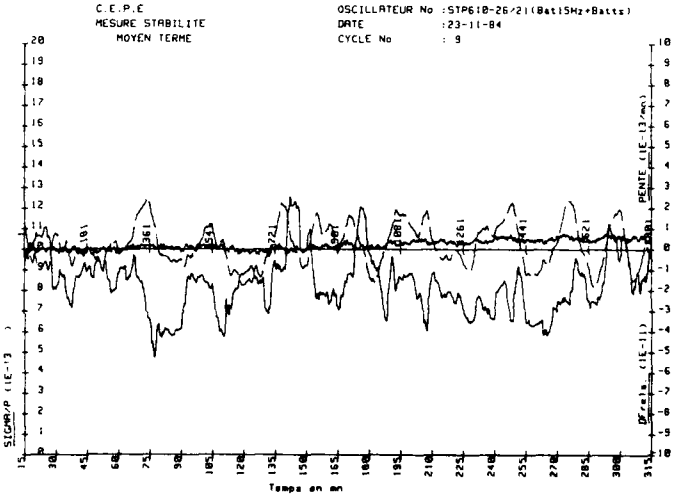


Figure 5c
Med term stability BVA 5th OT

Figure 6a
Spectral density QAS 5th OT

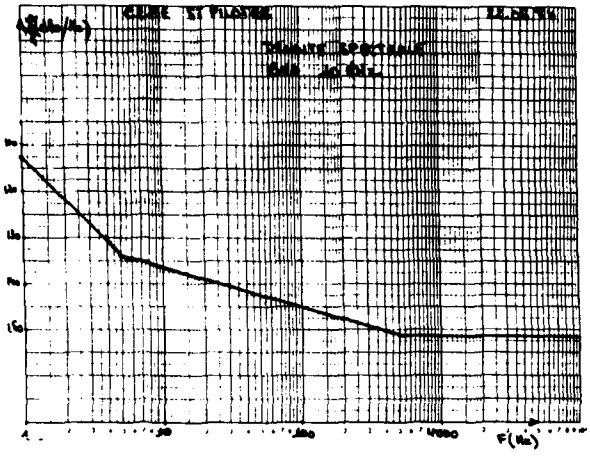


Figure 6 b

Spectral density BVA 5th OT

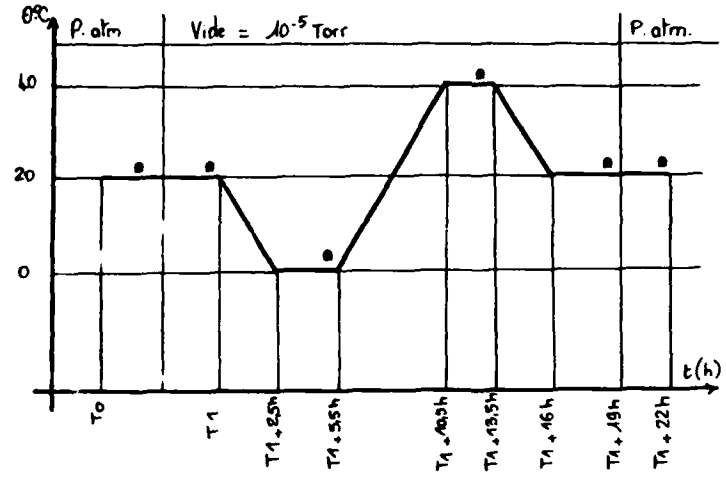


Figure 7

Thermal vacuum tests

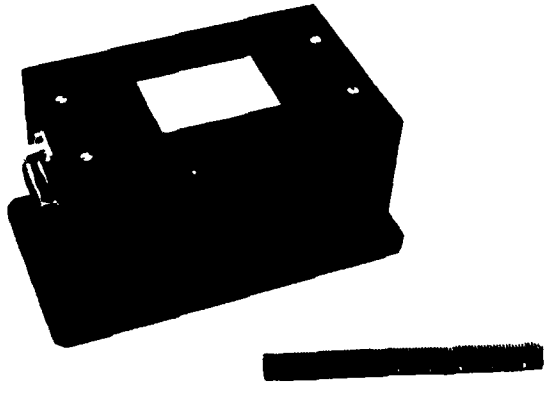


Figure 8

M I DORIS Oscillator

AN OSCILLATOR FOR THE G P S PROGRAM

By G. MAROTEL - G. CARET and J.P. AUBRY

C E P E - Compagnie d'Electronique et de Piézo-Electricité

95100 - ARGENTEUIL - FRANCE

ABSTRACT

It is well known that the G P S program requires oscillators with very high technical performances. Basic specifications call for g-sensitivity lower than 3.10^{-10}/g, and for a fast warm-up (10^{-7},10^{-8} within 5 or 10 minutes at low temperature), in a wide temperature range with a low power consumption.

In the past two years, QAS design resonators have shown their capability to reach low g-sensitivity.

Additional results about the pressure sensitivity of these resonators ($\Delta f/f < 3.10^{-9}$ between ambient and 10^{-5} torr), g-sensitivity, retrace and turn over temperature are presented.

CEPE has also started studies to meet the warm-up requirements. The use of hybrid technology and an accurate design of the thermal dissipation have allowed reduction of size and consumption. Details of the oscillators design and preliminary results are presented.

INTRODUCTION

The NAVSTAR Global Positionning System is devoted to radionavigation, positionning and time transfer dissemination. This system uses a spread spectrum technique on two L band carriers [1].

The various measurements performed (Doppler Shift, Time of Arrival from 3 or 4 satellites) are exploited to lead to 3 directions positionning, velocity, clock offset and clock drift of the user's receiver [2,4].

The accuracy of the system strongly depends on the performance of the local oscillator used in the receiver.

To avoid Navigation errors, jitter on demodulation, jamming sensitivity, the short-term stability, spectral density of phase noise of the oscillator must be as low as possible under the nominal environmental conditions of the user.

Because of the g-sensitivity, phase noise of an oscillator is degraded by dynamic acceleration [2,3].

For a given vibration defined by its frequency fv and acceleration level Ap, the output signal of an oscillator (frequency Fo , g-sensitivity Γ) will show induced side bands at Fo$\pm$ fv , the level of which being :

$$\mathcal{L}_v^1 (fv) = 20 \text{ Log } \frac{\Gamma Ap F_o}{2 fv} \quad (dBc/Hz)$$

$$\text{if } \beta = \frac{\Gamma Ap F_o}{fv} < 0.1$$

After a high rank of frequency multiplication (i.e. 154 Fo or 120 Fo) the modulation index $\beta = \frac{N\Gamma Ap Fo}{fv}$ can become so high that induced side bands can be a few dB only below the carrier.

In particular, for $\beta = 2,4$ (i.e. N = 154, $\Gamma = 3\ 10^{-9}$/g, Ap = 15 g , Fo = 10,230 MHz and fr = 30 Hz). The output signal energy is distributed among the side bands and there is no longer a carrier at NFo !!

When an oscillator is used as a time clock, clock error resulting from the application of a constant acceleration A (applied during the time ta) would be :

$$\Delta t = \Gamma A \, ta$$

It appears clearly that the key parameter of a GPS oscillator is the **g-sensitivity in all the directions of space.** .

In fact, in all the GPS receivers which have been described, critical parameters of stability are lightly different from a manufacturer to another, from a user type to another (military, commercial) and from application to others (manpack, land mobile, shipboard, aircraft, space...).

For military GPS receivers, g-sensisivity of oscillators are generally required lower than $3\ 10^{-10}$/g, at least along one direction [2].

On the other hand, most of the oscillators required by the GPS receivers can be considered as general purpose devices. Disregarding a given GPS receiver designed for a given user, the GPS oscillators would be usable (directly or through minor modifications) in many areas of application.

We present in this paper some oscillators which are extrapolated from standard CEPE products.

The introduction of QAS design resonators, either in HC 40 can (10 MHz, 3rd OT, SC) or in HC 37 can (20 MHz, 3rd OT, SC) and some light modifications in mechanical and electrical design, allow CEPE to provide various ruggedized oscillators which meet the military GPS specifications.

TENTATIVE SYNTHESIS OF GPS SPECIFICATIONS

All over the world, they are as many specifications as specific use of oscillators. In the GPS program only, if we take into account the various receiver's manufacturers, the different user's areas (Space - Military - Industrial - Commercial), and the various military applications (manpack - land mobile - shipboard aircraft, missiles...) the number of specifications is quite high.

CH2186-0/85/0000-0212$1.00©1985IEEE

For other military use, the situation is quite so complicated because of the shape, size and performances which are asked for.

This is not the topic of this paper to answer to a given specification, but it is mainly to describe what are the basic segments of oscillators (resonator/oscillator/oven/mechanical design) which can be brought together to solve a given problem.

For the GPS receiver, one can try to summarize to date specifications on **table I.**

The precision of the informations delivered by the receiver depends on the accuracy of TOA and Doppler Shift measurements under user's nominal operating conditions. These conditions can be for exemple sequential sets of measurements under high J/S ratio and high vibration conditions.

From the user side, the most critical oscillator parameters are :

- fast warm-up (for quickness of data acquisition)
- low spectral phase noise (RF sensitivity, data demodulation, high J/S immunity)
- med term stability (sequential operation, J/S immunity, measurement consistency for sequential navigation sets)
- short term stability (Doppler Shift, navigation accuracy, acceleration performance).

G-sensitivity is the key parameter of the receiver accuracy under vibrations. Through degradation of short-term stability g-sensitivity can induce navigation errors. Because of possible phase noise degradation, g-sensitivity can induce jitter in data demodulation, and degrade RF sensitivity and anti jamming ability.

Carrier and code offset can also occur because of the vibration induced oscillator errors [2, 4].

From the specifications in **Table I**, it comes that the most important characteristics of GPS oscillators are :

- temperature range and f (T) stability
- size and DC power consumption
- warm-up time or time needed to reach operational stability
- dynamic g-sensitivity
- pressure sensitivity
- short term and medium term stability.

Depending on the user type receiver the most stringent parameters are :

Military oscillators

- temperature range up to 85 °C : frequency stability between $\pm$ 1 10^{-8} and $\pm$ 5 10^{-10}.
- warm-up time : as low as 5' to reach 10^{-7} or 10' to reach $\pm$ 10^{-8} *in the worst temperature conditions* ($\theta \simeq -$ 54 °C)
- g-sensitivity : lower than 3 10^{-10}/g
- DC power consumption : can be das high as 10 or 20 watts during warm-up, but it must be as low as possible after warm-up.

Non military oscillators

In GPS commercial receivers specifications, standard oscillators are asked for. Cost and size would be the main parameters.

STATE OF THE ART OF RESONATOR DESIGN

For general specifications, military oscillators require low g-sensitivity. In the resonator itself non linear effects must be as low as possible to minimize pertubations such as thermal dynamic behavior, short term stability, stress sensitivity or drive level stability. Turn over temperature are required between 77°C (oscillator temperature range up to 71 °C) and 92°C (oscillator range up to 85 °C).

Most of these parameters are often incompatible. The g-sensitivity is a field which has been intensively studied, both theoretically and experimentally. CEPE has engaged strong efforts in this area. Since 1982, experimental results have shown yearly improvements [5, 6]. The QAS design must now be recognized as a key device for very low g-sensitivity purpose.

In 1984, QAS design was shown to lead to S/G lower than 2 10^{-10}/g on industrial basis (80 % per batch).

We do not have yet a complete model describing g-sensitivity, but only some theoretical works dealing with phenomenon which are only a part of the g-sensitivity. Then the improvements of QAS technology must be obtained through experimental datas.

The overall sensitivity can be expressed as :

$S = f(\varphi, \theta, n, e, \emptyset 1, \emptyset 2, \emptyset 3, j, \psi i, li, \omega i, \emptyset m, \ldots)$

where :

φ, θ	: are the actual cut angle.
n, e	: overtone and thickness.
$\emptyset$ 1, ...	: are diameters of :
	. *vibrating disc*
	. inner ring
	. outer ring
j	: number of bridges (2 to .....)
ψi	: location of the bridge i (i = 1, j) versus plate coordinates
li, ωi	: length and width of bridge
$\emptyset$ m	: diameter of metallisation (which define the area of interest for the strain developed on the crystal resonator).

Such a relation has not yet be completely theoretically developed and optimization must lie :

- by theoretical analysis restricted to a few parametes
- by step by step experimental improvements

One can find elsewhere details on the QAS technology, manufacturing process and test methods together with the basic analysis of some phenomenon involved in g-sensitivity [6].

The actual S/G performance is shown by the histogram on figure 1.

This curve is the sum of the results obtained with three different batches of resonators manufactured at different period of time, following a given design.

The actual performance lies within approximately 100 % of resonators with a g-sensitivity better than $2\ 10^{-10}$/g in all the direction of the space (and 50 % with $S/G \leq 1\ 10^{-10}$/g).

On the left side of the distribution curve (figure 1) measurements are limited by the resolution of the equipement used for 2 G tip-over tests of resonators only. The lowest measurable value is approximatively $7\ 10^{-11}$/G. Below this limit, complete oscillators must be used.

Up to now, the best g-sensitivity was obtain with cut angles close to SC.

Despite the fact that g-sensitivity is one of the main problem on resonators, some other aspects are of great influence. They are mainly :

- turn-over temperature
- warm-up retrace (even at cold temperature)
- short term stability
- ageing.

To adjust all of these parameters simultaneously, the design variables one can work with are :
- size and shape of the resonator (frequency, overtone...)
- resonator design (QAS, classical).

Table II summarizes the actual QAS design results compared to the goals of GPS oscillators.

From this table, a surprising result is shown. One can relax on g-sensitivity (which was, till recent ly, the most stringent parameter) but one has to increase turn-over temperature.

By moving the φ angle lightly around SC, the inflexion temperature is changed. Then it becomes possible to achieve turn-over temperature higher than 92 °C, with a reasonable (i.e. industrial) tolerance on both φ and θ angles measurements ($\Delta\varphi \simeq 1'$ and $\Delta\varphi \simeq 15''$)

As far as φ angle is moving from SC, g -sensitivity is degraded. The φ angle deviation versus SC is limited by S/G equals to $5\ 10^{-10}$/G.

Furthermore, to achieve fast warm-up it is preferable to use HC 37 can crystals rather than HC 40. The smaller the volume to ovenized is, the smaller the required energy is, the quicker the thermal equilibrium is reached.

It has been already shown [6, 8] that QAS design resonators have mechanical resonances higher than 2 000 Hz and that they show very little frequency variation under vacuum : $\Delta f/f \simeq 1$ to $5\ 10^{-9}$ between ambient atm and 10^{-5} torr

Classical or QAS resonators, with modified or not SC cut, in HC 40 or HC 37 cans have been built up and tested. Performances of these resonators are shown on **Table III.**

Among this set of crystal results, it is easy to find the best resonator to use to meet any given GPS oscillator specifications.

QAS design resonators must be recognized as the key item for GPS oscillators, because it allows to reach simultaneously :

- low short term stability ($\leq 1\ 10^{-12}$)
- low g-sensitivity ($< 5\ 10^{-10}$/G)
- high turn-over temperature (> 90 °C)
- low pressure sensitivity ($< 3\ 10^{-9}$)
- mechanical resonances : > 2 000 Hz).

In addition it would be wrong to think that QAS type resonators are still at the laboratory stage. Some kinds of QAS design resonators are now under industrial control and the capacity, manufacturing processes, delay and cost are not too much different from classical SC cut.

BASIC OSCILLATOR DESIGN

Basically, the high stability oscillators required to meet the GPS or other military requirements, can be developed from three types of standard oven oscillators :

- PMT P
- PMT 730
- OUS.

These basic oscillators are designed for different temperature stabilities (from $\pm\ 10^{-10}$ to $\pm\ 10^{-8}$) with AT cut crystal resonators. Standard environmental conditions are 10 G/500 Hz. They are shown on figure 2.

Electronic design oscillator are clapp type, and they have fine electronic adjustement through varactors.

PMTP Oscillator

This oscillator is shown on figure 3. It has been used from years with AT cut fifth overtone or SC cut third overtone, between 5 and 20 MHz. Resonators used in PMTP oscillator are encapsulated in HC 40 cans. Outer dimensions are 67 x 60 x 40 mm and total volume is 160 cm³. It has been designed for direct mounting on printed circuit board.

The overall mechanical and thermal design is given by three successive metallic boxes, one into another.

The inner box is 20 cm³ in volume. It contains, on a PCB, the resonator, the oscillator and the thermal sensitive elements of the electrical adjustment circuitry

Two heating transistors and thermistors (CTN) are welded on the outside of the box which will be ovenized. DC power is equally distributed among the heating transistors.

The inner box is fixed on a PCB which receives also:
- voltage regulator
- DC power control
- isolator amplifier
- output amplifier.

A second metallic box is fixed on the PCB. It covers the inner one, the heating devices and some thermal sensitive components.

The PCB is fixed to the outer can by insulating supports.

Thermal insulation between the inner can and outside world is performed by the volume of gaz which is trapped between each can.

PMT Oscillator (PMT 730)

The design of the PMT 730 oscillator is the basis of some ruggedized oscillators described later.

The mechanical frame of this oscillator looks like the PMTP one.

Overall dimensions were reduced to 51 x 41 x 25 mm (total volume 50 cm^3). In that case, only two metallic envelopes were used.

The inner can does contain the HC 37 resonator and only some thermal sensitive elements of the oscillator. The volume of this inner can is reduced to 4.1 cm^3.

Only one power transistor is needed (even for fast warm-up) to provide thermal regulation of this small volume.

This inner assembly is fixed on a PCB on which there are the rest of oscillator circuitry, voltage regulator, DC power control, isolator and output amplifier.

O.U.S. Oscillator

(O.U.S. is for Oscillator Ultra-Stable).

This oscillator is devoted to very high performances either for ground or space applications [7.9]. HC 40 can resonators are used in this oscillator.

To provide vacuum operation, thermal insulating performed by gaz is prohibited.

The resonator is located inside a metallic cube (Al or Cu). The heating transistors are welded on the sides of this cube, together with oven control elements (hybrid package).

This metallic box is fixed by thermal insulators to PCBs which are around the box and which receive the rest of electronic circuitry.

This frame is fixed by thermal insulators to the outer can.

The oven control is PID type to reduce thermal transcient effects.

The mechanical design was completed to meet the vibration level specifications of space and military use.

Performances of these three standard oscillators are recalled on **Table IV**, together with those obtained with new oscillators.

RUGGEDIZED OSCILLATORS

Some kinds of the resonators described in § "State of the art of resonator design" have been implanted in modified PMTP or PMT 730 oscillators. These ones have been tested in various environmental conditions.

The main parameters one has measured are :
- g-sensitivity
- f (T) range stability
- retrace and warm-up
- $\mathcal{L}(f)$
- ageing.

Mechanical assembly of PMT 730 oscillator have been modified in order to meet mechanical environmental specifications.

For thermal and electronic point of view, one can mixed the following parameters to meet specifications of **Table I.**

- QAS resonators :
  10 MHz HC 40 in PMTP
  20 MHz HC 37 in PMT 730.
- Turn-over temperature of crystals (modified SC cut)
- DC power and thermal conductivity of inner can
- electronic design oscillator
- thermal insulation.

Typical performances of oscillators including some of these improvements are shown on **Table IV**, to be compared to standard oscillators results.

The warm-up results are given on figure **4** and **5**, (classical PMT 730 and modified PMT) and on figure **6** and **7**, (classical and modified PMTP).

Warm-up measurements have been measured either at 25 °C or at - 54°C (in pulsed air) with modified PMT 730 oscillators. (HC 37 10 MHz 3rd OT resonator).

These measurements have been performed with different DC heating power limitation, to determine the minimum power required to meet warm-up performances.

These warm-up results are summarized in **Table V.**

Warm-up time as low as 2 minutes 30" to reach 10^{-7} of final frequency at -54°C with a DC power consumption less than 15 watts is measured.

Warm-up results of PMTP and modified PMTP oscillators are summarized in **Table VI.**

An interesting result can be observed from these results. The flatter the f(T) curve is around the turn-over temperature, the better warm-up and retrace results are . This is the reason why we use modified SC cut rather than cuts which are far from SC, such as IT cut.

Typical ageing results obtained with these improved oscillators are shown on figure **8, 9 and 10.** Depending on temperature range, AT cut fifth overtone 5 MHz can show daily ageing between 1 10^{-10} and 5 10^{-11} for day, classical SC third overtone around 1.2 10^{-10}/day and QAS SC around 5 10^{-11}, 1 10^{-10}/day (both at elevated temperature).

Retrace lower than $1\ 10^{-9}$ (vs frequency before turn on) (typical 2 to 6 10^{-10}) are obtained with oscillators using QAS resonators (figure 11).

Frequency v.s. temperature stability ranges from $\pm$ $5\ 10^{-10}$ for PMTP QAS SC oscillator to $\pm\ 5\ 10^{-9}$ for PMT 730 SC oscillator in a wide temperature range (-40, 71°C) the PMTP oscillator has 3 metallic envelopes (instead of 2 in PMT 730), has a larger volume 160cm³ to 50 cm³) but it is 10 times better in temperature stability.

The OUS oscillator is mostly devoted to high performance or space application. It is designed for a lower temperature range (- 10, 60 °C).

The performances of this oscillator are :

- f v.s. temp. stability :< $5\ 10^{-10}$ (peak to peak)
- frequency retrace (at 25 °C) :
  - . $1\ 10^{-10}$ (24 hours after turn on)
  - . $3\ 10^{-10}$ (3 hours after turn on)
  - . $3\ 10^{-9}$ (30 minutes after turn on)
- ageing :< $1\ 10^{-10}$/day (figure 12)
- Pressure sensitivity (a/m/10^{-5} torr) : $1\ 10^{-9}$
- g-sensitivity : $1\ 10^{-10}$
- phase noise : (figure 13)
  - . 110 dBc/Hz 1 Hz offset from carrier
  - . 135 dBc/Hz 7 Hz offset from carrier
  - . 137 dBc/Hz 10 Hz offset from carrier
  - . 143 dBc/Hz 100 Hz offset from carrier
  - . 148 dBc/Hz 1 kHz offset from carrier
  - . 153 dBc/Hz 10 kHz offset from carrier

CONCLUSION

There are as many GPS specifications as receiver's manufacturers and user's configuration. Common requirements are low sensitivity, fast warm-up, operating temperature range, spectral density...

We have shown in this papet some basic configurations of such oscillators.

The introduction of the QAS design resonators in 1983 allow CEPE to propose a set of resonators suitable for specific use. These resonators are now economically practical for production.

The basic construction of oven oscillators also provide a set among which we can choose the best compromise regarding :
- size and f (T) stability
- warm-up and retrace
- spectral density (quiescent and under vibrations)
- cost.

The highest stability (frequency temperature, $\sigma_y(\tau)$ & (1) g-sensitivity) is obtained with an OUS oscillator working with a QAS SC 10 MHz, 3rd OT, resonator. Vacuum operation is allowed.

The best g-sensitivity oscillator is a ruggedized PMTP working with a QAS SC 10 MHz, 3rd OT, resonator. g-sensitivity value as low as $1\ 10^{-10}$/G have been measured, with a warm-up time around 15' in -54, 71 °C temperature range. Improvement of temperature range is under development.

The lowest warm-up time was obtained with a modified PMT 730 oscillators : the $\pm\ 1\ 10^{-7}$ window is reached within 2 mn 30 s for an ambient temperature of - 54°C. The 10 MHz resonator used in this oscillator provides g-sensitivity around $8\ 10^{-10}$/g.

A modified PMT 730 oscillator, working with a QAS SC cut 20,460 MHz, will meet both g-sensitivity and warm-up time requirements, in a very small size : 51 x 41 x 25 mm.

REFERENCES

[1] Capt. D.M. TENNANT
Proc. of 35th Ann. Freq. Control Symp. (1981)
p. 532

[2] M.D. YAKOS and E.H. HIRT
Proc. of 35th Ann. Freq. Control Symp. (1981)
p. 537

[3] R.L. FILLER
Proc. of 35th Ann. Freq. Control Symp. (1981)
p. 31

[4] D.L. HESSICK and W.C. EULER
Proc. of 38th Ann. Freq. Control Symp. (1984)
p. 341

[5] A. DEBAISIEUX - J.P. AUBRY and J.GROSLAMBERT
Proc. 15th PTTI (1983)

[6] J.P. AUBRY and A. DEBAISIEUX
Proc. of 38th Ann. Freq. Control Symp. (1984)
p. 190

[7] C. BEAUVY - G. MAROTEL and P. RENOULT
Proc. 15th PTTI (1983)

[8] J.P. AUBRY
Proc. of. seminary "Les Etalons de Frequence"
Besançon (1985) - p M1.

[9] A. DEBAISIEUX and J.P. AUBRY
Proc. of 39th Ann. Freq. Control Symp. (1985)
To be published.

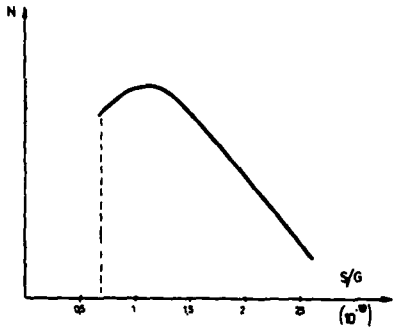


FIG 1 s/g distribution

Receiver-type	User	Operating temperature range (OTR)	Frequency stability v.s. OTR	Ageing	Med. term stability	σ y (τ)	G sensitivity	P sensitivity	Warm up		DC Power		Size (inches) (mm)
									t	L	warm up	stable	
Rockwell-Collins [2]	Mil	- 54,85 °C	$\pm 1 \cdot 10^{-8}$	$5 \cdot 10^{-10}$ /day	$\pm 10^{-9}$ 20'	10^{-10} rms	$5 \cdot 10^{-10}$ 1 axis $2 \cdot 10^{-9}$ 2 axis		$\pm 10^{-8}$ $\pm 10^{-9}$	30' 5'			
Magnavox [4]	mil @ 5	- 40,81 °C			$5 \cdot 10^{-13}$ 1 day	$5 \cdot 10^{-12}$	$5 \cdot 10^{-10}$ G 4-100 Hz		10^{-7}	5'	20 W	5 W	28 x 22,4 x 2" 71 x 57 x 51 mm 228 cm³
Magnavox IV	Space	- 30,63 °C					$5 \cdot 10^{-9}$ G			30'			
Magnavox III	mil	- 54,43 °C					$5 \cdot 10^{-10}$ G			5'			
Magnavox III A	Spacecraft	- 40,23 °C					$5 \cdot 10^{-9}$ G			5'			
Magnavox II	Space	- 30,63 °C					$1,5 \cdot 10^{-9}$ G			2,5'			
Magnavox I	Commercial	- 20,4 °C					$5 \cdot 10^{-9}$ G			10'			
Wide-Mid				$5 \cdot 10^{-10}$ /day		$5 \cdot 10^{-12}$	$5 \cdot 10^{-10}$ G			30'		0,5 W	< 2 in³
Specifications	mil	- 54,85 °C	$5 \cdot 10^{-8}$	$5 \cdot 10^{-11}$ /day			$5 \cdot 10^{-10}$ G	$5 \cdot 10^{-9}$	$2 \cdot 10^{-8}$	10'	8,5 W	2 W	28 x 51 x 102 mm 345 cm³
		- 54,85 °C	$\pm 5 \cdot 10^{-8}$				$5 \cdot 10^{-10}$ G		$\pm 10^{-8}$	5,5'		2 W	50 x 30 x 68 54 cm³
		- 54,85 °C	$\pm 5 \cdot 10^{-10}$				$5 \cdot 10^{-10}$ G	$5 \cdot 10^{-10}$	0,1 T $\pm 10^{-11}$	30'			3 x 3 x 3" 440 cm³
		- 54,85 °C - 54,85 °C	$\pm 5 \cdot 10^{-9}$				$5 \cdot 10^{-10}$ G		10^{-7} $\pm 7 \cdot 10^{-9}$	5' 20'			3 x 3 x 3" 294 cm³

(1) : From a 6 hr cold soak at - 54 °C , after 5' = 1 2 L

TABLE I - TENTATIVE SYNTHESIS OF GPS OSCILLATOR SPECIFICATIONS

Parameter	GAS Best Results	GPS goal	Frequency Results / goal
			better
			better
			lower
Warm-up			
			lower
			lower
Power			
			better
			better
Physical			
Size			

TABLE II - ACTUAL MASER RESULTS VERSUS GPS GOALS

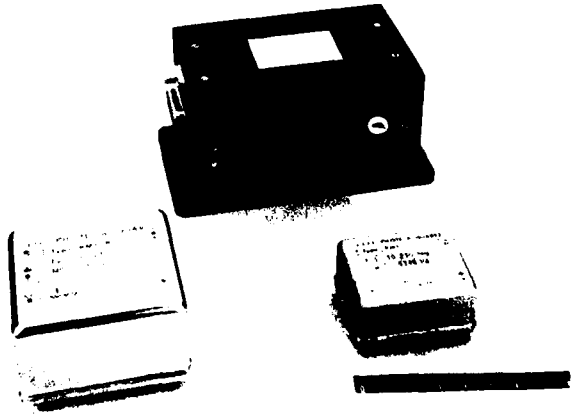


figure 2 - 3

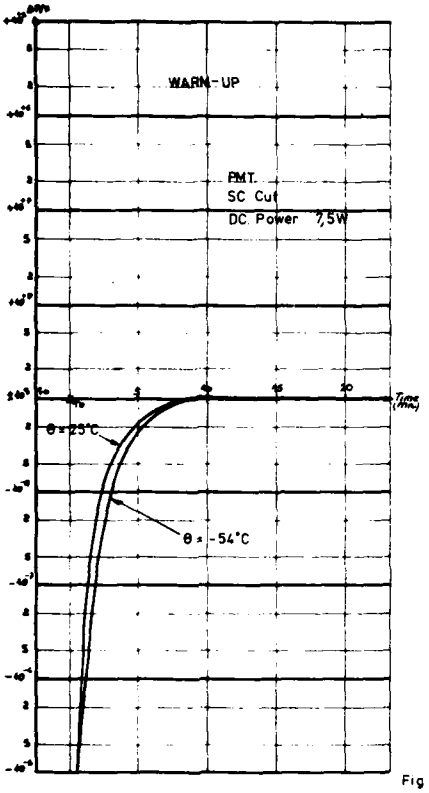


Fig 4

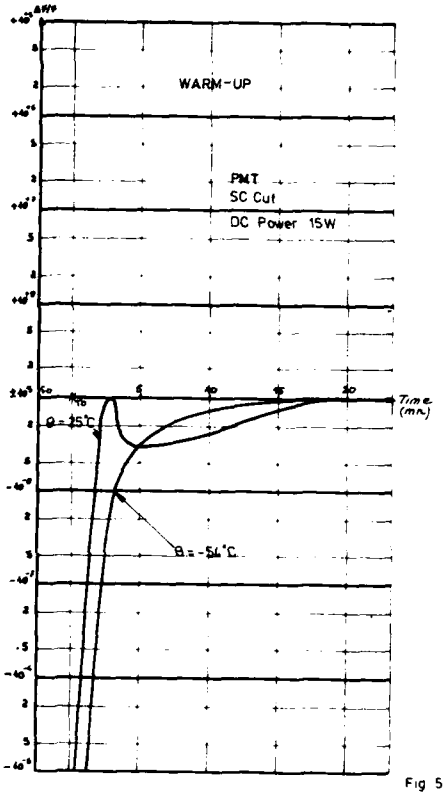


Fig 5

TABLE III — UP TO DATE BAS AND MODIFIED SC CRYSTAL RESONATORS RESULTS

Lab	Technology	Turn Over Temperature	α/G	Status

TABLE V — WARM UP PERFORMANCES OF MODIFIED PMT 310 OSCILLATORS

(Size : 51 x 41 x 25 mm — Volume : 50 cm³)

Osc. Crystal S.D.S.	Warmup Conditions	Max. DC Power	T i m e 10^{7}	t a 10^{8}	R e s o l n 10^{-9}/s

		WARM-UP	AGEING	S/G (Worst Direction)	SHORT TERM.	PHASE NOISE	STABILITY θ -40 +70 °C
PMTP 160 cm³	AT 5 MHz 5th OT (HC 40)	See curves	$1\ 10^{-10}$ / $5\ 10^{-11}$/day	$5\ 10^{-9}$/g	1 ms $2\ 10^{-10}$ 10 ms $2\ 10^{-11}$ 100 ms $5\ 10^{-12}$ 1 s $2\ 10^{-12}$ 10 s $5\ 10^{-12}$ 100 s $1\ 10^{-11}$	1 Hz - 100 dB/Hz 10 Hz - 125 dB/Hz 100 Hz - 135 dB/Hz 1 kHz - 140 dB/Hz 10 kHz - 145 dB/Hz 50 kHz - 145 dB/Hz	10^{-8} p-p.
	SC 3rd OT 10 MHz (HC 40)	See curves	1-2 10^{-10}/day	1,5 10^{-9}/g	1 ms $1\ 10^{-10}$ 10 ms $1\ 10^{-11}$ 100 ms $5\ 10^{-12}$ 1 s $1\ 10^{-12}$ 10 s $2\ 10^{-12}$ 100 s $5\ 10^{-12}$	1 Hz - 100 dB/Hz 10 Hz - 130 dB/Hz 100 Hz - 135 dB/Hz 1 kHz - 145 dB/Hz 10 kHz - 150 dB/Hz 50 kHz - 150 dB/Hz	10^{-9} p-p.
	QAS SC 3rd OT 10 MHz (HC 40)	See curves	$5\ 10^{-11}$ / $1\ 10^{-10}$/day	$1\ 10^{-10}$/g			
PMT 50 cm³	SC 3rd OT 10 MHz (HC 37)	See curves	1-2 10^{-10}/day	$8\ 10^{-10}$/g	1 ms $2\ 10^{-10}$ 10 ms $2\ 10^{-12}$ 100 ms $5\ 10^{-12}$ 1 s $2\ 10^{-12}$ 10 s $3\ 10^{-12}$ 100 s $8\ 10^{-12}$	1 Hz - 100 dB/Hz 10 Hz - 130 dB/Hz 100 Hz - 135 dB/Hz 1 kHz - 145 dB/Hz 10 kHz - 150 dB/Hz 100 kHz - 150 dB/Hz	$1\ 10^{-8}$ p-p.
	SC 3rd OT 20 MHz (HC 37)		2-5 10^{-10}/day	$5\ 10^{-10}$/g			$1\ 10^{-8}$ p-p.
OUS	SC 3rd OT 10 MHz (HC 40)		$1\ 10^{-10}$/day	$1,5\ 10^{-9}$/g	1 ms $1\ 10^{-10}$ 10 ms $1\ 10^{-11}$ 100 ms $5\ 10^{-12}$ 1 s $1\ 10^{-12}$ 10 s $1\ 10^{-12}$ 100 s $2\ 10^{-12}$	1 Hz - 100 dBc Hz 10 Hz - 130 dBc Hz 100 Hz - 135 dBc Hz 1 kHz - 145 dBc Hz 10 kHz - 150 dBc Hz 50 kHz - 150 dBc Hz	- 110,50 dBc $< 5\ 10^{-10}$
	SC 3rd QAS 10 MHz (HC 40)	See curves	$5\ 10^{-11}$/day	$1\ 10^{-10}$/g	1 ms $1\ 10^{-10}$ 10 ms $1\ 10^{-11}$ 100 ms $1\ 10^{-12}$ 1 s $5\ 10^{-13}$ 10 s $5\ 10^{-13}$ 100 s $5\ 10^{-13}$	1 Hz - 110 dBc Hz 10 Hz - 135 dBc Hz 100 Hz - 145 dBc Hz 1 kHz - 150 dBc Hz 10 kHz - 150 dBc Hz 50 kHz - 155 dBc Hz	

TABLE IV - PERFORMANCES OF STANDARD AND RUGGEDIZED OSCILLATORS

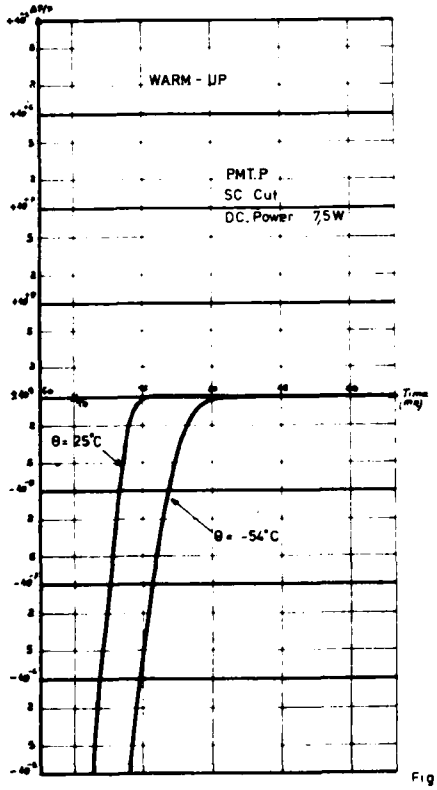


Fig 7

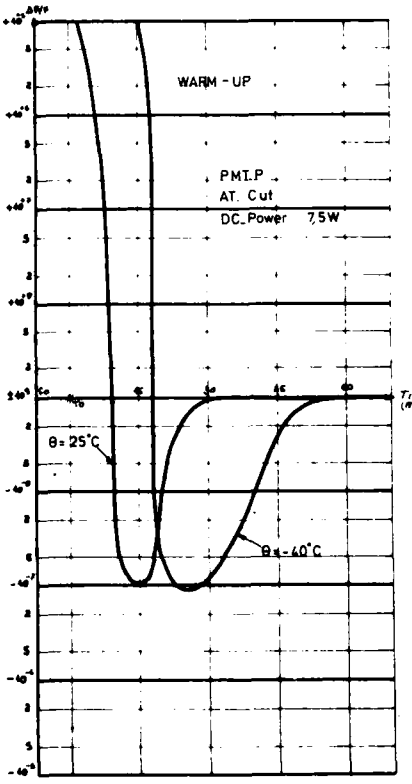


Fig 6

C.E.P.E
44,Av de la glaciere
95130 Argenteuil
Tel : 982.89.45

Le 17/5/85

PILOTE No : PMT305C
Fo : 10000000 Hz

Cahier des charges : — — — (: 5E-09 / Mois)
Pente du pilote : ———
(En 1E-09)

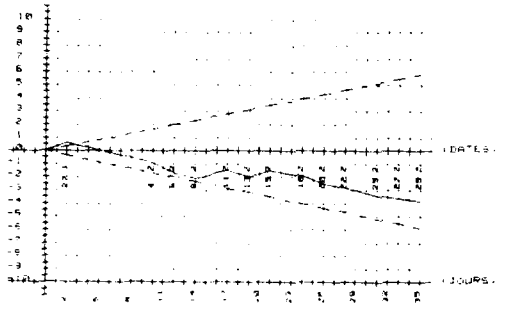


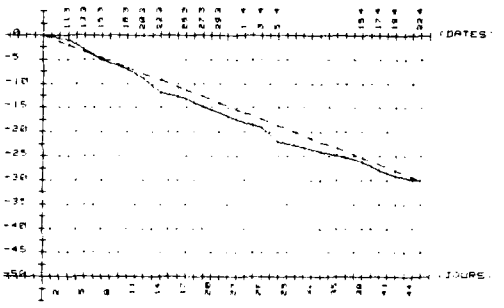
Figure 8 10MHz PMT P3T ageing

C.E.P.E
44,Av de la glaciere
95130 Argenteuil
Tel : 982.89.45

Le 17/5/85

PILOTE No : PMTPSC0AB
Fo : 10000000 Hz

Cahier des charges : — — — (: 2E-09 / Mois)
Pente du pilote : ———



(En 1E-10)
Figure 9 10MHz (GAS) PMT P-SC ageing

C.E.P.E
44,Av de la glaciere
95130 Argenteuil
Tel : 982.89.45

Le 19.05.85

PILOTE No : PMTPAT
Fo : 5000000 Hz

Cahier des charges : — — — (: 1E-09 / Mois)
Pente du pilote : ———
(En 1E-10)

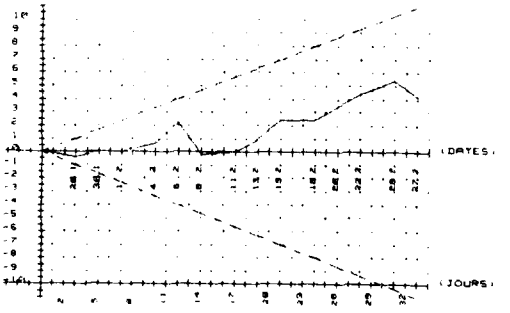


Figure 10B 5MHz PMT P-AT ageing

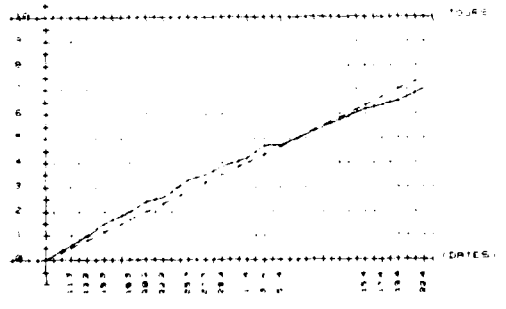


Figure 10A 10MHz PMT P-SC ageing

Osc. type	Warm-up Conditions	Max. DC Power	Time 10^{-7}	Lo 10^{-8}	Reach 10^{-9}/s
PMT P	25 °C ambient air	7,5 W	7 : 10"	20'	10'
	- 40 °C pulsed air	7,5 W	18	40'	20'
PMT P QAS SC	25 °C ambient air	7,5 W	8'	10'	5'
	- 55 °C pulsed air	7,5 W	17'	20'	17'

TABLE VI - WARM-UP PERFORMANCES OF MODIFIED PMT P OSCILLATORS

(Size : 67 x 60 x 40 mm - Volume : 160 cm³)

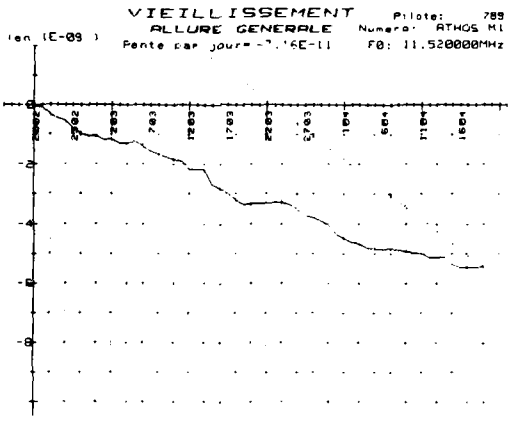
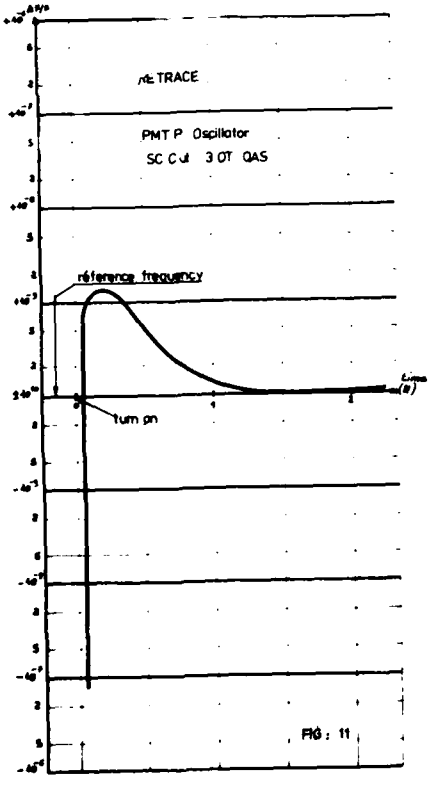


Figure 12 QAS - SC 11520 MHz OUS oscillator - ageing -



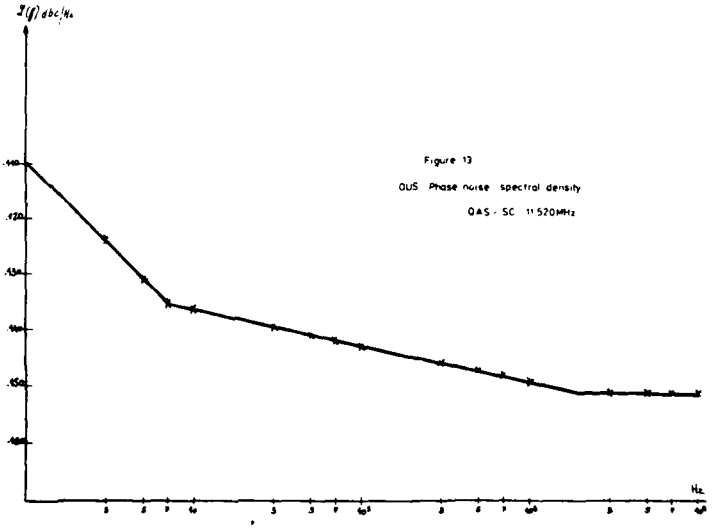


Figure 13

OUS Phase noise spectral density

QAS - SC 11520MHz

CHARACTERISTICS OF NATURAL, SWEPT NATURAL, AND CULTURED X- AND Z-GROWTH QUARTZ MATERIAL IN HIGH TEMPERATURE, HIGH STRESS APPLICATIONS

John A. Kusters and Gwen S. Kaitz

Hewlett Packard Company
Santa Clara, CA 95051

Abstract

Preliminary work associated with the development of an extended range BT-cut pressure transducer was discussed last year [1]. At that time, results which led to operation of the pressure transducer over extended temperature ranges were discussed, including preliminary results on hysteresis measurements.

Results of continued parametric and performance testing conducted on a variety of materials, including natural, swept natural, cultured Z-growth, and cultured X-growth quartz crystalline material are discussed.

Introduction

Unique use of crystalline quartz has a long tradition at Hewlett Packard. Quartz crystals form the time base for many of the instruments produced since the introduction of the first 10 MHz counter, the 524A, in the 1950's. Oven controlled crystal oscillators are used extensively in frequency counters, synthesizers, frequency standards, and other products.

Quartz crystal resonators also find use in metrology transducer applications such as quartz thermometry and quartz pressure gauges. Traditionally, the quartz pressure gauge has been processed from unswept natural quartz crystalline material.

At the time that this project was started, the 2813B Quartz Pressure Gauge operated to a maximum temperature of 150°C. and 11,000 psi [2,3].

A desire to improve the performance of the basic unit and recent threats to our sources of raw material have prompted an investigation into alternate materials and sources of supply.

Investigation

The quartz transducer in the 2813B is a 5MHz, 3rd overtone, BT-cut crystal fabricated into a geometric shape that permits hydrostatic pressure to be applied uniformly to the perimeter of the sensing element. The transducer is manufactured in three parts; a center cylinder with an integral BT-cut resonator, and two matching end caps. Because of anisotropic effects, the crystallographic orientation of the center cylinder and end caps are identical. A photo of the current transducer design is shown in Figure 1.

Because the BT-cut, in general, has a high temperature sensitivity, first order temperature compensation is achieved by beating the transducer output frequency against an essentially identical reference BT-cut crystal that sees only temperature changes, not pressure changes.

The BT-cut was chosen for implementation as a pressure gauge because of its high linear pressure sensitivity under radial hydrostatic compression.

Because of the generally favorable properties of our original design and the sensitivity of the BT-cut crystal in pressure measurements, the study was limited to looking at possible alternate materials. A redesign of the sensor crystal was not considered feasible at this time.

Materials of interest were confined to varieties of crystalline quartz. Studies were made of:

1. Unswept natural quartz (used as an experimental control).
2. Swept natural quartz.
3. Cultured X-growth quartz.
4. Cultured Z-growth quartz.

The initial project also planned to investigate swept cultured X-growth material. Particulate inclusion count in the swept bodies received was too high to proceed further in experimental testing.

All natural quartz material was imported directly from Brazil. As a result, we were able to exercise far better control on material quality. In fact, the material quality is among the best we've ever received.

Cultured material of all types were obtained from Sawyer Research and cut for us by Crystal Systems, a division of Sawyer. Some natural material was also cut by Crystal Systems.

Electrodiffusion sweeping of natural quartz material was performed by both Hewlett Packard [1] and Crystal Systems. Sweeping of cultured material was done by Crystal Systems.

Confirmation of sweeping completeness was done by irradiating each sample with 2 MRad of gamma radiation as discussed previously in [1].

After receipt of the material, and initial sweeping and irradiation where necessary, all transducer bodies were processed using our normal production process. All bodies received identical treatment in all steps of the production process.

Because previous results indicated that the type of material used for the end caps was not critical to the performance of the unit, and to avoid further complication of the experimental process, all trans-

CH2186-0/85/0000-0223$1.00©1985IEEE

ducer units were fabricated using unswept natural quartz material for end caps.

Parametric Testing Results

All fabricated units were tested for standard crystal parameters such as resistance, Q, motional capacitance and temperature coefficients. There were no measureable differences between the experimental materials for any of the parameters except the variation of resistance with temperature.

Cultured X-growth shows only a minor change in resistance at temperatures above 150°C. Unswept natural material shows a substantial increase in series resistance above 150°C., becoming unusable at temperatures exceeding 170°C. Swept natural material performs essentially the same as cultured X-growth, showing little increase in resistance even at 200°C. Cultured Z-growth shows resistance increases above 150°C. similar to that of unswept natural. About 15% of the unswept natural units show resistance-temperature characteristics similar to that of swept natural. See Figure 2 for a comparison of the resistance-temperature performance of the various materials

Performance Testing Results

Performance testing of finished transducers concentrated on two major areas, non-linear effects and material strength.

1. Material Strength

The original work done in 1968 confirmed that the material strength using unswept natural material was sufficient to maintain structural integrity of the design to pressures exceeding 90,000 psi. Because the mechanical design had not changed, all pressure testing was confined to ranges within the capability of our present test systems, 21,000 psi.

All pressure cycling occurred in a special temperature controlled test vessel, designed for maximum pressures of 30,000 psi., but restricted for safety reasons to 21,000 psi. by a diaphram blow-out port. All testing was done at 200°C. Pressure was provided using a Ruska motor-driven pressure pump. Automatic controls permitted repeated pressure cycling over any desired time period.

X-growth material, swept natural material, and unswept natural material were subjected to repeated cycles from atmospheric pressure to 20,000+ psi and back. No failures were seen.

Z-growth material in 1968, and again during this project, showed incipient cracks, especially in the region of particulate inclusions, at pressures as low as 11,000 psi.

This confirms an experimental observation that when cultured material fractures, it almost always occurs in the Z-growth region [4].

2. Non-linear Effects

Repeated pressure and temperature cycling of many samples of cultured X-growth material has shown that unacceptable hysteresis levels exist in every unit. Our specification limit for hysteresis is 0.6 psi. at any pressure and temperature. Experimental results of the hysteresis measured during repeated pressure and temperature cycling to 15,000 psi. and 200°C. are shown at three different temperatures for four samples of X-growth cultured quartz (Figures 3-5). Although two of the transducers are acceptable at 150°C., all are unacceptable at 175° and 200°C. One of the units at 200°C. exceeds specifications by a factor of 25.

Similar data is shown for three samples of swept natural quartz (Figures 6-8). All units meet specifications at 150° and 175°C. One exceeds specifications at 200°C by a factor of 5. Even so, the worst swept natural unit is still 5 times better than the worst cultured unit.

Although not enough data has yet been obtained to make a definitive statement, the few unswept natural units that are still functional to 200°C. show little or no hysteresis (Figure 9).

Also of interest, but not explained at the present, is the non-linear response that swept natural units show at 100°C. In all cases seen to date, hysteresis peaks at 100°C. in these units even though it usually remains within the specification limits. Unswept natural units show no such preference.

Pure Z-growth material shows a hysteresis pattern very similar to swept natural material. Because of material strength and resistance vs. temperature problems, testing of the Z-growth units was restricted to 11,000 psi. and 150°C.

Two of the X-growth transducers which showed much higher hysteresis than normally seen for this material were irradiated after pressure testing. These were found to contain regions of Z-growth material and clearly defined X-Z growth boundaries.

An attempt was made to determine any correlation between inclusion density and hysteresis in X-growth material. No correlation was seen.

All of the above does not lead to any firm conclusions as to the source of the non-linear effects seen. Units with growth boundaries show the largest hysteresis. This implies that internal lattice defects might be a major cause. However, lattice dislocations can also be expected to be high in the region of particulate inclusions. The results do not show that high-inclusion density material is any more hysteretic than low-inclusion density material. Data which lends credance to lattice defects as the primary source of hysteresis is the difference between swept and unswept natural material.

Another clue might be in the strong temperature dependence seen. In X-growth, the maximum hysteresis occurs at the higher temperatures. In swept natural units, the maximum usually occurs at 100°C. In unswept natural units, the maximum occurs at different temperatures from unit to unit.

Every experimental test which might uncover the fundamental cause of the pressure hysteresis has been inconclusive. All of this is reminiscent of the temperature hysteresis seen in the LC-cut [5], also still unexplained.

Conclusions

Sweeping of natural quartz material improves the resistance-temperature characteristics above 150°C. Sweeping also increases pressure hysteresis in natural material.

Cultured X-growth material shows good resistance-temperature characteristics, but an unacceptable level of pressure hysteresis.

Cultured Z-growth material shows an unacceptable resistance-temperature response, hysteresis similar to swept natural material, and poor mechanical strength.

The basic cause of the observed effects are unknown at this time. Although negative conclusions of this type are generally unsatisfactory, this may be an important practical problem whose solution may well come from other independent studies of the basic properties of crystalline quartz.

Acknowledgements

The authors wish to acknowledge helpful discussions with Mr. Kelly Scott of Crystal Systems, and Jerry Leach, one of the original developers of the quartz pressure gauge, now at Hewlett Packard, Andover, MA.

References

[1] Kaitz, Gwen S., "Extended Pressure and Temperature Operation of BT-cut Pressure Transducers," Proceedings of the 38th Annual Symposium on Frequency Control, 1984, pp. 245-250.

[2] Leach, Jerry G., "5 MHz BT Cut Resonators," Proceedings of the 24th Annual Symposium on Frequency Control , 1970, pp. 117-125.

[3] Miller, G.B., Seeds, R.W.S., and Shira, H.W., "A New, Surface Recording, Down-Hole Pressure Gauge," Paper number SPE 4125, presented at the 47th Annual Fall Meeting of the Society of Petroleum Engineers, San Antonio, Texas, Oct. 8-11, 1972.

[4] Private communication, Kelly Scott, Crystal Systems, Conroe, Texas, a division of Sawyer Research Inc.

[5] Hammond, Donald L., Adams, Charles A., Benjaminson, Albert, "Hysteresis Effects in Quartz Resonators," Proceedings of the 22nd Annual Symposium on Frequency Control, 1968, pp. 55-66.

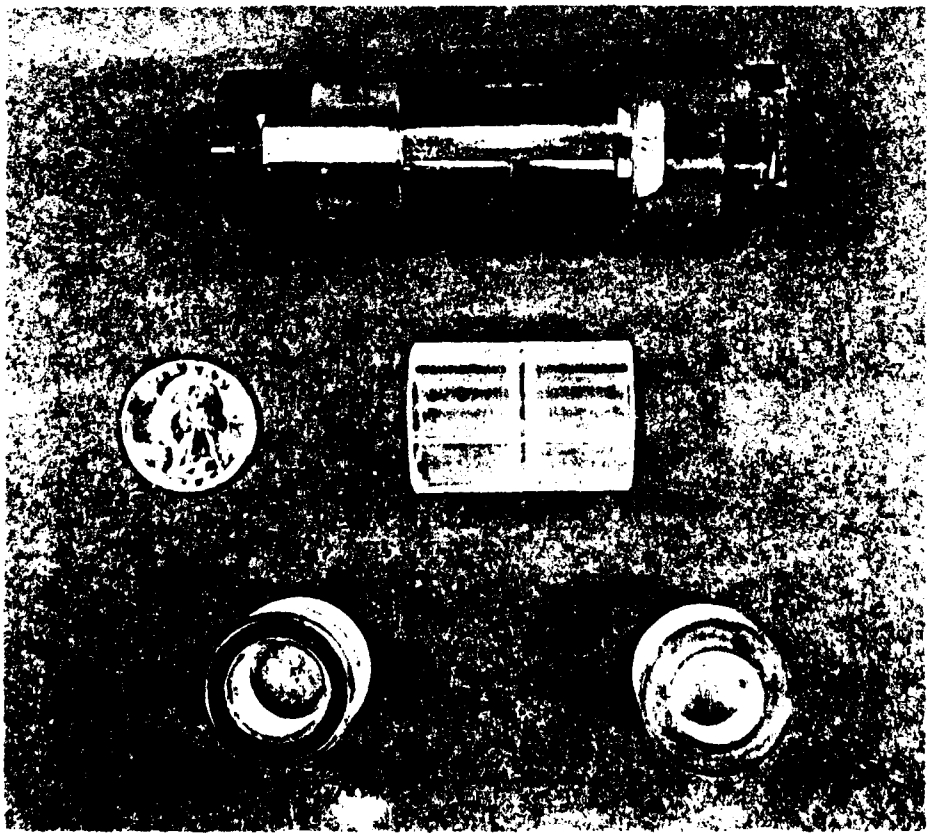


Figure 1. Quartz Pressure Transducer body and end caps before sealing and fininshed transducer after sealing

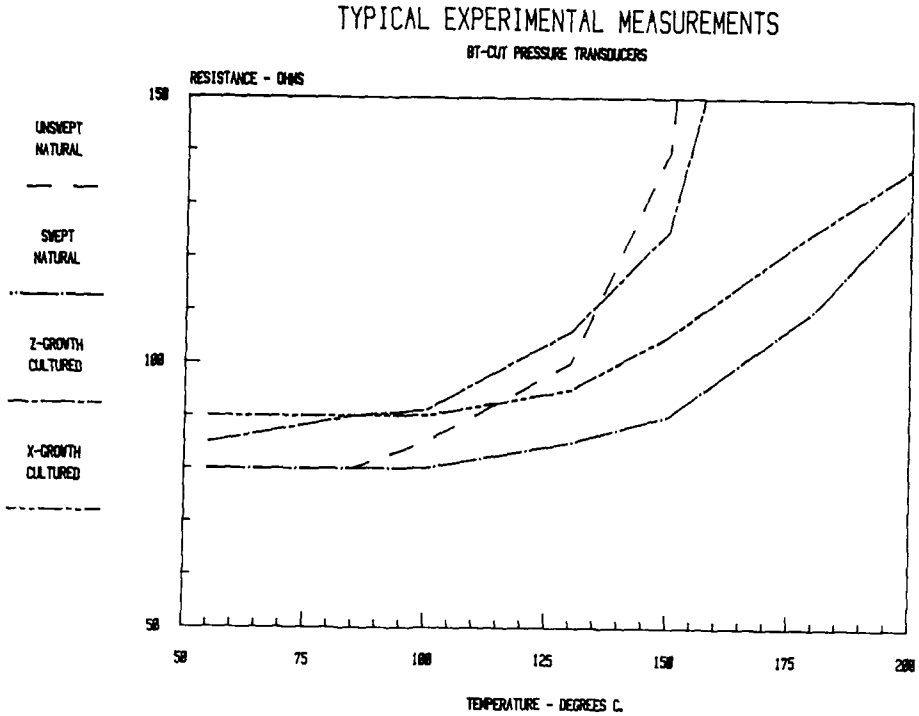


Figure 2. Comparison of Resistance-Temperature Performance for Various Types of Quartz

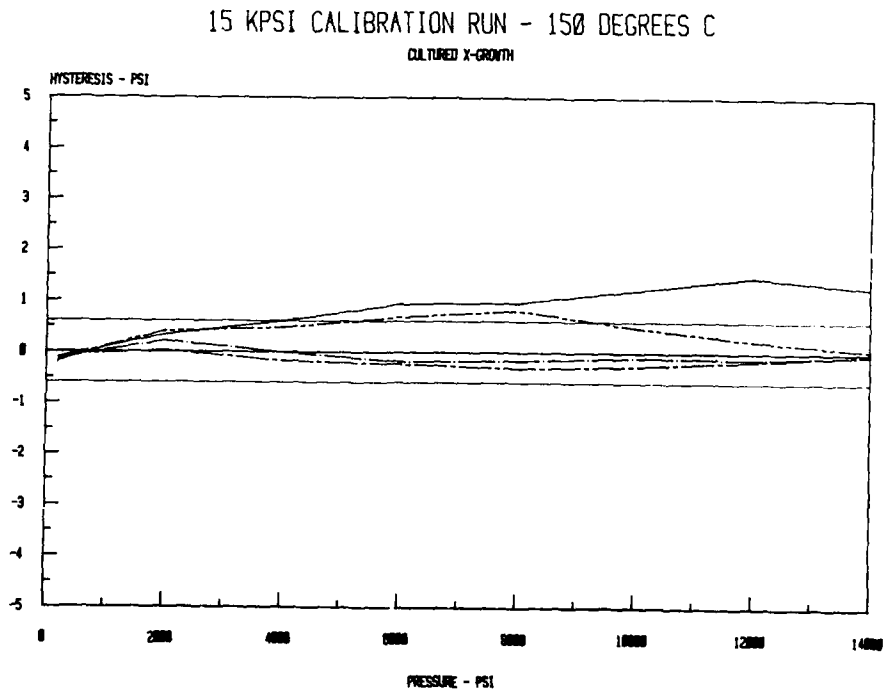


Figure 3. 150° C. Pressure Hysteresis - Cultured X-Growth

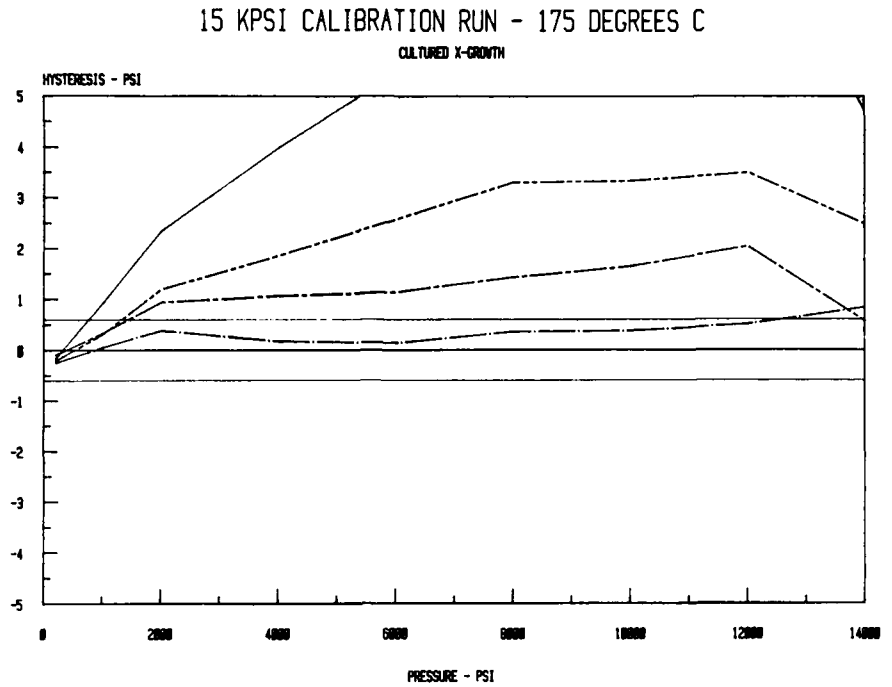


Figure 4. 175° C. Pressure Hysteresis - Cultured X-Growth

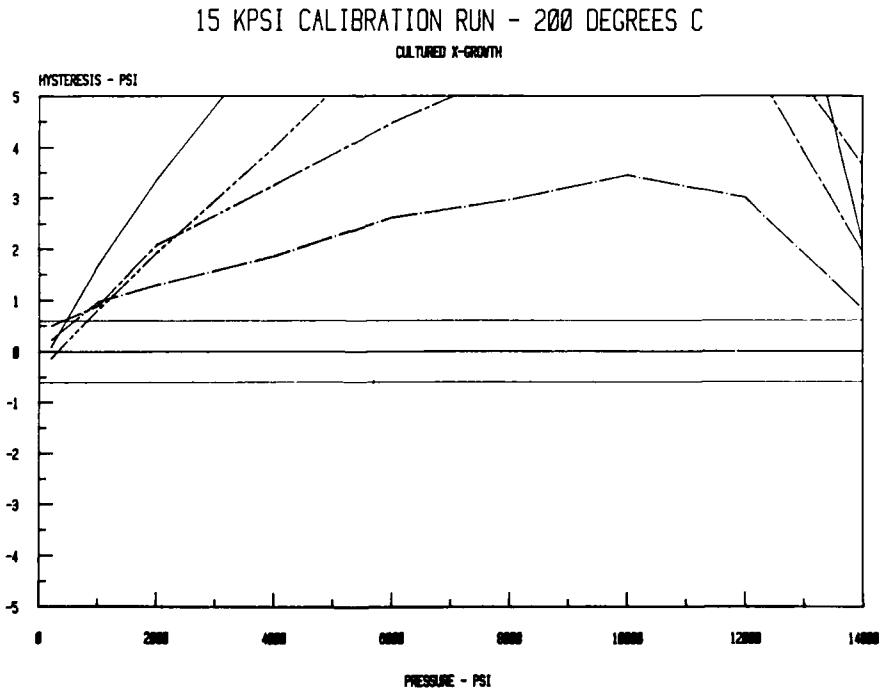


Figure 5. 200° C. Pressure Hysteresis - Cultured X-Growth

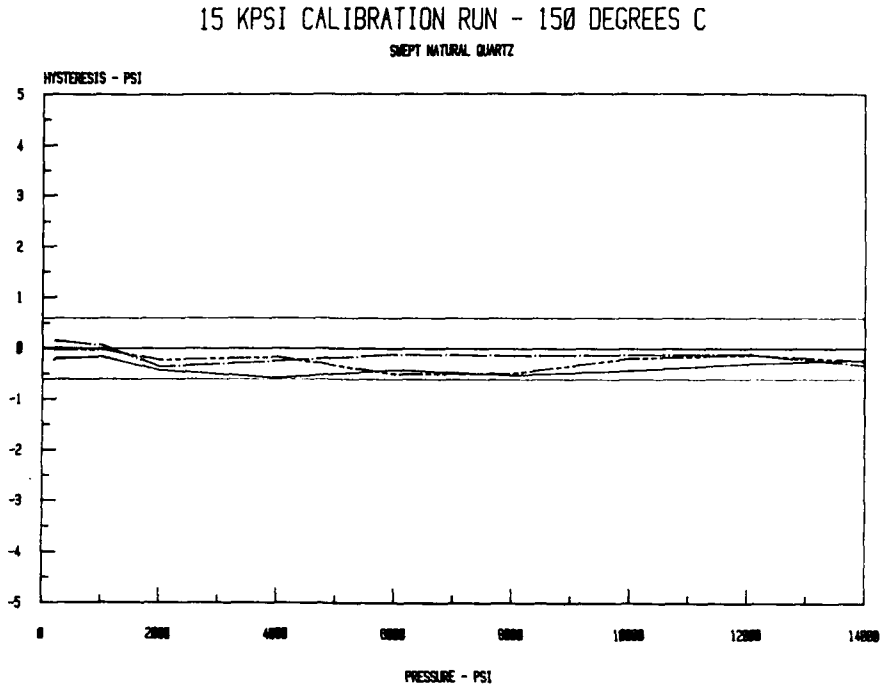


Figure 6. 150° C. Pressure Hysteresis - Swept Natural

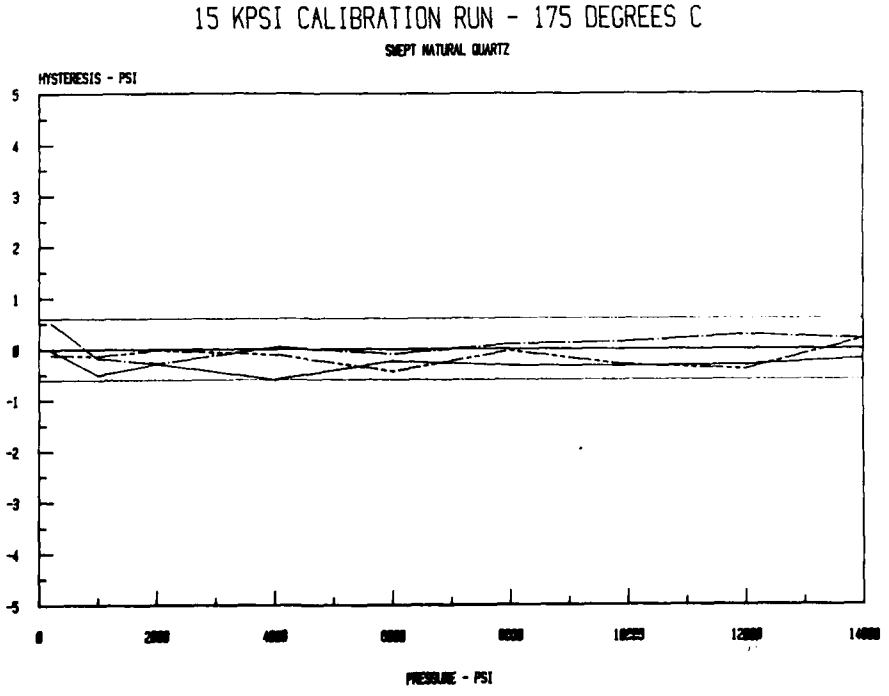


Figure 7. 175° C. Pressure Hysteresis - Swept Natural

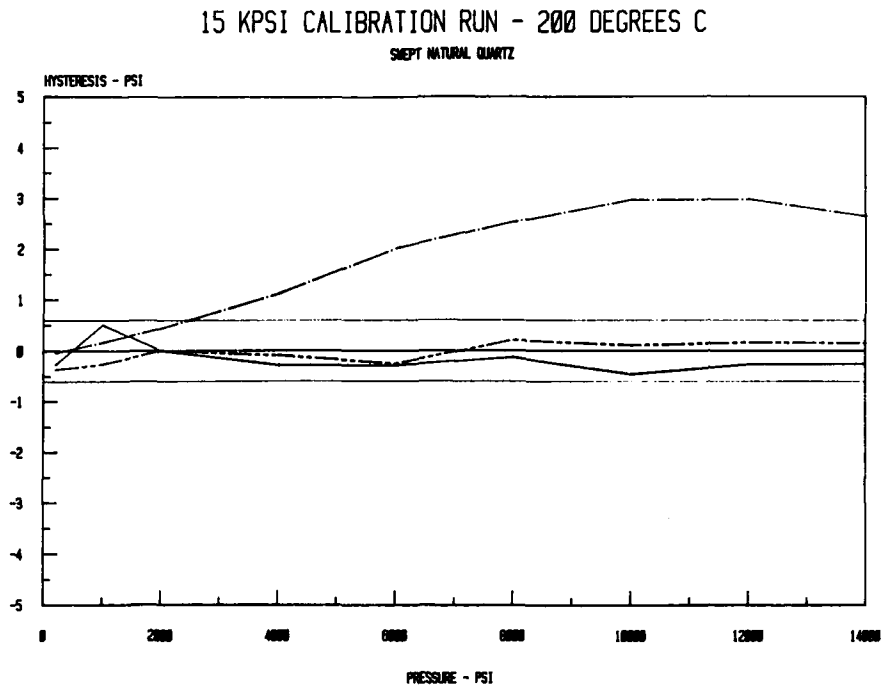


Figure 8. 200° C. Pressure Hysteresis - Swept Natural

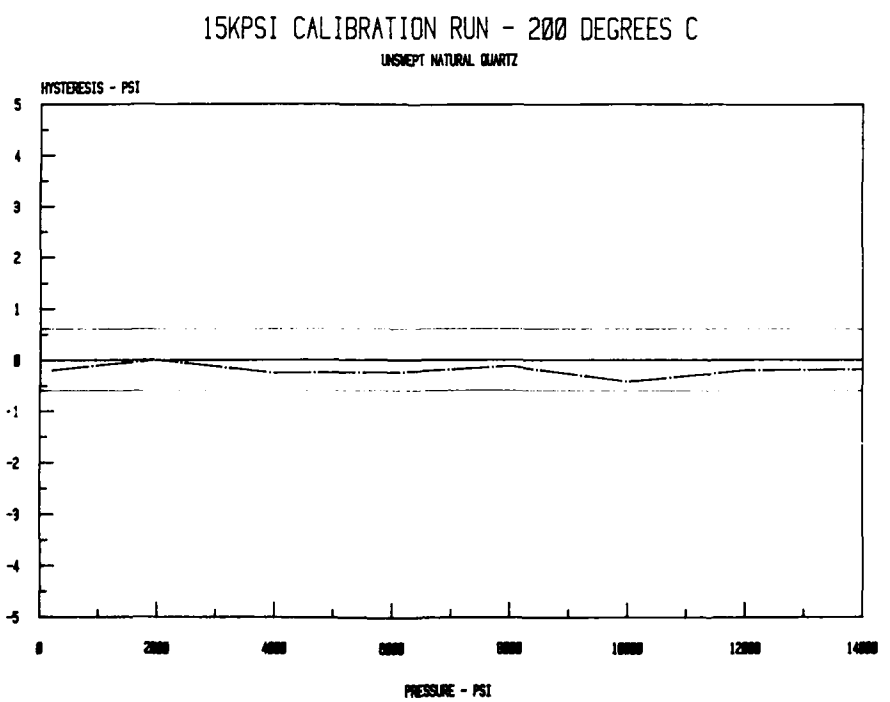


Figure 9. 200° C. Pressure Hysteresis - Unswept Natural

THE GROWTH OF HIGH QUALITY QUARTZ IN COMMERCIAL AUTOCLAVES

Alton F. Armington
Electromagnetic Materials Technology Branch
Solid State Science Division
Rome Air Development Center
Hanscom AFB, MA 01731
and
Joseph F. Balascio
Motorola, Inc.
Components Division
Quartz Operations
Carlisle, PA 17013

Abstract

Hydrothermal growth runs were performed in both commercial and research autoclaves in an effort to achieve higher purity alpha quartz with a minimum of dislocations. These runs employed both sodium hydroxide and sodium carbonate as mineralizers. It was found that the impurity concentration of aluminum can be routinely reduced to below 0.5 parts per million without the use of a liner. The average dislocation count can be reduced by at least an order of magnitude with the use of special seeds.

Introduction

Low aluminum content and low dislocation alpha quartz is limited in supply because of the high demand for lower grades of alpha quartz and the lack of understanding of all the necessary criteria to produce it. As part of a program to improve frequency and timing devices, a project on the growth of improved quartz was initiated to meet this need. The goals of this program were to develop a method for the production of high Q, low aluminum and low etch channel density quartz. The systems employed in this work have been previously described and consist of autoclaves ranging from one to ten inches in diameter; all under computer control.[1]

The principal impurity in alpha quartz is aluminum. It is prevalent and difficult to remove because it substitutes readily for silicon. The role of aluminum in quartz has been well documented and this impurity ion appears to greatly affect the properties of quartz.[2] The alkali impurities (Li, Na, K) that combine with aluminum in the crystal to form electrically neutral impurity sites are a cause of frequency offset since they are easily dissociated from the aluminum by relatively low energies. The use of sweeping (solid state electrolysis) can replace the alkali impurities with hydrogen, which is more strongly bound thus, essentially eliminating these impurities. Present work indicates the iron has little effect on frequency offset.

Dislocations appear to be related to etch channel formation. Bundles of dislocations may be the source of channel formation. Since these sites possess high free energies, they are locations for the impurity ions.

Several methods of analysis have been employed to determine the concentration of impurities in alpha quartz. Atomic absorption spectroscopy was the first method used and its detection limit is about 0.5 ppm. The second technique is EPR which is considered the most reliable technique below 0.5 ppm. The third technique is inductively coupled mass spectroscopy which is believed to be accurate to the parts per billion range. In this technique, a plasma is used to excite the species to be measured. These ions are then passed into a mass spectrometer where they are separated and analyzed. A photograph is shown in Figure 1. The reported limits of detection are shown in Figure 2. Some results for aluminum were analyzed using low temperature infrared. A comparison of these techniques is shown in Figure 3. It can be seen that the reproducibility in this case, at least, is quite good.

Results and Discussions

In the previous paper[1] we discussed exploratory results in both research and full size autoclaves. In this paper we will extend the results on the effects of liners, scaling up, nutrient and growth rate effects, seeds and sweeping.

Liner Effects

Hydrothermal growth runs were conducted in autoclaves up to four inches in diameter. Some of these runs were carried out in lined (usually silver) autoclaves. Data measured in each run were infrared Q and aluminum levels. Q values ranged from 2.3×10^6 to 3.4×10^6. Impurity concentrations were determined by EPR and atomic absorption spectroscopy. Figure 4 compares these data between runs conducted in lined and unlined autoclave runs. The results indicate that high purity, low aluminum alpha quartz can be obtained with or without the use of a liner during growth. These results were obtained using Cultured Quartz III as the nutrient supply and are the best values we have obtained. The best results were in the range of 30 to 50 parts per billion and were obtained from crystals grown from a carefully cleaned unlined autoclave. In general, the Cultured Quartz III nutrient produces crystals with a part per million or less. No attempt has been made to use liners in commercial size autoclaves.

Growth Rate Effects

The effect of growth rate at higher aluminum levels has been previously reported.[3] In general it has been shown that lower growth rates produce alpha quartz crystals with a lower concentration of aluminum, at least. Employing higher grade starting material (Cultured Quartz III), we have extended this work to lower impurity levels as shown on Figure 5. These results were obtained from runs performed in 3" to 4" ID autoclaves. The previous results[3] indicated that the impurity level increased more rapidly when the Z growth rate exceeded 20-25 mils/day. The results shown here, employing higher purity starting material, indicate that high purity quartz

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

can be grown at an equivalent growth rate. The circles indicate runs employing a liner and the crosses are for unlined runs.

Nutrient Effects

Three nutrients will be discussed here. First, I, is Cultured Quartz I nutrient which is grown from a natural quartz supply (lascas). Next, II, is Cultured Quartz II nutrient which is grown from a Cultured Quartz I supply and III, is Cultured Quartz III which is Cultured Quartz II nutrient with the X regions removed. It has been shown that the X regions contain higher concentrations of impurities (about an order of magnitude) than the pure Z region. We have ppm results for these nutrients both in small autoclaves and commercial units. Some of these results are shown in Figure 6. It is apparent that the highest purity (in terms of aluminum) is obtained using the Cultured Quartz III nutrient, which has the highest purity. The first three results were obtained from crystals grown from a sodium hydroxide solution. It is interesting to note that apparently one obtains the same purity in the sodium carbonate mineralizer runs as in the hydroxide mineralizer but, the growth rate is approximately cut in half. These are averages of several runs. The comparison between the hydroxide and carbonate mineralizers is based upon runs utilizing Cultured Quartz I as the nutrient.

A comparison of the purity between runs conducted in large and research autoclaves indicate that using the same nutrient and growth rate, somewhat higher purity can be produced in the smaller autoclave (Figure 7). This may be more of a case of preparation rather than an effect of autoclave size. The difference in purity is probably not significant.

Etch Channel Density

Dislocations generally emanate from the seed and propagate through the whole growth region. A small number may also start at inclusions and progress to the crystal surface. Five different seed pretreatment techniques have been employed during the course of this investigation. In addition to the three techniques previously reported,[1] hydrothermally polished seeds and unetched seeds were also added. Regardless of the pretreatment employed, etch channel densities varied. All pretreatments had yielded quartz with channel densities in the 100-200cm^{-2} range. There appears to be a wide latitude in the solution of an etchant for seed preparation without greatly affecting the resultant etch channel density. However, none of these pretreatments resulted in a consistently lower channel density compared to the others. It has been found, however, that the use of a seed cut from the +X region of the crystal will produce a material with a minimum of dislocations.[1] Seeds usually employed in the growth process are cut from the Z growth region parallel to the X and Y axes. These seeds often etch during the warm-up cycle often forming etch channels that penetrate into the growth region. An example of this is shown in Figure 8. An alternative to this is to use seeds cut entirely from the +X region of the crystal where the dislocations project in a different direction than those produced from the seeds cut from the Z region of the crystal. An example of this is also shown in Figure 8. The resulting dislocation density is much less in this case. Some etch channel densities have also been determined on mated quartz material from each growth run (Figure 9). All GC run numbers listed in this figure were runs conducted in commercial size autoclaves. The first three runs (GC20 to GC26) employed pure Z seeds and the last two crystals were grown upon +X-Z oriented seeds. Q values, determined by the infrared technique, range from 2.4×10^6 to 3.1×10^6. The count varied from 0-25 per square centimeter. The average value was 14cm^{-2} for the worse case. This is at least an order of magnitude less than that measured in unswept crystals grown in commercial autoclaves. The etch channel density of the material shown in Figure 8 varied from 0-0.5 per square centimeter. The etch channel density in the swept mated sections was reduced approximately tenfold. No channel density data had yet been measured on the swept section of GC32. There are two average etch channel values listed for this run. It is interesting to note that the etch channel density is the same whether a +X seed is used or a first generation seed cut from the Z region of the crystal. The reason for this is still under investigation. Resonators, both swept and unswept, have been made from crystals grown upon these seed types. The quartz was grown from high purity nutrient and possessed about 0.5 ppm aluminum. These crystals are being evaluated in devices and these results will be presented when they are available.

Summary

It has been found that the use of high quality nutrient can produce high purity quartz with an aluminum content less than 0.5 ppm without the use of a liner. It has also been found that the use of seeds cut from the +X region of a quartz crystal can reduce the etch channel density by at least an order of magnitude.

Acknowledgements

The authors would like to thank M. Harris; J. O'Connor; J. A. Horrigan and E. Cormier of RADC and T. Walker of Motorola for their help during this work. The atomic absorption work was done by A. Fluesmeier, Eagle Picher Research, Miami, OK and the EPR work was done by L. Halliburton of Oklahoma State University.

References

1. Armington, A. F. and Balascio, J. F., Proceed. 38th Freq. Control Sympos., p. 3 (1984).
2. Fraser, D. B., Physical Acoustics, Vol. 5, Academic Press, N.Y., p. 61 (1968).
3. Martin, J. and Armington, A. F., J. Crystal Growth, 62, p. 203 (1983).

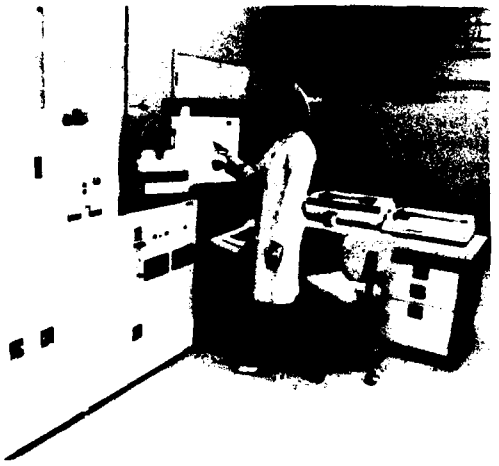


FIGURE 1

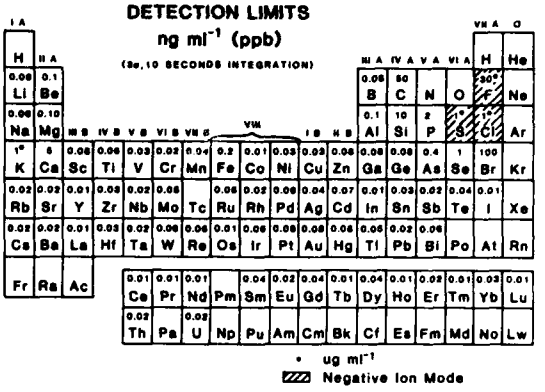


FIGURE 2

COMPARISON OF ANALYTICAL TECHNIQUES*

AA-0.5

EPR-0.56-0.48

IR- 0.5

ICP/MS-0.38-.51

*ACCURACY OF PARTS PER MILLION - ATOMIC

FIGURE 3

RESULTS OF LINED AND UNLINED RUNS *

RUN		ALUMINUM CONTENT (EPR)	IRQ
X 46	UNLINED	0.3	
X 52	"	0.13	
X 55	"	0.03 - 0.05	
QA 33	SILVER LINER	0.42	
QA 37	"	0.6	
QA 39	"	0.05	
QA 40	"	0.20	2.67
QA 43	"	0.12	
QA 38	"	0.5 (AVE 6 SAMPLES)	2.89

•PARTS PER MILLION ATOMIC

FIGURE 4

EFFECT OF GROWTH RATE ON PURITY
(EPR RESULTS)

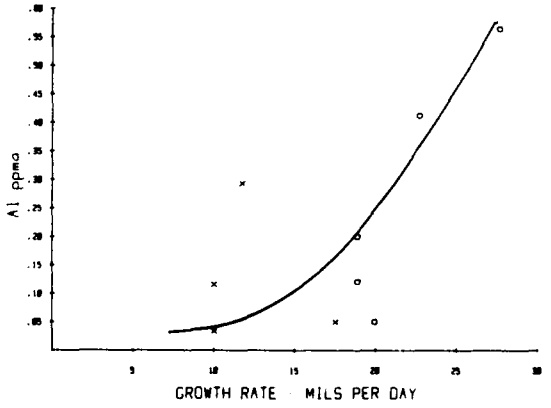


FIGURE 5

EFFECT ON NUTRIENT ON ALUMINUM CONCENTRATION

NUTRIENT	GROWTH RATE MILS/DAY	MINERALIZER	ALUMINUM (PARTS PER MILLION)
1	35	OH	3.3
2	33	OH−	1.7
3	26	OH−	0.5
1	18	CARBONATE*	3.2

*AVERAGE OF 6 RUNS

FIGURE 6

AUTOCLAVE DIAMETER		ALUMINUM CONTENT (PPM)
3"	—	0.20*- 0.3
4"	—	0.3 - 0.4
12"	—	0.5 - 1

*NOT OF SUFFICIENT SIZE FOR FOR RESONATOR FABRICATION

FIGURE 7

X-RAY TRANSMISSION TOPOGRAPHS

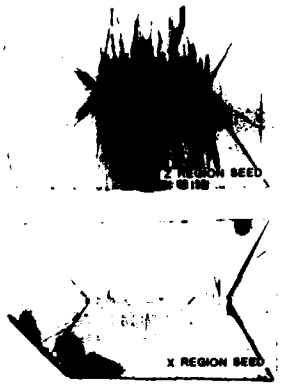


FIGURE 8

ETCH CHANNEL DENSITIES

RUN#	GC 20	GC 26	GC 22	QA 38*	GC 32*
UNSWEPT	380	253	247	0.5	13/14
SWEPT	38	20	39	0	- -

*X + SEEDS = MOTHER AND FIRST GENERATION

FIGURE 9

CRYSTAL GROWTH, PHYSICAL CARACTERIZATION AND BAW DEVICES APPLICATIONS OF BERLINITE

J. Détaint[1], E. Philippot[2], J.C. Jumas[2], J. Schwartzel[1]

A. Zarka[3], B. Capelle[3], J.C. Doukhan[4]

(1) Centre National d'Etudes des Télécommunications - PAB/BAG/MCT 92220 BAGNEUX-FRANCE
(2) Laboratoire de Physico-Chimie des Matériaux Inorganiques - Université de MONTPELLIER 34000-FRANCE
(3) Laboratoire de Minérologie et de Cristallographie Université de PARIS VI 75006-FRANCE
(4) Laboratoire de Structures et Propriétés de l'Etat Solide Université de LILLE 59655-FRANCE

ABSTRACT

The feasibility of large crystals of Aluminium phosphate has already been demonstrated so was the interest of this material to obtain B.A.W. and S.A.W. devices with attractive properties. Our purpose in this study was :

a) to synthetize crystals with low acoustical losses
b) to specify the useful crystal orientations more accurately
c) to design devices that take in account, all the specificities and advantages of this material.

With a view to obtain high perfection and high Q crystals, three hydrothermal crystal growth methods were investigated in H3PO4 solvent, compared and improved with the assistance of crystal characterization techniques and BAW devices measurements.

Experimental conditions used most frequently with the slow heating, the reverse vertical gradient and the horizontal gradient methods are rapported together with specific features of each method as applied to grow berlinite.

X ray topography has shown that the best crystals have a dislocation density of 10 to 100 dislocations per cm^2. X ray topography examination above room temperature (25°-150°C) has revealled temperature dependant quasi reversible localized variations of strain in water containing crystals.

Transmission electron microscopy was used after high temperature precipitation of water of assess the H_2O content of crystals as a function of some growth parameters and to determine the distribution of this impurity within crystals.

Thickness shear resonators of several Y rotated cuts were measured to compare the growth methods. Some experiments with recently obtained crystals have demonstrated the feasibility of very high Q crystals and the possibility of obtaining superior thermal behaviour from this material.

Experiments concerning the AT cut has demonstrated the possibility to obtain band-pass of filter of shift of oscillators twice that of quartz. Specifity of energy trapping in this material is then discussed.

We conclude to the major interest of waterless berlinite for devices application.

INTRODUCTION

Since the work of Chang and Barsch (1) many works were devoted to grow and characterize Berlinite crystals and to modelize and experiment bulk and surface wave devices made with this material. It was successively shown that this material was a more piezoelectric analog of quartz (2), (3), that possess similar cuts with zero frequency temperature coefficients (4) (5), either for bulk and surface wave propagation.

In the past ten years, many efforts were made to study hydrothermal methods that can produce the large crystals of high perfection needed for piezoelectric applications. Last year, the feasibility of large crystal and the interest lateral excitation were demonstrated (5). More recently the importance of reducing the water solubility in crystals was, emphasised in two papers (6) (7) that show the large influence of this phenomenon on devices properties (insertion losses, or Q factor, temperature coefficients.

In France, studies concerning this material were resumed at the initiative of CNET LANNION with the participation of four University Laboratories and one industrial company the SICN.

The studies are aimed at the production of high Q factor crystals permitting the obtention of bulk and surface wave devices for Telecommunications applications.

Here we report work done to improve growth methods, and to understand the phenomenon of water solubility in this material. Comparison of growth methods will be made with the help of characterization of crystals by X ray topography, Transmission Electron Microscopy (TEM) and resonators measurements. Recently obtained results will demonstrate the feasibility of high Q crystals and their unique properties. This will permit us to conclude to the major interest of water-less berlinite for devices application.

I - CRYSTAL GROWTH STUDIES in H_3PO_4

NUTRIENT SYNTHESIS : Berlinite in small crystals form can be obtained by hydrothermal reaction of H_3PO_4 with different chemicals containing aluminium (8) (9) (10). After try, with different kinds of commercially available aluminium phosphate, the use of high purity alumina was prefered, because of the critical importance of nutrient purity on crystals quality (10) (11) (12) (13). The crystallization reaction :

CH2186-0/85/0000-0234$1.00©1985IEEE

$$Al_2O_3 + 2\ H_3PO_4 \rightarrow 2AlPO_4 + 3H_2O$$

is obtained in a teflon lined autoclave that is rotated while heating at 2 RPM. Conditions for this operation are :

Concentration of starting products determined to obtain a solution 6.5 M in H_3PO_4 ; 2.5 M in $AlPO_4$
Filling 80 %
Temperature cycled between 150° and 200°C
Number of T° cycles depending of wanted granularity

Alumina from different sources were used, their purity was determined by mass spectrometry at CNET LANNION. Results for two of the best alumina are given in table I together with analysis of crystals obtained from them.

TABLE I

CHEMICAL ANALYSIS OF ALUMINA A&B AND CRYSTALS *

(in p.p.m. weight)

	Si	Fe	Ti	Na
Alumina A	57	17	2443	37
Berlinite from Alumina A	21	20	517	9
Alumina B	30	90	4	280
Berlinite from Alumina B	12	31	5	24

* ANALYSIS MADE AT CNET LANNION

OBTENTION OF SEEDS : For the initial obtention of seeds a modified version of the slow heating method was used (14). In this modification, an horizontal glass autoclave filled with a quasi saturated solution, is heated for 25 days at the rate of 1°C/day from 160°. A symetrical horizontal gradient (5 to 6°C) is maintened while heating. This set up allows to obtain spontaneously nucleated crystal of good quality, 5 to 8 mm large.

SEEDS PREPARATION AND SELECTION : Crystals are sawn after orientation in plates 1mm thick. The plates are lapped and polished with Nalco, then are slightly etched. A selection is made by binocular inspection. Twinned zones are removed by abrasion. Plates with insufficient quality are rejected.

CRYSTAL GROWTH BY THE SLOW HEATING METHOD : This method previously described in (11) (15) was mainly used for improving the quality of seeds and to enhance size of crystals in the initial phase of this study with conditions similar to those used to obtain nucleations. Typical growth rates for differents seeds orientations are given in Table II. Crystals 25 to 30 mm long are obtained with X or Y seeds in two cycles. They present no appearent defects. With Z seeds, growth rate is too high and liquid inclusions normal to the seed plane can be observed.

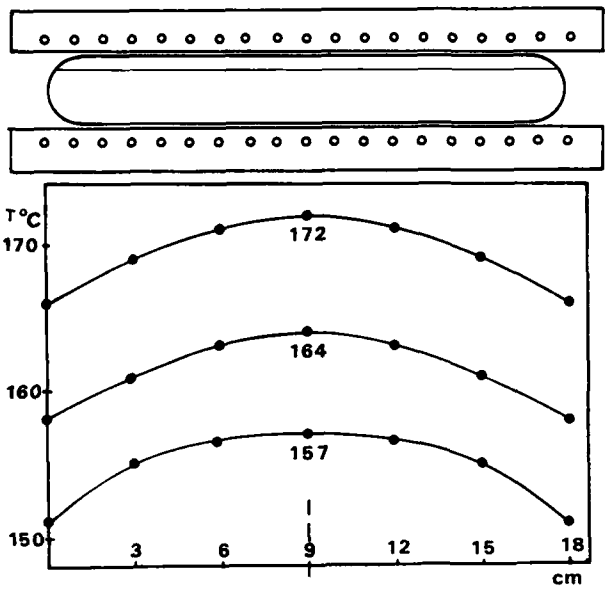


FIGURE 1:PRINCIPLE OF CRYSTALLIZATION AUTOCLAVE

TABLE II

MEAN GROWTH RATE FOR DIFFERENT SEEDS ORIENTATIONS (in mm/day/face)

METHOD / SEED	S.T.R.	V.T.G.	H.T.G.
X	0.15	0.12	0.12 ($10° \leq \Delta T \leq 15°C$)
Y	0.10	0.08	0.10 ($15° \leq \Delta T \leq 20°C$)
Z	0.50	0.40	0.20 ($3° \leq \Delta T \leq 4°C$)
r	0.08	0.06	0.08 ($15° \leq \Delta T \leq 20°C$)

A major drawback of this method is the impossibility to obtain crystals of sufficient size in one operation due to the limited quantity of aluminium phosphate available in the solution.

GROWTH WITH THE REVERSE TEMPERATURE GRADIENT METHOD : Only improvements to this method will be described here (11) (14) (15). This method was set up using two 375 cc platinium lined autoclaves. The following points were the object of in depth investigations :

1) The internal disposition was optimized considering the temperature distribution, fluids circulation and nutrient dissolution. Two disposition were retained, one is displayed in Figure 2.

2) Initial conditions of growth : To insure good initial growth it is important to insure a proper saturation of the solution from start. This was achieved by a blank experiment with rejected seeds for 24 hours in order to reach equilibrium. The real growth operation is made with the quenched solution after having installed the good seeds and adjusted the filling. It was experimentally found that, optimal crystal quality was obtained when temperatures in the second experiment were reduced by about 2°C.

3) Initial seeds dissolution is avoided by reaching the equilibrium temperature very quickly (preheating of the furnace).

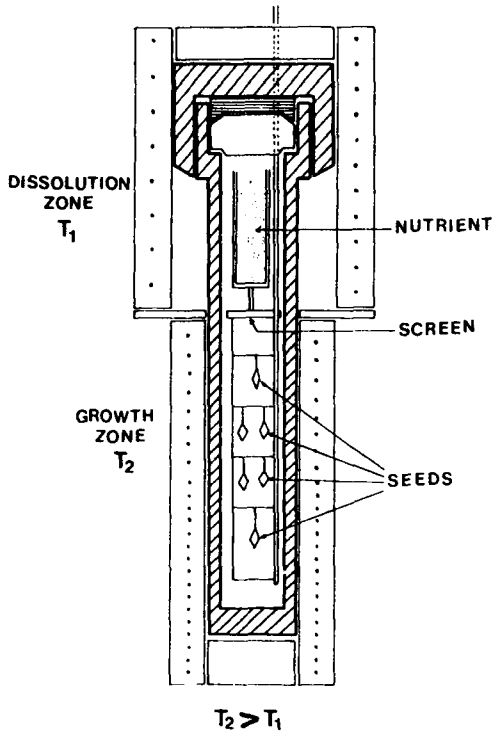


FIGURE2: REVERSE TEMPERATURE GRADIENT GROWTH OF BERLINITE

The influence of filling on growth rate was studied for values ranging from 80 % to 94 %, with X,Y,Z, seeds. Results are given in Figure 3. It can be seen that a maximum occurs for 92 % this correspond most probably to the complete filling of the autoclave at the operating temperature.

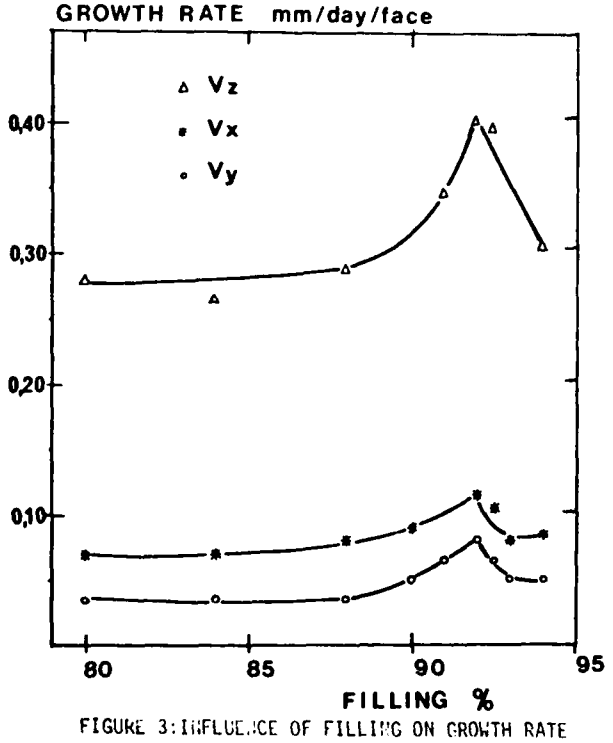


FIGURE 3: INFLUENCE OF FILLING ON GROWTH RATE

After experimentation, the following conditions were retained for most of the growth operations :

Autoclaves with hot/cold zone ratios of 5/2 and 3/1
Solution $H_3PO_4/H_2O/AlPO_4$ (6.5 M/1.1 M)
Temperature of hot zone = 170°C (-2°C/blank experiment)
Vertical gradient 4° ΔT 8°C
Time to reach thermal equilibrium : 1 hour
Boring ratio of the baffle 8.5 % (room T°)
Nutrient mesh 2 mm - 4 mm
Filling 92 %

Observed growth rate in theese condition are given in table II. The growth rate for Z seeds are too high so that crystals obtained with them display liquid inclusions. In the other cases very good results are obtained. This method was used to grow crystals 30 mm long.

COMPOSITE GRADIENT METHOD : In order to obtain more flexibility in adjusting the growth rates for Z seeds (with R.V.T.G. the unworkable value of 1°C would be required for the gradient), and to avoid the preliminar work of crystal growth needed to obtain nutrient with sufficient size (that does not cause crystallites drop on crystals), a modification of the horizontal gradient method originally proposed by Krauss and Lehmann (17) was used (18). The principle of the autoclave retained used for the first experiments is given on Figure 4.

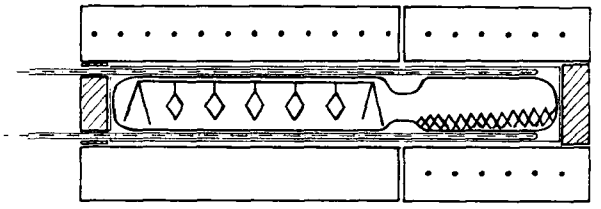


FIGURE 4:
PRINCIPLE OF HORIZONTAL GRADIENT OPERATION

The nutrient is put in the shorter part (cold region), while seeds are installed in the larger part ; theese two regions are separated by a constriction of the glass vessel. An horizontal temperature gradient is established between the two zones. A second vertical gradient is established in the growth zone to improve fluids circulation and to avoid parasitic crystallization. The following growth conditions were choosen after try and trial :

Silica glass autoclaves 100 cc diameter 30 mm
Ratio hot/cold zone 2/1
Solution $H_3PO_4/H_2O/AlPO_4/PO_4/$ (6.5 M/1.1M)
Growth temperature 170°C
Horizontal gradient $3° \leq T \leq 30°C$
Vertical gradient $2° \leq T_2 \leq 5°C$
Constriction 10 mm
Pulverulent nutrient
Filling 80 %

the following features of this method as applied to grow berlinite were observed :

- growth rate similar to those of R.V.T.G. are obtained with horizontal gradients two to three times higher. This gives much more flexibility to adjust the growth rates and gives the possibility to grow crystals inclusions free from Z seeds.
- Nutrient in pulverulent form can be used.

- Crystal Quality is similar to those obtained by S.H. or R.V.T.G. (much better for Z growth).

The observed growth rates, obtained with the reported conditions, are given in table II. Considering the advantages of this method, we have decided to realize a metallic prototype of much larger size with the support of CNRS/ANVART (18). This autoclave is now being tested.

Typical crystals obtained from the different growth methods are shown in Figure 5 and Figure 6.

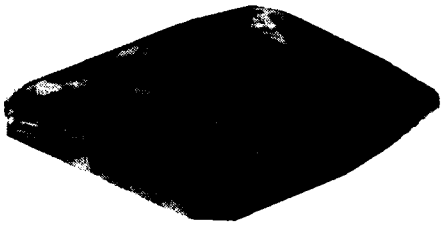


FIGURE 5a: CRYSTAL GROWN BY THE S.H. METHOD.

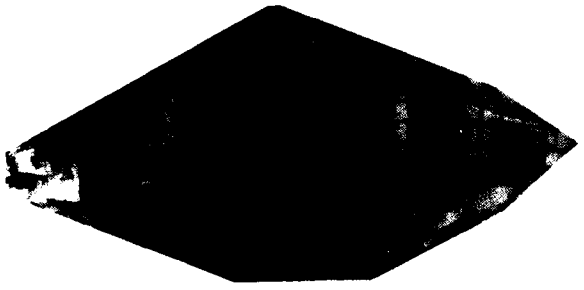


FIGURE 5b: CRYSTAL GROWN BY THE V.G. METHOD

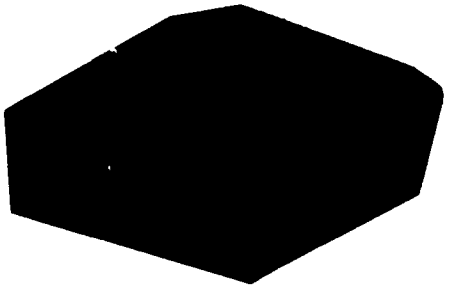


FIGURE 6 : CRYSTAL GROWN BY THE H.G. METHOD

II CHARACTERIZATION BY X Ray TOPOGRAPHY

ASSESSEMENT OF EXTENDED DEFECT DENSITY : At the begining of this study many Lang topographs of crystals grown under different conditions were take to determine the nature and the density of extended defects and to study the influence of the growth conditions. They have shown that the most important factors of crystal perfection are the seeds quality and initial growth conditions. Screening of seeds and optimization to the low value of 10 to 100 per cm^{-2} of growth conditions have permit to reduce the density of dislocations to the low value of 10 to 100 per cm^2.

The topograph of Figure 7 display a typical situation encountered while improving the crystalline quality.

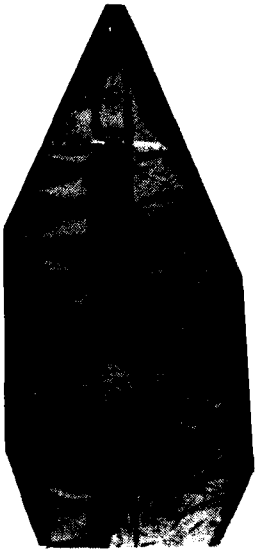


FIGURE 7: TOPOGRAPH OF A Y PLATE

The sample is a Y plate grown by the slow heating method ; the diffraction vector g is in the X direction (211.0 reflection), Cu $K\alpha_1$ radiation was used. The seed, easily discernible in this Figure was an X plate itself cut in a Z grown crystal. It can be noticed that the regions G_1 of the seed which contain a little density of defects induce only a few dislocations, whereas the perturbated regions of seeds induce a high density of dislocations (bundles of dislocations). Near the end of growth, a modification of growth conditions, which results of the nature of the slow heating method, has produced a large density of defects near the rhomboedral faces.

No fundamental differences were found, by X ray topography, between the gradient (H. and V.) methods when good seeds and optimized conditions are used. Moreover, contrary to the case of quartz growth, it was demonstrated that, with berlinite, it was possible to enhance the crystal quality while increasing the size of crystals. This fact is of major interest.

OBSERVATION OF CONTRASTS VARIATIONS WITH TEMPERATURE : Topographs of a Y plate grown at 170°C one year ago, were made at different temperatures in repeated thermal cycles between room temperature (RT) and 423K. The 112.0 reflexion was used with MoK_α radiation. A selection among this topograph are presented in Figure 7 (19).

In the first RT topograph (Figure 8a) the crystal exhibits a high density of defects : dislocations and growth bands oriented parallel to (211) : The whole contrast remains quite similar from RT to about 330.5K (Figure 8b) where some zones show slightly different contrasts. These changes are more obvious at 382.8K (Figure 8c) : the contrast of the growth band marked A vanishes and the intensity in the zone marked B is clearly enhanced. On decreasing the temperature the sample shows contrast behaviour very similar to that obtained during the temperature increase. We note in Figure 8d (322.5K) the inversion of contrast between zones A and B (in comparison with the contrast obtained in Figure 8c). At RT (end of the first cycle of temperature) the sample exhibits (Figure 8e) quite similar contrasts to those obtained at the begining of the temperature cycle Figure 8a.

During the second cycle the topographs show the same phenomenon with inversion of the contrast between the zones A and B (compare the contrasts obtained at RT in Figure. 8e 1e and those obtained at 396.2K in Figure 9a. At the end of this cycle (Figure 9b) the contrast of the growth band A is more accentuated and the whole sample seems to present a more perfect aspect than in the beginning of the study.

These first results can be interpreted in term of localized variations of the deformation in the sample with temperature. These temperature changes do not affect the configuration of the dislocations but the contrasts of the growth bands varies. Other recent studies (20) on another sample have confirmed this point and this localized change of contrasts seems very important in relation to the anomalies observed in propagation loss and frequency-temperature characteristics. These anomalies have been attributed (21) (7) to water impurities in these hydrothermally grown crystals. This explanation is qualitatively available and in good agreement with the change of deformation contrast during the temperatures cycles. The contrast due to the water impurity content is observed in particular zones and is predominantly located at growth bands which corresponds to the absorption of water on some growing faces during the crystal growth. The contrast changes observed during the cycle are roughly similar on recycling the crystal.

A combination of optimizing growth conditions to reduce the water content and also heat treatment to control the water distribution could be a great interest for improving the crystals for piezoelectric devices applications.

III - TRANSMISSION ELECTRON MICROSCOPY STUDIES (Water solubility)

Like for quartz, hydrothermally grown berlinite can dissolve water in its lattice. In the case of quartz it has been recently shown (22) (23), that these water associated defects are substitutional point defects $(4H)_{Si}$. It can thus be argued that in berlinite the corresponding water associated defects would be $(4H)_{Al}$ and $(4H)_{P}$. As both material exhibit identical crystal structure and very similar physical properties, one can also assume that the water associated defects should have similar behaviour. To test this assumption, "wet" berlinite synthetic single crystals have been annealed at various temperature ranging from 370° to 830°C. In "wet" synthetic quartz such a thermal treatment produces water precipitation in small bubbles (24) (diameter in the range 100 A to micron depending upon initial water content, duration and temperature of annealing). Bubbles with a diameter comparable to wave length of visible ligth generate diffusion of light, which gives to annealed quartz its characteristic milky color.

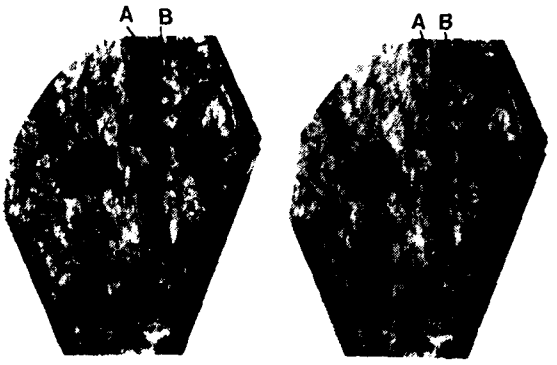


FIG. 8a:ROOM TEMPERATURE FIG. 8b:T=330.5 K

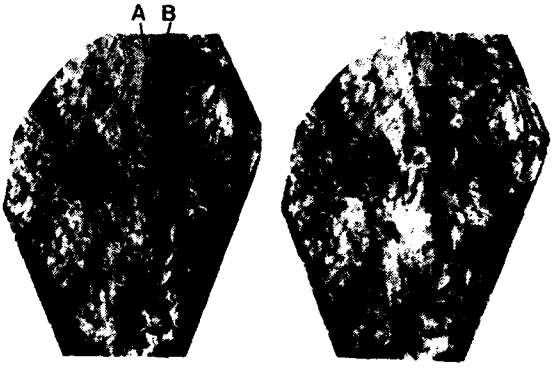


FIG. 8c:T=382.8 K FIG. 8d:T=322.5K

FIG. 8e:ROOM TEMPERATURE

FIGURE 8 : FIRST THERMAL CYCLE

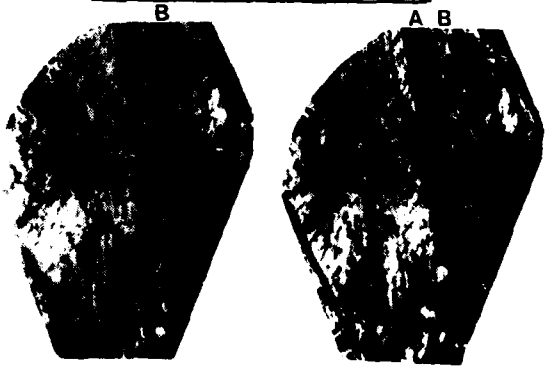


FIG. 9a:T=396.2 K FIG. 9b:ROOM TEMPERATURE

FIGURE 9 : SECOND THERMAL CYCLE

A similar phenomenon has been also observed in berlinite. For high enough annealing temperature berlinite becomes milky and in transmission electron microscopy one observes a large density of bubbles with a mean size of about 1000 A ; while in untreated -as-grown-berlinite no such bubbles are visible. Further more, the growth of these bubbles is accompanied by a severe increase of the dislocations density. In the as-grown berlinite crystals, the dislocations density is of the order of $10^3 cm^{-2}$ or much less but in the heat treated crystals this density reaches $10^9/cm^{-2}$. An increase of at least 6 orders of magnitude is produced like for wet quartz which exhibit the same dramatic dislocation multiplication. This situation can be observed in Figure 10 where bubbles with a mean diameter of 250 A can be observed together with a very high density of dislocations.

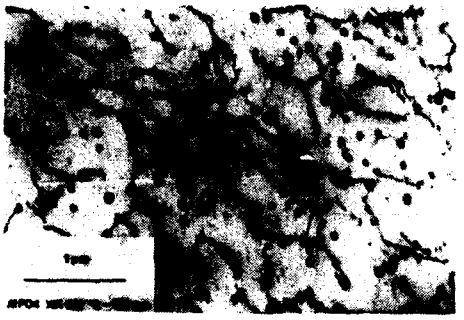


Figure 10 : Larges bubbles and connected dislocations

This phenomenon can be understood as follows (24) (25) : as supersaturated water precipitates during the heat treatment, pressure in the bubbles would increase up to extremely high values ; it is thus necessary to increase the bubble diameter by removing material which becomes the extra half plane of sessil dislocations loops, each being connected to a bubble. This mechanism is represented on Figure 11.

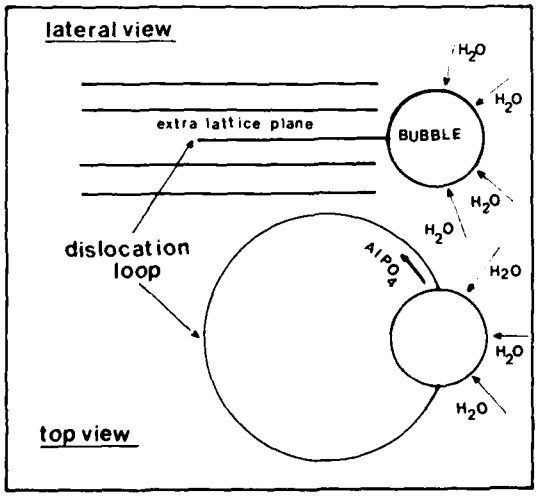


Figure 11 : Growth of bubble by creation of a dislocation

The energy necessary for the formation and the development of these dislocation loops is provided by the gain of Gibbs energy associated with the decrease of concentration of the water associated points defects i.e.

$\Delta G = kT\ Ln(C/C_0)$ per defect precipited in the bubble (C is the actual defect concentration, C_0 the equilibrium concentration). If this Gibbs energy gain is not large enough to allow the initial stage of dislocation multiplication, the bubble cannot grow and the precipitation of water produces a huge density of very tiny bubbles (diameter of 100 to 300 A), the size of which is limited by the internal pressure. This process is probably governed by the nucleation rate of the bubbles. In the opposit, if ΔG allows dislocations loops to be formed, these latter ones grow by pipe diffusion of $AlPO_4$ in the dislocation core, while the growth of the bubbles themselves and the increase of molecular water in them, proceeds by bulk diffusion of the water associated point defects to the bubbles.

On Figure 12 this situation can be observed is some zones (A,B) with much lower water concentration (precipitation of bubbles of very small size without creation of dislocation).

Figure 12 : Precipitation in small (A,B) and large bubbles

Such a mechanism has already been observed in synthetic quartz (25) ; these experiments clearly show that a certain amount of water is incorporated in Berlinite during growth, as point defect. This has also been observed in synthetic quartz years ago. Whereas now, rather "dry" (a few 10 ppm) quartz crystals are grown ; the difference being that growth conditions for quartz were change in a manner that have reduced equilibrium water concentration. Similar work is the relevant step to do, now, concerning berlinite.

IV - PIEZOELECTRIC CHARACTERIZATION OF CRYSTALS

Crystals obtained by the three growth methods were cut into resonators of different orientations -mostly Y rotated cuts-. Plane energy trapped resonators were experimentaly designed to have absent or very reduced anharmonics response by choice of electrode diameter and plating. Transmission measurements were made with network analysers (a HP3575 up to 10 MHz, a R&S ZPV, and in some instance a HP 3577 above) to determine with precision the resonant and antiresonant frequencies, equivalent resistance, and Q factor. The resonant and antiresonant frequencies and resistance were measured in function of temperature from 0° to 150°C. The 1^{st}, 2^{nd}, 3^{rd} order frequency temperature coefficients (F.T.C.) were extracted by least square polynomial fitting of fr(T) and fa(T). Coupling coefficient were computed from fr and fa of the fundamental (energy trapped) mode. In Table III results for three Y rotated cuts are given and compared to results computed (one dimensional model (9). with the data of Bailey et al

TABLE III

MODULUS FOR SOME Y ROTATED CUTS

	Y+21°20' (V.G.)		Y-30°05' (V.G.)				Y-38°30' (S.H.)	
	measured*	computed	measured*	computed			measured*	computed
$N_{f_{(m.s^{-1})}}$	1962	1955	1445	1415			1444	1448
$N_{f_{a(m.s^{-1})}}$	1969	1964	1453	1441			1448	1452
TCf_a $(10^{-6}\theta_C^{-1})$	39.1	40.6	ftal 5.2	ov.5 2.9	ftal 1.7	ov.5 -.23	-38.5	-31.9
TCf_a $(10^{-6}\theta_C^{-1})$	32.6	37.9	2.3	.74	-.47	-.47	---	-33.1
K (%)	9.7	9.8	10.1		10.1		8.4	8.0
Q (at 5 MHz)	4400	---	4100		---		2250	---

* without a mass loading correction of # 3% for N_{f_r} and N_{f_a}

In Table III it can be seen that the frequency constants ($N_i = f_i.2h$), the 1st order FTC and the coupling coefficents are similar the computed values. The observed Q factors are low. Although these values are not intrinsic Q factors of the material used, they indicated a high acoustic dissipation in the crystals grown by vertical gradient method at 180°C (V.G.), and by the slow heating method at 170°C. As previously indicate, (6) this attenuation is to be related to the water content of the crystals (500 to 1000 ppm) (26). The case of the Y-30°05' cut which is very close to the AT cut of berlinite will be further discussed in part V.

COMPARISON OF Y RESONATORS : About an hundred Y resonators were cut from crystals grown by the three methods with the conditions given in part I and also from other experimental crystals. Typical results for each growth method (mean values on the plates cut from one crystal) are reported in table IV together with results obtained with an experimental crystal and the computed ones.

TABLE IV

COMPARISON OF Y PLATES FROM DIFFERENT GROWTH METHODS

SAMPLES*	Y (H.G.)	Y (S.H.)	Y (V.G.)	Y (V.G.)	Y experim.	Y computed
	Y:IGH9.084	EURY2.084	PP20	PP26	Experimental	---
N_{f_r} $(m.s^{-1})$	1703	1697	1697	1698	1714	1688
N_{f_a} $(m.s^{-1})$	1711	1705	1706	1708	---	1677
Q (at 5 MHz)	3550	3140	4650	4110	65400 ftal 290000 3ov.	---
TCf_r $(10^{-6}.°C^{-1})$	65.	56.	58.	58.	61.9	63.6
TCf_a $(10^{-6}.°C^{-1})$	65.	53.	55.	54.	---	59.5
TCf^2_r $(10^{-9}.°C^{-2})$	-226 to -712	-268 to -537	-238 to -396	-395 to -507	16.2	---
K (%)	11.0	11.3	12.6	12.7	13.1	12.5

*Nota:NF without any correction for mass loading (# 3%)
K measured from fa-fr of energy trapped resonators.

In this table, it can be seen that crystals obtained at 170-180°C by Horizontal Gradient (H.G.). Slow Heating (S.H.) or Vertical Gradient (V.G.) have similar properties close to the computed values. However an attentive examination would revealed some differences with computed results that indicates that the values of C_{66} and e_{11} given by Bailey (2) may be a little underestimated. Whatever it may be, the computed results give coupling coefficient for infinite plates that are, in many instances, lower than experimental values for energy trapped resonators.

In column 5 of Table IV we give results obtained recently with a crystal grown in experimental conditions by the slow heating method ; small, but significatives differences with conventional crystals, exist for Nf and k. A major difference exists for the Q factor and the second order FTC.

Frequency responses of resonators insertion losses : On Figure 13 are given the typical frequency responses of Y resonators with similar design (plate thickness ≈ 0.3 mm electrode diameter 3.0 mm platting ≈ 3 %) cut from, crystals obtained at 170°-180°C by the different growth methods. It can be seen that they are very similar with insertion losses on the range 14 to 17.5 dB this corresponds to resistances in the range 400 to 650 Ω and to Q factors in the range 4000 to 2500. These low values of Q factor which are not the intrinsic values (that can be only measured with plano-convex resonator in this frequency range and are at least twice) result principally of the water dissolved in the samples (6).

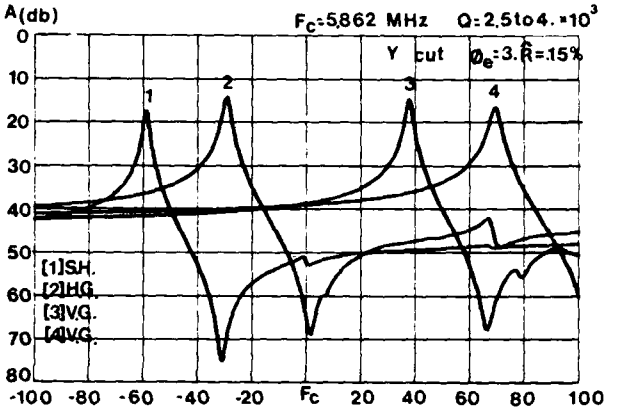


FIGURE 13:
Y CUT RESONATORS (different growth methods)

On Figure 14 is given the frequency response of one plate of an experimental crystal. It can be noticed that the Q factor is greater by an order of magnitude than for resonators of Figure 13. In this figure is also given the frequency response of Y quartz plate of similar design whose Q factor is not so different and whose coupling coefficient is smaller. This is in contradiction with the computed results which indicate a larger one for Y quartz (k=13.4 % against 12.5 % for berlinite).

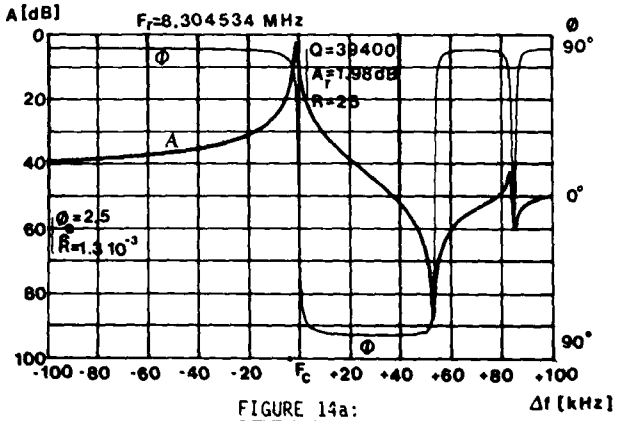


FIGURE 14a:
Y CUT PLANE FUNDAMENTAL RESONATOR

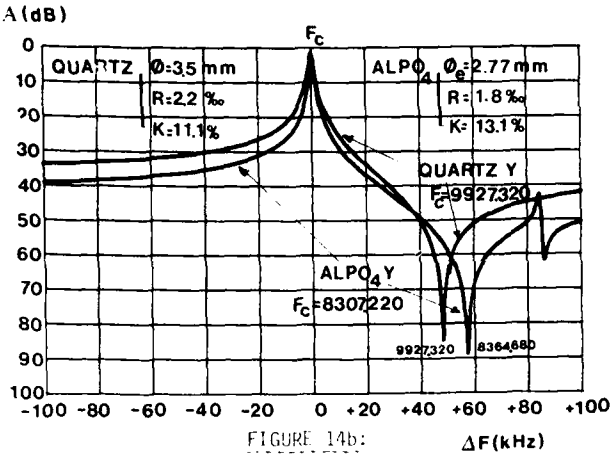


FIGURE 14b:

COMPARISON OF Y QUARTZ AND Y BERLINITE RESONATORS

The same plate was measured on 5th overtone : the Q factor was 52187 that reduced to 5 MHz corresponds to $4.34.10^5$. This crystal contains a few tens ppm of water (26). Account taken of the under optimal design of this resonator (plane lapped plate of small area with too thick electrods) these results lead us to the conclusion that "dry" berlinite resonators will be as Q factor is concerned, very similar to quartz. The exact intrinsic Q factor is still to be determined.

Frequency temperature behaviour : Although quite similar 1st order FTC are measured with crystal of good crystalline quality grown at 170°-180°C in orthophosphoric acid, the higher order temperature coefficients measured with different crystals and even with different plates of the same crystal can present a large dispersion (see table IV). In all cases, they have values much larger than for quartz resonators of the corresponding orientation. On Figure 15 is displayed the frequency - temperature behaviour of Y plates cut in a crystal that is known to contain a large amount of water. (~ 600 to 1000 ppm). The observed variations of $[fr(T)-fr(20°C)]/fr(20°)$ are much more dispersed than the Q factors of the same plates.

It was observed by T.E.M. or simply by precipitation of dissolved water and optical observation (10) that the concentration of dissolved water can have large variations accross a crystal or a plate. Thus, the values of 2nd order and 3rd order FTC appears to be much more sensitive to high water concentration than Q factor and IR transmission.

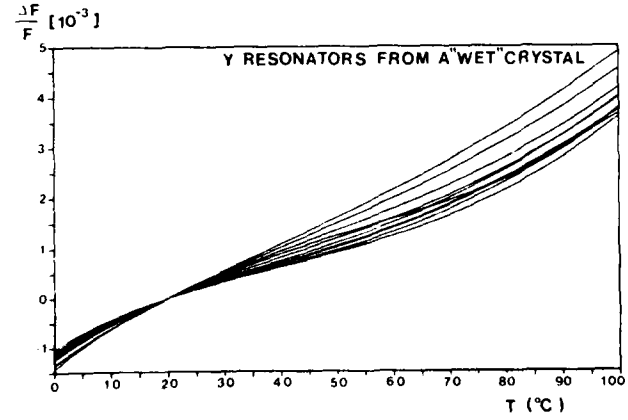


FIGURE 15: FREQUENCY-TEMPERATURE BEHAVIOUR OF A "WET" CRYSTAL

The variations of fr(T) of water containing crystal can be understood as the sum of two contributions : one being the intrinsic value for dry berlinite (which is quasi linear for the Y cut) ; the other being a function of temperature very alike to :

$$-C\ th(T-T_c)$$

That possesses an inflexion at temperature T_c (near 50°C for most cases), C being a function of water concentration.

For plates of figures 15 the 2nd order FTC were in the range - 268 to $-537.10^{-9}°C^{-2}$ where as 3rd order FTC were in the range + 3700 to + 6800 $10^{-12}°C^{-3}$. This values are typical of crystal with a high water content. As indicated in part III, with such resonators some aging ($\sim 10^{-5}$) can be observed after the 2 or 3 first thermal cycles (0-150°C).

For the best crystals grown at 170-180°C, with the conditions given in part I and seeds of great perfection, lower and less dispersed 2nd order FTC are observed (-180 to -60 $10^{-9}°C^{-2}$) but the 3rd order FTC remain high (a few 1000 $10^{-12}°C^{-3}$). A typical case of frequency - temperature behaviour for Y plate of such crystal is given in figure 16.

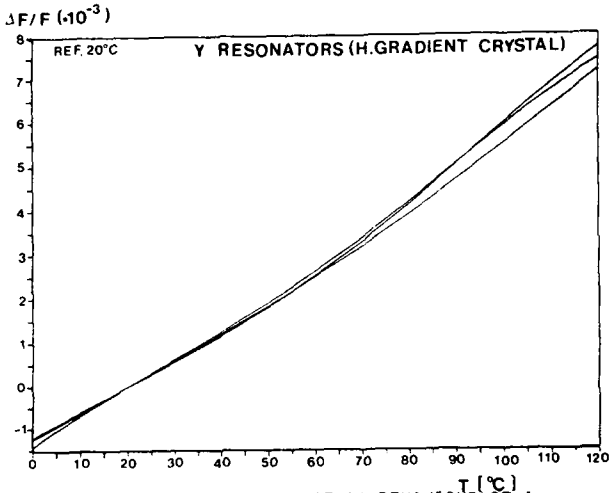


FIGURE 16: FREQUENCY TEMPERATURE BEHAVIOUR OF A H.G. CRYSTAL GROWN AT 170°C

TABLE V: COMPARISON OF FTC (Low Q, High Q crystals, Quartz)

SAMPLE	1st ORDER FTC 10^{-6} °C^{-1}	2nd ORDER FTC 10^{-9} °C^{-2}	3rd ORDER FTC 10^{-12} °C^{-3}
Y CUT 170° - 180°C	55 to 65	-60. to -180.	+1000 to +7000
Y CUT Crystal n°1	61.9	16.	1.4
Y CUT Crystal n°2	81.2	47.	-6.
QUARTZ Y (B.B.L. (32))	92.	58.	6.
Y-21°12' Crystal n°3	47.6 to 47.9	14. to 21.	21. to 32.
QUARTZ Y-25 (Computed)	48.	55.	90.

The frequency - temperature behaviour of plates with very low water content are very different : the values of the 2nd and 3rd order FTC are now smaller by at least one order of magnitude. They are generally smaller than the values corresponding to similar cuts of quartz. Some exemples are given in Table V.

Nota : The dispersion on values of the 2nd and 3rd order FTC in this table results for a large part of insufficient reproductibility and absolute accuracy in resonant frequency measurements (at least some 10^{-6} accross the temperature range).

The frequency temperature behaviour of a Y plate of experimental crystal n°1 is displayed on figure 17.

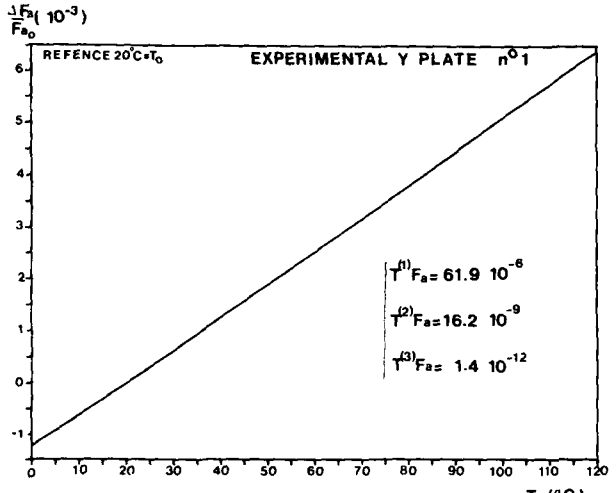


FIGURE 17: BEST RESULT OF THERMAL BEHAVIOUR (Y PLATE)

The figure 18 gives the thermal behaviour for two Y plates of experimental crystal n°2 ; these two resonators have different design and different resonant frequencies ; the curves represent the variations of $\Delta f/f$ for the resonant frequencies of fundamental mode. It was found that the 1st order FTC for antiresonant frequencies was about 2 ppm lower than for resonant frequency. This indicates that the temperature coefficient of the coupling coefficient is small and negative.

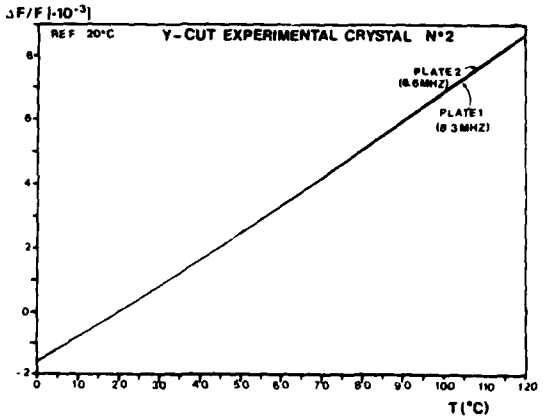


FIGURE 18: THERMAL BEHAVIOUR OF HIGH Q Y PLATES

On figure 19 are displayed the thermal behaviour of two similar Y-21°22' plates (z faces of rhombohedron) either for fundamental resonant frequency or for 3rd overtone.

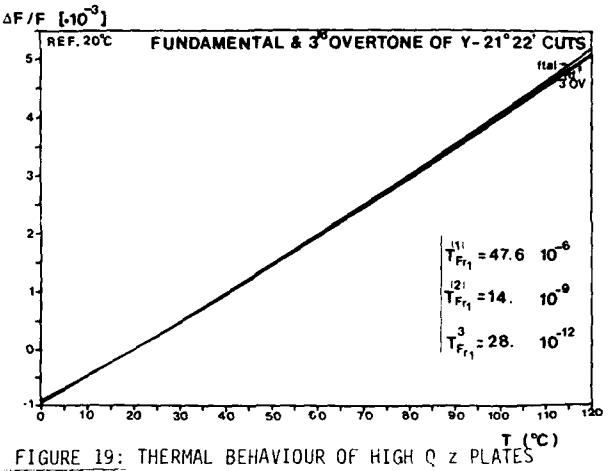


FIGURE 19: THERMAL BEHAVIOUR OF HIGH Q z PLATES

VARIATIONS OF INSERTION LOSSES WITH TEMPERATURE : As already noticed acoustical losses in crystals grown at 170°-180°C are high and have large variations with temperature. Insertion losses of many resonators were measured from 0° to 150°C for the zero phase resonant frequency : some measurements were made from 77 K to 550 K. For all crystals grown at 170°-180°C, a maximum of loss was observed near 50°C (frequencies 5 to 10 MHz), in some instances, a second relative maximum was also observed near 77 K.

For the crystals with very low water concentration, the insertion losses are very low and have no significative variations with temperature. On figure 20 typical exemples are shown for the two types of crystals.

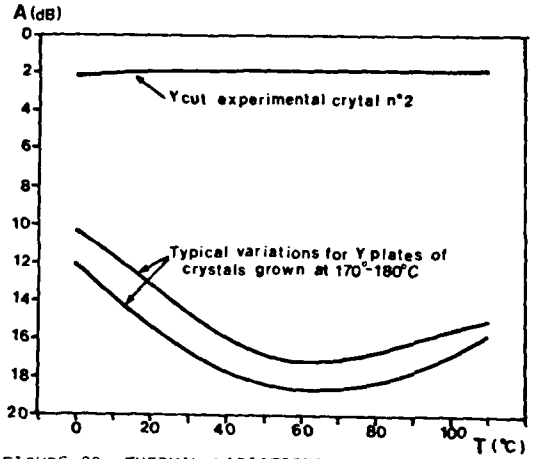


FIGURE 20: THERMAL VARIATIONS OF INSERTION LOSSES.

CONCLUSIONS DRAWN FROM PIEZOELECTRIC MEASUREMENTS : No determinant differences were found between the properties of crystals grown at 170°-180°C by the three different methods. But it was demonstrated that, the water dissolved in this crystals, induces low values of Q factors and deteriorates significantly the intrinsically very favorable thermal behaviour of this material. It was also shown that, the coupling coefficients that can be computed with the presently most reliable constants of this material (2) underestimate the values that can be experimentally obtained.

V - EVALUATION OF BULK WAVE DEVICES FOR FREQUENCY GENERATION AND FILTERING

AT CUT OF BERLINITE : The angular position of AT cut was computed from the data of Bailey et al. (2). On figure 21 are displayed the main computed properties of resonators near the AT cut (1st order FTC, phase velocity, coupling coefficient for the C mode).

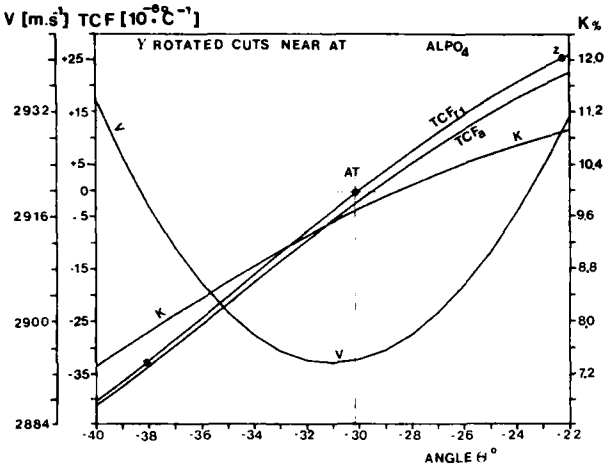


FIGURE 21: COMPUTED PROPERTIES NEAR AT CUT

Experimentally, this cut was found to be at about 30° 05', which is very close to the computed one. Resonators with filter type response were made with crystals grown by V.G. and S.H. methods. In Figure 22 are displayed the frequency responses of resonators made from plates of similar thickness (.190 mm), same electrode diameter (3mm) and different platings (about 1.5 ‰ to 1 %).

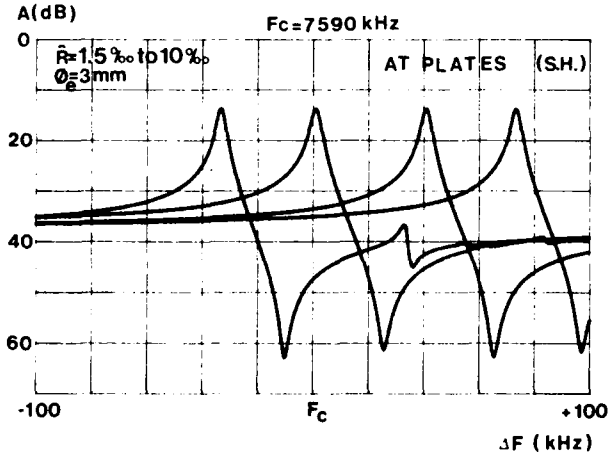


FIGURE 22: FREQUENCY RESPONSE OF AT PLATES

On this figure it can be seen that insertion losses are in the range 13 to 14 db (resistances 350-400 Ohms) and that anharmonics responses are abnormally low for the highest plating (resonator of lowest resonant frequency). This fact, frequently observed with water containing plates, was interpreted as the result of an abnormally high attenuation of the thickness-shear and twist waves accross these plates that prevents the anharmonics to have a normal amplitude. The frequency temperature behavior of these plates where measured, and again some dispersion was observed. The best result observed for the fundamental mode is shown in figure 23. The corresponding 1st order FTC is 2.38 $10^{-6}°C^{-1}$, the 2nd and 3rd order FTC are $-70.10^{-9}°C^{-2}$ and $+853.10^{-12}°C^{-3}$.

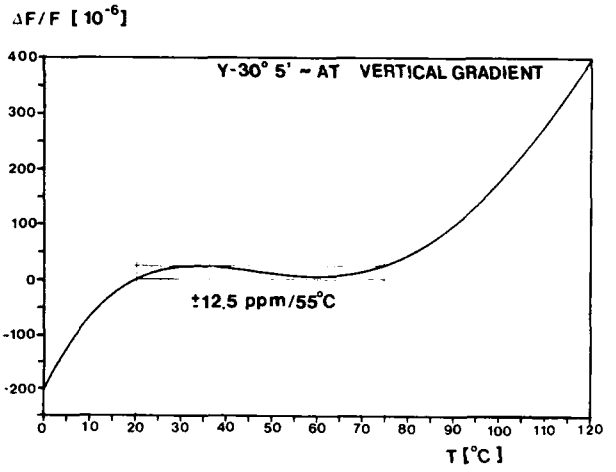


FIGURE 23: THERMAL BEHAVIOUR OF FUNDAMENTAL MODE

The best results for 3rd overtone is shown in figure 24. The value of 1st order FTC is 2.93 $10^{-6}°C^{-1}$; the values of 2nd and 3rd order FTC are -140 $10^{-9}°C^{-2}$ and +1400 $10^{-12}°C^{-3}$.

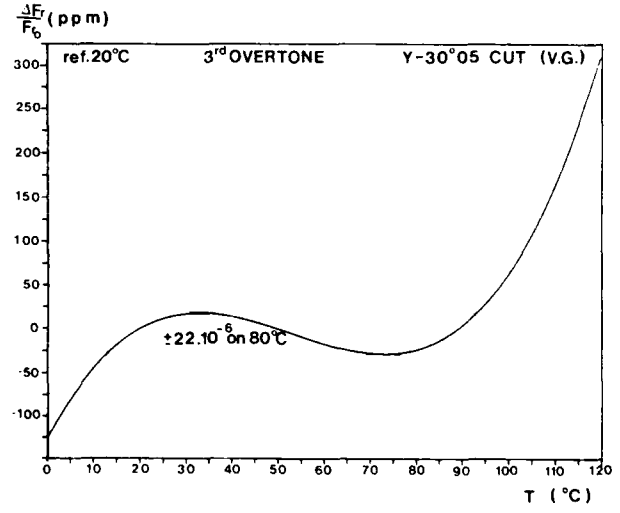


FIGURE 24:THERMAL BEHAVIOUR OF 3rd OVERTONE (AT CUT)

Although this results may seem good, we can induce, from results obtained with waterless crystals (part IV table V) that, with such crystals, the values of 2nd and 3rd order FTC of the AT cut of berlinite will be reduced by more than an order of magnitude and most probably much lower than for AT cut of Quartz. We cannot exclude, for dry crystals, the possibility of some change in the angular position of this cut (this was already seen for water containing synthetic quartz), and of some shift of the inflexion temperature.

FILTER EXPERIMENTATION : The coupling coefficients measured with AT resonators with filter type response were slightly above 10 %. With some optimization of the design the value of 11 % will probably be achieved. In terms of k^2, or in terms of ratio of the capacitors of the equivalent scheme, this means that a value about twice of AT quartz is obtained (7.4 % is a common value of k for a filter type AT quartz) ; and, this allows, with same design rules, to obtain filters with 2 times larger bandwidth.

To demonstrate this a 4 poles Jaumann filter was contructed. This filter was designed to use no coupling the capacitor between the two sections (other than the plates to case capacitances of the resonators 4x1.35 pF). Its relative bandwidth is .58 %. Insertion loss is 4.5 dB one dB of this being due to the transformers. The response of this filter is given in figure 25 where it can be seen that the Q factor of the resonators are somehow insufficient to approach enough the .1 dB Chebitcheff theoretical response (corner of bandpass).

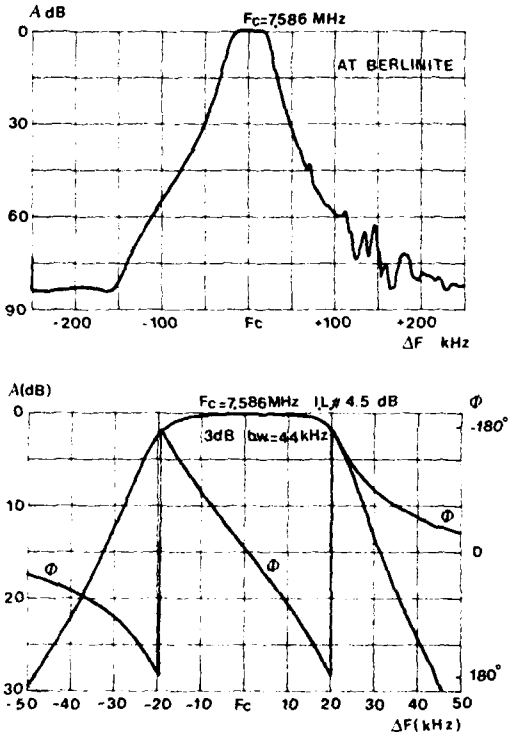


FIGURE 25:
4 POLES AT BERLINITE JAUMANN FILTER

ENERGY TRAPPING IN BERLINITE RESONATORS : Experiments with plates cut in crystals obtained at 170°-180°C can lead to the erroneous conclusion that this material is much less sensitive to energy trapping parameters than quartz as far the amplitude of anharmonic modes is concerned (Exemples of Figure 22). More sophisticated experiments with theses plates, making use of static capacitance compensation, would reveal that these modes do exist, but have very reduced amplitude because of the large acoustic attenuation. Such an examination would also reveal that, for a similar design, much more anharmonic modes are excited than in a quartz plate. Experiments with the first high Q resonators have confirmed these observations, and the existence of an important difference with quartz.

This difference can easily be understood by using the two dimensional theory of Beckmann numbers (for thickness-shear -ts- and thickness-twist -tt-) established by Mindlin years ago (27) (28) (notations of these references). To establish a comparison let us consider the case of AT quartz and AT Berlinite resonators at the same frequency and with the same plating (.005).

$$\text{from } F_r = \frac{1}{4h}\sqrt{\frac{\bar{C}_{66}}{\rho}} * \left(1-R-\frac{4k^2}{\pi^2}\right) \quad (2h = \text{thickness of plate}) \quad [1]$$

$$\text{it comes} \quad \frac{2h_q}{2h_b} \# 1.157 \qquad [2]$$

The Bechmann number for 2D thickness twist is

$$B_{tt} = \frac{L_{tt}}{2h} = \left[\frac{4\gamma_{55}}{\bar{C}_{66}\left(2R+\frac{8k^2}{\pi^2}\right)\left(1+R-\frac{8k^2}{\pi^2}\right)}\right]^{1/2} \qquad [3]$$

In theses conditions (same F_r, R, one trapped mode) the ratio of electrodes dimensions for quartz to berlinite is :

$$\frac{L_{tt_q}}{L_{tt_b}} = \left[\frac{\gamma_{55q}}{\gamma_{55b}} * \frac{\rho_b}{\rho_q} * \frac{\left(2R+\frac{8k^2_b}{\pi^2}\right)}{\left(2R+\frac{8k^2_q}{\pi^2}\right)}\right]^{1/2} \# 1.32 \qquad [4]$$

similarly for thickness shear direction we have (27) :

$$\frac{L_{ts_q}}{L_{ts_b}} \# 1.24 \qquad [5]$$

The principal consequences of relations [2],[4] are [5] the reduction of dimensions (thickness and lateral dimensions) of berlinite resonators as compared to quartz (about by 1.3 for electrodes and lateral dimensions). Most of this difference arises from differences in values of elastic constants (those of quartz being 1.2 to 1.3 time higher) and in coupling coefficient (k^2 of AT berlinite is about twice that of AT quartz).

Similary it can be established that in these conditions the static capacitance of a berlinite resonator will be reduced by a factor of about 1.35 as compared to quartz. Since C/C_0 of AT berlinite is about twice that for AT quartz, at same frequency, a filter type AT berlinite resonator will have a motional capacitance about 2.70 times that of an AT quartz.

To obtain more precise design of high Q berlinite resonators, a 3D computation of modes and electrical admittance is being made (31) using a solution for flat plates with circular electrodes of the equation given by Tiersten for TT and TS (29) (30). Experimental observation of modes by X ray topography of high Q resonators have confirmed the need, at equal plating, to reduce electrode diameter as compared to quartz.

They have also shown, that the phenomenon of flexure coupling with thickness shear will most probably be more important in berlinite. The topograph of Figure 26 display the fundamental mode of vibration in a Y-21°22' plate with R=3 ‰ and ϕ/2h=16 (suboptimal value).

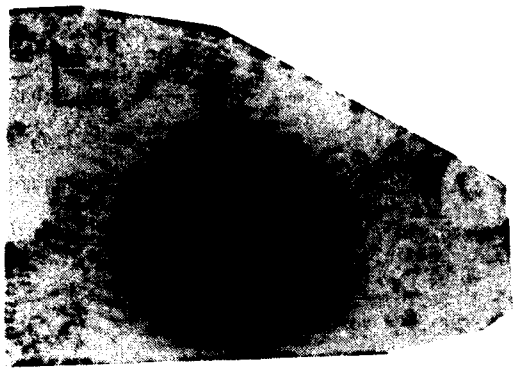


FIGURE 26 : VIBRATION MODE IN A Y -21°22' PLATE

OTHER COMPENSATED CUTS : In Figure 27 are given the computed values of the 1st order FTC for all doubly rotated orientations, as previously found by A. Ballato (5), several locus of compensated cut exist in this material. The interest of the other cuts will probably be dependant of either lower higher order FTC or better non linear properties than for quartz.

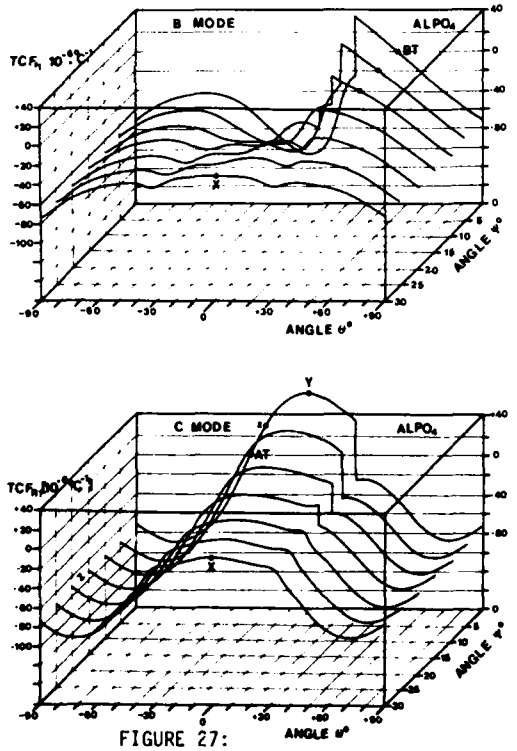


FIGURE 27:
COMPUTED THERMAL BEHAVIOUR OF BERLINITE
(B&C THICKNESS SHEAR MODES)

A TENTATIVE COMPARISON WITH QUARTZ : On the basis of previously known properties and of new results given in this paper, the most probable advantages of berlinite devices over corresponding quartz devices are summarized in table VI.

TABLE VI : A TENTATIVE COMPARISON WITH QUARTZ *

(Thickness Y rotated resonators at same frequency)

USEFUL COUPLING COEFFICIENT (AT filter trapped resonators)	ENHANCED BY 1.4 - 1.5
SHIFT OF OSCILLATORS or BANDWIDTH OF FILTERS (AT cut)	TWICE
ANGULAR SENSITIVITY (first order F.T.C.)	REDUCED
THERMAL STABILITY** (higher order F.T.C.)	BETTER
Q FACTOR-PROPAGATION LOSSES	ALREADY SUFFICIENT (may be comparable)
THICKNESS OF PLATES (AT)	REDUCED BY 1.15
ELECTRODES DIMENSIONS (AT) (same plating , 2D TT or TS)	REDUCED BY 1.32 (TT) REDUCED BY 1.24 (TS)
NON LINEAR PROPERTIES	TO BE DETERMINED

*Hypothesis:"Dry" Berlinite has similar C,E,EPS constants, 1st order T.C. and much reduced higher order T.C.

The most important advantage is the possibility of obtaining bandpass of filters or shift of oscillators two times larger. The advantage concerning the increased frequency versus temperature stability may be also of large importance.
Although some preliminary experiments have indicated a lower influence of exciting level on resonant frequency (anisochronism) than for a similar cut (Y) of quartz, the nonlinear properties of Berlinite remain to be evaluated.

CONCLUSION

In this contribution three growth methods were described, among them the vertical gradient method and a variant of the horizontal gradient method have interesting properties in view of the obtention of large crystal of high perfection. It was demonstrated that, at growth temperature near 170°-180°C, an imporant intrinsic solubility of water exist ; and that the dissolved water is retained in the form of point defects much similar to those existing in quartz. Some feature of water rich crystal were found or confirmed by X ray topography or resonators measurements. Among them the most important for devices applications are the reduction of the Q factors, a large defacement of thermal behaviour and an increase of aging. The possibility to obtain crystals with as low water content as quartz was demonstrated. Several properties of major interest of waterless crystal were shown : their very high Q factors, their superior thermal behaviour and the ability of this material to give devices with smaller size than quartz. The larger coupling coefficient of this material gives the possibility to obtain filters with twice the bandwidth of quartz or to obtain two times larger shifts of oscillators.

This unique set of properties makes of "dry" berlinite a material of major interest for piezoelectric applications.

ACKNOWLEDGMENT : The authors wish to acknowledge the skilful assistance of C. Joly, L. Jenselme, L. Astier, A. Jeanne-Michaud, Y. Bernard for experimental work and measurements and the dextrous aid of A. Jouanin, R. Chenebault and G. Le Thuaut for cutting and polishing the many samples used in this study. The authors are indebted to Mr. P. Amstutz for the theoretical design of berlinite filters and helpful discussions and also to Professor Maurin for valuable comments. They wish to acknowledge the expert advices of Y. Toudic of CNET/LAB/ICM whose work on crystal growth was one of the basis of the present study. Special thanks are due to M-Christine Mourier for preparing this manuscript.

Several parts of this work were supported by DAII, DRET and ANVAR/CNRS contracts.

REFERENCES

(1) Y.P. Chang, G.R. Barsch
IEEE Trans. sonics ultras vol.Sa 23 n°2 p127 (1976)
(2) D.S. Bailey, J.C. Andle, D.L. Lee, W. Soluch
J.F. Vetelino, B.H.T. Chai
Proc. 1983 ultrasonics Symposium p.335 (1983)
(3) J. Hénaff, M. Feldmann, M.A. Kirov
Ferroelectrics vol.42 pp.161-185 (1982)
(4) J. Détaint, M. Feldmann, J. Hénaff, M. Poignant, Y. Toudic
Proc. 33rd Frequency control Symposium p.70 (1979)
(5) A. Ballato, E.R. Match, M. Mizan, B.H.T. Chai
R.S. Tilton, T.J. Lukaszek
Proc. 38th Frequency control Symposium p (1984)
(6) R.S. Narayan, B.H.T. Chai, J.F. Vetelino, W. Soluch
R.S. Falconer, B. Chick
Proc. 1984 IEEE Ultrasonics Symposium
(7) R.F. Steinberg, M.K. Roy, A.K. Estes, B.H.T. Chai,
Proc. 1984 IEEE Ultrasonics Symposium p.
(8) J.M. Stanley
Industr. Eng. Chem. Vol. 46 p.1684 (1954)
(9) B.H.T Chai, M.L. Shand, E. Buehler, M.A. Gilleo
Proc. IEEE Ultrasonics Symposium p.577 (1979)
(10) K. Nagai, T. Okawa, J. Okuda, J. Asahara
Proc. 1st Int. Symposium on Hydrothermal Reactions
Tokyo 1982 p.496
(11) J. Détaint, M. Poignant, Y. Toudic
Proc. 34th Frequency Control Symposium p.93 (1980)
(12) B.H.T Chai, E. Buehler, J.J. Flynn
Us Patent 4 324 773, 13 April 1982
(13) E.D. Kolb, R.A. Landise
J. Crys. Growth vol.43 p.313 (1978)
(14) A. Goiffon, J.C. Jumas, R. Astier, E. Phillipot
J. Crystal Growth to appear (1985)
(15) D.F. Croxall, T.R.A. Christie, B.J. Isherwood
A.G. Todd, J. Birsch
European Conf. on Crystal Growth Lancaster
University 1979
(16) E.D. Kolb, R.L. Barnes, J.C. Grenier, R.A. Landise
Proc. 33rd Frequency Control Symposium p.80 (1979)
(17) U. Krauss, G. Lehmann
Z. Naturforsch Vol. 30a p.28 (1975)
(18) E. Philippot, J.C. Jumas, A. Goiffon, R. Astier, M. Maurin
Brevet Français 8401911
(19) A. Zarka, B. Capelle
to appear in J. Appl. Crystallography
(20) A. Zarka, B. Capelle
To be published
(21) J.P. Bachheimer, B. Berge, G. Dolino,
P. Saint-Grégoire, C.M.E. Zeyen
Solid State communication vol.51 p.55 (1984)
(22) R.H.D. Nuttal, J.A. Weil
Solid State communications vol.33 p.99-102 (1980)
(23) J.A. Weil
Phys. and Chem. of Minerals vol.10 p.149-165 (1984)
(24) A.C. Mc Laren, R.F. Cook, S.T. Hyde, R.C. Tobin
Phys. and Chem of Minerals vol.9 p.79-94 (1983)
(25) J.C. Doukhan, L. Trepied
Bull. Minerologie vol.108 p.97-123 (1985)
(26) E. Philippot. Y. Toudic, J.C. Doukhan, J. Detaint
To be published
(27) R.D. Mindlin
J. Acoust. Soc. Am. Vol.43 n°6 p.1329 (1968)
(28) R.D. Mindlin
J. Acoust. Soc. Am, Vol.41 n°4 p.969 (1967)
(29) H.F. Tiersten, R.C. Smythe
J. Acoust. Soc. Am. Vol.65 n°6 p.1455 (1979)
(30) H.F. Tiersten
Proc. 36th Frequency Control Symposium p.37 (1982)
(31) J. Detaint, E. Philippot, J.C. Jumas
To be published
(32) R. Beckmann, A.D. Ballato, T. J. Lukaszek
Proc. IRE Vol.50 p.1818 (1962)

THE INFLUENCE OF TEMPERATURE AND ELECTRIC FIELD ON THE ETCH-CHANNEL DENSITY IN SWEPT CULTURED QUARTZ

John G. Gualtieri

US Army Electronics Technology and Devices Laboratory (ERADCOM)
Fort Monmouth, New Jersey 07703-5302

ABSTRACT

It is known that sweeping can reduce the etch-channel density in quartz. Various electrode types may be used for sweeping quartz. Usually employed are platinum-foil electrodes under pressure or evaporated metallizations to the z-surfaces with pressure or welded contacts. The reduction in etch-channel density varies with the electrode type and with the temperature and applied field conditions. In the present work, emphasis was on studying sweeping with gold metallized z-surfaces. The temperature and field values which reduce etch-channel density to zero were investigated.*
Under certain values of temperature and field, gold has been found to diffuse into the quartz. This diffusion correlates with the reduction of etch-channels to zero.

Using dark-background light-scattering microscopy, it was found that etch-channels were originating at the surface intersection of straight growth tunnels as previously reported by Katz, et al.,[1] and at irregularly curved defect lines (possibly precipitate decorated dislocations) which invariably connect the growth tunnels. After sweeping, metal decoration of both defect types was also observed using the light-scattering method.

INTRODUCTION

Defects in natural and cultured quartz have been studied by many investigators. X-ray topography and etching were used to reveal etch-channels.[2-5] In particular, Nielson and Foster have suggested that etch-channels were formed by leaching of impurities which had precipitated along dislocation lines.[6] Also, Katz, et al. reported that etching broadens existing tunnels which were formed during growth of the crystal, making them observable.[1] Both diffusion and electrodiffusion of metals into quartz has been previously reported with decoration of dislocations[7] and with no apparent connection between diffusion and dislocations or precipitates.[8]

Vig, et al., reported that swept quartz exhibited a lower incidence of etch-channels than did unswept quartz, and that one variety of vacuum-swept quartz was free of etch-channels.[9] Martin, et al.,[10] reported that sweeping reduced the etch-channel density by approximately 75%. The variability in etch-channel density, reported by investigators using different sweeping methods, was the motivation for this study. The goal was to determine the sweeping conditions that minimize etch-channel density. A parametric study was performed in which various combinations of sweeping temperature and electric field strength were correlated with etch-channel densities subsequent to sweeping.

*Zero etch channels means we cannot find any using 60X magnification and diligence.

EXPERIMENTAL

The cultured quartz used in this investigation was obtained from three different domestic growers. The quartz was lumbered into y-bars or SC-bars cut from z-growth sectors after seed removal.[11] Evaporated gold electrodes were applied to the sawn surfaces after cleaning. The electrodes consisted of 100 Å of chrome, followed by 50 Å of chrome-gold mixture, and finally, 1000 Å of gold. An Auger electron spectroscopy (AES) depth profile of atomic concentration of the electrode surface, before sweeping, is shown in Fig. 1. Gold-plated molybdenum leads were welded to both z-surface electrodes after the selected bonding area was electroplated with gold to improve weld adhesion. The sweeping was performed in laboratory-air atmosphere. The following sweeping procedure was used in each run:

1. The full DC voltage was applied at the beginning and was maintained throughout the entire run.
2. The temperature was automatically increased to the maximum temperature at a rate of 20°C/hr.
3. The temperature was held constant at the maximum for at least 24 hours.
4. The temperature was automatically decreased at 20°C/hr. down to room temperature.

During the entire run, data consisting of temperature, current, and time were automatically recorded at 6 minute intervals. At the completion of each run, a computer plot of current density and temperature vs. time was generated. A typical plot is shown in Fig. 2. AT-cut plates, usually 1 mm thick, were cut from the swept bars and subjected to etching in a saturated solution of ammonium bifluoride at 75°C for 2 hours. The plates were then optically examined under 60X magnification in order to determine etch-channel densities.

 "US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

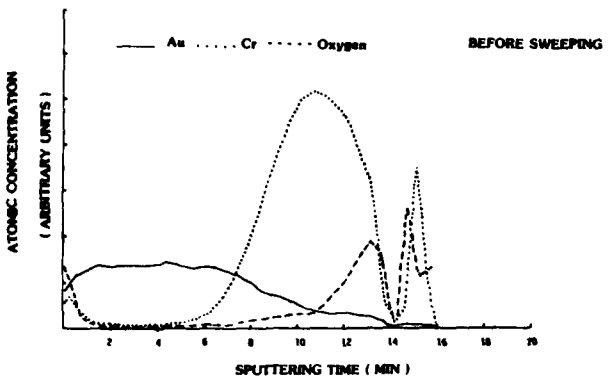


Fig. 1

Auger Electron Spectroscopy (AES) depth profile of a Cr-Au metallized z-surface before sweeping. Atomic concentrations are in arbitrary units for each element.

In order to view defects in bars or plates using scattered light, samples were polished on two perpendicular surfaces. On y-bars, for example, an x-surface and a y-surface were polished. This permitted viewing in either the x- or y-direction while illuminating with high intensity, narrow beam white light at right angles to the viewing direction. To create a dark background, the opposite, sawn, x- and y-surfaces were coated with a black cement**. Alternatively, the opposite surfaces may also be polished and placed above a dark background. High intensity, narrow beam, white light, and a dark background are essential for viewing and photographing inclusions on the order of 1 μm in size.***

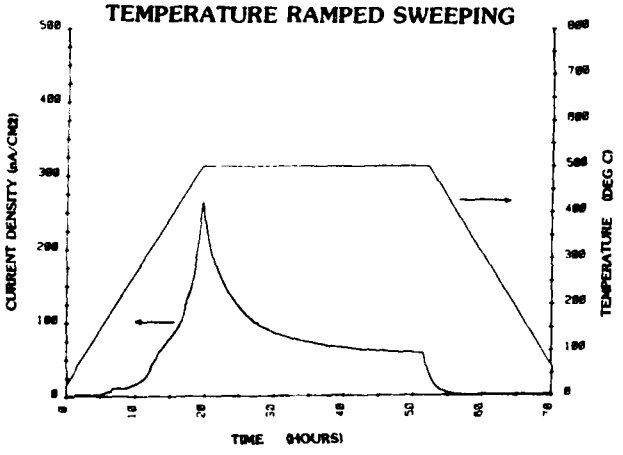


Fig. 2

A typical computer plot of the current density and temperature vs. time for a sample which was swept using evaporated gold electrodes.

EXPERIMENTAL OBSERVATIONS

One of the differences between sweeping with evaporated gold electrodes vs. platinum foil electrodes applied directly to the quartz is the decreased initial resistivities when using evaporated gold. The resistivities at the beginning of the constant temperature period are approximately 5 times smaller. However, the final stabilized resistivities, those at the end of the constant temperature period, are similar. Table I compares the initial and final resistivities and the transported charge for two different sample types, using either evaporated gold or Pt-pressure electrodes.

TABLE I

SAMPLE	INITIAL RESISTIVITY (OHM-CM)	FINAL RESISTIVITY (OHM-CM)	TOTAL CHARGE TRANSPORTED (COULOMBS)	ELECTRODE MATERIAL
1	1.51×10^8	1.79×10^9	0.733	EVAP. Cr + Au
	8.62×10^8	2.27×10^9	0.271	Pt-FOIL
2	2.38×10^8	2.50×10^9	0.650	EVAP. Cr + Au
	1.09×10^8	2.27×10^9	0.175	Pt-FOIL

Note: Samples were heated to 537°C at a rate of 20°C/hr, followed by a constant temperature period of 24 hours, and then cooled at a 20°C/hr rate. The initial and final resistivities were the field/current-density values at the beginning and end of the constant temperature period. The transported charge is the integral of the current over time, the field was 250 V/cm. Sample types had equal dimensions.

Another difference is that only when sweeping with evaporated gold electrodes have we found certain combinations of temperature and field which eliminate etch channels. Those conditions which yield zero etch channels correlate with the observation of cathodic depletion of gold. The combinations are indicated in Table II. Whenever cathodic depletion was not clearly evident, etch channels were observed.

**The black cement was obtained from Ernest F. Fullam, Inc., Latham, NY 12110. It is No. 14820 Low Resistance Contact Cement (a suspension of fine carbon particles in amyl acetate).

***Tunnel and precipitate inclusions usually required 15 min. photographic exposures using 3000 ASA film.

For example, etch channels were found when y-bars were swept after one end of the bar had been cut off at the AT-angle. Fig. 3 illustrates this situation and shows that there is no anode region directly opposite the no-depletion region.

TABLE II

AT-PLATE ETCH-CHANNEL DENSITIES

COMPARISON OF SWEPT/UNSWEPT (CM^{-2})

USING Cr/Cr + Au/Au ELECTRODES

MAX TEMP OF SWEEP (DEG C)	CONSTANT SWEEPING FIELD (V/CM) THROUGHOUT RUN						
	0	50	125	250	500	1000	2000
550	10/36	25/36 *	0/36 *	–	–	–	–
500	–	–	12/30	0/30 *	0/30 *	0/30 *	0/30 *
400	–	–	–	–	–	7/46	0/151 *
300	–	–	–	–	–	7/46	71/151
200	–	–	–	–	–	44/46	–

* DEPLETION OF GOLD AT THE CATHODE WAS OBSERVED.

Note: The temperature-time program includes a 20°C/hr heat-up, a constant (maximum temperature) 24 hour period and a 20° C/hr cool-down.

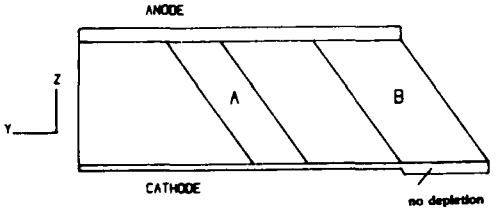


Fig. 3

Outline drawing of a quartz y-bar cut at the AT-angle (∿35°). There is no anode directly opposite the region of no depletion. Therefore, region B contains electric field lines which are not parallel to paths of easy migration of gold. Etch-channels were observed in AT-plates cut from region B, but were not observed in those cut from region A.

An attempt was made to etch AT-plates, which were etch-channel free, in aqua regia to determine if removal of gold from defects in the quartz would cause etch-channels to reappear after etching again with ammonium bifluoride. An increase in etch-channel density was observed after etching at 75°C for 2 hours. The results are given in Table III.

TABLE III

AT-Plate Etch Channel Densities

SWEEPING CONDITIONS		DENSITY AFTER AMMONIUM BIFLUORIDE	
FIELD	MAX TEMP.	BEFORE AQUA REGIA	AFTER AQUA REGIA
250 V/cm	500 C	0 CM-2	6.9 CM-2
500	500	0	5.6
2000	500	0	2.4

Note: AT-plate samples were subjected to a saturated solution of ammonium bifluoride at 75°C for 2 hours. Etching in aqua regia was at 65°C for 1 hour. AT-plates cut from control bars had etch-channel densities in the range 30-40 cm-2.

Gold electrodes were examined before and after sweeping, using AES. Depth profiles of the metallized surfaces showed differences depending on sweeping conditions. Figs. 4 (a, b, c, and d) show some features of these surfaces on two bars in which no etch-channels were observed. The following features are noteworthy:

1. At the cathode, gold appears to be diffusing toward the quartz interface, depleting the amount of gold at the metal surface. Depletion is apparent in visual inspections of swept bars.

2. At the anode, an increase of gold is evident at the quartz-metal interface, especially for the case of 500 V/cm. An increase in the thickness of gold at the anode may also be discerned by visual inspection of swept bars.

3. Both electrodes show diffusion of chrome to the surface, with subsequent oxidation.

Concerning this last feature, significant Cr migration has been reported at annealing temperatures of 250° and 450°C. Oxygen, presumably in the form of Cr_2O_3, had diffused to the surface.[12]

Since diffusion of gold into the quartz was suspected, dark-background light-scattering microscopy was used to observe possible colorations in Au-swept bars that exhibited gold depletion at the cathode. Fig. 5 (a and b) are photomicrographs of an x-growth region of a y-bar swept at 50 V/cm and 550°C. The light colored regions shown are numerous planar regions, oriented parallel to yz planes, which scatter blue light. The more intense light-scattering regions were colored purple and transmitted yellow light. This is an indication of the presence of colloidal gold. Neutron activation analysis (NAA) of Au-swept samples showed about 10 ppb Au, while in unswept samples, only 0.05 ppb

gold was detected. The detectivity limit for gold, using NAA is 5 X 10^{-13}. Most of the bar was from the z-growth region which exhibited no decorated planar regions. However, large conical inclusions were sometimes observed near the cathode in the z-growth region (see Fig. 6).

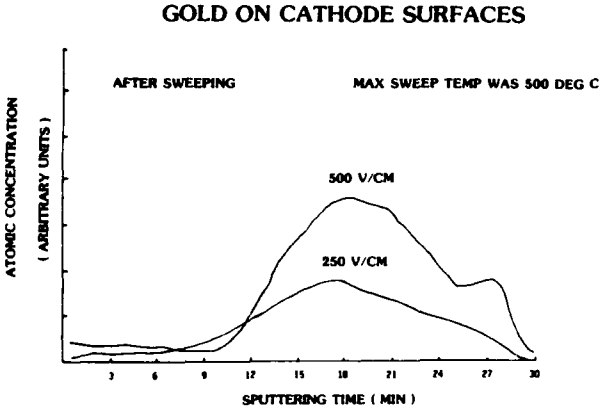


Fig. 4a

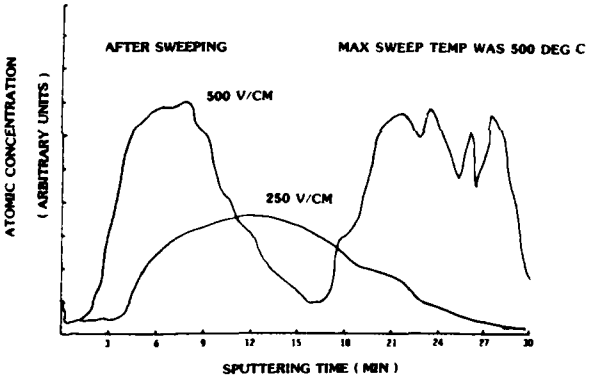


Fig. 4b

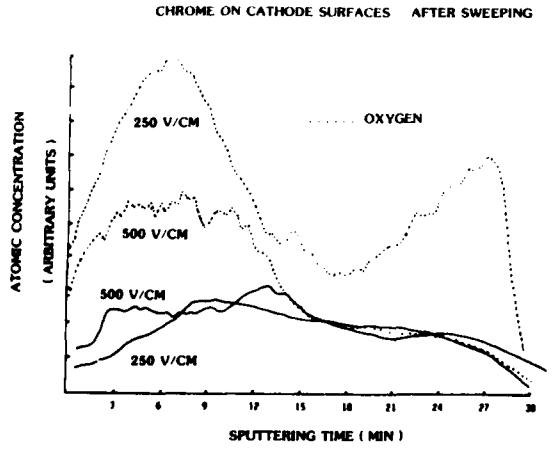


Fig. 4c

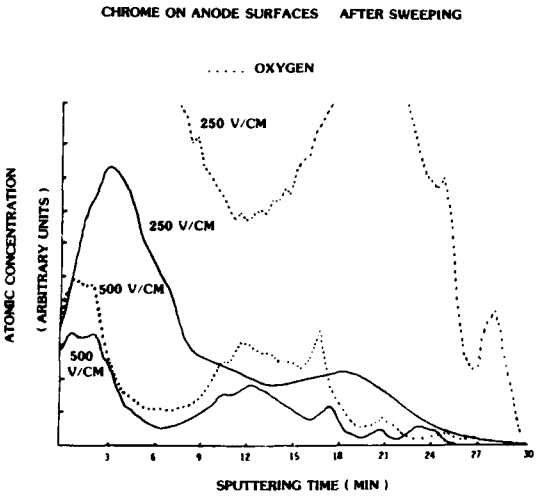


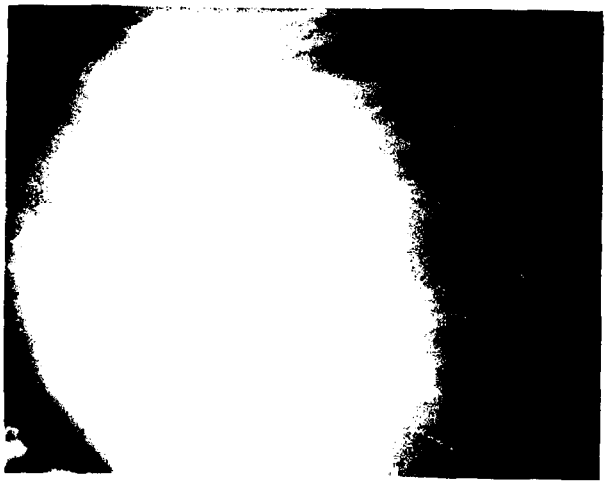
Fig. 4d

AES depth profiles a) gold on cathode surfaces, b) gold on anode surfaces, c) chrome on cathode surfaces, and d) chrome on anode surfaces. Atomic concentrations are in arbitrary units.

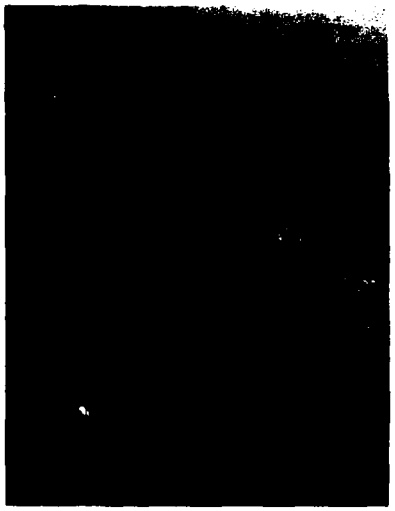
Intense blue-colored inclusions were observed only in the portion of the bar which had received the gold electroplating. Also, gold coloration of inclusions could readily be observed in the z-growth region. After observation of coloration in Au-swept quartz, samples were routinely studied before and after sweeping, using dark-background light-scattering microscopy. Particular types of inclusions were found in bars that were unswept, which correlated with the formation of etch-channels in AT-plates removed from the bars. The inclusions can best be described as long, straight tunnels (see Fig. 7) and irregularly curved lines resembling precipitate decorated dislocation lines (see (Fig. 8). We do not know what impurity is decorating these inclusions - they could be voids. In some cases, the tunnels appear as a series of dashed lines seemingly unconnected (see Fig. 9 (a)). However, on closer examination, it was found that irregularly curved, precipitate-decorated lines connect the straight tunnels (see Fig. 9 (b)). In Au-swept samples, these inclusions were found to scatter blue and red light. In Pt-pressure electrode-swept samples, gray planes were also observed in x-growth regions. Precipitate decorated inclusions scattered more light where they passed through planar regions (see Fig. 10). In unswept samples, tunnels and irregularly curved inclusions were white. If either type of inclusion intersected a surface of a sample to be etched, it was found that, in almost every instance an etch-channel appeared at the point of intersection.

Another inclusion type was identified after sweeping with gold electrodes, they are near-circular plates approximately 1 μm thick with diameters in the 1-2 mm range. These inclusions have a transparent appearance much like the wings of an insect. At the center, there is usually a white conglomeration

showing traces of gold coloration (see Fig. 11 (a)). Similar platelike inclusions were previously reported by Vanfleet et al.,[13] who electrodiffused gold and silver into quartz. Reexamination of the inclusion in Fig. 11(a) six weeks later revealed that the inclusion had a different appearance. The massive central portion and other features of the plate have changed (see Fig 11 (b)). During the six week period, the sample was stored at room temperature.



(a)



(b)

Fig. 5

Light-scattering photomicrograph of a x-growth region of a y-bar swept using evaporated gold electrodes. the light regions are planar bands which are colored blue a) 21 X, b) 200 X.

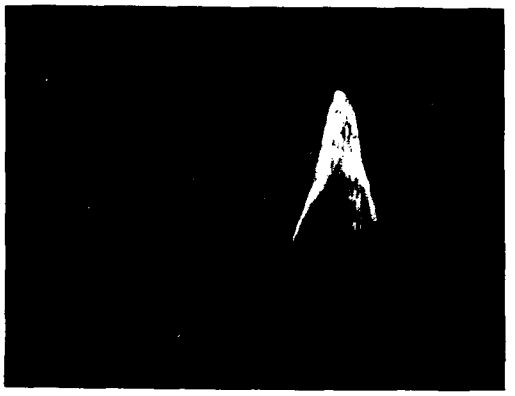


Fig. 6

Photomicrograph of a large conical inclusion which scatters blue light. These inclusions were found near the cathode of a Au-swept sample. 18.75 X.

Fig. 7

Light-scattering photomicrograph of a straight growth tunnel using 60 X magnification.

DISCUSSION OF OBSERVATIONS

The lower initial resistivities observed, using evaporated gold electrodes for sweeping, suggest that the electrodiffusion of gold is contributing to the conduction current. Approximately three times as much charge was transported using evaporated gold electrodes as compared with using Pt-foil electrodes under pressure. Some platinum must also be electro-diffusing, since gray planar regions were observed in Pt-swept quartz. However, the concentrations must be much lower than for evaporated gold, since diffusion is possible only at those points where the Pt-foil actually touches the quartz. Because of sodium interference the detectivity limit for Pt in quartz is 20-40 ppb, using NAA. Pt was not positively detected in any of the samples. Since the final resistivities are about the same, and the activation energies at

the end of sweeping are in the 1.5 to 2.0 eV range, either method is capable of sweeping impurities. The observed activation energy is typical of H^+ ion conduction; therefore, other impurities appear to have been fully swept before the cooldown phase.

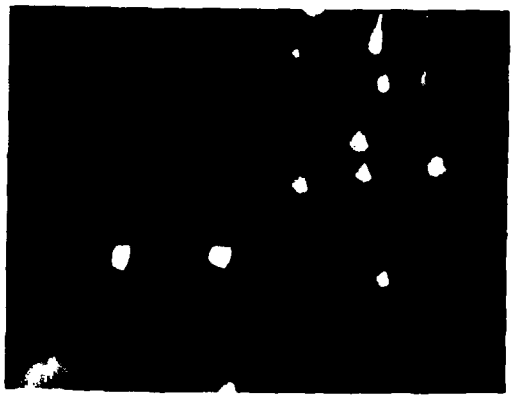


Fig. 8

Light-scattering photomicrograph of irregularly-curved (precipitate decorated) dislocation line leading to a bright spot (over-exposed etch-channel) on the surface. 52.5 X. (Ignore other bright features.)

If gold is electromigrating from cathode to anode, it must be negatively charged. Normally, electrochemical deposition of metals occurs at the cathode. If the charge on a cluster of gold atoms or a molecular-ion complex is negative, the charge could be removed by anodic oxidation and the gold species could be expected to deposit at the anode. This type of behavior has been found to occur with colloidal gold particles in aqueous solution.[14] The negative charge on a colloidal particle in quartz would probably be the result of adsorption of some oxygen-containing species, probably OH^- ions. An oxygen-containing molecular-ion complex is also a possibility. In either case, the discharge process is probably oxidation of these species with the formation of Cr_2O_3, or oxygen, or both. The size of colloidal particles (normally >10 A) diffusing through quartz is difficult to explain in terms of a perfect lattice. However, it is known that the strained region of a dislocation provides a channel of easy diffusion. Dislocations also provide sites for nucleation of precipitates. The resulting decoration of dislocations allows them to be optically observable. The easy migration of a large negatively charged metal colloidal particle or molecular-ion complex would require clusters of vacancies near a dislocation line. A similar explanation has been suggested for the electromigration of negativley charged lead complexes in KCl.[15]

Observation of straight tunnels in quartz was previously reported by Nielson and Foster[6] and by Katz, et al[1]. who also determined that the tunnels were the sites of etch-channels. Nielson and Foster suggested that growth tunnels started at the seed. Katz and Halperin

(a)

(b)

Fig. 9

Light-scattering photomicrograph of a series of 3 growth-tunnels seemingly unconnected in view (a) two shown connected by a fine (precipitate decorated) line in the overexposed view, (b) 52.5 X (ignore other bright features).

extended this hypothesis to include tunnels originating at other defects, such as inclusions in the crystal. The explanation for straight growth tunnels connected by irregularly curved dislocation lines may be related to the advancing growth front, the z-surface, containing macro-steps. Duckett and Lang found that in a solution-grown crystal, where the advancing growth front contained macro-steps, dislocations were deflected as a consequence of the dislocation taking a minimum energy path through each new growth increment.[16]

The conical inclusions which resemble computer plots of 3-dimensional surfaces are similar to the cones of color-center darkening reported by Brown and Thomas.[17] The long thin lines making up the surface of the cone may be gold-decorated dislocation lines. The plate-like inclusions, which change their appearance

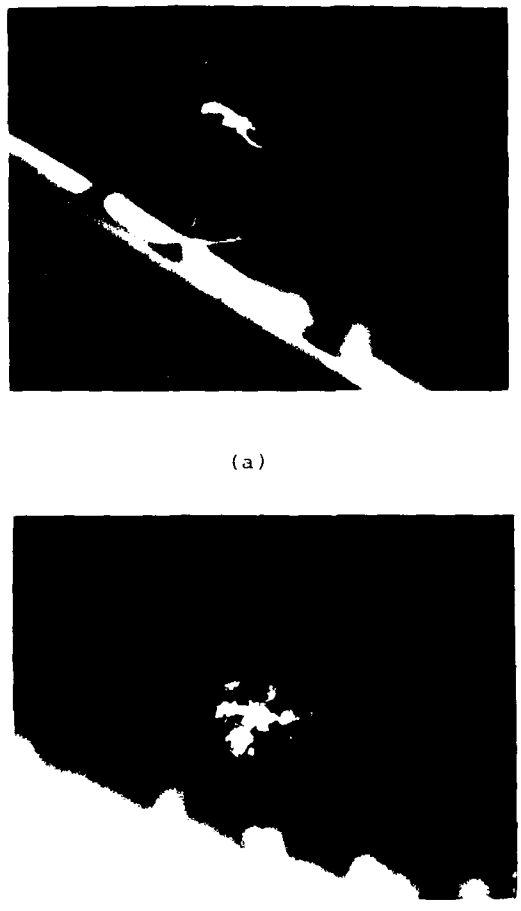
with time, suggest that considerable diffusion with agglomeration is taking place at room temperature and may be causing a fracture to spread out. If gold or other impurities are diffusing at room temperature, this could have an influence on aging in resonators using gold electrodes. Belser and Hicklin have reported that 8 ppb Au diffuses into evaporated Au-plated resonators using a 200°C plating temperature.[18] The adhesion of gold electrodes on resonators fabricated more than 20 years ago appears to be just as weak as it is on resonators fabricated recently.[19] The readily peeled-off gold electrodes do leave a "ghost" behind (i.e., an imprint of where the electrode perimeter had been). Therefore, gold diffusion probably occurs at defect-and high-strain-sites only. Long-term aging tests are underway on resonators cut from our Au-swept samples.

Fig. 10

Precipitate-decorated dislocation inclusions passing through planar bands, which are horizontal, in the x-growth region of a y-bar which was swept using Pt-foil pressure electrodes. The scattered light appeared gray.

CONCLUSION

Sweeping probably lowers the etch-channel density, not only from removal of impurities, but also from the electromigration of impurities into growth tunnels and clusters of vacancies bound to dislocations. Both defect types are responsible for the etch-channels. Some of the swept impurities probably are trapped at these sites. Ernsberger[20] has offered an explanation of rapid etching in quartz. The 100-fold increase in etching rate of basal planes over prism faces was explained on the basis of the number of oxygen atoms in a silica tetrahedron, exposed to the surface, which are not bound to the substructure. The basal plane has two, whereas the prism faces have only one such atom not bound to the substructure. If this is the case, then, a metal-complex swept into a potentially favorable etching site, such as a growth tunnel or a cluster of vacancies associated with a dislocation, may become bound to one or more of these non sub-structure-bound oxygen atoms and render the site unreactive to the etchant.



(a)

(b)

Fig. 11

A large platelike inclusion oriented parallel to y- and z-axes, (a) 22.5 X, and(b) the same inclusion 6 weeks later, 27 X.

Sweeping under conditions of field and temperature, which lead to electro-migration of gold into the quartz, is apparently more effective in modifying defect sites and can reduce the density of subsequent etch-channels to zero. Although we have not yet investigated etch-channels in vacuum swept quartz, the etch-channel-free vacuum-swept quartz reported by Vig et al.[9] may have been etch-channel-free not because of the vacuum, but because of the electromigration of the electrode material into the quartz during the sweeping.

More information may be forthcoming when experiments with other electrode types, and the use of diffusion barrier films to prevent

migration of the electrode metal are completed. No differences were found in the three domestic varieties of cultured quartz used in this investigation. In the future, we will use other domestic and foreign varieties, as well as natural quartz, in our studies.

ACKNOWLEDGMENTS

The author thanks William Washington for the cutting and polishing of samples, Ronald Brandmayr for the etching work, Donald Boyce for the gold evaporations, Donald Fox for the Auger analysis, and Donald Eckart for photographic assistance. Thanks are also due to John Vig for many suggestions and stimulating conversations.

A special acknowledgment to 1st Lt. P. E. Biscaye, whose in-house Memorandum Report, filed more than 25 years ago, on etching and sweeping studies provided ideas and corroborative support.

REFERENCES

1. S. Katz, A. Halperin, and M. Schieber, Proc. 36th AFCS (1982) pp. 193-196.

2. A. R. Lang, J. Appl. Phys. 30 (1959) p. 1748.

3. T. Hanyu, J. Phys. Soc. Japan 19 (1964) p. 1489.

4. A. R. Lang and V. F. Miuscov, J. Appl. Phys. 38 (1967) p. 2477.

5. Fumiko Iwasaki, J. Cryst. Growth 39 (1977) pp. 292 - 298.

6. J. W. Nielson and F. G. Foster, Amer. Min., 45 (1960) pp. 299 - 310.

7. A. A. Shternberg, L. A. Gordienko, and L. I. Tsinober, Soviet Phy. Crystallography 14 (1970) pp. 728 - 732.

8. S. Keerti and A. R. Lang, J. Appl. Cryst. 5 (1972) pp. 72 - 78.

9. J. R. Vig, J. W. LeBus, and R. L. Filler, Proc. 31st AFCS (1977) pp. 131-143.

10. J. J. Martin, R. B. Bossoli, L. E. Halliburton, Brinda Subramaniam, and J. D. West, Proc. 37th AFCS (1983) pp. 164-168.

11. J. G. Gualtieri and J. R. Vig, Proc. 38th AFCS (1984) pp. 42 - 49.

12. J. K. Hirvonen, W. H. Weisenberger, J. E. Westmoreland, and R. A. Meussner, Appl. Phys. Lett. 21 (1972) pp. 37 - 39.

13. H. B. Vanfleet, G. S. Baker, and P. Gibbs, J. Appl. Phys. 34 No. 4 pt. 1. (1963) pp. 891 - 897.

14. A. R. Despic and M. G. Pavlovic, J. Electroanal. Chem. 180 (1984) pp. 31-40.

15. W. J. Fredericks and A. B. Scott, J. Chem. Phys. 28 (1958) p. 249.

16. R. A. Duckett and A. R. Lang, J. Cryst. Growth 18 (1973) pp. 135 - 142.

17. C. S. Brown and L. A. Thomas, J. Phys. Chem Solids 13 (1960) pp. 337 - 343.

18. R. B. Belser and W. H. Hicklin, Proc. 23rd AFCS (1969) pp. 132 - 142.

19. J. R. Vig, private communication, 1985.

20. F. M. Ernsberger, J. Phys. Chem. Solids 13 (1960) p. 347.

DEFECT CENTERS IN IRRADIATED AND SWEPT QUARTZ

Alfred Kahan and Herbert G. Lipson

Solid State Sciences Division
Rome Air Development Center
Hanscom Air Force Base, Bedford, MA 01731

Abstract

Quartz samples are characterized by low-temperature Fourier spectroscopy, and infrared peaks associated with as-grown hydroxide (OH) and aluminum hydroxide (Al-OH) point-defect centers are measured. Crystals were ^{60}Co irradiated to saturation and subsequently swept (electrodiffused) in an air atmosphere. For some crystals the combination of as-grown and other internal hydrogen-ion sources is sufficient to compensate all radiation-induced dissociating aluminum-alkali-metal defect centers. For other samples the concentration of aluminum-ion centers exceeds that of the available hydrogen ions. Radiation-induced dissociated hydrogen ions from as-grown OH defects do not migrate to remote sites to compensate dissociating aluminum-ion centers. The non-destructive infrared technique can be used to determine the substitutional aluminum-ion impurity profile, and for swept crystals the completeness of sweeping, on samples large enough to fabricate high precision resonators. The sample is measured as-received, after sweeping, and after irradiation. Al-OH distributions after sweeping establish the Al^{3+} profile, and those of the swept and subsequently irradiated crystal indicate sweeping effectiveness.

Introduction

In previous publications[1-3] we discussed point-defect models applicable to as-grown, irradiated, and swept (electrodiffused) cultured quartz. A set of defects are formed by an Al^{3+} impurity ion substitutional at a Si^{4+} site. The aluminum ion is compensated either by an interstitial monovalent cation, hydrogen, lithium, or sodium, or by removing an electron from a nonbonding orbital of an oxygen anion and forming a hole. For simplicity these defects are designated as Al-OH, Al-Li, Al-Na, and Al-h (aluminum-hole), respectively. Another set of as-grown defects are hydrogen-impurity-associated centers, and we designate this family as the as-grown OH defect centers. Models for this defect include a water molecule incorporated during crystal growth,[4] or a silicon vacancy compensated with four hydrogen atoms.[5] Both hydrogen-ion-related centers, Al-OH and as-grown OH, give rise to near infrared absorption bands. In Refs. 1-3 we applied low-temperature Fourier infrared absorption spectroscopy and determined the spatial distribution of hydrogen-ion centers along the crystal z-growth axis for as-grown, irradiated, and air or vacuum atmosphere swept quartz.

All results reported in Refs. 1-3 are consistent with the model that ionizing radiation dissociates as-grown OH and forms Al-OH. However, we do not find a one-to-one correspondence between these processes. We observe a substantial increase in Al-OH with irradiation dose after as-grown OH is depleted. This implies the existence of an infrared inactive hydrogen-ion source. The data also indicates that the additional hydrogen-ion defect is nonuniform along the crystal z-growth axis. The samples also exhibit the phenomenon that at two crystal sections the strengths of the corresponding radiation-induced Al-OH bands differ for identical initial as-grown OH concentrations and radiation-induced decreases. A possible explanation was suggested that radiation-released hydrogen ions migrate over large distances, on the order of centimeters, and compensate nonuniformly distributed aluminum ions.

Radiation effects on resonators fabricated from quartz are reduced by minimizing the substitutional aluminum ions during crystal growth. Radiation effects are also reduced by sweeping the crystal in an air or vacuum atmosphere. Sweeping in an air atmosphere provides an unlimited external hydrogen-ion source. During sweeping, the interstitial lithium and sodium impurities of the Al-Li and Al-Na centers are replaced by hydrogen ions to form Al-OH. It is implicitly assumed that for a sample swept in an air atmosphere, for a long time period, the Al-OH distribution corresponds to the substitutional Al^{3+} impurity profile. Sweeping in a vacuum atmosphere removes both alkali-metal and internal hydrogen ions and forms Al-h.

It is then of interest to develop non-destructive characterization techniques and quality control procedures to determine the Al^{3+} profile and sweeping effectiveness, for crystals large enough to fabricate high precision 5 or 10 MHz resonators. It is suggested that low temperature Fourier spectroscopy can be utilized for such purposes. In this paper we continue to investigate hydrogen-ion defect centers related to irradiation and sweeping by examining radiation-saturated crystals after sweeping them in an air atmosphere.

Experimental Procedures

Detailed experimental procedures are described in Refs. 1-3. Quartz samples were irradiated at room temperature in the Rome Air Development Center (RADC) ^{60}Co source. In Ref. 1, we reported accumulated dose effects of 0.2, 0.4 and 0.8 Mrad, and in Ref. 3 we extended the doses to 1.6, 3.2 and 6.4 Mrad. Sweeping was performed at RADC at 500 ^{o}C, in an air atmosphere, with an electric field strength of 1000 V/cm and the sample clamped between platinum foil electrodes placed on the z-faces.[2,6] The sweeping direction is towards the crystal seed, designated as the -z-face, with the anode placed on the +z-face. Infrared transmissions were measured with a Nicolet 170SX Fourier spectrophotometer, with the sample mounted inside a dewar cooled to 85 K. The dewar was moved across the spectrophotometer beam axis and data recorded for several sample positions between the +z and -z-faces. The most prominent infrared bands associated with as-grown OH are at 3348, 3396, 3438, and 3581 cm^{-1}, and those associated with Al-OH are at 3306 and 3366 cm^{-1}. For our purposes we select the peak absorption coefficients of the 3581 and the 3366 cm^{-1} bands to monitor changes in as-grown

 "US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

OH and Al-OH defect centers, respectively. The band strengths are measured with the unpolarized optical beam aligned normal to the z-axis and are weaker than values measured with the beam aligned along the z-axis.

Results and Discussion

Figure 1 shows peak absorption coefficients of the 3581- as-grown OH and 3366-cm^{-1} Al-OH bands for Premium-Q crystal D14-45 as a function of position along the z-growth axis. The curves indicate sequential experiments. The crystal was irradiated to a total dose of 6.4 Mrad, swept in an air atmosphere for one day, and re-irradiated with 6.4 Mrad. The as-received crystal has a relatively low aluminum-ion concentration, <1 ppm as determined from EPR experiments, and low as-grown OH, band peak height between 0.07 and 0.11 cm^{-1}. Figure 1, Curve 1, shows data for the as-received unirradiated crystal, and Curves 2 are results after the 6.4 Mrad irradiation. These curves are reproduced from Ref. 3, Fig. 4. Irradiating the as-received crystal to 0.8 Mrad depletes as-grown OH over most of the sample.[1] Between 0.8 and 3.2 Mrad irradiation the crystal exhibits a substantial increase in Al-OH band strength, indicating the existence of an additional hydrogen-ion source.[3] This implies that hydrogen ions are trapped at various defect sites, and that in addition to "as-grown OH" which gives rise to characteristic infrared bands, the crystal contains other hydrogen-ion centers which are either infrared inactive or vibrate in other spectral regions. Curve 2 also shows that the region 8 to 14 mm from the z-face contains residual as-grown OH, hydrogen-ion centers which are insensitive to radiation.

Irradiating crystal D14-45 to a total dose of 6.4 Mrad saturates Al-OH absorption.[3] This saturation may be due to a scarcity of internal hydrogen-ion sources or may indeed reflect the Al^{3+} profile. In order to clarify this ambiguity the sample was swept for one day in an air atmosphere. Figure 1, Curves 3, show as-grown OH and Al-OH band strengths after sweeping. As-grown OH is restored and band peaks slightly exceed unirradiated values. Measurements on another irradiated D14-45 sample show that even without sweeping as-grown OH bands are re-formed by thermal annealing, 500 °C for 25 minutes, or even at lower temperatures. This is consistent with results of Subramaniam, Halliburton, and Martin.[7] The Al-OH band formed after sweeping extends from the anode to the cathode, and its strength is very close to the 6.4 Mrad irradiation-saturated value. This indicates that although as-grown OH became depleted at 0.8 Mrad, the additional internal hydrogen-ion source is sufficiently strong to compensate all dissociating Al-Li and Al-Na centers. However, the data does not indicate whether the additional hydrogen-ion source itself becomes exhausted during the compensation process. For this particular crystal, radiation-saturation and air-sweeping give similar results, and both reflect Al^{3+} substitutional impurity distribution along the crystal growth axis. The similarity of Al-OH band strength after radiation saturation and air sweeping also implies that for this sample Al-h centers are probably not formed by irradiation.

After sweeping, sample D14-45 was re-irradiated to 6.4 Mrad, Fig. 1, Curves 4. As-grown OH remains essentially unchanged over most of the crystal and the slight decrease in band strength is within experimental error. Al-OH changes are more difficult to interpret. In principle, irradiating a completely swept crystal, in which all Al-Li and Al-Na defects are compensated with hydrogen ions, should not affect Al-OH band strength. Irradiating a partially swept sample should locally increase Al-OH and decreae as-grown OH. The data shows a decrease in Al-OH over the entire crystal, but no corresponding increase in as-grown OH. We are unaware of any defect center, or mechanism, which results in a decrease in Al-OH with irradiation in a swept crystal.

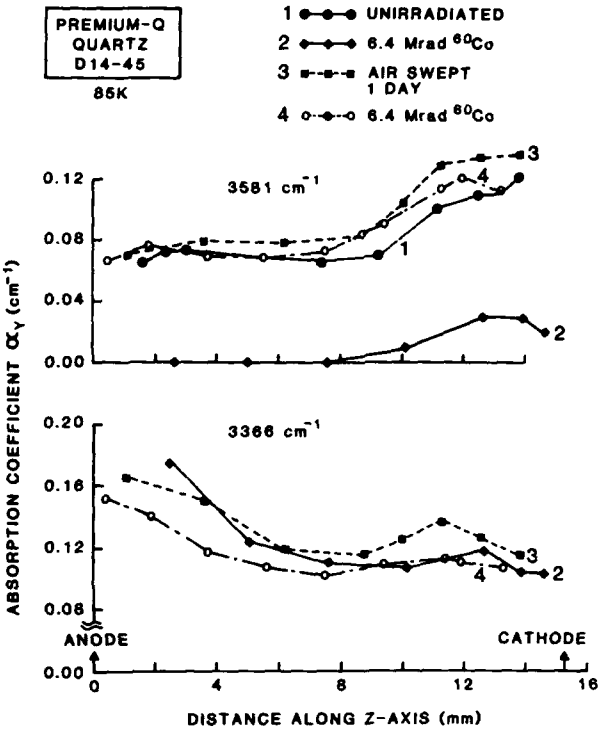


Figure 1. Peak absorption coefficient values for the 3581- and 3366-cm^{-1} bands as a function of position along the z-axis for Premium-Q sample D14-45. The curves indicate sequential experiments. The crystal was irradiated to saturation, swept, and re-irradiated after sweeping.

High-Q crystal E42-21 is a sample with a relatively high aluminum-ion concentration, 6-8 ppm, but low as-grown OH strength, peak heights between 0.08 and 0.12 cm^{-1}. The continued increase of Al-OH after as-grown OH depletes again indicates the presence of an additional hydrogen-ion source.[1] The dissociation sequence of the hydrogen-ion sources is unclear. The additional hydrogen-ion source dissociates either simultaneously with as-grown OH or only after as-grown OH is depleted. We also noted considerable differences in the ratio of Al-OH production to as-grown OH reduction at two different crystal positions for the same initial strength and decrease of as-grown OH. For example, as-grown OH absorption decreased from 0.8 to 0.2 cm^{-1} but Al-OH band strength was 0.12 cm^{-1} at one end and 0.20 cm^{-1} at the other end of the crystal.

One possible interpretation for local variations in Al-OH production is that the governing radiation-induced mechanism is the dissociation of the alkali-metal compensated aluminum-ion center. Accordingly, hydrogen-ion sources dissociate in proportion to Al-Li and Al-Na, and at any stage of irradiation Al-OH distribution is proportional to the Al^{3+} profile. However, this interpretation is inconsistent with the uniform decrease in as-grown OH observed for the initial irradiations, and with the reversal in Al-OH distribution with increasing radiation dose.[3] This data may be explained in terms of radiation-induced diffusion of hydrogen ions. Hydrogen ions dissociate uniformly across the crystal and migrate to compensate nonuniformly distributed Al^{3+} sites. Another possibility is that the governing radiation mechanism is the dissociation of the hydrogen-ion sources, and, as a consequence, at any stage of irradiation Al-OH reflects the combined nonuniformities of the hydrogen-ion sources.

In order to clarify some of these considerations the irradiation-saturated High-Q crystal E42-21 was swept in an air atmosphere for one day. Figure 2 shows the peak absorption coefficients of the 3581- and 3366-cm^{-1} bands as a function of position along the z-growth axis, radiation dose, and sweeping. Curve 1 for the as-received and Curves 2 for the 6.4 Mrad irradiation-saturated crystal are those from Ref. 3, Fig. 5. Curves 3 are results subsequent to sweeping the crystal for one day. As-grown OH is restored across the entire crystal and values exceed initial concentration strengths. This is consistent with the previous observation that radiation effects anneal below 500 ^{o}C and that after sweeping there is a slight increase in overall as-grown OH band strength.

Al-OH band distribution after sweeping for one day indicates that this sweeping time is insufficient to dissociate alkali-metal compensated aluminum-ion centers across the entire crystal. Absorption coefficients in the region up to 6 mm from the anode range between 0.50 and 0.55 cm^{-1}, and decrease gradually to zero between 6 and 10 mm. This is consistent with results of Ref. 2, where it was demonstrated that the Al-OH leading edge moves between the anode and cathode as a function of sweeping time and impurity concentration. The increase in the completely swept portion of the crystal, from the irradiation-saturated value of 0.28 cm^{-1} to 0.55 cm^{-1}, shows that the combined internal hydrogen-ion sources that were available during irradiations were insufficient to compensate all aluminum-ion impurities.

In Ref. 3 we also reported results for Premium-Q crystal BH-A, a sample with relatively high aluminum-ion and as-grown OH concentrations, 11 ppm and band peak heights between 0.3 and 0.4 cm^{-1}, respectively. Based on irradiation and sweeping experiments we concluded that for crystal BH-A residual as-grown OH was due to an excess of as-grown OH compared to Al^{3+}. We thus find that crystal BH-A has an excess of as-grown OH compared to Al^{3+}, D14-45 has sufficient, while E42-21 has insufficient internal hydrogen-ion sources to compensate the radiation-induced dissociating aluminum-alkali-metal centers.

The partially swept sample E42-21 can be used to proof, or disprove, the hypothesis of radiation-induced hydrogen-ion migration over large distances.

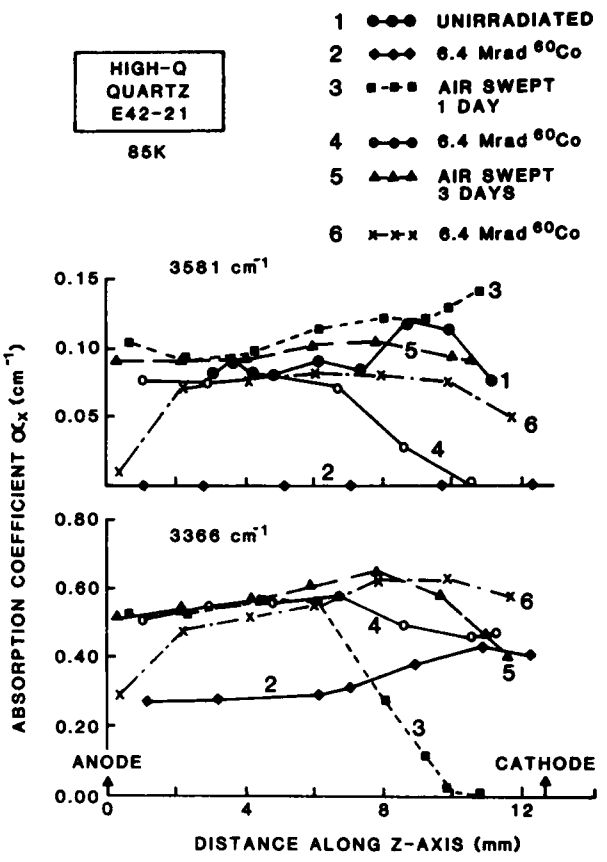


Figure 2. Peak absorption coefficient values for the 3581- and 3366-cm^{-1} bands as a function of position along the z-axis for High-Q sample E42-21. The curves indicate sequential irradiations and sweepings.

Curves 3 show that hydrogen-ion sources are restored to their full as-received values. Al-OH band strength has reached its maximum over one-half of the crystal, while in the unswept and partially swept regions residual lithium and sodium ions have recombined with aluminum ions. If this partially swept sample is now irradiated, and radiation induces hydrogen-ion migration over large distances, then hydrogen ions from the region where Al-OH has reached its maximum should diffuse to the incompletely swept portions of the crystal. In the swept portion of the crystal the results of this mechanism would be indicated by a decrease in as-grown OH and in the incompletely swept portion by an increase in Al-OH in excess of that observed in Curve 2.

Figure 2, Curves 4, show results of the 6.4 Mrad irradiation for the partially swept crystal. As-grown OH is essentially unchanged in the swept portion of the crystal, decreases in the partially swept transition region, and is depleted near the cathode. Al-OH shows no change in the swept portion of the crystal, decreases from 0.55 to 0.45 cm^{-1} in the partially swept transition region, and remains at 0.45 cm^{-1} between 10 mm and the cathode. In the transition region Al-OH band strength is higher than the level obtained for the first 6.4 Mrad irradiation. In the unswept portion of

the crystal Al-OH is essentially at the same strength as that obtained after the first irradiation. These results imply that hydrogen ions do not migrate to remote sites.

In the previous discussion of hydrogen-ion migration over large distances we assumed that Curve 4 for Al-OH does not reflect the Al^{3+} profile, and that there are additional Al-Li and Al-Na centers available for compensation between the center of the sample and the cathode. In order to verify this assumption the sample was swept for an additional 3 days. Curves 5 show that as-grown OH is restored to close to initial values, and that the Al-OH leading edge continues to progress toward the cathode with an increase in band strength over that of Curve 4. This is then consistent with our assumption of residual aluminum-alkali-metal centers in this sample region.

In interpreting the sweeping and irradiation results for sample E42-21 we are faced with a serious difficulty. The differences between of Curves 4 and 5 indicate that after irradiating the sample swept for one day the transition region still contains residual as-grown OH and aluminum-alkali-metal centers. We then expect that in the transition region the 6.4 Mrad irradiation of Curve 4 should locally deplete as-grown OH and form additional Al-OH. Al-OH band strengths should be closer to the distribution depicted in Curve 5 rather than that of Curve 4. These results show that more complex models are necessary to describe defect center interactions associated with sweeping and irradiation.

After the 3 days of sweeping, sample E42-21 was re-irradiated to 6.4 Mrad. Figure 2, Curves 6, show that near the cathode as-grown OH decreases and Al-OH increases to approximately 0.60 cm^{-1}. We assume that the Al-OH distribution of Curve 6 corresponds to the Al^{3+} profile. Overall Al-OH level is high, 0.5 to 0.6 cm^{-1}, but is fairly uniform. Curve 6 also shows that after the last irradiation both as-grown OH and Al-OH decrease close to the anode. The shape of the curves near the anode after irradiation are similar to those obtained for unirradiated crystals swept in a vacuum atmosphere.[2] After irradiating swept crystal D14-45, we also observed a small decrease in Al-OH near the anode. We have difficulty explaining these effects, but it seems that these particular sweepings introduced precursors near the anode which upon irradiation reduce both as-grown OH and Al-OH.

Quality Assurance Methods

After sweeping, the average Al-OH band strengths of samples D14-45, BH-A, and E42-21 are approximately 0.13, 0.55, and 0.55 cm^{-1}, respectively. We assume that these values are proportional to the relative Al^{3+} concentrations. Independent EPR measurements on samples cut from the same crystals yield <1, 11, and 6-8 ppm Al^{3+}, respectively, results which are inconsistent with ratios obtained from infrared measurements. For sample BH-A one day of sweeping was sufficient to produce an Al-OH band across the sample,[3] whereas for E42-21 sweeping is incomplete even after four days. This would imply that the aluminum-ion content of E42-21 is at least equivalent to that of BH-A. These discrepancies between infrared and EPR measurements may be due to the fact that impurity concentrations in quartz are highly nonuniform, and EPR sample sizes are considerably smaller than ones used for infrared investigations.

Critical resonator performance characteristics are determined by the strength and distribution of aluminum-ion-impurity defect centers, and their compensators, along the crystal growth axis. It is then important to establish non-destructive procedures for determining the Al^{3+} profile, and the completeness of sweeping for samples swept in an air atmosphere. Using as-grown OH strength as the complete characterization technique is an insufficient test for crystal quality. The sequence of room and low-temperature irradiations and EPR measurements devised by Halliburton, Koumvakalis, Markes, and Martin[8] for determining the absolute aluminum-ion concentration of unswept and swept quartz can not be performed on samples large enough to fabricate high precision resonators.

We suggest the low-temperature Fourier spectroscopy infrared scanning technique measuring as-grown OH and Al-OH as a possible quality assurance test for determining the spatial variations of these defects. The sample is measured as-received, after sweeping, and after irradiation. The distributions after sweeping establish the Al^{3+} profile, and those of the swept and subsequently irradiated crystal indicate completeness of sweeping. The advantages of this procedure is that it is non-destructive, and that it can be performed on samples large enough to fabricate high precision resonators. For some crystals, if the available hydrogen-ion sources exceed the aluminum-ion concentration, the Al^{3+} profile can be determined from irradiating the as-received crystal. This suggests an intermediate step of irradiating and measuring the sample before sweeping. However, when as-grown OH is depleted with the initial saturating irradiation it is necessary to sweep the sample as the formation of Al-OH may be inhibited by lack of internal hydrogen-ion sources.

References

1. H.G. Lipson and A. Kahan, Proceedings of the 38th Annual Frequency Control Symposium, 1984, [IEEE Document No. 84CH2062-8, IEEE, 445 Hoes Lane, Piscataway, NJ 08854], p. 10.

2. H.G. Lipson and A. Kahan, IEEE Trans. Nucl. Sci. NS-31, 1223 (1984).

3. H.G. Lipson and A. Kahan, J. Appl. Phys. 58, 963 (1985).

4. A. Kats, Phillips Res. Rep. 17, 133 (1962).

5. R.H.D. Nuttall and J.A. Weil, Solid State Commun. 33, 99 (1980).

6. The authors would like to thank A. Armington and M.T. Harris of RADC for performing the sweepings.

7. B. Subramaniam, L.E. Halliburton, and J.J. Martin, J. Phys. Chem. Solids, 45, 575 (1984).

8. L.E. Halliburton, N. Koumvakalis, M.E. Markes, and J.J. Martin, J. Appl. Phys. 52, 3565 (1981).

RADIATION EFFECTS IN QUARTZ: LOW DOSES AND DEFECT PRODUCTION MECHANISMS

L. E. Halliburton, C. Y. Chen, and S. D. Tapp
Department of Physics, Oklahoma State University
Stillwater, OK 74078 USA

Abstract: The production of point defects by low doses of radiation has been investigated in a series of commercially available quartz samples. Electron spin resonance (ESR) was used to identify and monitor the paramagnetic defects produced by doses ranging from approximately 700 rads up to several Mega-rads. The radiation sources were a Cobalt-60 gamma cell (5.5 rads per second) and a Van de Graaff accelerator (1.7-MeV electrons). Both unswept and hydrogen-swept quartz was included in the study, and comparisons were made between samples irradiated at 77 K and at room temperature. From our data, it appears that there is no simple correlation between the response of quartz to low and high levels of radiation. In conclusion, we describe possible mechanisms (both non-thermal and thermally activated) which lead to the radiation-induced dissociation of hydrogen and alkali ions from aluminum sites.

Introduction

Results obtained from low-dose radiation tests on quartz resonators have been recently summarized by Norton et al.[1] Of their conclusions, two held implications as to fundamental defect properties within the bulk of the quartz material. These were (1) the lack of correlation between aluminum content and radiation-induced frequency shifts for low dose level and (2) the failure to predict the low-level radiation sensitivity of a resonator by extrapolating from the high-level radiation data. Although details of resonator processing may be a major factor, these observations nonetheless prompted us to extend our investigations of radiation-induced point defects to lower dose levels in an effort to better understand the responsible mechanisms.

The experimental technique we have used in this investigation is electron spin resonance (ESR) and we have concentrated on the aluminum impurity. We believe that the production of $[AlO_4]^0$ centers will be closely related to the radiation response of a quartz resonator for two reasons, both of which are related to the acoustic loss spectrum of the resonator. First, formation of $[AlO_4]^0$ centers is related to destruction of $[AlO_4/Na^+]^0$ centers. (This latter defect is often referred to as the Al-Na^+ center.) The $[AlO_4/Na^+]^0$ center has a very intense acoustic loss peak at low temperature which has been described in considerable detail by King[2] and by Martin.[3] Second, the $[AlO_4]^0$ center itself has associated acoustic loss peaks.[2,3] Radiation-induced changes in these various loss peaks are directly translated into frequency shifts of the resonator.

Experimental Conditions

Samples

The quartz crystals used in this investigation were obtained from two sources; Thermodynamics of Shawnee Mission, Kansas and Sawyer of Cleveland, Ohio. Table I summarizes the information available about each sample, including aluminum content, growth region, sweeping status, and grower. Of the four Thermodynamic samples involved, two were cut from the z-growth region of the original stone and two were cut from the x-growth region. The six Sawyer samples were z-growth Premium Q material; they were cut from two separate bars (labeled PQI and PQK).

The aluminum content, determined by the ESR method,[4] varied from 3.6 to 14.5 ppm for the z-growth samples and was approximately 36 ppm for the two x-growth samples. This increase in aluminum content for the x-growth region was expected. Two of the Sawyer samples (PQI7 and PQI8) had been swept in a hydrogen atmosphere at Oklahoma State University by Professor Joel J. Martin. These latter two samples contained significantly more aluminum than the other two samples taken from the same PQI bar. However, the swept and unswept samples were cut from quite different locations in the PQI bar and, presumably, the large differences in aluminum content simply reflect a nonuniform distribution of this impurity within the bar.

TABLE I. Description of samples used in this investigation.

SAMPLE	ALUMINUM CONTENT (ppm)	GROWTH REGION	UNSWEPT (U) OR SWEPT (S)	SOURCE
TD5	6.0	z	U	Thermodynamics
TD6	7.5	z	U	
TD7	36.0	x	U	
TD8	36.9	x	U	
PQI4	3.6	z	U	Sawyer
PQI6	4.0	z	U	
PQI7	14.5	z	S	
PQI8	12.4	z	S	
PQK2	8.3	z	U	
PQK4	4.8	z	U	

CH2186-0/85/0000-0259$1.00©1985IEEE

Apparatus

ESR data were obtained from an IBM Instruments (Bruker) ER200D spectrometer. The microwave frequency was 9.45 GHz and the modulation frequency was 100 kHz. All the spectra described in this paper were taken at 77 K. After irradiation, either at 77 K or room temperature, the sample was placed in a finger Dewar of our own design and this, in turn, was inserted in an IBM (Bruker) 4102ST retangular cavity. Sample dimensions were nominally 2 x 3 x 7 mm^3. Defect concentrations were determined by comparison with an $Al_2O_3:Cr^{3+}$ crystal obtained from the National Bureau of Standards.

Two radiation sources were used. The less intense of these was a Cobalt-60 gamma cell with a dose rate of 5.5 rad(Si) per second. The other source was a Van de Graaff accelerator operating at 1.7 MeV in the electron mode with a current of approximately 0.2 microamps/cm^2 on the sample. During a 77-K irradiation, the sample was immersed directly in liquid nitrogen. Then, following the irradiation, it was transferred to a spring-loaded Delrin sample holder while under the liquid nitrogen and then quickly placed in the finger Dewar without any intermediate warming.

Background Information

Defect Models

Aluminum is one of the more pervasive impurities in quartz and also is an important hole trap.[5] However, the role of this impurity is complicated by the fact that a number of aluminum-associated hole centers can be formed in quartz by radiation.[6-8] The more important of these centers are illustrated in Fig. 1. In all cases, they consist of a hole (i.e., a missing electron) trapped in a non-bonding p orbital of an oxygen ion located adjacent to the substitutional aluminum ion. The defect represented in the upper center portion of

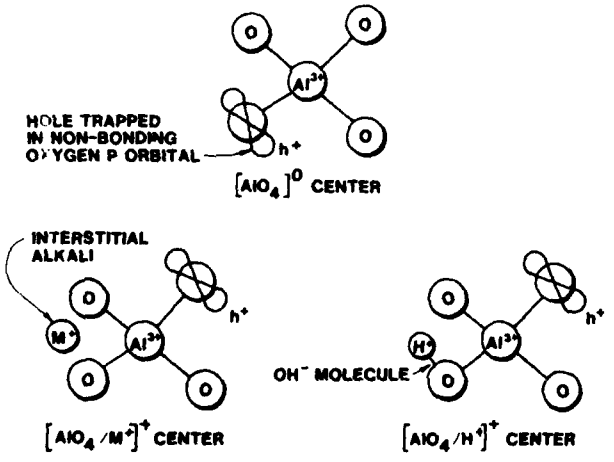


Figure 1. Models of aluminum-hole centers.

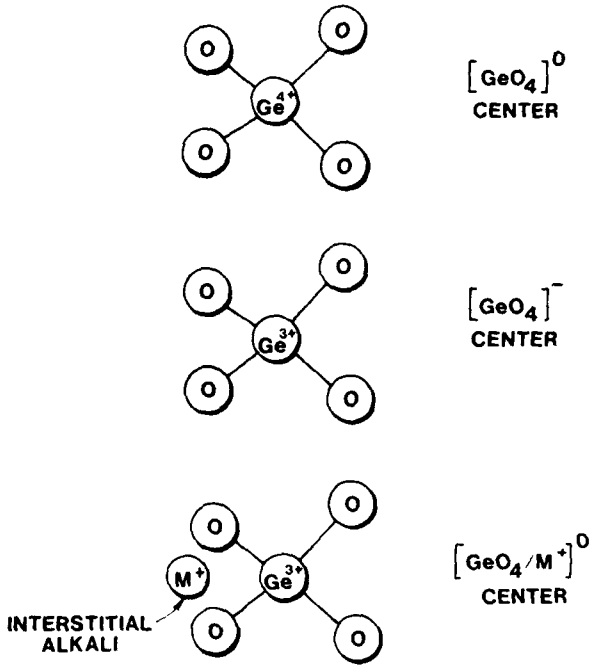


Figure 2. Models of germanium centers.

Fig. 1 has a net charge of zero and is referred to as the $[AlO_4]^0$ center, or more simply as the aluminum-hole center. The defect depicted in the lower left portion of the figure has a net charge of +1 and is known as the $[AlO_4/M^+]^+$ center where M^+ represents either Li^+ or Na^+. The defect depicted in the lower right portion of the figure also has a +1 effective charge; it is known as the $[AlO_4/H^+]^+$ center. These latter two types of defects are commonly referred to as perturbed aluminum-hole centers and they are thermally stable only up to temperatures near 200 K.

Germanium is another impurity known to be an active charge trapping site in quartz.[9,10] However, instead of holes, the germanium ions trap electrons. Figure 2 schematically illustrates three of the more common germanium defect models. The upper diagram represents the neutral substitutional germanium ion and is labeled the $[GeO_4]^0$ center. The defect in the middle diagram is labeled the $[GeO_4]^-$ center and it consists of an extra electron trapped at a substitutional germanium. The lower diagram in Fig. 2 has both an alkali interstitial ion, either Li^+ or Na^+, and the extra electron trapped at the substitutional germanium site. This latter defect is labeled the $[GeO_4/M^+]^0$ center.

It was earlier thought by the authors that commercially available high-quality quartz contained no detectable concentration of germanium; however, the present work has shown that this is untrue and that there is probably up to 0.1 ppm of germanium in most commercial quartz material. Even though the germanium is present, we note that the aluminum remains the dominant substitutional impurity in all quartz and is at least one-to-two orders of magnitude greater in concentration than the germanium.

ESR Spectra

The ESR spectra which we have monitored during the present investigation are shown in Fig. 3. A derivative of the microwave absorption is represented by the vertical axis while the horizontal axis corresponds to the linearly increasing "static" magnetic field. In all cases, this static magnetic field was aligned parallel to the c axis of the crystal in order to simplify the resulting spectrum. The upper trace in Fig. 3 is the ESR spectrum of the $[AlO_4]^0$ center. It is spread over a 30-gauss region and consists of a large number of hyperfine lines, all of which arise from the interaction of the unpaired electron with a single aluminum nucleus (100% abundant, I = 5/2). The middle trace in Fig. 3 is the ESR spectrum of the $[AlO_4/H^+]^+$ center and the lower trace is the ESR spectrum of the $[AlO_4/Li^+]^+$ center. A six-line hyperfine pattern arising from the aluminum (i.e., I = 5/2 nucleus) is easily seen in the latter two spectra. Also, the ESR spectra of the $[GeO_4]^-$ center are present in the middle trace; they consists of the two weaker lines at the high field (i.e., the right) side.

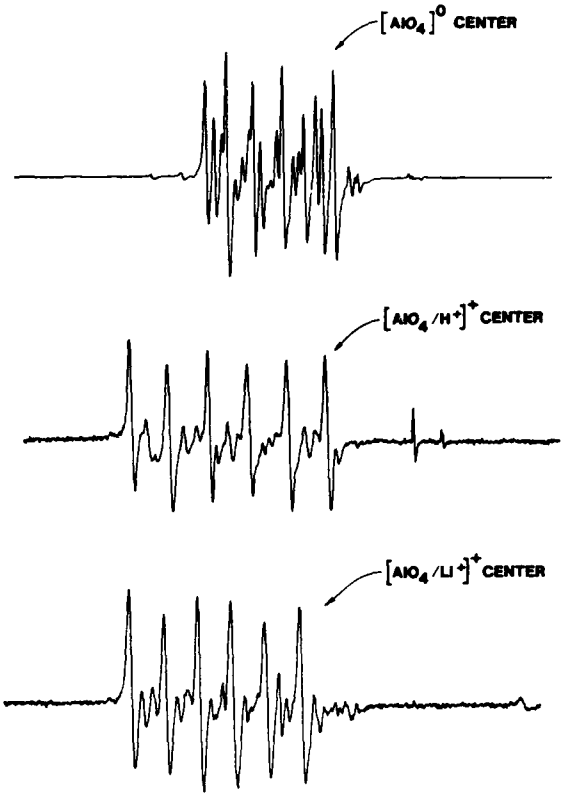


Figure 3. Aluminum-hole center ESR spectra.

Experimental Results

We present a series of defect production curves in this section. In each case, the sample was subjected to a sequence of irradiations; after exposure to each increment of radiation, the sample was placed in the ESR spectrometer where the defect concentration was monitored. Thus, each of the data points in our production curves represents the net defect concentration resulting from the total dose accumulated at that particular stage in the irradiation sequence. For most samples, this sequential course of events proceeded until saturation of the defect's concentration was observed. When the irradiations were done at 77 K, the sample was not allowed to warm to room temperature until the entire sequence was completed.

77-K Irradiation

Figure 4 shows the defect production curves for the unswept z-growth sample TD6 as a function of irradiation at 77 K in the ^{60}Co gamma cell. The growth curves for the $[AlO_4]^0$ centers and the $[GeO_4]^-$ centers are closely approximated by single saturating exponentials. Both defects saturate at about 30 krad. However, the production of $[AlO_4/Li^+]^+$ centers is quite different. This latter defect exhibits a very rapid growth during the first 3 krad of radiation, then reaches a maximum, and finally drops about 30% to an equilibrium concentration as the radiation accumulates. At the present time, this unique 77-K production behavior of the $[AlO_4/Li^+]^+$ center is not understood in terms of discreet atomic processes.

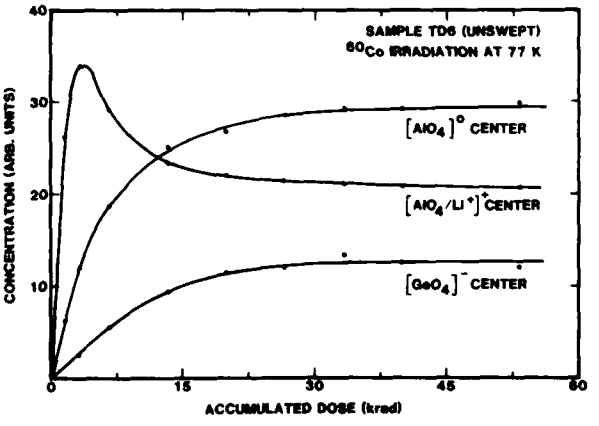


Figure 4. Production of defects in unswept quartz by ^{60}Co radiation at 77 K.

The production of $[AlO_4]^0$ centers at 77 K by ^{60}Co radiation is compared in Fig. 5 for four unswept quartz samples. A single saturating exponential provides a good description

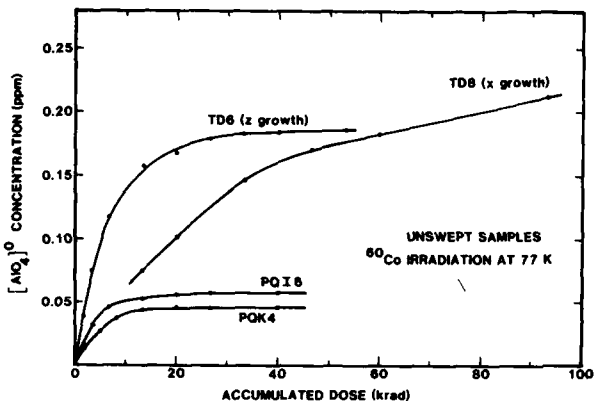


Figure 5. Comparison of defect production at 77 K in four unswept quartz samples.

of the growth curves for the three z-growth samples. This, however, does not appear to describe the x-growth sample. Also, we found that the saturation concentrations of the $[AlO_4]^0$ centers did not scale directly with the known aluminum contents for the three z-growth samples. This lack of correlation with aluminum is even more dramatic when the x-growth sample (TD8) is compared with the z-growth sample (TD6) in the dose range up to 40 krad. Despite having nearly five times more aluminum, the x-growth sample has about half the concentration of $[AlO_4]^0$ centers over this range. Above 40 krad, the $[AlO_4]^0$ center concentration continues to grow for the x-growth sample but remains constant for the z-growth sample. The results presented in Fig. 5 suggest that other factors, such as the nature and concentration of unobserved electron traps, may be as important as the aluminum content in determining the production characteristics of $[AlO_4]^0$ centers during 77-K irradiations.

Room-Temperature Irradiation

The production of $[AlO_4]^0$ centers as a function of ^{60}Co radiation at room temperature is shown in Fig. 6. Four unswept samples are compared in this figure; three of these represent z-growth material and one was taken from x-growth material. The $[AlO_4]^0$ centers grow rapidly at lower dose in each of the samples, but the final saturation values for the defect concentrations are quite different for the four samples. In contrast to the irradiations at 77 K described earlier in Fig. 5, the room-temperature irradiations result in a correlation, at least to first order, between the aluminum content of a sample and its saturation concentration of $[AlO_4]^0$ centers reached after lengthy irradiation. This is evident from Fig. 6 where the aluminum content is included after the sample label for each of the curves.

A surprising result from these room-temperature irradiations was the low values of the saturation concentrations for each of the samples. In the case of the three z-growth samples in Fig. 6, the final concentration of $[AlO_4]^0$ centers is only of the order of 6% or less of the total aluminum content. In previous experiments[5] where defect production was achieved by means of high-energy electrons from a Van de Graaff accelerator, the saturation value of $[AlO_4]^0$ centers was between 10% and 70% of the total aluminum.

Figure 7 is an enlargement of the low-dose portion of Fig. 6. These expanded scales provide a more detailed view of the initial growth of the $[AlO_4]^0$ centers. Again, we see that the total aluminum content is not a good predictor of the initial radiation sensitivity of a given sample. For example, sample TD7 has a factor of ten more aluminum than sample PQI4, but the $[AlO_4]^0$ center concentrations for these particular samples are only a factor of two different during the first three krads of accumulated radiation.

A comparison of defect production in swept and unswept samples is presented in Fig. 8. Both samples were cut from the same bar and were exposed to ^{60}Co radiation at room temperature. The initial growth rate for the $[AlO_4]^0$ centers is significantly greater in the unswept sample PQI4 than in the hydrogen-swept sample PQ17. This effect is even more striking when one considers that PQI7 has a factor of four higher aluminum content than PQI4. These results are in agreement with earlier suggestions that interstitial alkali ions can act as electron traps.[11]

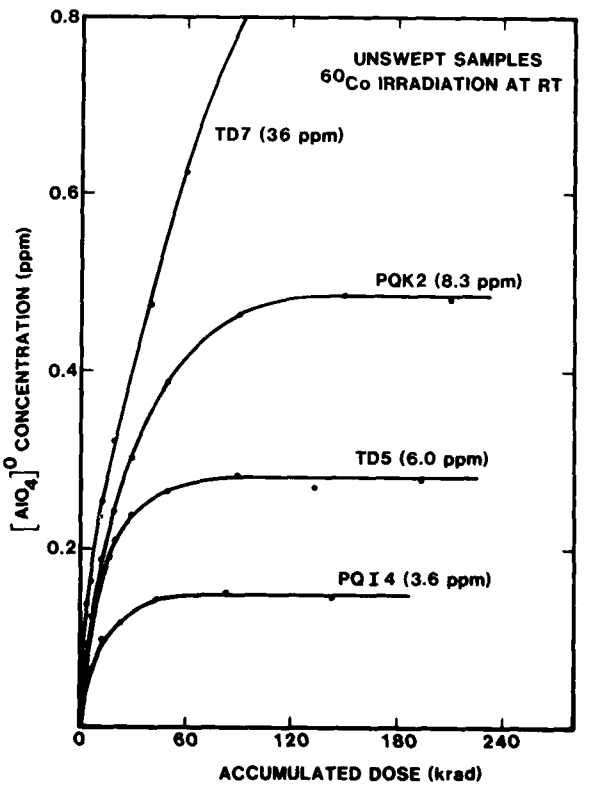


Figure 6. Comparison of defect production at room temperature in four unswept samples.

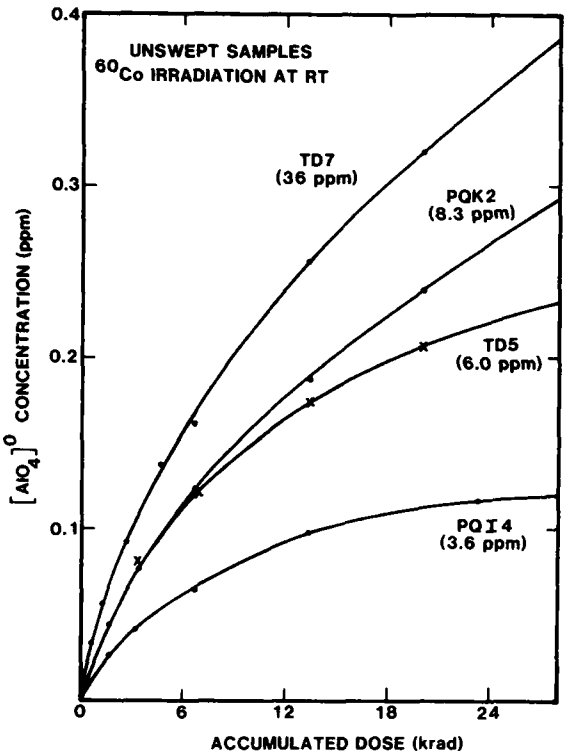


Figure 7. Expanded view of the low-dose data shown in Figure 6.

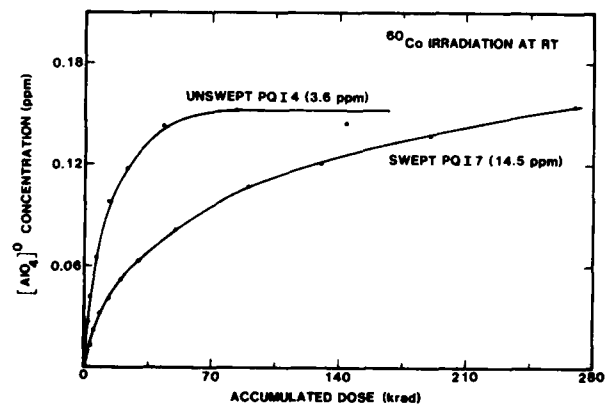


Figure 8. Comparison of defect production at room temperature in swept and unswept samples.

Discussion

Radiation Response of Oscillators

It is well known that the formation of $[AlO_4]^0$ centers by radiation is directly related to changes in the acoustic loss spectrum of quartz resonators. Thus, the production curves for $[AlO_4]^0$ centers that are presented in this paper should provide insight to the expected frequency shifts of quartz resonators exposed to low radiation doses. In general, we have found that extreme care must be exercised in predicting the radiation-induced concentrations of $[AlO_4]^0$ centers and other defects when the quartz is exposed to low doses. This is especially true for doses below approximately 10 krads. It appears that knowledge of the total aluminum content of a sample is not sufficient information from which to reliably predict the final saturated value of the $[AlO_4]^0$ concentration after irradiation, and that other factors such as the nature and concentration of available electron traps are equally important.

The data in this paper also suggest the possibility that a dose-rate dependence may be a feature of defect production in quartz. However, much additional work needs to be done before reaching such conclusions. If there is a dose-rate dependence, then care will need to be exercised in the design of radiation testing procedures for resonators.

Defect Production Mechanisms

The effects of 77 K and room temperature irradiations are compared in Fig. 9. The two samples, PQI7 and PQI8, had similar amounts of aluminum and they were both hydrogen swept. In this experiment, the Van de Graaff accelerator was used instead of the ^{60}Co cell. The dose rate of the accelerator was approximately 40 rad/sec and the arbitrary dose units in Fig. 9 are nearly equal to kilorads. The growth rate for the $[AlO_4]^0$ centers is much less at 77 K than at room temperature, and this suggests that temperature-dependent defect production mechanisms may be involved. Since both of these samples were hydrogen swept, the radiation must lead to the dissociation of hydrogen from aluminum sites in the quartz.

The most interesting of the radiation-related questions in quartz concerns the mechanisms for dissociation of either the alkalis or the protons from the aluminums. In the remaining paragraphs, we will outline

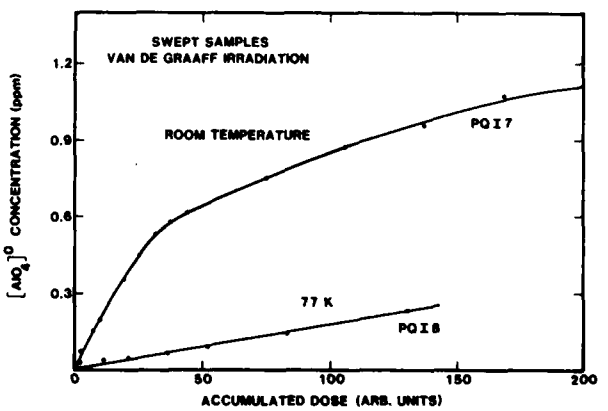


Figure 9. Comparison of defect production at 77 K and room temperature for swept samples.

possible sequences of events that lead to the separation of the aluminums and their charge compensators. Figure 10 provides a schematic representation of the thermal processes. The alkali case is considered on the left side of the figure. Initially, the Li^+ and the Al^{3+} ions are bound, but isolated, in the lattice. Then, during irradiation, a hole becomes trapped at an oxygen adjacent to the aluminum and forms the $[AlO_4/Li^+]^+$ center, as illustrated at the top left in Fig. 10. The middle and lower left portions of Fig. 10 show the potential wells for the Li' and the hole. We assume the hole is in a deeper well than the alkali. Below 200 K, there is insufficient thermal energy for either the alkali or the hole to escape from their potential wells. At 200 K, sufficient thermal energy is available for the alkali, but not the hole, to start escaping. At room temperature, either the alkali or the hole can thermally escape from the aluminum site.

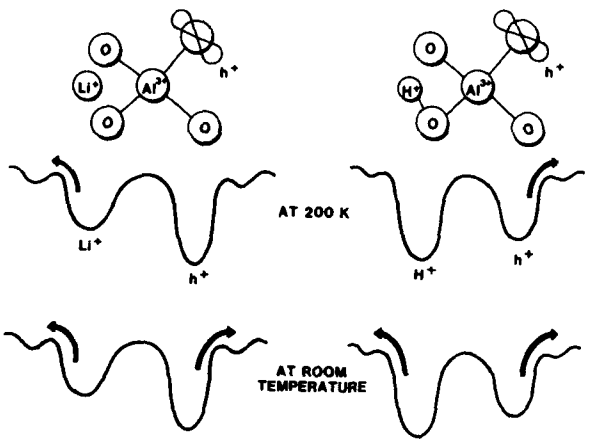


Figure 10. Thermally activated mechanisms for dissociation of alkalis and protons from substitutional aluminum.

The right side of Fig. 10 shows how the hydrogen can escape from the aluminum. As in the case of the alkali, the hydrogen and the aluminum are initially bound, but isolated, in the lattice. Then radiation forms the $[AlO_4/H^+]^+$ centers, as illustrated at the top right of Fig. 10. The potential wells for the proton and the hole are shown in the middle and lower right portions of the figure. Now, we assume that the well for the proton is deeper than that for the hole. At 200 K, there is sufficient thermal energy for the hole to escape from the aluminum, but not the proton. However, at room temperature either the hole or the proton can thermally escape. As the proton thermally diffuses away, we are left with an $[AlO_4]^0$ center. It is important to note that this last mechanism is only operative near or above room temperature.

Since the 77-K radiation produced $[AlO_4]^0$ centers in the swept sample (see Fig. 9), there must be a non-thermal mechanism for dissociating the proton from the aluminum. For this purpose, we propose the process described in Fig. 11. Basically, we are starting with the $[AlO_4/H^+]^0$ center in line (a). Then, upon irradiation, the $[AlO_4/H^+]^+$ center is formed as in line (b) of the figure. However, some of the electrons will return to this defect as illustrated in line (c). Most of these returning electrons will recombine with the hole and restore the original $[AlO_4/H^+]^0$ centers. In some cases, the returning electron will go to the OH^- side of the defect and form an OH^{2-} molecule, as shown in line (d). This OH^{2-} molecule will be very unstable and will immediately dissociate into a hydrogen atom and an oxygen ion as in line (e). The hydrogen atom will thermally diffuse away and the final result will be the $[AlO_4]^0$ center, as shown in line (f) of Fig. 11.

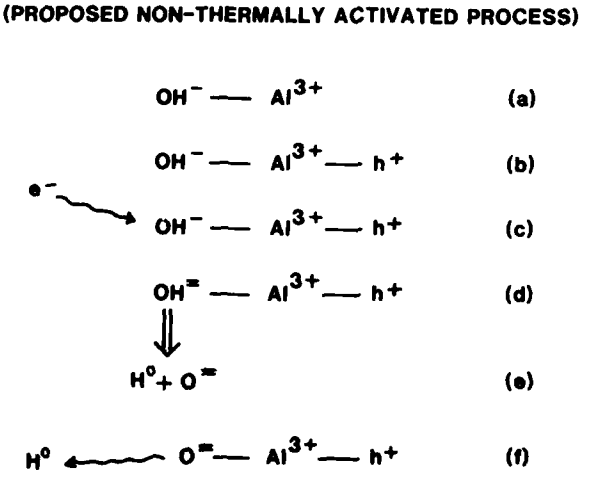


Figure 11. Proposed non-thermally activated process for dissociation of protons from aluminum.

Acknowledgements

The authors wish to thank Thermo Dynamics, Inc. for generously supplying quartz bars for use in this investigation and also to thank Professor Joel J. Martin for sweeping assistance. Helpful discussions with Dr. M. G. Jani concerning interpretations of the data are especially appreciated. This work was supported in part by the U. S. Air Force under Contract No. F19628-83-K-0007 with the Rome Air Development Center, Solid State Sciences Division, Hanscom Air Force Base, and monitored by Dr. Alton F. Armington.

References

[1] J. R. Norton, J. M. Cloeren, and J. J. Suter, Proceedings of the 38th Annual Symposium on Frequency Control, p. 63, 1984.

[2] J. C. King, Bell Syst. Tech. J. 38, 573 (1959).

[3] J. J. Martin, J. Appl. Phys. 56, 2536 (1984).

[4] M. E. Markes and L. E. Halliburton, J. Appl. Phys. 50, 8172 (1979).

[5] L. E. Halliburton, N. Koumvakalis, M. E. Markes, and J. J. Martin, J. Appl. Phys. 52, 3565 (1981).

[6] R. H. D. Nuttall and J. A. Weil, Can. J. Phys. 59, 1696 (1981).

[7] R. H. D. Nuttall and J. A. Weil, Can. J. Phys. 59, 1709 (1981).

[8] R. H. D. Nuttall and J. A. Weil, Can. J. Phys. 59, 1886 (1981).

[9] J. H. Mackey, Jr., J. Chem. Phys. 39, 74 (1963).

[10] J. Isoya, J. A. Weil, and R. F. C. Claridge, J. Chem. Phys. 69, 4876 (1978).

[11] M. G. Jani, L. E. Halliburton, and E. E. Kohnke, J. Appl. Phys. 54, 6321 (1983).

RADIATION EFFECTS IN THE ACOUSTIC LOSS SPECTRA OF AT-CUT QUARTZ CRYSTALS

J. J. Martin, Ho B. Hwang, and H. Bahadur
Department of Physics
Oklahoma State Universty
Stillwater, OK 74078

SUMMARY

As-grown synthetic quartz contains substitutional aluminum which requires charge compensation. The charge compensators are intersitial Li^+ and Na^+ ions. Protons trapped at as-yet unidentified sites form the OH^- related growth-defects which are responsible for the infrared absorption bands observed at low temperatures. Of these centers, only the Al-Na^+ center with a strong loss peak at 53 K and a weak peak at 135 K produce observable acoustic losses at temperatures below 400 K. If the sample is irradiated at temperatures above about 200 K, the alkali ion is released from the aluminum site and is subsequently trapped elsewhere in the sample. The Al-hole and Al-OH^- centers are formed with the proton coming from the growth-defect sites. The acoustic loss spectrum then shows a reduced 53 K Al-Na related peak and additional loss peaks at 23 K , 100 K and 135 K. These additional loss peaks decay with the Al-hole center upon annealing. We have measured the reduction of the Al-Na center and the production of the 23 K loss peaks as a function of radiation dose in unswept and Li-swept AT-cut crystals fabricated from synthetic quartz. Irradiations were carried out using either ^{60}Co γ-rays or 1.75 MeV electrons from a Van de Graaff accellerator. The Al-OH^- center was tracked by parallel infrared absorption measurements on optical samples taken from the same bars of quartz. The 53 K Al-Na^+ center decreased rapidly; by 0.5 Mrad it was lowered to less than 5 % of its initial strength. The initial production rate of the Al-OH^- center was significantly larger than the growth rate of by the 23 K loss peak. The Al-OH^- center approached saturation near 0.2 Mrad while the 23 K peak saturated near 1 Mrad. Preliminary EPR measurements of the production of the Al-hole show that it grows much more rapidly than the defect responsible for the 23 K acoustic loss peak. A new radiation-induced alkali-related loss peak was observed. This peak occurs at 305 K in Li-swept samples and near 340 K in samples which contain significant amounts of sodium. The 340 K peak is also present in as-Na-swept samples. The height of both the 305 K Li-related and 340 K Na-related peaks reaches a maximum near 50 krad and then decreases; consequently the trap involved in this defect center cannot be the final alkali trap. We have not yet been able to identify the defect center responsbile for these two peaks.

INTRODUCTION

Alpha-quartz is used in a wide variety of precision electronic devices where radiation sensitivity is undesirable. It is well known that crystal controlled oscillators may exhibit transient and steady-state frequency and Q shifts when exposed to ionizing radiation[1-4]. Early results obtained by King[5] and other investigators[6-9] showed that these effects were associated with the presence of impurities. As-grown synthetic quartz contains substitutional aluminum which is charge compensated by intersitial Li^+ and Na^+ ions. Protons trapped at as-yet unidentified sites form the OH^- related growth-defects which are responsible for the infrared absorption bands observed at low temperatures. Of these centers,only the Al-Na^+ center with a strong loss peak at 53 K and a weak peak at 135 K produce observable acoustic loss at temperatures below 400 K[10]. Irradiation at temperatures above about 200 K releases the alkali ion from the aluminum site and it is subsequently trapped elsewhere in the sample[11-14]. The Al-hole and Al-OH^- centers are formed with the proton coming from the growth-defect sites. The acoustic loss spectrum then shows a reduced 53 K Al-Na^+ related peak and additional loss peaks at 23 K, 100 K, and 135 K. These additional loss peaks decay with the Al-hole center upon annealing[10]. The intersitial alkali ions are in the relatively large c-axis channel and at high temperatures can move along the channel under an applied electric field. King[5], and later Kats[15] and Fraser[16] used this technique to "sweep" hydrogen and specific alkalis into quartz. Replacing the original alkalis with hydrogen has been shown to improve the radiation hardness of quartz oscillators.[8]

The identification and production of both growth- and radiation-induced defects which affect the performance of quartz resonators is an important part of our project. We report here a production study of the 23 K loss peak center and the Al-OH^- center and the reduction of the Al-Na^+ center as functions of radiation dose in unswept and Li-swept 5 MHz AT-cut crystals fabricated from synthetic quartz. IR absorption mesurements were used to track the Al-OH^- center. We also report on a new radition-induced alkali-related loss peak first reported by Koehler and Martin[17]. This peak occurs at 305 K in Li-swept samples and near 340 K in samples which contain a significant amount of sodium. The 340 K peak is also present in as-Na-swept samples.

EXPERIMENTAL PROCEDURE

Samples for this study were 5 MHz 5th overtone AT-cut Warner design[18] crystals fabricated from a pure-Z growth Sawyer Premium Q bar given an in-house designation PQ-E and 5 MHz 3rd overtone crystals fabricated from a Toyo Supreme Q designated SQ-B. Samples from both bars have been extensively study at Oklahoma State University using ESR[11,13], IR, and acoustic loss [10,19]. These investigations show that both bars are of high-quality but they contain somewhat more aluminum (10-15 ppm) than the average Premium Q material (5-8 ppm). Consequently, aluminum related effects are more readily observed. The as-received Toyo material has a substantial amount of intersitial sodium while the Sawyer material has almost no sodium[10,19]. The acoustic loss was determined from the equivalent series resistance of the crystal which was measured using a transmission technique. The log-decrement method was used to study the loss in the 300 K to 370 K tempearature region. The samples were irradiated by placing the crystal in a 20 krad/hr ^{60}Co γ-cell. A Y-plate optical sample was cut from bar SQ-B for the production study of the Al-OH^- center. All IR measurements were made at liquid nitrogen temperature. Later, a section of the optical sample was given to Chen and Halliburton for EPR studies of the Al-hole center production.

CH2186-0/85/0000-0266$1.00©1985IEEE

RESULTS AND DISCUSSION

Figure 1 shows the acoustic loss of the unswept SQ-B sample in the as-received condition, after 50 krad, and after 652 krad irradiations. Approximately one third of the 10 to 15 ppm aluminum in SQ-B material is charge compensated by an interstitial Na^+ ion with the remainder compensated by a Li^+. The large loss peak at 54 K is caused by this large Al-Na^+ center concentration. As the sample is irradiated the Al-Na^+ (and Al-Li^+) centers are converted into Al-OH^- and Al-hole centers. Figure 1 shows the growth of the 23 K, 100 K and 135 K radiation induced loss peaks at the expense of the 54 K Al-Na^+ center. The 100 K peak was first reported by King and Sander[1] who assigned it to the Al-hole center. Later, Martin[10] reported that all three radiation induced peaks followed the same annealing curve as the Al-hole center. Martin also showed that the anneal of the Al-hole center is accompanied by the partial return of the alkali ions to the aluminum site. The small radiation produced 135 K peak should not be confused with the different 135 K peak associated with the Al-Na center. Since the radiation related peak is also present in Li-swept crystals it must have a different origin.

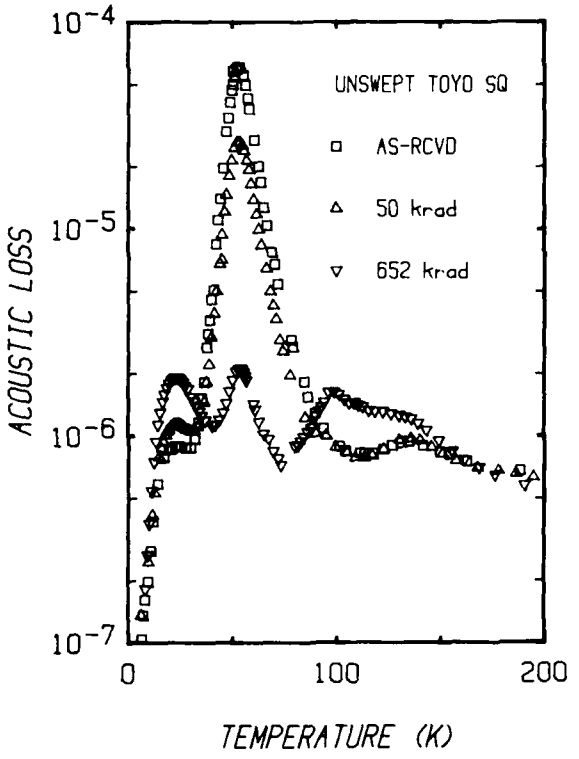


Fig. 1. The acoustic loss versus temperature curves are shown for the unswept SQ-B crystal in the as-received condtion and after several irradiations.

Figure 2 shows the production curves for the Al-OH^- center and the 23 K loss peak and the reduction of the Al-Na^+ peak versus radiation dose. By 0.2 Mrad the Al-Na^+ center has decreased to 12 % of its initial concentration while the Al-OH^- has grown to about 90 % of saturation; the 23 K peak has reached 55 % of its final height. At about 0.2 Mrad the 3581 cm^{-1} OH^- growth-defect IR band disappeared into the background; the Al-OH^- continued to grow slowly out to 1 Mrad. This result is similar to those of Kahan and Lipson[20]

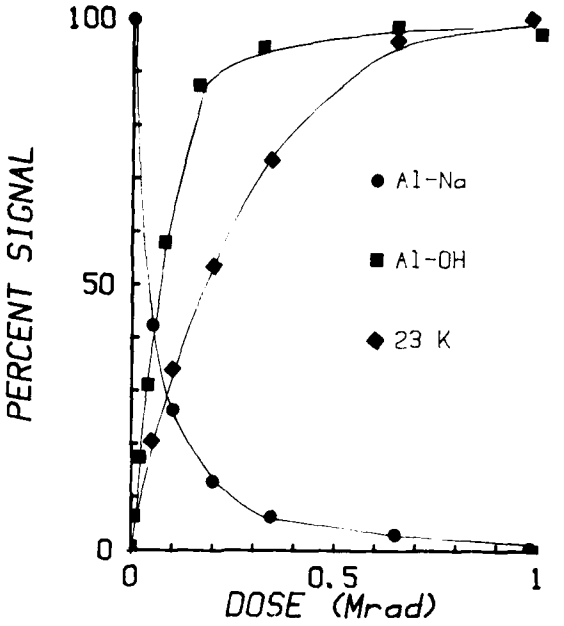


Fig. 2. The production of the 23 K loss peak center and the Al-OH^- center are shown for the SQ-B sample. The reduction of the Al-Na^+ center is also shown.

who reported the contiued growth of the Al-OH^- center after the depletion of the growth-defect bands. Since the strong Al-Na^+ center masked the initial growth of the 100 K loss peak we did not track it in this study. The curve shown in Fig. 2 are for a γ-irradiation; a repeat of the study using 1.75 MeV electrons gave essentially the same results. The growth curves for boththe Al-OH^- and 23 K loss peak centers seem to follow the equation

$$n = N(1 - \exp(KD))$$

where n is the concentration, N is the saturated value, K values of 9/Mrad and 5/Mrad respectively, and D is the dose in Mrad. Figure 4 compares the growth of the 23 K loss peaks in the unswept and in the Li-swept SQ-B crystals. Both samples show the same general behavior except that the peak strength is somewhat larger in the Li-swept than in the unswept sample. The larger 23 K peak in the Li-swept sample may be caused by the presence of lithium or it may be due to a variation in sample impurity content such as aluminum. The EPR results on the Al-hole center of Chen and Halliburton[21] are also shown in Fig. 3. Their results as well as those of Halliburton, Chen and Tapp[22] suggest that the Al-hole center grows much more quickly upon irradiation than does the center responsible for the 23 K loss peak. While these results must be considered preliminary it appears that the Al-hole center is not responsible for the 23 K acoustic loss peak.

Figures 4 and 5 show the 305 K and 340 K radiation-induced loss peaks first reported by Koehler and Martin[17] in Li-swept and unswept Toyo Supreme Q quartz respectively. Our earlier studies[12,14] seem to have missed these two loss peaks because of their unusual radiation behavior. Figure 6 shows the acoustic loss of a Na-swept Sawyer Premium Q sample PQ-ER11 of the 300 K to 370 K temperature range. Here the 340 K loss peak is present in the as-sodium swept condition. The results for a series of irradiations are also shown in Fig. 6; the loss peak grows quickly reaching a maximum strength at 40 krad then falls off

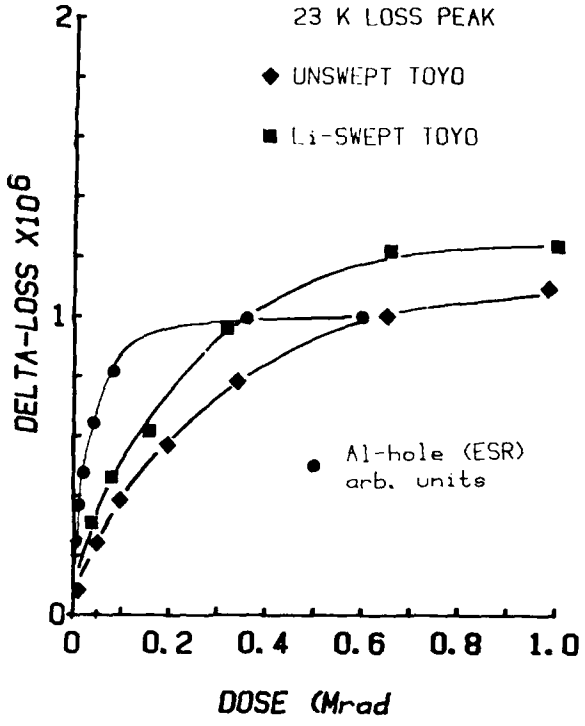


Fig. 3. The production of the 23 K loss peak center in unswept and Li-swept SQ-B crystals is compared with the production of the Al-hole center. The Al-hole center grows more quickly than the 23 K peaks.

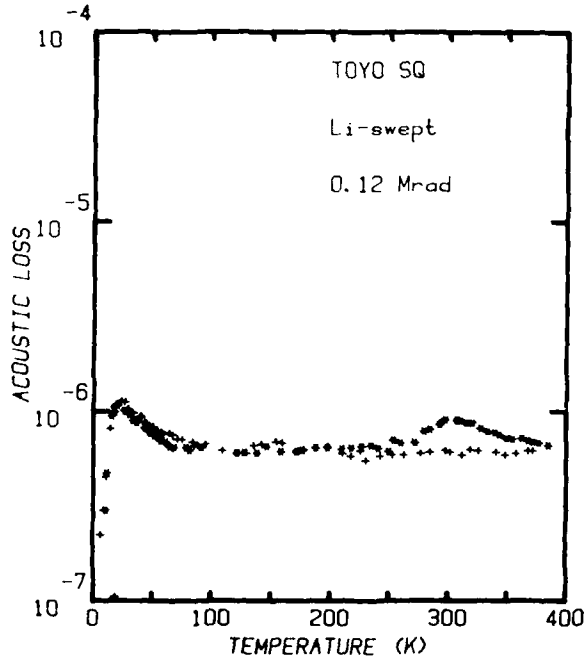


Fig. 4. The acoustic loss spectrum of a Li-swept SQ-B crystal is shown in the as-swept condition and after irradiation. The irradiation produced Li-related peak at 305 K.

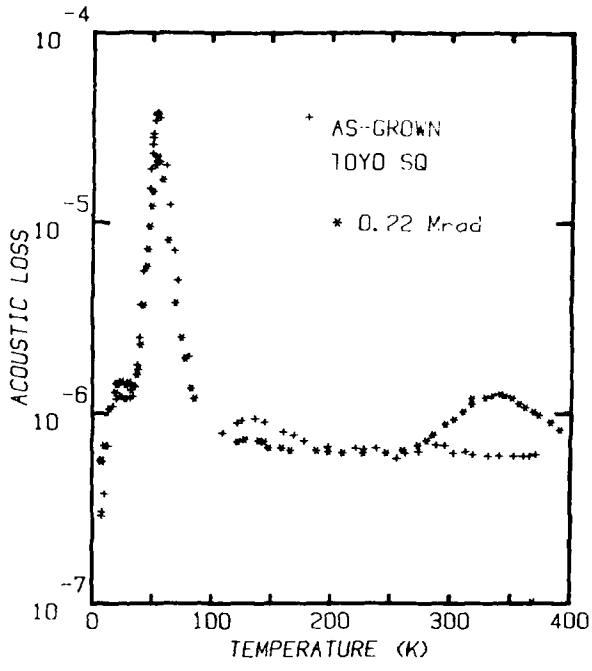


Fig. 5. The acoustic loss spectrum of an unswept SQ-B crystal is shown in the as-grown condition and after irradiation. The irradiation produced the Na-related peak at 340 K.

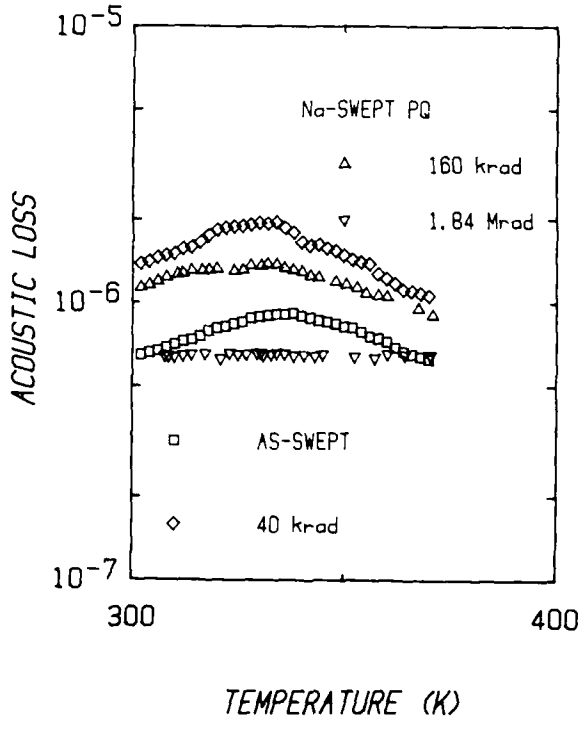


Fig. 6. The acoustic loss spectrum of a Na-swept PQ-E sample is shown for a series of irradiations. Here the 340 K peak is present in the as-swept condition.

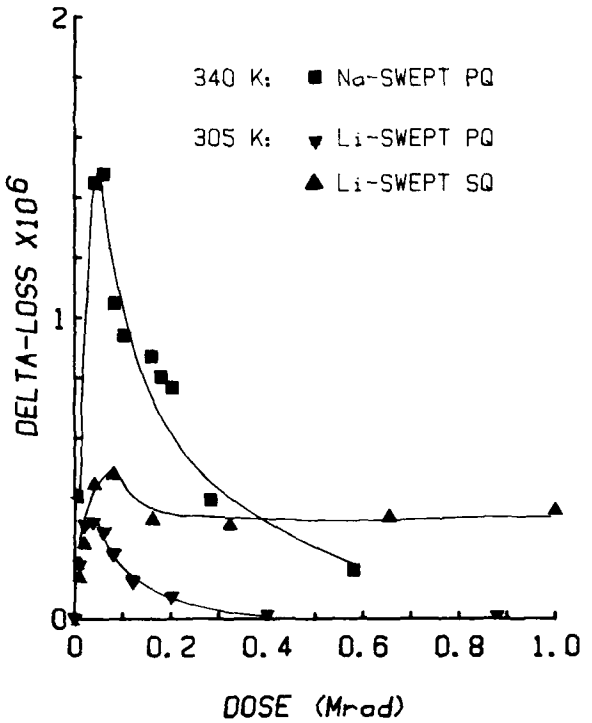


Fig. 7. Production curves for the 305 K L-related and 340 K Na-related loss peaks are shown.

for higher doses. A 1.84 Mrad dose has completely removed the 340 K peak in this Na-swept sample. A similar study was made on a Li-swept Premium Q sample. While the 305 K peak was too weak to observe in the as-Li-swept condition it did grow rapidly upon radiation. Figure 7 shows the production curves for the 340 K peak in the Na-swept Premium Q sample and for the 305 K peak in Li-swept Premium Q and Supreme Q samples. Most likely the same defect center is responsible for the 305 K and 340 K loss peaks; the 305 K peak is caused by the center trapping a Li^+ and the 340 K peak by trapping a Na^+. During irradiation, the alkali ions are released from the aluminum site and migrate along the c-axis channel until they are trapped at some other site. While the defect center causing the 305 K and 340 K loss peaks must be one of these alkali trapping sites it cannot be the final trap since a strong irradiation removes the alkali from the center. Since an alkali ion moves into the site during the initial early irradiation one would suppose that a different positive ion such as a proton moves out of the trap. However, we have not seen any OH^- related IR bands grow in or out during the irradiation sequence that would correlate with the trapping or release of an alkali ion from this intermediate trap.

CONCLUSIONS

The radiation-induced defect that causes the 23 K loss peak grows more slowly than either the $Al\text{-}OH^-$ or the Al-hole center. Thus, the earlier assignment of the Al-hole center as the origin of the 23 K loss peak is in doubt. Loss peaks at 305 K in Li-swept material and at 340 K in Na-containing material are caused by an intermediate alkali trap.

ACKNOWLEDGEMENTS

This work was supported by the Solid State Sciences Division of the Rome Air Development Command and by Sandia National Laboratories.

REFERENCES

1. J. C. King and H. H. Sander, IEEE Trans. Nucl. Sci. NS-19, 23 (1972).

2. P. Pellegrini, F. Euler, A. Kahan, T. M. Flanagan, and T. F. Wrobel, IEEE Trans. Nucl. Sci. NS-25, 1267(1978).

3. T. J. Young, D. R. Koehler, and R. A. Adams. Proc. of the 32nd Annual Symposium on Frequency Control, p.34 (1978).

4. D. R. Koehler and J. J. Martin, Proc. of the 37th Annual Symposium on Frequency Control, p.130(1984).

5. J. C. King, Bell System Technical J.,38, 573 (1959).

6. R. A. Poll and S. L. Ridgway, IEEE Trans. Nucl. Sci. NS-13, 1. ember, 1966).

7. T. M. Flanagan and T. F. Wrobel, IEEE Trans. Nucl. Sci. NS-16, 130 (December 1969).

8. B. R. Capone, A. Kahan, R. N. Brown, and J. R. Buckmelter, IEEE Trans. Nucl. Sci. NS-17, 217 (December 1970).

9. T. M. Flanagan, IEEE Trans. Nucl. Sci. NS-21, 390 (December 1974).

10. J. J. Martin, J. Appl. Phys. 56, 2536 (1984).

11. L. E. Halliburton, N. Koumvakalis, M. E. Markes, and J. J. Martin, J. Appl. Phys. 52, 3565 (1981).

12. S. P. Doherty, J. J. Martin, A. F. Armington and R. N. Brown, J. Appl. Phys. 51, 5449 (1980).

13. M. E. Markes and L. E. Halliburton, J. Appl. Phys. 50, 5449 (1979).

14. W. A. Sibley, J. J. Martin, M. C. Wintersgill, and J. D. Brown, J. Appl. Phys. 50, 8172 (1979).

15. A. Kats, Philips Res. Repts. 17, 133 (1962).

16. D. B. Fraser, J. Appl. Phys. 35, 2913 (1964).

17. D. R. Koehler and J. J. Martin, J. Appl. Phys. In press.

18. A. W. Warner, Bell System Technical J. 40, 1193 (1960).

19. J. J. Martin, Proc. 38th Annual Symposium on Frequency Control, p 16. (1984).

20. A. Kahan and H. G. Lipson, Proc. 39th Annual Symposium on Frequency Control, in press.

21. C. Y. Chen and L. E. Halliburton, private communication, 1985.

22. L. E. Halliburton, C. Y. Chen and S. D. Tapp, Proc. 39th Annual Symposium on Frequency Control, in press.

QUARTZ FOR THE NATIONAL DEFENSE STOCKPILE

R.A. Laudise

AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

The U.S. Strategic and Critical Materials Stockpile Act provides for stockpiling critical materials which are deemed essential in the event of National emergencies. Quartz is one of the more than 60 materials presently stockpiled. At the request of the Federal Emergency Management Administration a committee was commissioned in 1984 to assess requirements, supply-demand trends and technology factors relating to needs for the stockpiling of quartz and to advise the U.S. Government concerning the size and composition of the strategic quartz stockpile. A report on the conclusions of this committee* was given at the 39th Annual Frequency Control Symposium. Among the important committee conclusions are: Quartz is considered to be an essential material into the foreseeable future; the United States has a domestic cultured quartz industry that uses domestically-available feedstock for the majority of present requirements. The industry is subject to vigorous foreign competition, and because of a rather thin industry-supported R&D effort, the United States' present strengths could erode rapidly. Present production capacity, based on current quality standards, exceeds projected mobilization demands for military and essential civilian needs. Identified leading-edge military and essential civilian requirements call for special materials that require longer cultured crystal-growth cycles that would effectively reduce total available capacity as the demands increase for these types of crystals.

It is recommended that the current stockpile of natural quartz crystals be considerably reduced, but that large crystals be retained as seeds for cultured crystal growth, and that some additional quartz be retained for special crystal requirements and as feedstock for unique essential silica glass applications. Finally, it is recommended that additional government support be given to R&D on the preparation and properties of leading edge material, and its scale-up into production.

In addition, the author discussed his ideas on R&D needs for research on leading edge material and how a vigorous R&D program can be a viable and perhaps a preferable alternative to stockpiling.

*The committee's final report is for sale by the National Technical Information Service, Springfield, VA, 22161; "Quartz for the National Defense Stockpile", report of the committee on cultured quartz for the National Defense Stockpile, National Materials Advisory Board, National Research Council, #NMAB-424, 1985.

CH2186-0/85/0000-0270$1.00©1985IEEE

ETCHING STUDY OF AT-CUT CULTURED QUARTZ USING ETCHANTS CONTAINING FLUORIDE SALTS, HYDROFLUORIC ACID, AND AMMONIUM BIFLUORIDE

Anthony J. Bernot

Sperry Corporation
Aerospace & Marine Group
Phoenix, Arizona 85036

ABSTRACT

This paper reports the results of a chemical etching study of AT-cut cultured quartz when the etchants hydrofluoric acid and ammonium bifluoride are modified with fluoride salt additives. These fluoride salts supply positive ions other than H^+ and NH_3^+ at adsorption sites on the quartz surface.

Results show that specific fluoride salts affect surface topography and etch defects. The addition of calcium, aluminum, and lithium fluoride reduce etch pits and improve surface topography when added to hydrofluoric etchants at 60°C. Addition of potassium fluoride reduces etch pit and etch channel formation when added to hydrofluoric acid or ammonium bifluoride etchants at 25°C or 60°C.

INTRODUCTION

In the fabrication of quartz crystal resonators, blanks are adjusted to appropriate crystal orientation and frequency by lapping resonator surfaces with abrasive particles in the 3- to 10-micron range. Due to the surface damage induced by this surface lapping, the resonator surfaces are normally chemically etched to improve the Q and stability of the resonator. Lapped surface damage can extend between .25D and .6D below the surface, depending upon the quality of the quartz. (D is the diameter of particles in lapping abrasive.)

An ideal etching process would remove this lapping surface damage, chemically polish the surface, and minimize material defects exposed by the etchant.

PREVIOUS WORK

Hydrofluoric acid and ammonium bifluoride are the most common etchants used for etching quartz surfaces. Dr. John Vig[1] introduced the concept of deep etching to provide chemically polished surfaces while removing lapped surface damage. He found that saturated ammonium bifluoride solutions chemically polished AT-cut surfaces whereas solutions of 11 percent hydrofluoric acid chemically polished SC-cut surfaces. He described the chemical polishing process as shown in Figure 1. An equilibrium surface is formed which depends upon the initial surface roughness. It was also reported that material defects are exposed in the form of etch pits and etch channels and that well swept AT-cut cultured quartz could be etched channel free.

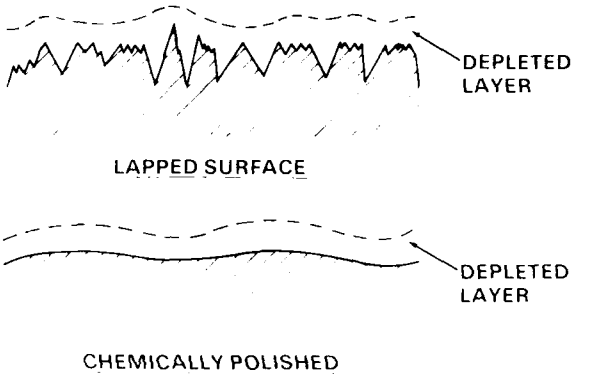


Figure 1

J.K. Vondeling[2] studied the effects of etching AT-cut quartz with the addition of several additives to both hydrofluoric acid and ammonium bifluoride. He reported that the addition of potassium fluoride to hydrofluoric acid reduced the occurrence of etch defects and that etch pit free surfaces could be achieved at 25°C when a 2 to 1 molar etchant of hydrofluoric acid and potassium fluoride were used. Vondeling postulated that the quartz etching process is also affected by positive ions other than H^+ and NH_3^+ adsorbed onto the quartz surface. This adsorption changes the electronic configuration around the silicon which makes possible the final dissolution by HF.

DESCRIPTION OF STUDY

It can be concluded from the reports of the preceding investigators that the parameters which control chemically etched surface topography and defects are

- Chemistry of the etchant
- Quartz material defects
- Crystallographic orientation of surfaces
- Temperature of etch bath
- Initial surface roughness

In this study, only the first two parameters were studied.

CH2186-0/85/0000-0271$1.00©1985IEEE

The chemistry of the etchant was studied by adding the following fluoride salts to hydrofluoric acid and ammonium bifluoride

- Aluminum Fluoride
- Calcium Fluoride
- Lithium Fluoride
- Potassium Fluoride
- Sodium Fluoride
- Zinc Fluoride

Material influence was studied by using cultured quartz bars from two suppliers: P.R. Hoffman Materials Processing Division and Motorola Quartz Operations. Only the Motorola material was swept prior to fabrication of blanks. Bars from both suppliers were fabricated into AT-cut, 5 MHz fundamental .410-inch square blanks. Final frequency adjustment and orientation of surfaces was performed by surface lapping with a 5-micron abrasive.

RESULTS OF STUDY

The effect of etchant chemistry upon unswept quartz is reported in Table 1. Etch pit and etch channel occurrence were measured under a microscope utilizing edge and top lighting. Chemically polished surfaces were analyzed by both surface profiling and SEM micrographs. Figures 2 and 3 are SEM micrographs of surfaces where etch pits and surface features associated with some of the etchants can be observed.

In the micrographs of Figure 4 the effect of potassium fluoride on etch channeling is observed for unswept quartz. Since the surfaces are AT-cut and etch channels occur along the crystallographic Z axis when edge lit, the channels appear as "light tubes." Short "light tubes" indicate that the etch channel has penetrated only the material below the surface (usually 10 to 20 microns). In summary, the addition of potassium fluoride at low concentrations limits the penetration of most etch channels to just below the surface. As observed in Table 1, highly concentrated KF + HF etchants do not provide chemically polished surfaces but do inhibit etch channeling completely. This was confirmed by helium leak testing using a Varian Leak Detector.

Figures 5, 6, 7, and 8 relate surface roughness angle (calculated from surface profiles) to depth of etch. Swept quartz was used to develop this data since minimal etch defects were desirable in determining equilibrium surfaces produced by etchants. In the surface roughness angle plots for aluminum fluoride and lithium fluoride minimums are observed. It is felt that these etchants are only useful for etch removals below 50 f^2 before requiring replenishment. The molar ratio of fluoride to hydrofluoric acid is sensitive and as the hydrofluoric acid evaporates the molar ratio is disturbed.

CONCLUSIONS

In recalling the role of an ideal chemical etching process, etchants which contain calcium, aluminum, or lithium fluoride and hydrofluoric acid reduce etch pit occurrence and size while improving surface topography. Since etch channeling is moderate with unswept

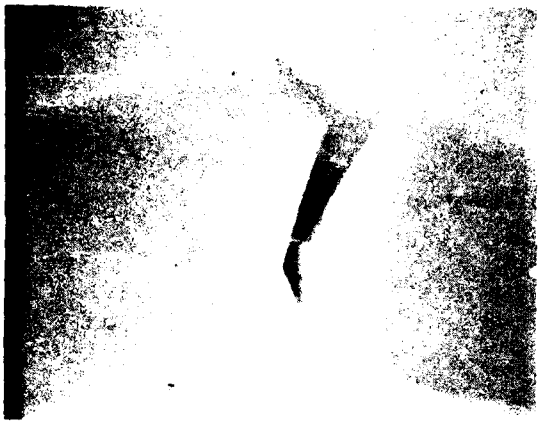
Table 1
Unswept Quartz

Etchant	Etch Rate f^2/min	Etch Temp °C	Material Removed f^2	Chemical Polish	Etch Pit		Etch Channel	
					Occurrence /cm²	Depth Micron	Occurrence /cm²	Development Size & Depth
Sat. $NH_4F \bullet HF$	.17	60	25	Good	1000	4-10	600	Severe
HF	.18	60	30	Poor	1000	2	600	Moderate
KF + HF	.007	25	8	Good	10	≤ 1	10	Slight
KF + HF	.026	60	22	Good	1000	8	400	Moderate
(KF + HF) Conc	.143	45	43	None	0	0	0	–
ALF_3 + HF	.112	60	34	Excellent	500	≤ 1	500	Moderate
CaF_2 + HF	.154	60	44	Excellent	200	≤ 1	500	Moderate
LiF + HF	.182	60	17	Excellent	500	≤ 1	500	Moderate
NaF + HF	.084	45	35	Poor	200	≤ 1	175	Slight
ZnF_2 + HF	.173	60	56	Poor	500	8	500	Moderate
KF + $NH_4F \bullet HF$	.169	60	30	Good	600	2-5	60	Slight
(KF + $NH_4F \bullet HF$) Conc	.170	60	30	None	0	0	0	–
NaF + $NH_4F \bullet HF$	.167	60	30	Good	800	8	500	Severe

SEVERE: CHANNELS PENETRATE THROUGH BLANK

MODERATE: APPROXIMATELY 50% OF CHANNELS PENETRATE THROUGH BLANK; SMALL IN DIAMETER

SLIGHT: MOST CHANNELS DO NOT PENETRATE BLANK (≈ 20 MICRONS BELOW SURFACE)

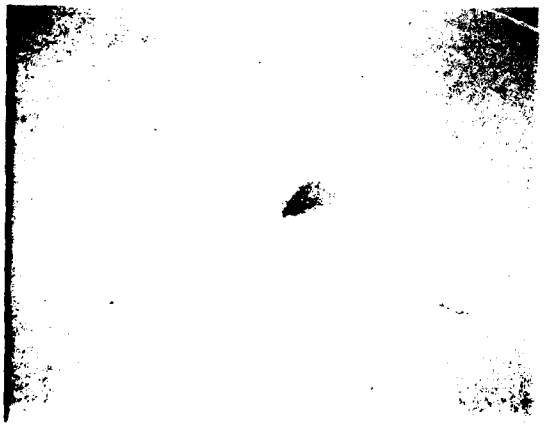


2a. $NH_4F \bullet HF$ @ 3700X Depth $25f^2$

2d. ALF_3 + HI @ 7200X Depth $34f^2$

2b. HF @ 3700X Depth $30f^2$

2e. ZnF_2 + HF @ 3700X Depth $50f^2$

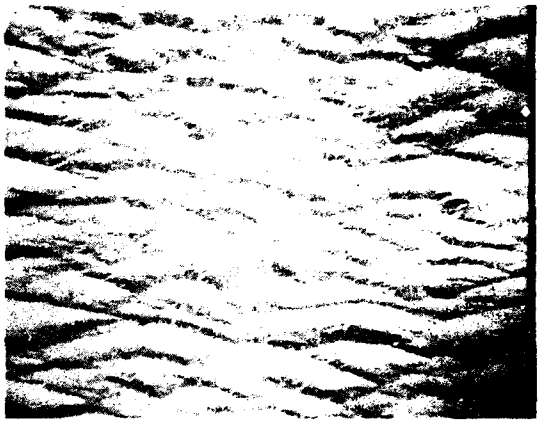


2c. CaF_2 + HF @ 7600X Depth $44f^2$

2f. KF + $NH_4F \bullet HF$ @ 3700X Depth $30f^2$

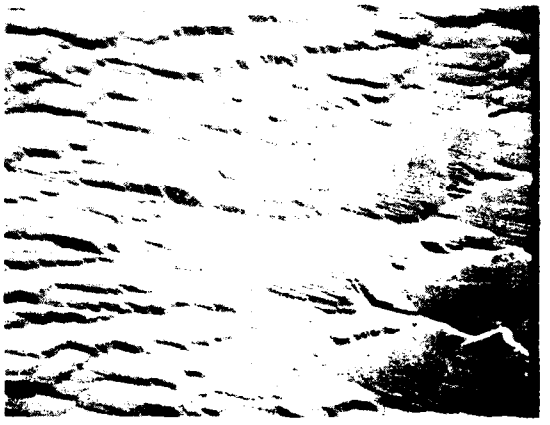
Figure 2

3a. $NH_4F \bullet HF$ @ 300X Depth $25f^2$



3d. ALF_3 + HF @ 400X Depth $34f^2$

3b. HF @ 300X Depth $30f^2$



3e. ZnF_2 + HF @ 400X Depth $50f^2$

3c. CaF_2 + HF @ 400X Depth $44f^2$

3f. KF + $NH_4F \bullet HF$ @ 300X Depth $30f^2$

Figure 3

4a. $NH_4F \bullet HF$ @ 100X Depth $25f^2$

4b. Channels KF + $NH_4F \bullet HF$ @ 100X Depth $30f^2$

Figure 4

CHEMICAL POLISHING STUDY — SWEPT QUARTZ

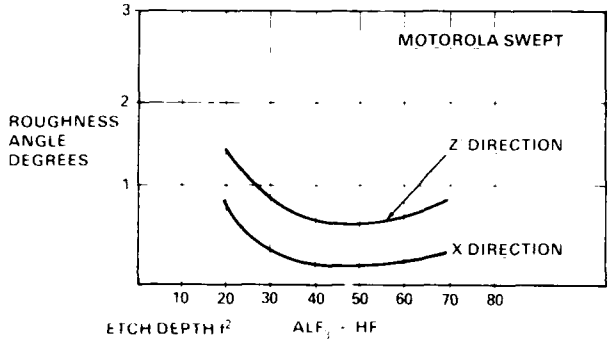


Figure 5

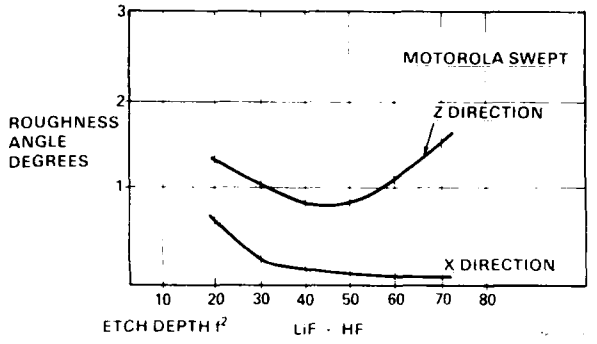


Figure 7

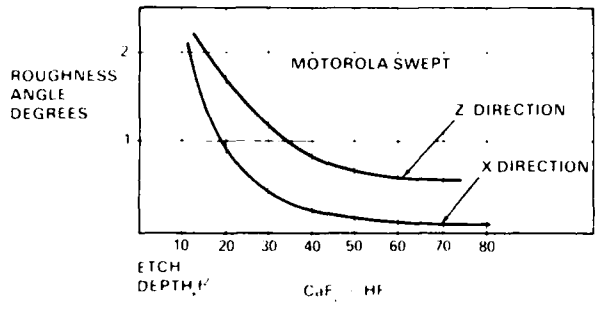


Figure 6

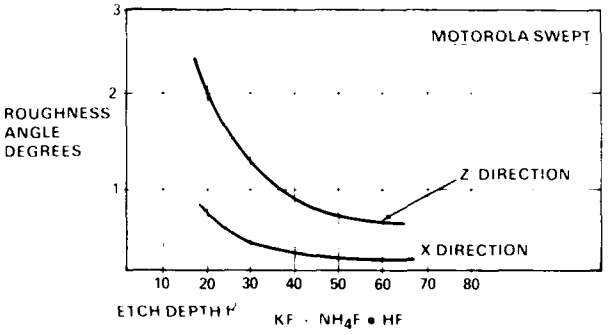


Figure 8

quartz these etchants would be more ideal for chemically polishing swept quartz.

Addition of potassium fluoride to hydrofluoric acid reduces etch pits and channels at 25°C but the etch rate is very low. In Table 1, the etch depth recorded took 21 hours of etching. Reasonable etch rate and reduction of etch pits and channels can be achieved by addition of potassium fluoride to ammonium bifluoride at 60°C making it most ideal for chemically polishing unswept quartz.

REFERENCES

1. J.R. Vig, R.J. Brandmayr, and R.L. Filler, "Proceedings of the 33rd Annual Frequency Control Symposium," Atlantic City, New Jersey, May-June 1979 (Electronic Industries Association, Washington 1979) pp 351-8.

2. J.K. Vondeling, "Fluoride-based Etchants for Quartz." Journal of Materials Science 18 (1983) 304-314.

39th Annual Frequency Control Symposium - 1985

CHEMICAL POLISHING IN ETCHING SOLUTIONS THAT CONTAIN SURFACTANTS

Ronald J. Brandmayr and John R. Vig

US Army Electronics Technology and Devices Laboratory (ERADCOM)
Fort Monmouth, New Jersey 07703-5302

ABSTRACT

The feasibility of chemical polishing a wide variety of quartz cuts has been demonstrated previously. Chemical polishing etchants have been identified for AT-, BT-, IT-, SC-, ST-, Z-cuts and others. It has also been found that, in order to attain a uniform polish with the various etchants, the surfaces must be extremely clean. Surface contaminants can act as etch resists. The etching of contaminated surfaces can result in a "blotchy" appearance. Elaborate and time-consuming cleaning procedures have been developed in order to ensure that the surfaces are uniformly polished. SC-cuts and Z-cuts are especially sensitive to surface contamination.

It has been found that adding certain surfactants to the etching solution can make the chemical polishing process virtually foolproof. The cleaning action and increased wettability produced by the surfactants minimize the need for precleaning. A short and simple cleaning procedure is all that is needed to ensure a uniform polish.

A commercially available premixed etching solution that contains a proprietary surfactant has been found to consistently produce a uniform chemical polish on SC- and AT-cut crystals after only a minimum of precleaning. Commercially available fluorochemical surfactants which the user can add to a variety of fluoride-type etchants, have also been found to produce uniformly polished surfaces on both AT-cut and SC-cut crystals. The results of the evaluation of various surfactants will be reported.

INTRODUCTION

The object of the work described in this paper was to simplify chemical polishing procedures and possibly, to improve the quality of chemically polished quartz surfaces through the use of surfactant additives in the etching solutions. The feasibility of chemically polishing a wide variety of quartz cuts has been demonstrated previously. Chemical polishing etchants have been identified for AT-, BT-, IT-, SC-, ST-, Z-cuts and others.[1-4] It has also been found that, in order to attain a uniform polish with the various etchants, the surfaces must be extremely clean. In particular, it has been found that SC- and Z-cuts are very sensitive to surface contamination.

A four-step cleaning process has been able to consistently give satisfactory results when chemically polishing SC- and Z-cuts in buffered HF solutions. This cleaning process consists of: 1) ultrasonic cleaning in WRS 200S surfactant cleaner, 2) plasma cleaning, 3) soak in Genesolv DES or DA solvent, and 4) soak in boiling H_2O_2. While this cleaning process is very effective, it is time-consuming and labor-intensive. It takes 17 minutes to complete the four steps. Since etching to a precise frequency is an iterative process, each time the blank frequencies are measured, the blanks must be recleaned to ensure a uniform polish. Therefore, cost-saving techniques, which can be used to simplify the surface cleaning before and during chemical polishing, are highly desirable.

The etching was performed using a Briskeat heating mantle which contained a Teflon beaker for holding the etchant, as reported by Vig et al.[2] NH_4F:HF = 4:1 was used to chemically polish SC-cuts, NH_4F:HF = 1:1 was used to polish Z-cuts, and saturated NH_4F·HF was used to polish AT-cuts. The etching temperature was 75°C throughout this study.

SURFACTANTS

Surfactants (short for surface-active agents) can perform a wide variety of useful functions and are, therefore, found in a wide variety of products. Solutions of surfactants exhibit one or more of the following functional properties[5]: detergency, foaming, wetting, emulsifying, solubilizing, dispersing, demulsifying and defoaming. Some of these properties are highly desirable in the etchants used in the chemical polishing of quartz crystals. In particular, the wetting and detergency properties are desirable in chemical polishing, especially if such properties make continuous, in situ cleaning during the etching process possible.

About five billion pounds of surfactants are used per year in products such as semiconductors, textiles, industrial cleaning, foodstuffs, household laundry, and agriculture. Surfactants are classified according to the type of charge they assume in solution and are in four classes: 1) anionic, which have a negative charge; 2) cationic, a positive charge; 3) nonionic, no charge; and 4) amphoteric; either a positive or a negative charge. The majority of surfactants may be chemically unstable in hot fluoride-type etchants. Some surfactants, however, notably

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

the fluorochemical ones, can be stable in such etchants. The cost of using surfactants in the solutions to chemically polish crystals is negligible because only very small amounts are required.

Buffered oxide etchants (BOE) that contain a surfactant additive have been used in the semiconductor industry.[6] According to the manufacturer of a line of "Superwet BOE" etchants[7]: "Superwet BOE buffered oxide etchants contain a surfactant additive which provides significantly greater uniformity of oxide layer etching than when no surfactant is included. The greater uniformity of etching across the wafer reduces the need for over-etching, minimizes the total etch time when excessive lateral etching may occur, reduces apparent undercut and improves the ability to use liquid etchants for manufacturing of high density circuitry with minimal dimensions... These Superwet BOE etchants have surface tensions in the range of 25 to 35 dynes/cm which is equal to or below the interfacial tensions required to uniformly wet patterned substrates. BOE etchants without surfactants have surface tensions of about 85-90 dynes/cm...Other etchant characteristics are similar to those of BOE etchants without the surfactant additive. Etch rate, crystallization temperature, and bath life remain essentially the same... The Superwet BOE etchants are used in standard oxide etching procedures. No residual surfactant contaminants are left on the substrate after typical wafer cleaning procedures are completed."

This description prompted the evaluation of a premixed BOE etchant which was close in concentration to the 4:1 etchant routinely used for chemically polishing SC-cut crystals. The results of the evaluation were highly favorable.

A search for surfactants that may be stable in HF-based etchants revealed that a class of fluorochemical surfactants exists which is likely to be stable. Such surfactants are available from several manufacturers. This paper is a report on the evaluation of various fluorochemical surfactants for potential application as etching solution additives. Table I lists the fluorochemical surfactants studied in this work. Also included is one hydrocarbon surfactant.

NAME	CLASS	CONCENTRATION	MANUFACTURER
FC-93	Anionic	100 ppm	3M
FC-95	Anionic	0.01%	3M
FC-99	Anionic	0.04%	3M
Lodyne S-100	Amphoteric	0.1%	Ciba-Geigy Corp.
Lodyne S-103	Anionic	0.1%	Ciba-Geigy Corp.
Lodyne S-107B	Nonionic	0.06-0.6%	Ciba-Geigy Corp.
Zonyl FSN	Nonionic	0.3%	DuPont
Zonyl FSC	Cationic	0.1%	DuPont
Zonyl FSO	Nonionic	0.01-0.05%	DuPont
Silvatol NP	Nonionic Hydrocarbon	0.03%	Ciba-Geigy Corp.
Superwet BOE 1235	Unknown	Premixed	Allied Chemical

Table I - Surfactants studied. All but the last two are fluorochemical surfactants.

In addition to a Superwet BOE etchant, a group of three FLUORAD Brand Fluorochemical Surfactants[8], the FC-93, FC-95, and FC-99, has been evaluated. The FC-93 is a solution which contains ammonium perfluoroalkyl sulfonates (25%), isopropyl alcohol (20%), and water (55%); the FC-95 is a powder consisting of potassium perfluoroalkyl sulfonates (100%); and the FC-99 is a solution which contains amine perfluoroalkyl sulfonates (25%) and water (75%). It is sufficient to use small quantities of these surfactants. For example, "... 100 ppm of active FC-93 is sufficient to give maximum surface tension reduction" in buffered HF solutions, from 78.1 dynes per cm to 19.1 dynes per cm, at 25°C.

Also evaluated were Lodyne S-100,[9] a fluoroalkyl amino acid type solution which contains: actives (25%), water (62%), hexylene glycol (10%), and tetramethylene sulfone (3%); Lodyne S-103, a fluoroalkyl sodium sulfonate type, which contains actives (45%), water (3%), hexylene glycol (20%), and magnesium sulfate (1%); and Lodyne S-107B, a fluoroalkyl polyoxyethylene type with actives (45%), isopropanol (34%), alkyl polyoxyethylene (18%) and water (3%).

The Zonyl[10] surfactants listed are of the perfluoroalkyl type. Silvatol NP[11], the hydrocarbon type, is an ethoxylated nonyl phenol.

A literature search, via Dialog data bases, uncovered several patents dealing with the use of surfactants in fluoride-based etching solutions. These patents are enumerated in reference numbers 12 through 20. No other relevant publications in the scientific literature were uncovered when using "surfactant," "etch," "fluoride" and "hydrofluoric acid" as the search terms.

EXPERIMENTAL RESULTS

SC-CUT CRYSTALS

In Table II, results are tabulated for experiments on the SC-cut. Some crystals were mechanically polished and some were lapped with 3 μm abrasives. Both natural and cultured quartz were used. In this table it can be seen that when the SC-cut is cleaned with the standard procedure, it polishes well, without smudging. When the cleaning procedure is re-

Surfactant Additive	Quick Water Rinse	Standard Cleaning
None	Smudged	Polished
FC-93, 100 ppm	Polished	-
FC-99, 0.04%	Polished	-
Lodyne S-100	Smudged	-
Lodyne S-103	Smudged	-
Zonyl FSO, 0.05%	Polished	-
Zonyl FSN	Smudged	-
Zonyl FCS	Smudged	-

Table II - Chemical Polishing Results For SC-Cut Crystals Using NH_4F:HF = 4:1 Etchant With Surfactants

duced to a quick water rinse, i.e., flushing under cold running tap water for one-minute, the crystals smudge badly after deep etching. When the FC-93 surfactant was added to the etching solution and the crystals were deep-etched after a similar one-minute rinse, the SC-cut blanks came out polished and smudge-free.

Figure 1 shows a photograph in which a comparison is made between SC-cut crystals etched in solutions containing surfactant FC-93 (clear), and no surfactant (smudged). A similar finding was achieved using the FC-99 surfactant in the etching solution. This result is shown in Figure 2. Zonyl FSO was also evaluated on the SC-cut and found to be effective. Zonyl FSO^{10} is a fluoroalkyl poly (ethylene oxide) ethanol.

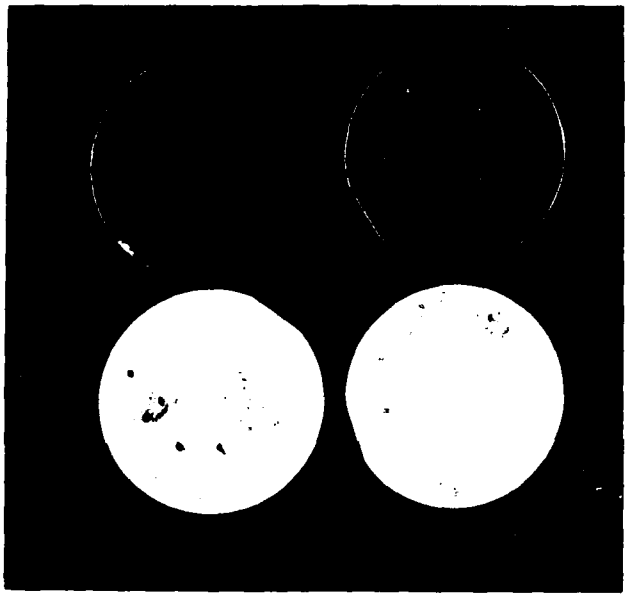


Figure 1. SC-cuts etched with surfactant FC-93 (clear) and without surfactant (smudged).

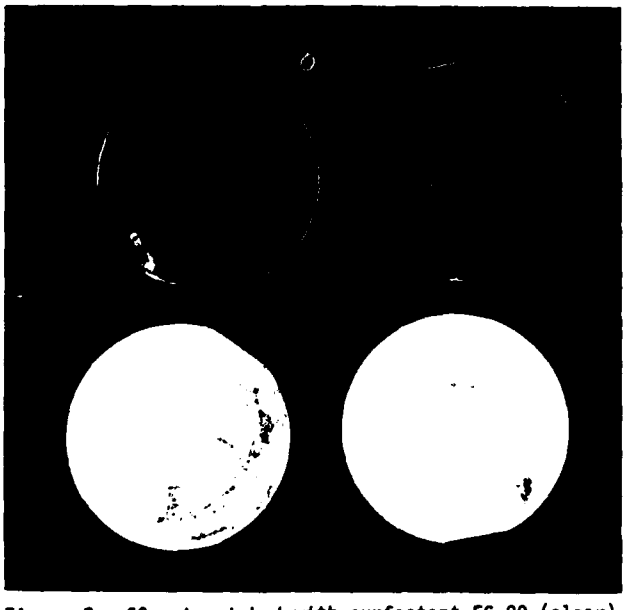


Figure 2. SC-cuts etched with surfactant FC-99 (clear) and without surfactant (smudged).

Of the three crystal cuts investigated, the advantage of using surfactants is most pronounced for the SC-cut. Two different grades of etching chemicals were used for the experiments with the SC-cuts. The first was J. T. Baker Chemical Co., VLSI grade Low Particulate Grade Chemicals. The second was Ashland Chemical Co., Electronic Grade. The results of etching with and without surfactants were equally striking, independent of the grade of the etchant.

The following surfactants also were tried on the SC-cuts: (1) Lodyne S-100, (2) Lodyne S-103, (3) Zonyl FSN and (4) Zonyl FSC. None of these was as effective as those discussed above, at the concentrations employed. (Of course, it is possible that these surfactants can be more effective at different concentrations.)

SC-cut crystals with a 3 μm lapped surface were chemically polished with the FC-99 surfactant in the etching solution and after a thorough cleaning with the standard four-step cleaning process. Surface profile measurements were then made using a Tencor Alpha-step profile meter. The average surface roughnesses were measured along the Z-directions and compared with the same for crystals chemically polished without using the surfactant. No significant differences were observed.However, when mechanically polished crystals were etched after only a quick water rinse, the polished surfaces were severely degraded on both sides of the crystals when no surfactant was used. Similarly prepared crystals etched with surfactant in the solution remained polished. An SEM of the surfaces of both sides of two crystals is shown in Figure 3. The right side of the figure shows two sides of an SC-cut crystal etched with the FC-99 surfactant in the solution, and the left side shows two sides etched without the surfactant. These crystals were both mechanically polished before etching. After etching, only the crystal which was etched with the surfactant in the solution maintained the polish.

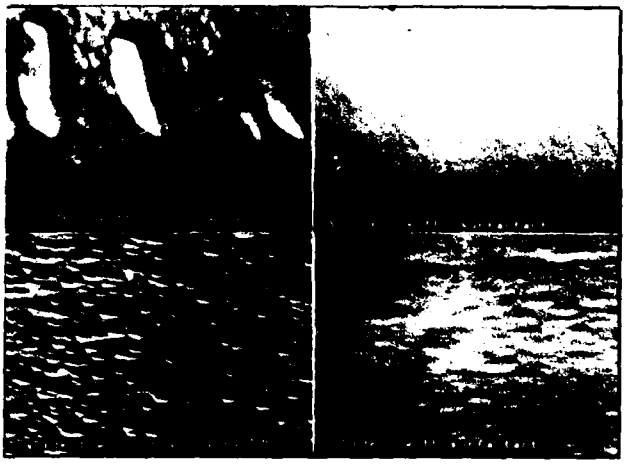


Figure 3. Two sides of an SC-cut crystal plate etched with surfactant (right) and without surfactant (left).

ETCH CHANNEL DENSITY

A set of unswept SC-cut crystals was deep etched in NH_4F:HF = 4:1 solution to which which surfactant FC-99 was added. The etch channel density of these crystals was determined by counting the total etch channels under the microscope, using a micrometer disc. These results were then compared with a similar set of crystals etched in a 4:1 solution but which did not contain a surfactant. No significant difference in the etch channel densities was observed.

AT-CUT CRYSTALS

Table III tabulates the experiments performed on AT-cut crystals with surfactant containing etching solutions. It can be seen from this table that use of the four-step cleaning procedure results in a well-polished, smudge-free crystal when deep etched without the use of a surfactant. When the cleaning procedure is reduced to merely rinsing under cold running tap water, the AT-cut crystals become smudged, but only slightly. The crystals used for this experiment were all 5 MHz fundamental mode, either mechanically polished or lapped with a 1 μm abrasive. The smudging of the crystals was not readily observable with the unaided eye, but became apparent when viewed under the microscope at 30X magnification. However, when the crystals were intentionally contaminated with fingerprints and flushed under cold running tap water only prior to etching, the smudging became readily observable after deep etching.

Intentional Contamination	Surfactant Additive	Quick Water Rinse	Standard Cleaning
No	None	Slightly Smudged	Polished
Yes	None	Smudged	-
No	FC-93, 100 ppm	Polished	-
Yes	FC-93, 100 ppm	Polished	-
No	FC-99, 0.04%	Polished	-
Yes	FC-99, 0.04%	Polished	-

Table III – Chemical Polishing Results for AT-Cut Crystals Using NH_4F·HF Etchant With Surfactants

Experiments were then performed using the surfactant additive FC-93. It can be seen from Table III that, when crystals were polished with FC-93 in the solution, there were no smudges when the precleaning consisted of only a quick water rinse, even after intentional contamination.

Finally, etching solutions with the FC-99 surfactant were prepared. In this experiment, polished crystals were produced as shown in Table III. In particular it should be noted that samples were also intentionally contaminated with fingerprints. After etching with the FC-99 surfactant-containing solution, the polished surfaces showed no degradation due to the fingerprints; this compares with the intentionally contaminated crystals which remained smudged after etching without a surfactant. Figure 4 shows a comparison between AT-cut crystals etched with surfactant FC-99 in the etching solution (clear), and with no surfactant (smudged). All of the four crystals shown had been intentionally contaminated with fingerprints. SEM analysis indicated that the surface roughness of the AT-cut crystal was degraded when the crystal was etched after the shortened cleaning procedure without the surfactant additive.

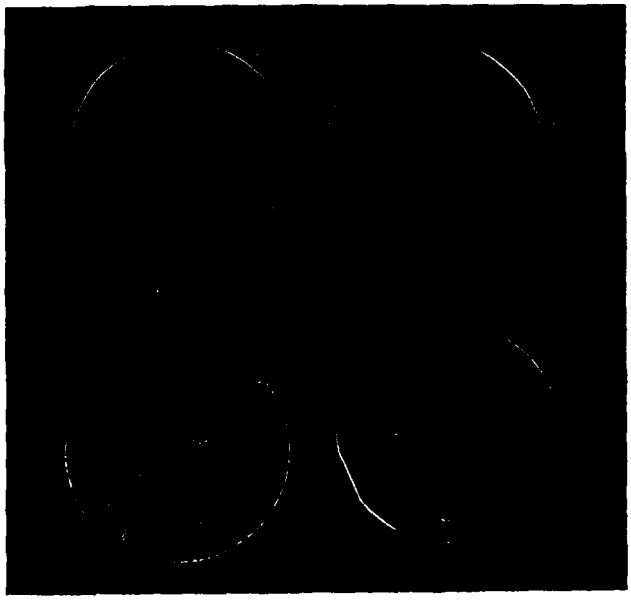


Figure 4. Comparison between AT-cut crystals etched with surfactant FC-99 (clear) and no surfactant (smudged).

For AT-cuts, the contrast between etching with and without surfactants was not as pronounced without intentional contamination as it was for the SC-cut.

Z-CUT CRYSTALS

Table IV shows a tabulation of etching experiments performed on Z-cut crystals. The crystal plates used were either mechanically polished or lapped with 3 μm aluminum oxide abrasive. The Z-cut is particularly contamination-sensitive. In Table IV it can be seen that Z-cuts cleaned by the standard process polished well. When the cleaning was simplified to one minute under cold running tap water, the crystals smudged badly. Z-cut crystals could be polished in an etching solution with FC-93 surfactant after a one-step cleaning in H_2O_2. The remainder of the table shows various attempts at chemical polishing using surfactant additives FC-93, FC-95, FC-99, Silvatol NP, FSC, FSO and S-107B in the etchant. In no instance was success achieved in obtaining a polished smudge-free Z-cut surface when only a cold water rinse precleaning

procedure was used. Included were several experiments using the NH_4F:HF = 4:1 solution. In these instances, the smudging was slow to develop but gradually worsened with prolonged etching.

Surfactant Additive	Quick Water Rinse	Standard Cleaning
None	Smudged	Polished
FC-93	Smudged	-
FC-95	Smudged	-
FC-99	Smudged	-
Silvatol NP	Smudged	-
Zonyl FSO	Smudged	-
Zonyl FSC	Partially Smudge-free	-
Ladyne S-1078	Smudged	-

Table IV – Polishing Results for Z-Cut Crystals Using NH_4F:HF = 1:1 Etchant With Surfactants

SUMMARY AND CONCLUSIONS

For the SC-cut, the results of using surfactants were striking. Both the FC-93 and FC-99 surfactants significantly reduce the need for precleaning this contamination-sensitive cut. A short cleaning procedure prior to etching is sufficient to ensure a uniformly etched surface.

Although the AT-cut is not a serious problem with respect to contamination sensitivity, the use of fluorochemical surfactant additives still reduces the need for cleaning the crystal surfaces prior to etching. The FC-93 and FC-99 surfactants are also effective for the AT-cut.

The Z-cut crystal is even more contamination-sensitive and difficult to polish without careful precleaning than is the SC-cut crystal. The surfactants tested thus far were not able to ensure a uniform polish when the crystals were cleaned by only a cold-water rinse prior to etching.

Commercially available premixed etching solutions are available which contain proprietary surfactants. One of these, the BOE Superwet 1235, has consistently produced a smudge-free chemical polish on SC-cut crystals, with a minimum of precleaning.

Of the three cuts investigated, the rate of etching during chemical polishing is by far the highest for the Z-cut. Therefore, the contamination sensitivity of the Z-cut and the lesser effectiveness of surfactants are probably a consequence of the fast etching rate, i.e., there is less time for surface contaminants to be removed before the masking effects of the contaminants result in uneven etching. The relative etching rates alone, however, cannot account for the observed contamination sensitivities. The SC-cut is more contamination sensitive than the AT-cut, even though the etching rate of the SC-cut during chemical polishing is slower than that of the AT-cut. The adsorption properties of the surfaces must also play a role.

Chemical polishing with surfactants offers the advantages of increased throughput, increased yield and, therefore, lower costs.

Acknowledgements

The authors thank Steven C. Moore for bringing to their attention the availability of the Superwet BOE etchants and the FC-93 surfactant. The authors also thank Donald Eckart for the photomicrographs of etched surfaces.

REFERENCES

1. J. R. Vig, J. W. LeBus and R. L. Filler, "Chemically Polished Quartz," Proc. 31st ASFC*, pp. 131-143 (1977); copies available from National Technical Information Service, AD A088221.

2. J. R. Vig, R. J. Brandmayr and R. L. Filler, "Etching Studies of Singly and Doubly Rotated Quartz Plates," Proc. 33rd ASFC, pp. 351-358 (1979); copies available from Electronic Industries Assoc., 2001 Eye Street, N.W. Washington, D.C. 20006.

3. R. J. Brandmayr and J. R. Vig, "Further Investigation of Etchants for Chemically Polishing SC-cut Quartz Crystals," R&D Technical Report DELET-TR-81-16, Sept. 1981; copies available from Nat'l Technical Information Service, AD A104903.

4. R. W Ward, "Update on Semiconductor Process Techniques for Crystals," Proc. 4th Quartz Crystal Conference, pp. 276-287, 1982; copies available from Electronic Industries Assoc., 2001 Eye Street, NW, Washington, DC 20006.

5. R. L. Camp, K. C. Scott, K. F. Scholne and R. R. Holland, "Success of Your Process may Depend on A Surfactant," Research and Development, March 1985, pp. 92-97.

6. Steven C. Moore, private communication, June 1984.

7. Allied Chemical Preliminary Information Bulletin entitled "Superwet BOE Etchants," Allied Chemical Company; Electronic Chemical Products, 20 Peabody Street, Buffalo, N.Y. 14210

8. 3M Co., Commercial Chemicals Division, 223-6SE 3M Center, St. Paul, MN 55101.

* Annual Symposium on Frequency Control.

9. Ciba-Geigy Corp. technical bulletin "Lodyne S-100 Fluorochemical Surfactant," January 1, 1979; technical bulletin "Lodyne S-103 Fluorochemical Surfactant," May 1984 and Technical Bulletin "Lodyne S107B Fluorochemical Surfactant," June 1984, New Products Venture Group, Ciba-Geigy Corp., Ardsley, New York, 10502.

10. DuPont Product Information Bulletin "Zonyl Fluorosurfactants," E.I. Du Pont de Nemours and Company, Chemicals and Pigments Dept., Wilimington, Delaware 19898.

11. Ciba-Geigy Corp., private communication, April 1985.

12. U.S. Pat. No. 4,243,473, "Method for Detecting Crystal Defects...," 1981.

13. U.S. Pat. No. 4,230,522, "PNAF Etchant for Aluminum and Silicon..." 1980.

14. U.S. Pat. No. 4,055,458, "Etching Glass with Fluorine - Containing Surfactant...," 1977.

15. U.S. Pat. No. 4,040,897, "Etchants for Glass Films on Metal Substrates...," 1977.

16. U.S. Pat. No. 3,844,859, "Titanium Chemical Milling Etchant...," 1974.

17. Japanese Pat., JP 68 25,515, "Etching and Marking Glass," 1975.

18. Japanese Pat., JP 59 56,482 [84 56,482], "Etching Compositions for Fabrication of of Semiconductor Devices," 1984.

19. Japanese Pat., JP 59 31,029 [84 31,029], "Etching Solution for Semiconductors," 1984.

20. Japanese Pat., JP 58 55,323 [83 55,324], "Etching Solution for Silicon and Its Oxide Films," 1983.

EFFECT OF CRYSTAL ORIENTATION ON THE SURFACE TEXTURE OF CHEMICALLY ETCHED QUARTZ PLATES, THE CASE OF CUTS CLOSE TO THE AT CUT.

C.R. Tellier

Laboratoire de Chronométrie, Electronique et Piézoélectricité

E.N.S.M.M., Route de Gray, 25030 Besançon Cédex, France.

Summary

A study has been made of the influence of the crystal orientation on the surface texture of quartz plates etched by a concentrated ammonium bifluoride solution. The case of quartz plates with angle θ equal to about 31°, 33° and 37° and labelled AT-31, AT-33 and AT-37 has been extensively investigated.

The kinetics of etching of the various quartz plates are studied. The behavior, in the high temperature region, of some AT-31 and AT-33 plates departs from the usual behavior of other plates which is described by an apparent activation energy of about 0.38 eV.

The changes in profilometry traces depend on the orientation. In particular the size of dissolution figures is very affected by prolonged etching only for some typical orientations (AT-37, AT cut). The decrease in the roughness parameters with the depth of etch may be understood, with respect to the orientation, in terms of stable or continuously moving etch patterns. Surface textures characteristic of the crystal orientation are revealed by the SEM micrographs which agree well with the surface profilometry traces.

Comparison of kinetics, topography and SEM data allows us to give some evidence for two different behaviors even if the orientation θ varies only of 6°. For certain orientations moving etch patterns with a moderately varying size develop on the surface whereas for other orientations stable etch figures which enlarge markedly with prolonged etching are produced. For these last orientations a knee occurs generally in the Arrhenius plot.

Introduction

Interest in chemical etching as a procedure to prepare high-frequency quartz resonator plates[1-13] with clean and smooth surfaces has been revived in the past few years. Effectively mechanical lapping results in the formation of a disturbed surface layer[3,12,14] which can affect markedly the performance of resonators whereas chemical etching presents the advantage to prevent from misorientations the surface layer of quartz resonators. Hence some workers[1-13] have investigated the possibility of a chemical polishing of quartz surfaces by immersion in $NH_4.FHF$ solutions or in NH_4F with HF mixtures. In particular some satisfactory results have been obtained on AT-cut[1,2,7,8,13] and SC-cut[5,6] quartz plates.

However some recent works[9-11,15] on differently oriented quartz surfaces have revealed that the figures produced by NH_4 F.HF solutions are connected with the orientation of the crystal plane on which they are formed. Consequently a more detailed examination of the effect of the crystal orientation on the surface texture of chemically etched quartz plates may be of some interest. In this paper the case of singly rotated cuts close to the most universally used cut, namely the AT cut, is examined. For simplicity and for separation of the roles played by the etchant composition and by the crystal orientation in the kinetics of etching and in the development of typical etch figures the quartz plates were etched in a concentrated ammonium bifluoride solution.

Experimental Procedure

The resonators used were planoconvex quartz plates. The quartz plates with the angle θ respectively equal to 31°, 33° and 37° and labelled for convenience AT-31, AT-33 and AT-37 were cut from similar synthetic crystals. Before etching the plates were lapped with a 9.5 μm abrasive so that the initial centre line roughness, R_a, of the various plates remained in the range 0.1 - 0.12 μm. These plates as well as AT cut plates were etched in a concentrated ammonium bifluoride solution (typical concentration : 10.5 mol l^{-1}) at a constant temperature in the range 290 - 360 K for various periods of time.

The surface textures were characterized by means of two different procedures. Firstly surface topography data were determined from traces given by a microprocessor-based surface profilometer. The roughness parameters used for characterizing the surface textures were the centre line average roughness, R_a, and the r.m.s. roughness, R_q[16]. The traces were made along two rectangular directions of the plane surface of resonators, one of these directions coïnciding with the crystallographic X direction.. In addition after each isothermal etching scanning electron microscopy studies were carried out on these non-conducting specimens by using a low accelerating voltage (1 kV).

Etching Rate versus Time and Temperature

Decrements in thickness, Δd, are evaluated from frequency measurements by means of the well-known formulae[2,7]

$$\Delta d = - K \frac{\Delta f}{f_i f_f}$$

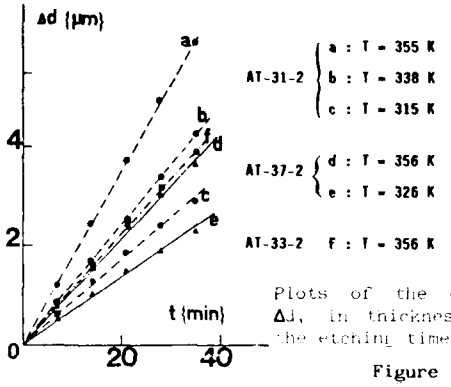
where f_i and f_f are the initial and final frequency. Values of the constant K which depends on the overtone of the vibration and on the orientation of the quartz plate are reported in table 1.

CH2186-0/85/0000-0282$1.00©1985IEEE

θ	overtone	k (MHz μm)
31°	3	49.746
33° 30'	3	49.747
35° 30'	5	83.9
37° 30'	3	50.018

Table 1 : The constant, K, for differently oriented quartz resonators.

Since during isothermal etchings the decrement in thickness, Δ d, varies linearly with the etching time,t, (figure 1) average etching rates can be easily determined from the temperature dependent slopes of the Δd versus t plots.



Plots of the decrement, Δd, in thickness against the etching time, t.

Figure 1

It is usual[17] to describe the general shape of the rate versus temperature plots in terms of an Arrhenius equation

$$R = A \exp\{ - \frac{E_a}{k_B T} \}$$

where E_a is an apparent activation energy.

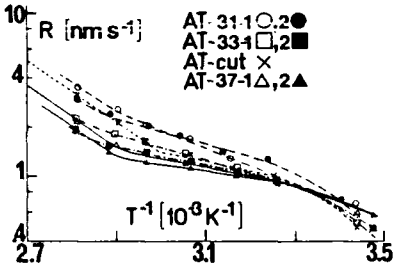


Figure 2

Plot of $\ell n\ R$ vs T^{-1} for various quartz plates.

The temperature, T, dependence of the etching rate, R, was measured for the four crystal orientations. Data are plotted in figure 2. The AT-37 and AT-cut plates exhibit a common feature. The data are fitted approximately by pairs of straight lines which correspond to the low and the high temperature regions. The slopes of the lines give in the high temperature region the apparent activation energies as shown in table 2. The behavior of some AT-31 and AT-33 plates is also interesting.

Figure 2 reveals that the knee occuring generally[7-11] in the Arrhenius plot seems to vanish or to take place at temperature higher than 348 K. However when it is possible the determination of the apparent activation energy in the high temperature regions leads to values which are almost independent of the orientation (table 2).

Quartz Plates		E_a (eV)
AT-31	AT-31-1	0.37
	AT-31-2	indefinite
AT-33	AT-33-1	indefinite
	AT-33-2	indefinite
AT-cut	AT-35-8	0.377
AT-37	AT-37-1	0.386
	AT-37-2	0.354

Table 2 : The activation energy, E_a, as evaluated in the high temperature region.

Surface Profilometry

The traces made along the two rectangular directions, namely the X direction and β_θ direction, were used to derive average values of the geometrical parameters R_a and R_q. The depth, Δd_s ($= \Delta d/2$), of etch dependence of the roughness parameters was measured for four crystal orientations. The data displayed in Figures 3 and 4 merit some comments.

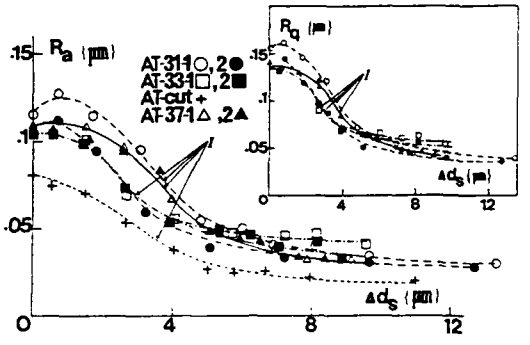


Figure 3

R_q vs Δd_s for X-traces and for various quartz plates. In the inset are shown R_q data.

1. Etching causes a rapid decrease in both roughness parameters measured along the X direction of all various quartz plates. The limiting values of roughness parameters as listed in table 3 do not clearly depend on the orientation.

2. We observe also a decrease in roughness parameters in the course of etching for traces made along the β_θ directions. The equilibrium values as shown in table 3 exhibit a clear dependence on the crystal orientation. In particular the changes in R_a and R_q with the depth of etch are considerably less marked for the AT-33 plates than for the other plates.

Examination of the changes in surface profiles produced by prolonged etching corroborates these observations. There are four particularly interesting features of profiles as presented in figures 5 to 8.

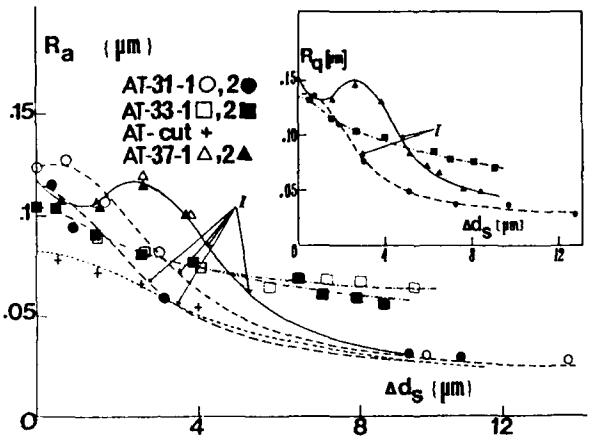


Figure 4

The roughness parameter, R_a, as a function of the depth, Δd_s, of etch for β_θ traces and for various quartz plates. In the inset are shown β_0 data.

1. In the X direction the surface profiles of deeply etched quartz plates are concave or roughly concave depending on the orientation. In particular the concave shape of the X profiles is well defined only for the AT-37 and AT-cut plates.

2. With prolonged etching the final shapes of β_θ profiles are characteristic to the crystal orientation as illustrated in figures 5a to 8a. Effectively the shape of the β_0 profile changes rapidly from convex to concave when the angle θ increases from 31° to 37°.

3. Repeated etchings on the AT-37 and AT-cut plates produce more stable surface profiles than the dissolution profiles formed on the AT-31 and AT-33 surfaces.

4. A marked enlargment of the dissolution profile with prolonged etching can be clearly distinguished only for the profilometry traces corresponding to the AT-37 and AT-cut quartz plates.

Scanning Electron Microscopy

The changes in dissolution figures on etching were also examined by scanning electron microscopy (SEM). The results (Figures 9 to 11) are in perfect agreement with those reported in the above section. The dissolution figures, all uniformly oriented and shaped on a given surface, show different final shapes characteristic to the crystal orientation.

1. The flat bottomed depressions formed on the surface of AT-37 plates extend slightly in a direction close to the X axis. They are more round than the pits which develop on AT-cut plates but their general direction seems to be the same.

2. The AT-33 surfaces are covered with asymmetric pits which are generally composed of three faces with more or less irregular shape. These pits are also elongated along a direction close to the X direction.

3. Repeated etchings on AT-31 surfaces give still different figures. Very often the contour of figures is indistinct, however the figures seem to be bounded by a curved plane which extends in a direction close to the X direction. The majority of these figures possesse some striations elongated in a direction which lies approximately at 60° to the X axis.

Moreover as indicated by the SEM micrographs successive isothermal etchings in the high temperature region induce an enlargment of the size of the dissolution figures formed only on the AT-37 surface in agreement with previous observations on AT-cut plates. In contrast SEM micrographs of AT-31 surfaces reveal figures whose shape still varies as the etching proceeds in the high temperature region and whose enlargment, when it exists, is less marked than for the other plates. The behavior of AT-33 surfaces as observed in Figure 10 stands between the behaviors corresponding respectively to AT-31 and AT-37 surfaces.

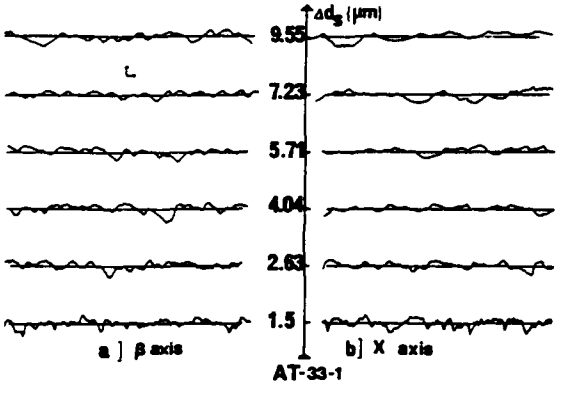
Discussion

The results presented in above sections will be discussed in terms of orientation effects and in terms of a possible one to one correspondence between some dissolution rate data and the evolution of the surface texture during the chemical attack.

The effect of orientation on the shape of etch figures remains evident. It is sufficient for example to compare the flat bottomed pits which develop on AT-37 surfaces with the somewhat convex and striated etch patterns exhibited by deeply etched AT-31 surfaces. It thus appears that the morphology of the etch figures varies rapidly with the orientation of the surface in which they lie. This observation agrees well with previous experiments [10-11] and confirms the statement which establishes a relationship between the morphology of etch figures and the crystal face on which they are produced [12-13]. Moreover the crystal orientation has also a marked influence on the stability of etch figures since, for example, for an increase of about 4° of the angle θ the concave background structure of AT-cut surfaces (figure 7b) disappears to be replaced by the continuously moving X profiles of AT-31 surfaces (figure 5b). At this point it may be of interest to mention that the enlargment of etch patterns which is connected to the formation of stable etch figures depends also on the orientation. In particular the enlargment of etch figures with prolonged etching is so much more marked that we are concerned with angles θ increasing from about 31° to about 37° (see for example Figures

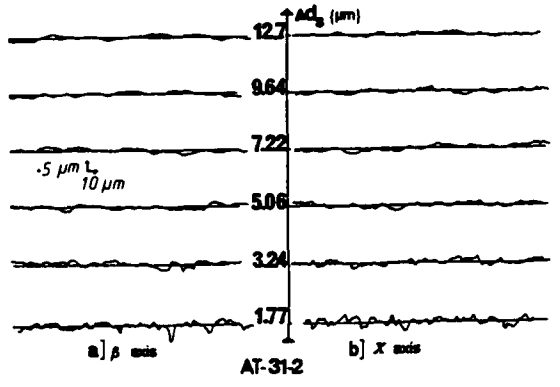
Quartz Plates	R_a vs Δd_s plot		R_q vs Δd_s plot	
	X-trace	β_θ trace	X-trace	β_θ -trace
AT-31-1	0.03	0.025	0.037	0.036
AT-31-2	0.03	0.025	0.035	0.029
AT-33-1	0.040	0.060	0.048	0.075
AT-33-2	0.034	0.053	0.040	0.068
AT-35-8	0.019	0.024	0.022	0.030
AT-37-1	0.028	0.028	0.038	0.040
AT-37-2	0.028	0.028	0.038	0.040

Table 3 : Equilibrium values of roughness parameters (in μm)



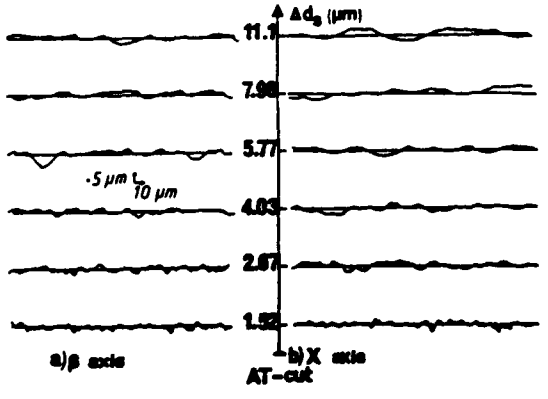
Changes in the surface profilometry traces with the depth of etch an AT-33 quartz plate.

FIGURE 6



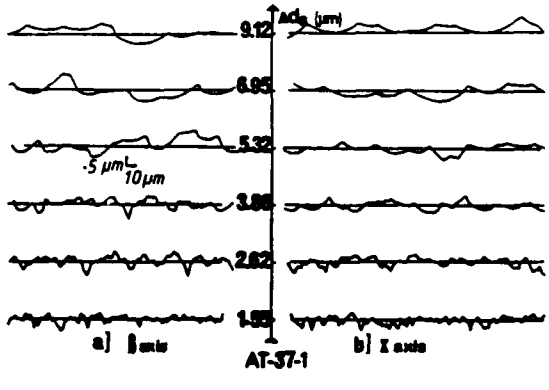
Changes in the surface profilometry traces with the depth of etch of an AT-31 quartz plate.

FIGURE 5



Changes in the surface profilometry traces with the depth of etch an AT-cut quartz plate.

FIGURE 7



Changes in the surface profilometry traces with the depth of etch of an AT-37 quartz plate.

FIGURE 8

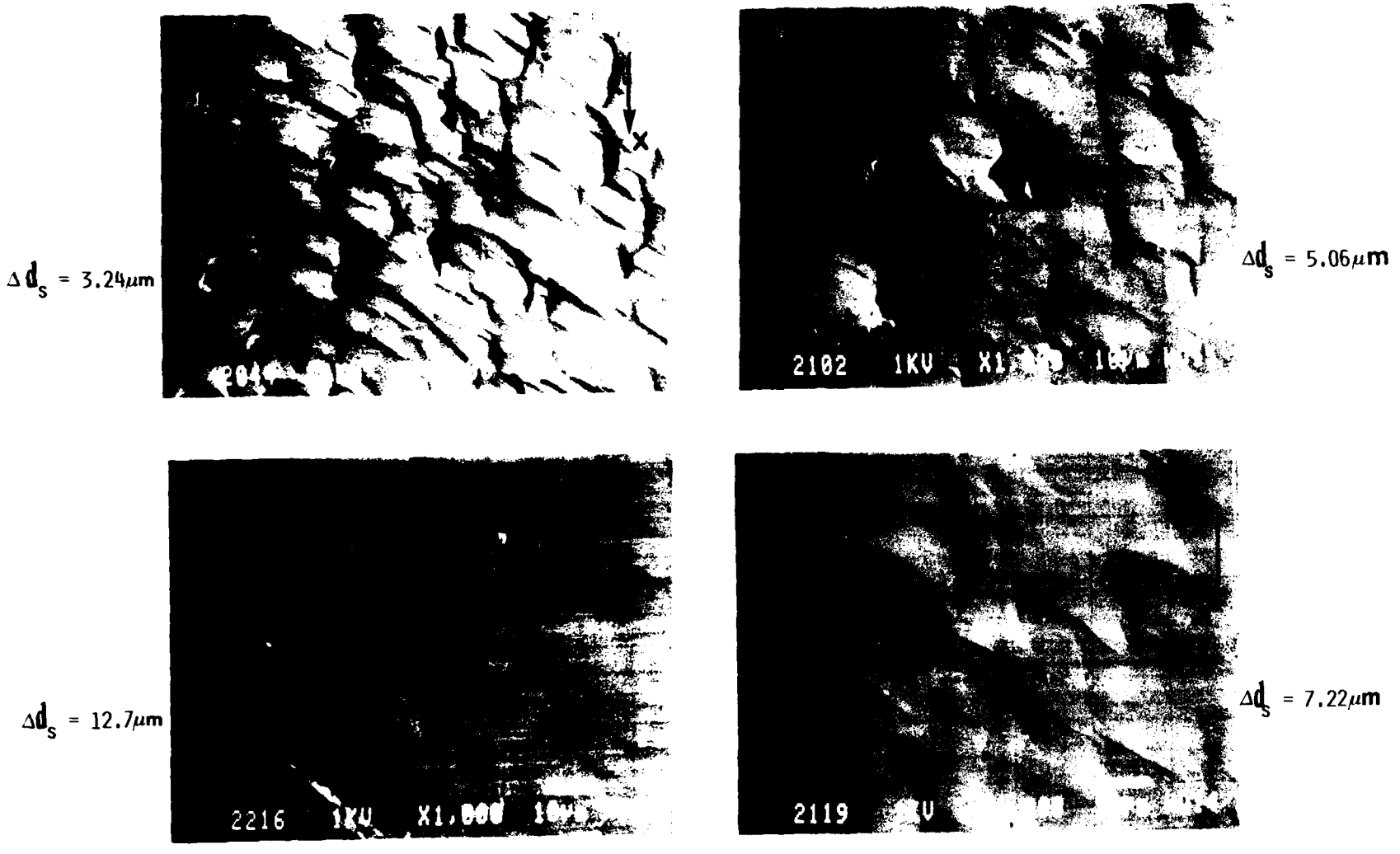


FIGURE 9

SURFACE TEXTURE OF AN AT-31 QUARTZ PLATE AS REVEALED BY SEM.

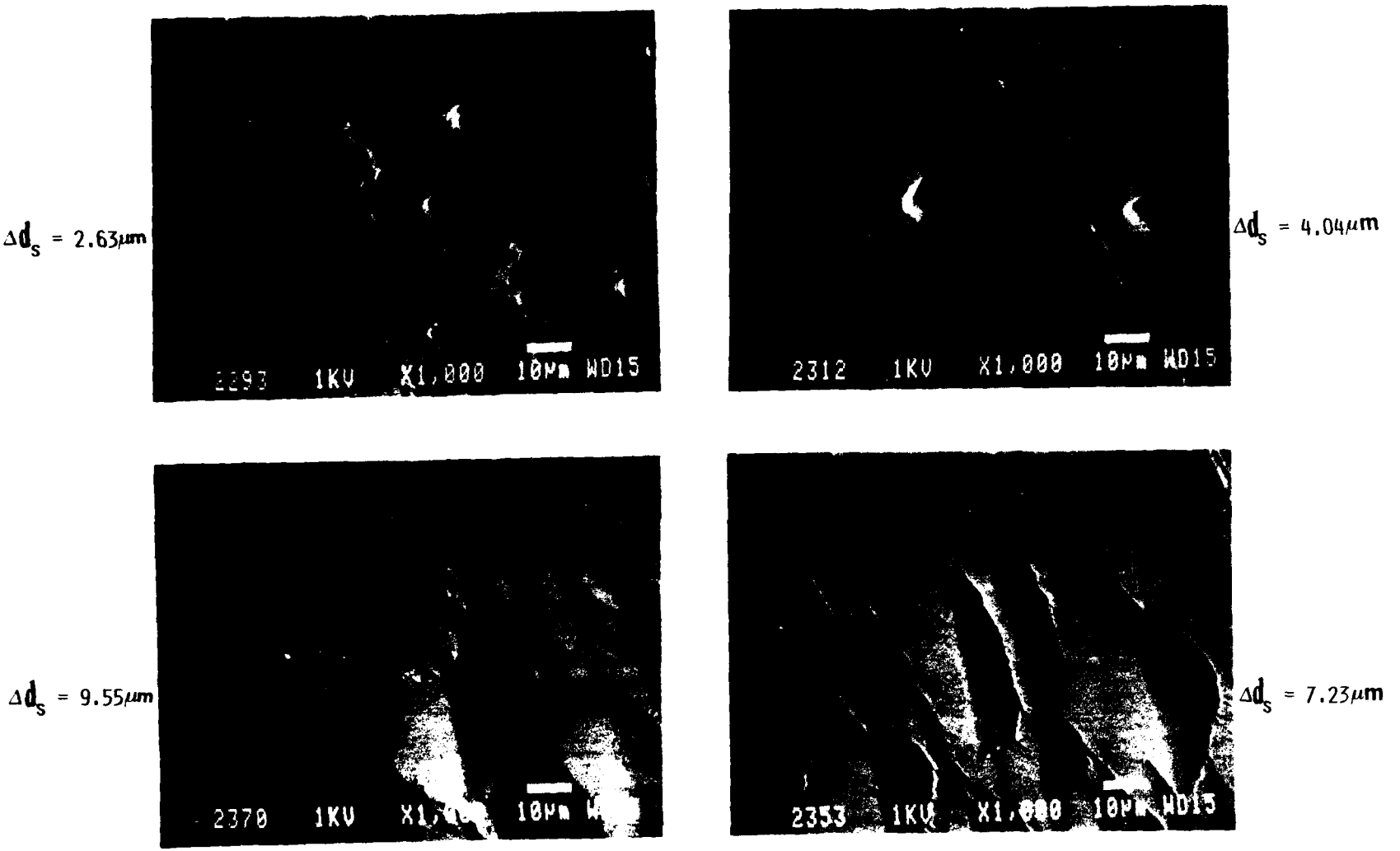


FIGURE 10

SURFACE TEXTURE OF AN AT-33 QUARTZ PLATE AS REVEALED BY SEM.

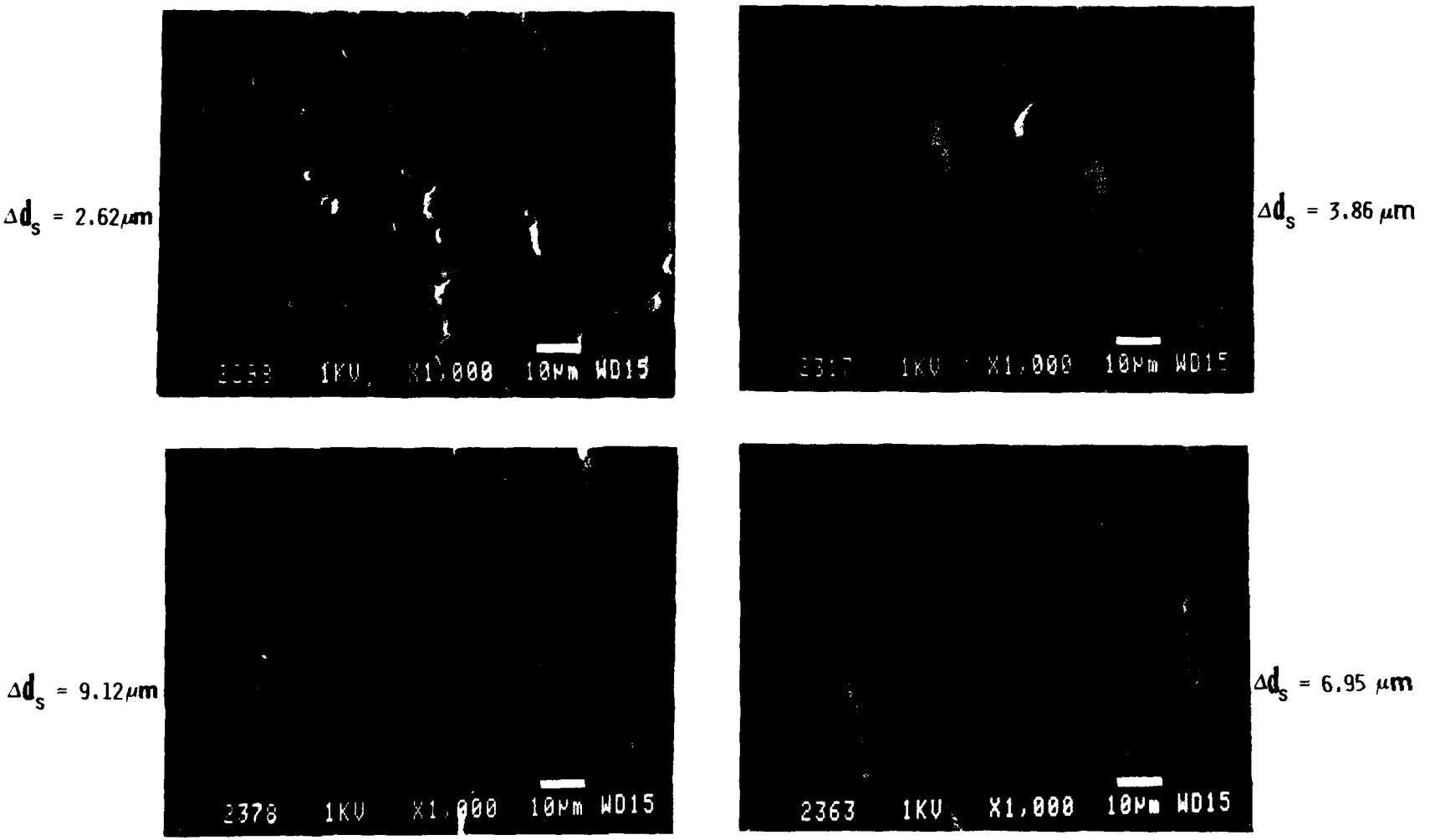


FIGURE 11

SURFACE TEXTURE OF AN AT-37 QUARTZ PLATE AS REVEALED BY SEM.

10 c,d and 11 c,d).

To try to relate certain dissolution rate data to the changes in surface texture on etching it is convenient to summarize results on the kinetics of etching and on the surface topography studies. Turning to the temperature dependence of the etch rate values for the depth of etch corresponding to the knee in the Arrhenius plot are estimated for the various orientations (table 4). Stable or roughly stable surface profiles or dissolution figures take place for critical values, Δd_{SC1} or Δd_{SC2} of the depth of etch which, for an easy comparison are also reported in table 4. With few exceptions the roughness parameters versus depth of etch plots exhibit for a critical value, Δd^*_{SC}, of the depth of etch depending on the crystal orientation, a point of inflexion I (figures 3 and 4). The results are shown in table 5.

Quartz Plate	Δd_s (µm) knee in R vs T^{-1} plot	Δd_{SC1} (µm) (surface profiles) X -trace	Δd_{SC1} (µm) (surface profiles) β_θ -trace	Δd_{SC2} (µm) SEM micrographs
AT-31-2	indistinct	≥ 5.06 (roughly concave)	≥ 5.06 (roughly convex)	≥ 5.06
AT-31-1	indistinct	≥ 4.8 (roughly concave)	≥ 4.8 (roughly convex)	
AT-33-2	$5.3 < \Delta d_s < 6.9$	≥ 5.3 (roughly concave)	≥ 5.3 (roughly convex)	
AT-33-1	indistinct	≥ 4.05 (roughly concave)	≥ 4.05 (roughly convex)	≥ 4.05
AT-35-B	$4 < \Delta d_s < 5.8$	≥ 4 (concave)	≥ 4 (roughly concave)	
AT-37-2	$5 < \Delta d_s < 6.5$	≥ 3.8 (concave)	≥ 3.8 (concave)	
AT-37-1	$5.3 < \Delta d_s < 6.9$	≥ 3.9 (concave)	≥ 3.9 (concave)	≥ 3.9

Table 4 : Critical values of the depth of etch as determined from kinetics, surface profilometry and S E M studies.

Quartz Plate	R_a vs Δd_s plot X -trace	R_a vs Δd_s plot β_θ -trace	R_q vs Δd_s plot X -trace	R_q vs Δd_s plot β_θ trace
AT-31-2	3	2.8	3	2.8
AT-31-1	3.8	3.6	3.6	3.4
AT-33-2	2.8	indefinite	2.8	indefinite
AT-33-1	2.8	indefinite	2.8	indefinite
AT-35-B	3.5	3.4	3.4	3.8
AT-37-2	3.8	4.2	3.8	4.5
AT-37-1	3.8	4.2	3.8	4.5

Table 5 : Typical values (in µm) of the critical depth of etch Δd^*_{SC}.

By examining the results presented in tables 4 and 5 there is some evidence that

1. For a given orientation and for surface with similar initial damage data related to various resonators are in close agreement. Such an agreement must be fulfilled to establish a clear correspondance between shape of pits and structure of the crystal plane.

2. The knee in the Arrhenius plots corresponding to AT-37 plates occurs in the final stages of etching. No knee may be distinctly observed for some AT-31 and AT-33 plates.

3. There is a good correspondence between the results of SEM and surface profilometry studies (table 4). The profiles related to AT-31 and AT-33 can only be roughly identified as convex or concave whereas AT-37 and AT-cut plates present particularly stable etch profiles.

4. The point of inflexion I of the roughness parameters versus the depth of etch plots corresponds to a value of the thickness removed from the AT-31 and AT-33 surfaces smaller than that required for the formation of stable etch profiles. In contrast for AT-37 and AT-cut plates the Δd_{SC1}, Δd_{SC2} and Δd^*_{SC} are all in close accord.

We thus suggest from remarks (2) and (3) that the fact that the knee in the Arrhenius plot remains indistinct for some specific orientations may be connected to moving surface profiles. Moreover for quartz AT-31 which reveals very stable dissolution profiles the knee always corresponds to a depth of etch greater than Δd_{SC1} or Δd_{SC2}. In previous papers[8-11] a correlation between the development of a relatively stable surface texture and the break in the Arrhenius plot has been proposed. This correlation is still satisfied by data on AT-31, AT-33 and AT-37 plates but if we consider the typical results for AT-37 plates it seems that the knee occurs generally when dissolution figures stable in shape and uniformly oriented are produced by prolonged etching but more precisely as soon as the enlargment of these dissolution figures becomes pronounced.

The discrepancy between values of the critical depth of etch corresponding to point, I, of inflexion and to the development of roughly stable etch profiles or dissolution figures on AT-31 surfaces can be interpreted as follows. The decrease in roughness parameters which, for similar initial surface damages, is more rapid for $\theta \tilde{\gtrsim} 31°$ than for $\theta \tilde{\gtrsim} 37°$ is essentially due to a decrease in the average peak to valley height. This description is still adequate for AT-33 plates even if a moderate increase in the roughness width is observed with successive etchings and if few pits associated with the history of crystal growth develop on AT-33 surfaces. In accord with description the deeply etched AT-31 plates are transparent and may be considered as chemically polished. In a different way the decrease in roughness parameters of AT-37 and AT-cut plates can be only attributed to a marked increase in the roughness and waviness spacings. Effectively, any decrease in the average peak to valley height is revealed by the surface profiles shown in figures 7 and 8. Since the formation of very stable dissolution figures is followed by an enlargment of etch pits[22] the various critical depths of etch Δd_{SC1}, Δd_{SC2} and Δd^*_{SC} must coïncide. As expected this condition is fulfilled by the data reported in tables 4 and 5.

Conclusion

This extensive study of the kinetics of etching and of the evolution of the surface texture during successive etchings by a concentrated $NH_4F.HF$

solution gives some evidence for two different behaviors associated with the orientation of the quartz plates.

By interpreting the rate data in terms of an Arrhenius equation it appears that they lie on two intersecting straight lines only for some specified orientations. Comparison of the rate data with the changes in surfaces textures with repeated etchings reveals that the absence of knee in the Arrhenius curve of some AT-31 and AT-33 plates seems to be connected with certain unstable etch profiles and etch figures which enlarge only moderately with prolonged etching. However the apparent activation energy as evaluated in the high temperature region is only slightly sensitive to the crystal orientation as expected for an attack mechanism which as suggested earlier by Ernsberger[23] involves a substitution of hydroxyl groups by fluoride ions.

But the reaction mechanism proposed by Ernsberger predicts also a dependence of the dissolution rate on the structure of the crystal plane. Thus it is not surprising that the shape and the stability of the etch patterns show orientation effects. Effectively the topography study allows us to arrange the behaviors of quartz plates in two groups. The first group corresponds to plates (AT-cut or AT-37) on which develop stable pits which enlarge with prolonged etching. The second group is illustrated by the particular behavior of AT-31 plates, the etch patterns produced on these plates present a somewhat indefinite shape and a slightly varying size.

Even if examination of the evolution of the surface texture during etching allows us to recognize two different behaviors the variations in the roughness parameters with the depth of etch seem roughly similar, i.e. with few exceptions a large decrease in roughness parameters in the course of etching. Hence care must be taken that when relatively moving etch figures develop on quartz surfaces this decrease is due to a decrease in the average peak to valley height whereas for surfaces covered with very stable dissolution figures the decrease in roughness parameters is produced by an enlargment of etch pits.

In conclusion the shape and the stability of etch patterns are in general correlated with the orientation of the crystal plane. The sensitivity to the orientation is so pronounced for cuts close to the AT cut that we give evidence for two different behaviors, the dissolution figures showing different form and varying stability as the orientation θ varies only from 31° to 37°.

References

1. J.R. Vig, W. Wasshausen, C. Cook, M. Katz and E. Hafner, "Surface preparation and characterization techniques for quartz resonators" in Proceedings of the 27th Annu. Symp. on Frequency Control, 1973, pp 98-112.

2. J.R. Vig, J.W. Lebus and R.L. Filler, "Chemically Polished Quartz", in Proceedings of the 31 st Annu. Symp. on Frequency Control, 1977, pp. 131-143.

3. H. Fukuyo and N. Oura, "Surface layer of a polished crystal plate", in Proceedings of the 30th Annu. Symp. on Frequency Control, 1976, pp 254-258.

4. D. Ang, "Design and implementation of an etch system for production use", in Proceedings of the 32nd Annu. Symp. on Frequency Control, 1978, pp 282-285.

5. J.R. Vig, R.J. Brandmayr and R.L. Filler, "Etching studies on singly doubly rotated quartz plates", in Report DELET-TR 80-5, Fort Monmouth, US Army Electrons Command, 1980, pp 1-8.

6. R.J. Brandmayr and J.R. Vig, "Further investigation of etchants for chemically polishing SC-cut quartz crystals", in Report DELET-TR 81-16, Fort Monmouth, US Army Electronics Command, 1981, pp 1-14.

7. C.R. Tellier, "Some results on chemical etching of AT-cut quartz wafers in ammonium bifluoride", J. Mater. Sci., vol 17, 1982, pp 1348-1354.

8. C.R. Tellier, "Surface texture of chemically etched AT-cut quartz plates, Surface Technology, vol 21, 1984, pp 83-89.

9. C.R. Tellier and C. Buron, "Etch figures on BT-cut quartz plates : evidence for asymmetrical etch rate distributions", Surface Technology, vol 22, 1984, pp 287-298.

10. C.R. Tellier, "Etch figures and etch rates in AT, BT, X and Y cut quartz plates" i. Proceedings of the 38th Annu. Symp. on Frequency Control, 1984, pp 105-112.

11. C.R. Tellier, "Caractérisation des surfaces de résonateurs à quartz de coupes X, Y, AT, BT et SC obtenues par usinage chimique", in Proceedings of the International Conference on Chronometry, Besançon, France, 1984, pp 115-120.

12. H. Fukuyo, N. Oura, N. Kitajima and H. Kono, "The refractive-index distribution normal to the polished surface of fused quartz measured by ellipsometry", J. Appl. Phys., vol 50, 1979, pp 3653-3657.

13. J.R. Vig, C.F. Cook, K. Schwidtal, J.W. Lebus and E. Hafner, "Surface studies for quartz resonators" in Proceedings of the 28th Annu. Symp. on Frequency Control, 1974, pp 96-108.

14. Y. Sekiguchi and H. Funakubo, "Strained surface layers of quartz plates produced by lapping and polishing and their influence on quartz resonator performance", J. Mater. Sci., vol 15, 1980, pp 3066-3070.

15. C.R. Tellier and F. Jouffroy, "Orientation effects in chemical etching of quartz plates", J. Mater. Sci., vol 18, 1983, pp 3621-3632.

16. T.R. Thomas, Rough Surfaces, London, Longman, 1981, ch. 4, pp 81-91.

17. H. Eyring and E.M. Eyring, Moderne Chemical kinetics, New York, Reinhold, 1965.

18. M.W. Wegner and J.M. Christie, "Chemical etching of deformation sub-structures in quartz", Phys. Chem. Minerals, vol 9, 1983, pp 67-68.

19. A.P. Honess, "On the etching figures of the dihexagonal alternating type, Am. J. Sci., vol XLV, 1918, pp 201-221.

20. A.P. Honess, "A study of the etching figures of the hexagonal alternating type of crystals", The American Mineralogist, vol 2, 1917, pp 57-74.

21. R.B. Heiman in J. Grabmaier (Ed.), Silicon Chemical Etching, Berlin, Springer, pp 175-220, 1982.

22. B.A. Irving in P.J. Holmes (Ed.), The Electrochemistry of semiconductors", London, Academic Press, pp 256-289, 1960.

23. F.M. Ernsberger, "Structural effects in the chemical reactivity of Silica and Silicates" J. Phys. Chem. Solids, vol 13, 1960, pp 347-351.

CHEMICALLY MILLED VHF AND UHF AT-CUT RESONATORS

J.R. Hunt and R.C. Smythe
Piezo Technology, Inc., 2525 Shader Rd.
Orlando, Florida 32804

Abstract

The highest frequency AT-cut fundamental resonators reported in the literature to date have been fabricated by Berte and co-workers [1] using non-reactive ion beam milling. Reactive ion beam milling has recently been investigated by Wang et al [2], and is potentially useful. Both techniques utilize elaborate equipment and require fine tuning of operating parameters to obtain satisfactory etching characteristics. In contrast, chemical etching, using various fluoride-containing etchants is standard practice throughout the quartz crystal industry. It is commonly used to remove surface and subsurface damage done by lapping processes and to adjust blank frequency prior to electrode deposition. The depth of etch for this purpose is normally limited to several microns. Vig and colleagues [3, 4] developed chemical polishing methods for quartz using fluorides, and suggested that they could be used to fabricate VHF resonators. Development work using this technique, repeated here, has resulted in AT-cut crystal resonators with fundamental frequencies up to 1.6 GHz, corresponding to a resonator plate thickness of just under 1 micron. The measured Q's of these units ranged from 73,000 at 100 MHz and 32,000 at 250 MHz to 5,000 at 950 MHz.

Introduction

The generation of high frequencies is required in a wide range of military and commercial VHF, UHF, and microwave systems including communications, navigation, radar, and next-generation high speed logic systems. High frequencies may be generated either directly or by frequency multiplication.

The direct generation of high frequencies offers well known advantages over frequency multiplication. First, direct generation is simpler, resulting in reduced size, weight, and power consumption [5]. Second, since frequency multiplication increases phase noise by 6 dB each time the frequency is doubled, direct generation allows improved phase noise to be obtained [6]. The availability of higher frequency resonators would in some instances allow frequency multiplication to be replaced with direct generation and in other instances would reduce the multiplication required.

Both bulk-wave and SAW resonators have uses in high frequency generation. SAW resonator advantages are maximum frequency range, power handling ability, and potentially low cost in high volume applications, due to batch processing.

High frequency bulk-wave resonators offer advantages over SAW resonators with regard to aging, temperature stability, vibration sensitivity, and acoustic noise. In addition, due to better Q, capacitance ratio (r), and figure of merit (M = Q/r), bulk-wave resonators afford circuit simplicity, improved phase noise, and high pullability. As an alternative to high fundamental frequencies, odd-order overtones of a lower fundamental may be utilized. A limitation is that the resonator impedance level increases as the cube of the overtone; consequently, practical applications are usually restricted to the first few overtones.

A limitation in the fabrication of conventional high frequency bulk-wave resonators is that the frequency is inversely related to the wafer thickness. Thus, a 100 MHz AT-cut resonator has a thickness of only 16.6 microns. Such a thin wafer is not only somewhat fragile but also impractical to fabricate by conventional means, which are normally limited to thicknesses of 30 to 35 microns (approximately 50 MHz for AT-cut resonators).

To overcome such restrictions ring-supported thickness-shear resonators were proposed by Guttwein, Ballato, and Lukaszek [7] and others. Figure 1 shows the essential features of the ring-supported structure.

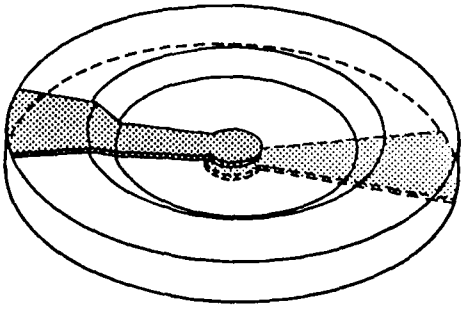


FIG. 1. RING-SUPPORTED RESONATOR

An outer ring provides a strong frame for an integral diaphragm of the required thickness. The mass of the electrodes confines acoustic energy to the central portion of the diaphragm so that the presence of the ring does not affect the resonance. Because the ring and diaphragm are a single homogeneous piece, it is possible to obtain the same frequency-temperature characteristic as for a conventional resonator.

The constraints imposed by parallelism and surface finish requirements make it impractical to fabricate such a device by mechanical lapping alone. Typically, the crystal blank is either lapped or lapped and polished to an intermediate frequency. The central

CH2186-0/85/0000-0292$1.00©1985IEEE

area is then selectively thinned by one of several techniques. The processes that have been used to accomplish this are chemical etching, reactive and nonreactive ion-beam milling, and plasma-etching. Of these, chemical etching requires the simplest process equipment and has been used as a standard crystal processing step for many years.

Chemical Milling

Chemical etching has been sucessfully used in the crystal industry to remove wafer damage caused by mechanical lapping. For this purpose, the depth of etch is typically limited to a few microns. To achieve practical benefits from chemical milling for the production of VHF and UHF resonators, up to 50 microns of material must be removed from a blank. Crystalline quartz is anisotropic and shows etch rates that vary by several orders of magnitude along the different crystallographic axes. Nevertheless Vig and co-workers [3] demonstrated that a significant reduction in surface roughness of lapped blanks could be accomplished by the prolonged etching of AT-cut quartz in saturated ammonium bifluoride solutions. The explanation given for this unexpected result was that a depletion layer exists at the etch solution/quartz interface. The factor which becomes the rate limiting step is the diffusion of etchant and reaction byproducts through this depletion layer. The diffusion is significantly slower than the inherent etch rate along any of the crystallographic axes. Further work by Vig and colleagues [4] showed that Syton-polished ST-cut quartz plates retained their polished surface after deep etching. Subsequent work in our organization has shown that it is possible using fluoride etchants to chemically mill an appropriately polished AT-cut quartz crystal with no apparent degradation of surface finish.

Normal production processes are only meant to etch several microns of material; etching, therefore, is normally done at temperatures close to ambient conditions. Figure 2 shows the observed etch rate of AT-cut quartz crystals versus temperature for saturated solutions of ammonium bifluoride.

Because the solution was allowed to go to saturation at each temperature, the resulting curve reflects a concentration increase with temperature as well as the thermally induced rate increases. In order to produce high frequency crystals in a timely manner, it was advantageous to control the etch solution at 70°C. This temperature is a compromise imposed by the physical difficulties of using fluoride-based etchants at elevated temperatures. The steep slope of the operating curve in this region requires the use of close temperature control of an etching system in order to achieve consistent results. For example, a variation in etch tank temperature of plus or minus 0.5°C, when etching from 30 MHz to 100 MHz, can result in an actual frequency ranging from 94 MHz to 107 MHz. The importance of close temperature control becomes obvious. Difficulties encountered in the handling and control of hot saturated solutions of ammonium bifluoride have led to the use of indirect heating techniques. Figure 3 is an illustration of a system used for etching crystals in hot ammonium bifluoride.

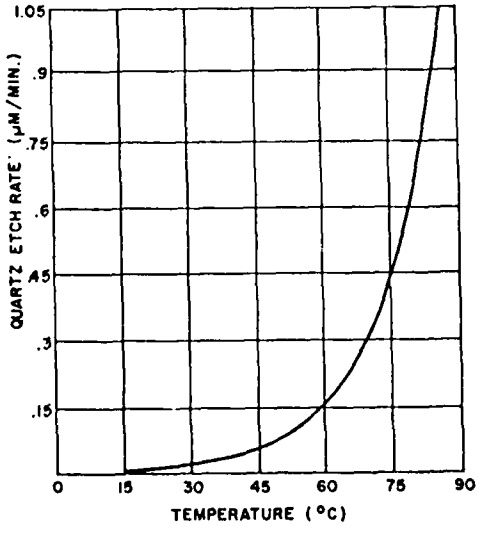
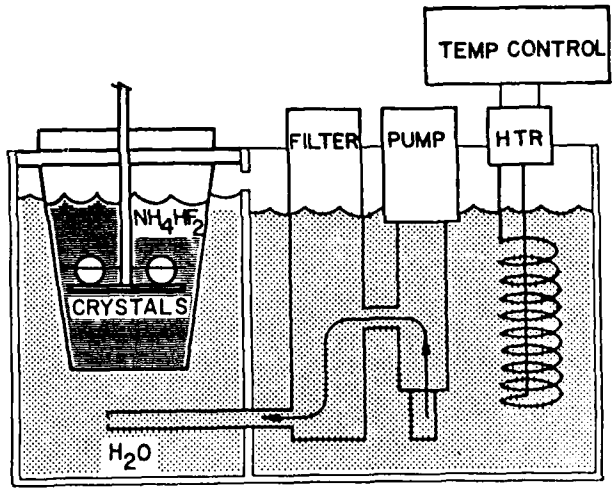


FIG. 2. ETCH RATE VS TEMPERATURE SATURATED NH_4HF_2.



AMMONIUM BIFLUORIDE ETCH

$$SiO_2 + 2\,H_2O \rightleftharpoons Si(OH)_4$$

$$Si(OH)_4 + 6\,HF \rightleftharpoons 2\,H^+ + SiF_6^= + 4\,H_2O$$

FIG. 3. AMMONIUM BIFLUORIDE ETCH SYSTEM

Fabrication Process

The starting materials and production processes used for the normal production of crystal resonators with fundamental frequencies below 50 MHz are not necessarily suitable for the production of VHF and UHF resonators fabricated by chemical milling. Early work revealed three factors which were crucial for the successful production of high frequency resonators. These were: starting material, surface finish prior to etch, and the etching process itself. Improvements were necessary in all three areas before VHF crystal production could be made practical.

Quartz

The quartz used for chemical milling must have low concentrations of impurities and low defect densities. Standard cultured quartz was found to be unacceptable as a result of the high incidence of etch channels and etch pits found after prolonged etching. The work reported here was done using both natural quartz and commercial premium Q cultured quartz which had been swept in an electrodiffusion process at PTI. Samples of unswept and swept quartz from the same lot were processed for a comparison of etch channel density. The two groups were fabricated into polished blanks approximately 55 microns thick. The blanks were then etched 32 microns under the same conditions and inspected visually for etch channel density. The unswept blanks exhibited a mean etch channel density of 481 channels/cm^2. Figure 4 is a micrograph showing a typical grouping of etch pits and etch channels found in the unswept material.

FIG. 4. UNSWEPT CULTURED QUARTZ, ETCHED 32 MICRONS, OPTICAL (80X)

The mean etch channel density for the swept blanks was 2 channels/cm^2. While all the unswept blanks had etch channels, half of the swept blanks had no etch channels at all. For comparison, the natural quartz used for this work typically had a mean etch channel density of 11 channels/cm^2. In the enlarged view of Figure 5, pairs of orthogonally oriented, fan-shaped etch pits, approximately 65 microns long by 22 microns wide, and 4 microns deep can be seen. One pit is located on the top face of the blank, and the second is on the underside. Each pair is connected by a small tubular etch channel, about 1.25 microns in diameter, visible as a dark line extending from the corner of one pit to the corresponding corner of its counterpart.

FIG. 5. UNSWEPT CULTURED QUARTZ, ETCHED 32 MICRONS, OPTICAL (200X)

No etch channels have been observed extending through the thickness of the blank without having a corresponding pair of etch pits. Numerous etch pits have been observed which are not associated with interconnecting channels. The work performed to date with swept cultured quartz indicates that it may no longer be necessary to use natural quartz in the production of VHF and UHF fundamental crystal resonators. This would allow better control over the quality of starting material and improve the consistency and repeatability of future VHF and UHF crystal production.

Surface Finish

Inadequate or improper surface preparation prior to etch results in unusable blanks, regardless of the quality of the starting material. The parallelism of the starting blanks must also be excellent, since it will not be improved by etching. Departures from parallelism affect unwanted mode performance and may degrade Q. Any surface defects produced by either the lapping or polishing process act as stress centers, lowering the effective activation energy necessary for chemical dissolution. This creates sites which preferentially etch at rates much higher than normal and may create holes completely through the membrane, Figure 6.

Blanks used for the initial etch work had been lapped with a 1 micron final abrasive. The blanks were then etched to 100 MHz fundamental frequency and fabricated into finished resonators. Etched surfaces resembled those reported by Vig and co-workers [3] and were essentially scratch-free. Figure 7 shows an SEM micrograph of the etched surface finish. A microscopically undulating surface can be seen, which remains even after etching to still higher frequencies.

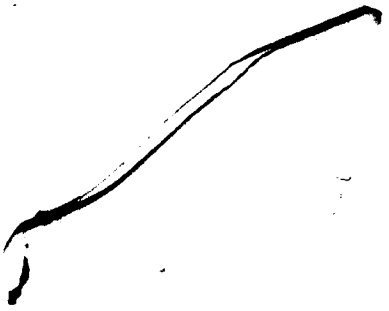


FIG. 6. ETCH FLAW IN A VHF RESONATOR.

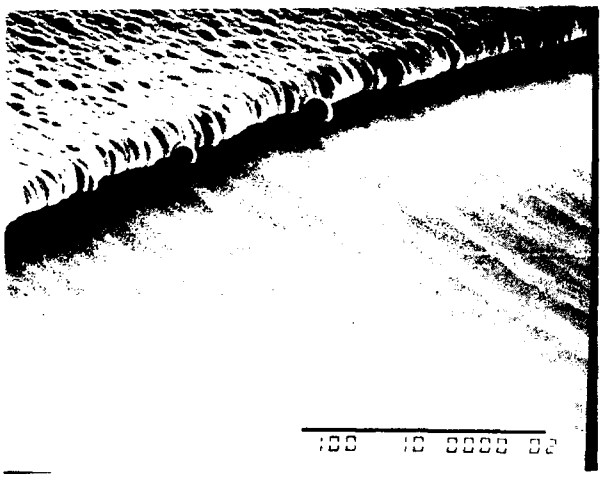


FIG. 7. NATURAL QUARTZ, LAPPED AND ETCHED TO 100 MHz, SEM (375X).

The highest observable Q on these 100 MHz resonators was 9000 with a motional resistance of 55.9 ohms. Figure 8 shows a mode scan typical of these early units. The low Q at these frequencies was attributed to the surface variations, although nonparallelism may also have been a contributing factor. (The surfaces produced were too microscopically rough to allow the observation of Haidinger's fringes under monochromatic light.)

While the standard production polishing techniques provided a surface finish which looked good to the naked eye, subsequent etching revealed much hidden damage which manifested itself as numerous fine scratches. Surface and subsurface defects which had been covered with a layer of amorphous silica and cerium polishing compound were revealed during the etching sequence. Modifications and refinements to the standard polishing process have eliminated this damage and provide surfaces with no hidden damage layers.

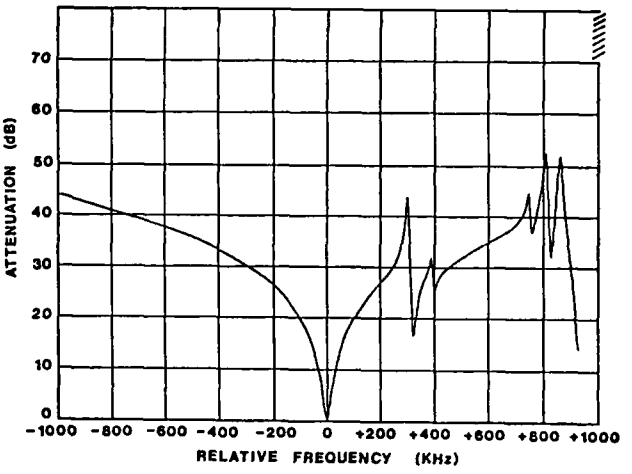


FIG. 8. TYPICAL MODE RESPONSE OF A 100 MHz FUNDAMENTAL RESONATOR FABRICATED FROM LAPPED BLANK.

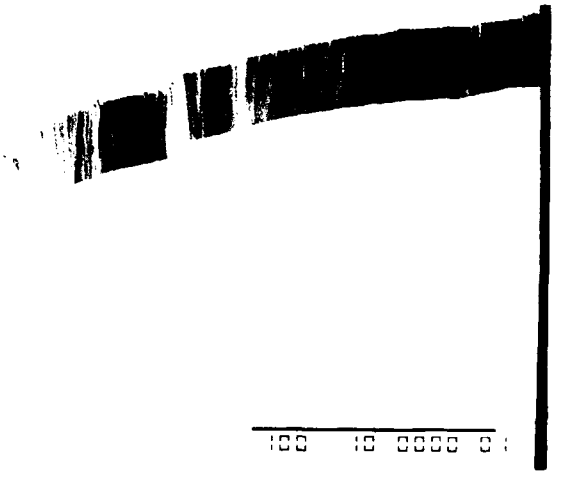


FIG. 9. NATURAL QUARTZ, POLISHED AND ETCHED TO 250 MHz, SEM (375X).

Figure 9 shows an SEM micrograph of the membrane of a 250 MHz fundamental resonator fabricated from an etched blank which had been polished using this technique. The etched surface at the bottom of the well appears as smooth and without detail as the unetched region in the upper lefthand portion of the photograph. AT-cut quartz crystals have been produced with excellent polished surfaces showing no apparent degradation even after etching to membrane thicknesses less than 1 micron.

To compare with the earlier lapped crystals, a group of carefully polished crystals was etched to a fundamental frequency of 100 MHz. The blanks were processed into finished resonators with measured Qs ranging up to 73,000 and motional resistances as low as 18 ohms. A mode plot typical of these units is shown in Figure 10.

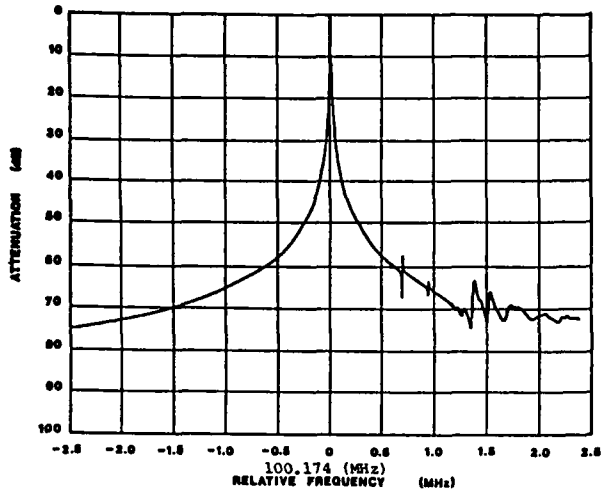


FIG. 10. FUNDAMENTAL MODE RESPONSE RESONANCE FREQUENCY 100.174 MHz, RESISTANCE 22.9 OHMS, Q 65K.

All subsequent etching of VHF and UHF resonators was performed with polished quartz blanks.

Etch Procedure

All chemical milling for this work was done using the apparatus illustrated earlier in Figure 3. A wide variety of chemical solutions was evaluated for etching use. These included various concentrations of ammonium bifluoride, ammonium fluoride, and hydrofluoric acid. Also, mixtures of hydrofluoric acid and ammonium fluoride, commonly referred to as buffered etches, were used. These were tried with and without various additives in attempts to improve one factor or another. Only the solutions strong in hydrofluoric acid proved unsuitable for use.

Prior to the etching step it was mandatory that all surface contamination be removed in a thorough cleaning sequence. Any film or debris left on the diaphragm area can act as an etch mask and cause non-uniform etching. Once this begins to occur, it cannot be corrected by subsequent washing. The crystal affected usually suffers from unwanted mode responses or a loss of crystal activity.

When etching a blank to frequency it is necessary to perform a rough etch, read the resulting frequency, and repeat this sequence until the desired end point is achieved. Often it is desirable to perform the final adjusting etches at a lower temperature to achieve better control while lowering the danger of overshooting.

Process Sequence

Figure 11 is a block diagram showing a simplified process sequence for VHF and UHF crystal fabrication.

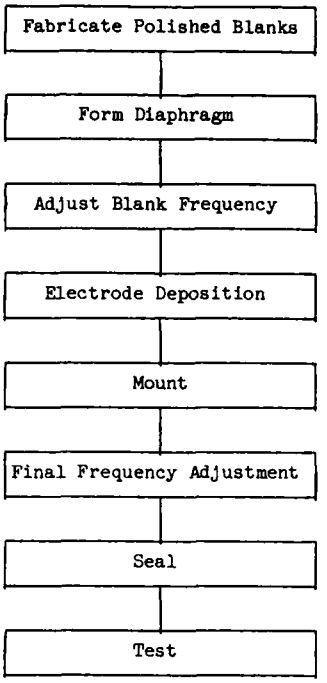


FIG. 11. SIMPLIFIED PROCESS SEQUENCE

The initial step is the selection of adequate starting material, either natural or suitably swept cultured quartz. The processing steps of x-ray, sawing, and rough lapping are the same as those used for the production of any resonator. It is after the lapping stages that the VHF and UHF production processes differ from those of lower frequency units. Blanks intended for chemical milling are carefully polished to remove all traces of previous lapping damage. They are then cleaned and resist is applied to provide for the selective etching of the diaphragm structure, leaving a strong outer ring for structural support. After etching, the resist is stripped from the blanks, they are cleaned, and electrodes are applied. Because of the very small electrode geometries, great care must be taken in the production and alignment of masks used for VHF and UHF resonators. The electroded blanks are mounted, adjusted to final frequency, and sealed. It is an indication of the inherent strength of the quartz that 65-80% of the units going into etch usually survive to become finished resonators. This includes crystals etched from 30 MHz to over 1 GHz by chemical milling.

Experimental Results

Chemical milling was successfully used to produce fundamental AT-cut resonators ranging in frequency from 100 MHz to 1655 MHz. Table 1 presents the equivalent circuit parameter values for fundamental modes of selected units, while Table 2 gives representative of third overtone measurements.

TABLE 1
Measured Parameters Of Fundamental Mode Resonators (*Denotes Estimated (Q Measurements)

F_s (MHz)	R_1 (Ohms)	L_1 (mH)	C_1 (fF)	C_o (pF)	Q ($x10^{-3}$)
100.387	18.01	2.08	1.2061	0.84	73
150.846	34.65	1.80	0.6188	0.73	49
244.889	37.19	0.73	0.5792	0.73	30
503.400	48.47	0.20	0.5049	1.70	13
620.121	33.22	0.08	0.8739	1.91	9
735.172	58.30	0.08	0.5904	1.11	6
843.145	27.65	0.03	1.1814	1.17	6
954.273	29.85	0.02	1.1461	1.02	5
1233.436					3.8*
1400.25					2.7*
1655.3					3.7*

TABLE 2
Measured Parameters Of Third Overtone Resonators

F_s (MHz)	R_1 (Ohms)	L_1 (mH)	C_1 (fF)	C_o (pF)	Q ($x10^{-3}$)
249.663	133.31	4.02	0.1011	0.77	47
450.258	288.42	2.72	0.0459	0.72	27
749.676	418.47	1.27	0.0356	0.78	14

Parameter measurements were made using an automatic measurement system utilizing the HP 4191A Impedance Analyzer [7] for all resonators up to the equipment limit of 1000 MHz. The units above 1000 MHz were measured with a Polarad ZPV vector analyzer in conjunction with an HP 8662A synthesized signal generator or a Tektronix 496P spectrum analyzer together with a Tektronix TR 503 tracking generator. In Figure 12 the best Q's achieved experimentally are compared with the estimated material Q [9, 10]. Figure 13 is presented as an aid in visualizing the dimensions of a 1655 MHz resonator fabricated on a .250 inch diameter blank. The actual dimensions of the diaphragm of this resonator are .050 inches in diameter and 0.9 microns thick. The cross-sectional view shows a scaled-up representation of the ring and diaphragm. Using this scale, the total blank diameter would be 60 feet. Figure 14 is a mode plot of a 150 MHz fundamental resonator, while Figure 15 shows the third overtone mode response of a crystal from the same process batch. Figures 16 through 21 are representative mode plots of some of the VHF and UHF AT-cut resonators produced. The electrode geometry used on many of the UHF models was not optimized either for maximum Q or suppression of spurious modes.

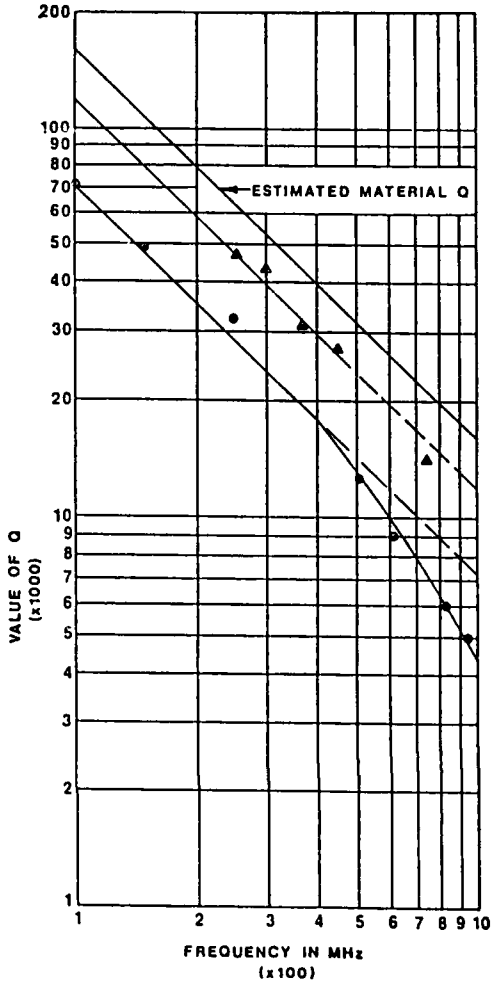


FIG. 12. MEASURED VALUES OF Q FOR FUNDAMENTAL (●) AND THIRD OVERTONE (▲) AT-CUT RESONATORS

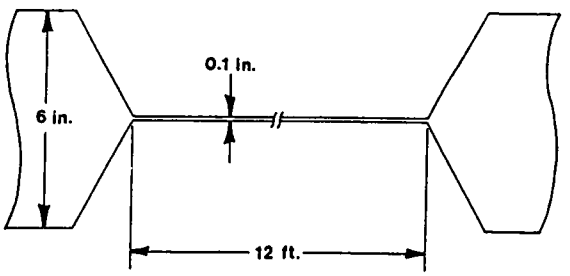


FIG. 13. CROSS-SECTIONAL SCALE DIAGRAM 1655 MHz FUNDAMENTAL RESONATOR.

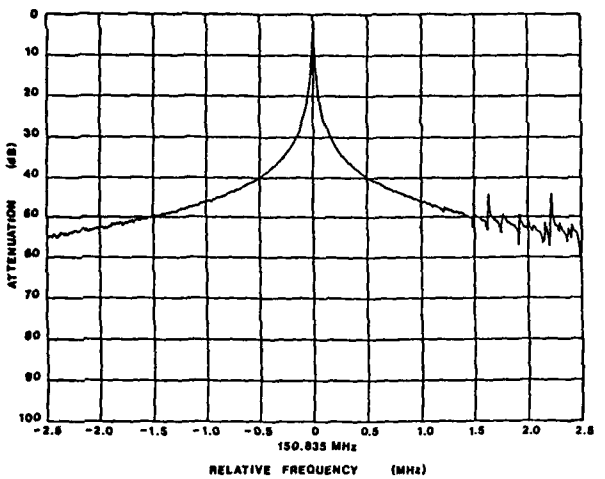


FIG. 14. FUNDAMENTAL MODE RESPONSE
RESONANCE FREQUENCY 150.834 MHz,
RESISTANCE 40.3 OHMS, Q 40K.

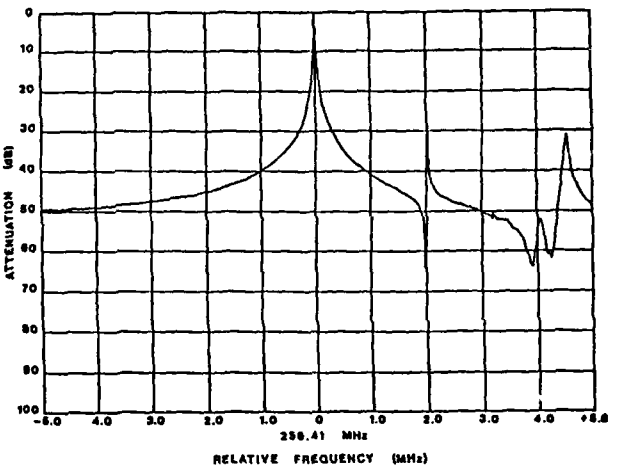


FIG. 16. FUNDAMENTAL MODE RESPONSE
RESONANCE FREQUENCY 256.41 MHz,
RESISTANCE 43.4 OHMS, Q 27K.

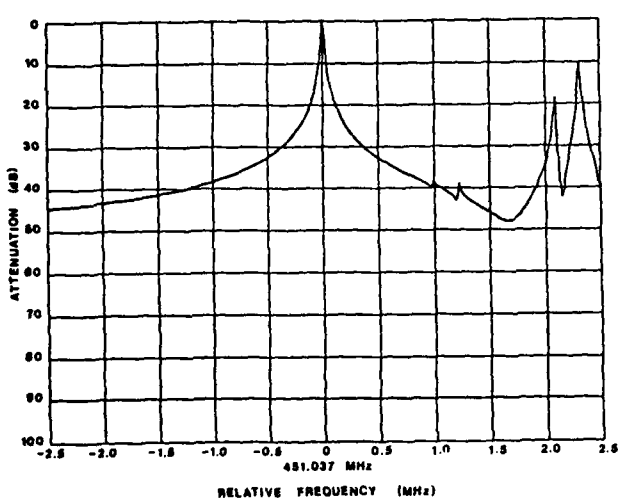


FIG. 15. THIRD OVERTONE RESPONSE
RESONANCE FREQUENCY 451.04 MHz,
RESISTANCE 323.0 OHMS, Q 25K.

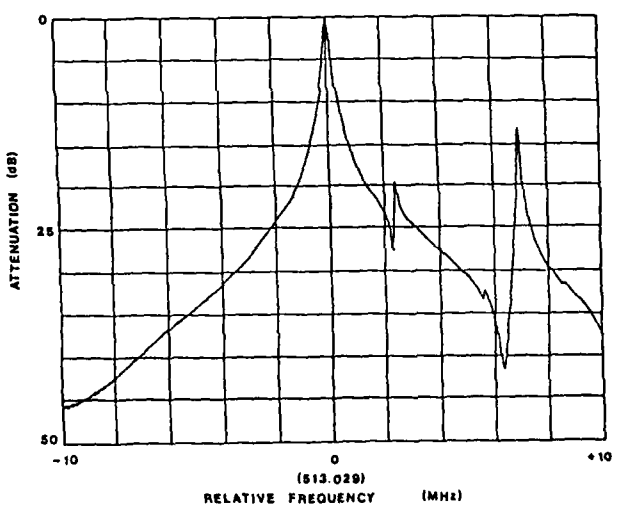


FIG. 17. FUNDAMENTAL MODE RESPONSE
RESONANCE FREQUENCY 513.029 MHz,
RESISTANCE 43.4 OHMS, Q 8K.

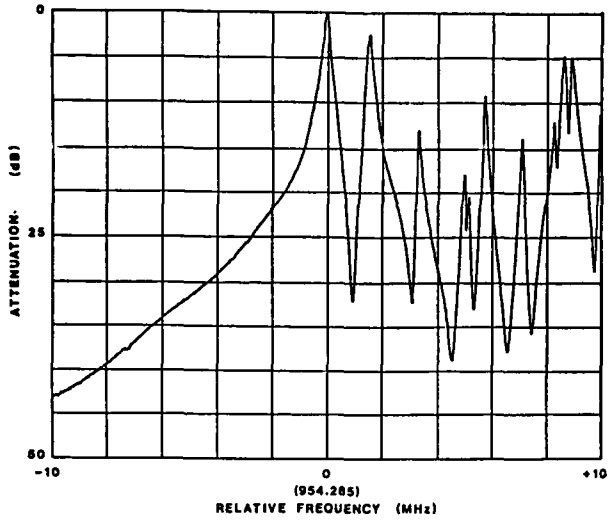


FIG. 18. FUNDAMENTAL MODE RESPONSE
RESONANCE FREQUENCY 954.285 MHz,
RESISTANCE 29.8 OHMS, Q 5K.

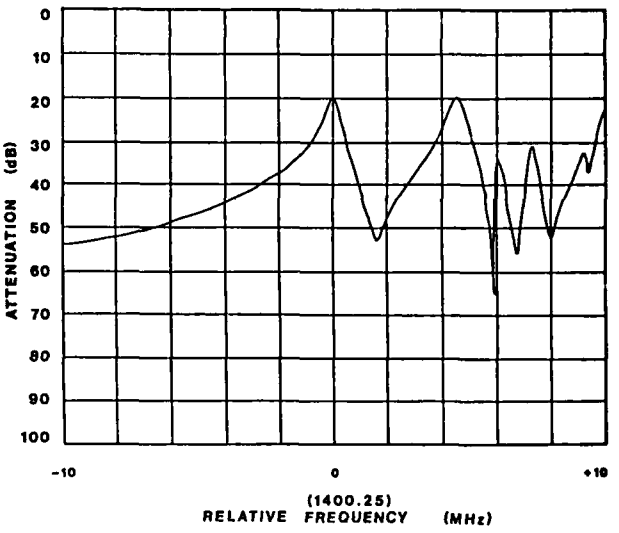


FIG. 20. FUNDAMENTAL MODE RESPONSE
RESONANCE FREQUENCY 1400.25 MHz,
ESTIMATED Q 1.6K.

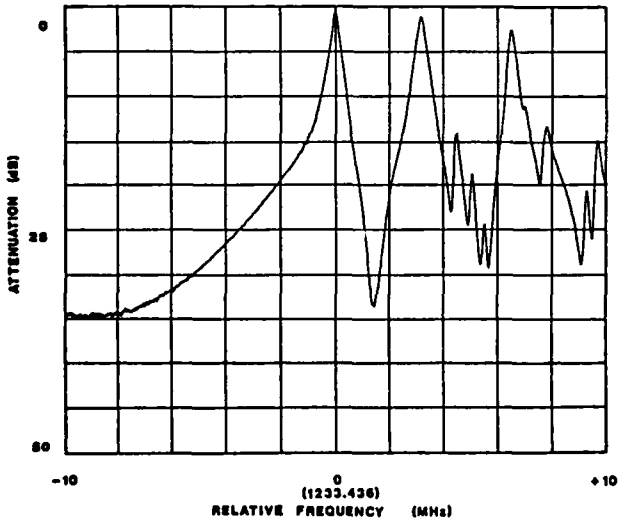


FIG. 19. FUNDAMENTAL MODE RESPONSE
RESONANCE FREQUENCY 1233.436 MHz,
ESTIMATED Q 1.4K.

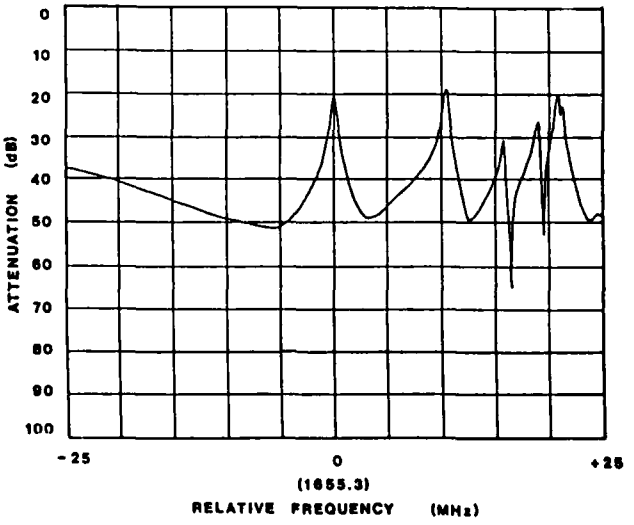


FIG. 21. FUNDAMENTAL MODE RESPONSE
RESONANCE FREQUENCY 1655.3 MHz,
ESTIMATED Q 2.2K.

Conclusion

High fundamental frequency AT-cut resonators are potentially useful for VHF, UHF, and microwave frequency generation as well as for VHF and UHF crystal filters. The work performed to date has demonstrated the feasibility of using wet chemical etching techniques to fabricate resonators at fundamental frequencies into the gigahertz range. The processes used lend themselves to practical manufacturing use. The work performed to date with swept cultured quartz indicates that it can be used for the production of VHF and UHF fundamental crystal resonators by etching. This will allow better control over the quality of the starting material and favorably improve the consistency and repeatability of future VHF crystal production.

Acknowledgements

The authors wish to thank E. McCluskey and V. Quach for their efforts in processing the resonators described in this paper, and to R. Yap, University of Central Florida, for providing SEM micrographs of the crystal surface finish.

This work was supported in part by USAF Rome Air Development Center contract F19628-83-C-0061, Hanscom AFB, monitored by Herbert G. Lipson.

References

[1] M. Berte, "Acoustic Bulk Wave Resonators and Filters Operating in the Fundamental Mode at Frequencies Greater than 100 MHz," in Proceedings of the 31st Annual Frequency Control Symposium, 1977, pp. 122-125.

[2] J.S. Wang, S.K. Watson, and K.F. Lau, "Reactive Ion Beam Etching for VHF Crystal Resonators," in Proceedings of the 38th Annual Frequency Control Symposium, 1980, pp. 101-104.

[3] J.R. Vig, J.W. LeBus, and R.L. Filler, "Chemically Polished Quartz," in Proceedings of the 31st Annual Frequency Control Symposium, 1977, pp. 131-143.

[4] J.R. Vig, R.J. Brandmayr, and R.L. Filler, "Etching Studies on Singly and Doubly Rotated Quartz Plates," in Proceedings of the 33rd Annual Frequency Control Symposium, 1979. pp. 351-358.

[5] R.D. Colvin, "UHF Acoustic Oscillators," Microwave J., v. 23, no. 11, pp. 22-23, November, 1980.

[6] C. Pegeot, and G. Sauvage, "UHF Oscillator Using SC-Cut Quartz Crystal with Low Noise Performance & High Long Term Stability," in Proceedings of the 34th Annual Frequency Control Symposium, 1980, pp. 233-236.

[7] G.K. Guttwein, A.D. Ballato, and T.J. Lukaszek, "VHF-UHF Piezoelectric Resonators," U.S. Pat. 3,694,677; 26 September, 1972.

[8] R.C. Smythe, "An Automated Resonator Measurement System Using a Reflection Coefficient Bridge," in Proceedings of the 35th Annual Frequency Control Symposium, 1981, pp. 280-285.

[9] H.E. Bommel, W.P. Mason, and A.W. Warner "Expermental Evidence for Dislocations in Crystalline Quartz," Phys. Rev., v.99, pp. 1894-1896; 1955.

[10] W.P. Mason, Physical Acoustics and the Properties of Solids. New York, N.Y.: D. Van Nostrand Co. Inc., 1958, pp. 287-289.

ETCH PROCESSING OF BULK AND SURFACE WAVE DEVICES

John Dowsett, Douglas F G Dwyer
and Francesca Stern

STC Components Ltd, Quartz Crystal Unit,
Harlow, Essex CM20 2DE, England

R A Heinecke and A H Truelove

STC Technology Ltd
Standard Telecommunication Laboratories

Introduction

Etch processing of bulk wave devices is a developing technique where the upper frequency limitation of the mechanical processes can no longer be applied. However, the use of chemical or plasma processes imposes other disciplines with regard to the quality of the material and to the mechanical processes preceding the final etch. This paper discusses the techniques and the results achieved in the manufacture of bulk wave resonators up to 150 MHz in the fundamental mode using the AT-cut.

The problems of material quality are related to the structure of the quartz crystals. The methods of achieving an adequate material quality in synthetic quartz includes the use of low growth rates, high quality "defect free" seed material and electrolysis (sweeping) to achieve a low level of tunnelling into the material by the etchant used.

The second part of the paper discusses plasma etch techniques for the production of SAW devices. Etch techniques are described which produce devices with 0.5µM features in 1000 Å aluminium. Graphs are included which show the electrical performance of 1 GHz resonators fabricated by such methods.

PART 1

Etch Processing of Bulk Wave Devices

Material and Specification

Any etch process in quartz, which removes amounts which exceed 2-3µM when processing plates of 60µM thickness or less, requires material with a high degree of purity and freedom from structural defects.

When work began on this process it became evident that most synthetic quartz crystal currently available is not suitable. Heavy etch tunnelling occurred and most material would not produce tunnel free characteristics. Typically, etch tunnels of 500-1000/sg cm were experienced. Figure 1 shows tunnelling within an area of a typical electrode for a standard commercially available synthetic quartz of minimum Q of 1.8×10^6.

Growing Methods

Process trials for the growth of suitable material were carried out and various factors investigated. A programme of growth runs has been carried out to establish the effect of growth rate, seed material, nutrient purity and autoclave cleanliness. It is established that slow growth rates produce high Q, and Q's of 3×10^6 are now being produced. However, this high Q material did not produce the low level of etch tunnelling required, although the tunnelling is reduced by about 2:1. See Figure 3.

Further investigations confirmed that a key element in producing lower dislocation density is the quality of the seed material.

The most successful material so far produced is that grown on natural seed. See Figure 5. This material is producing etch tunnel rates of less than 25/sg cm, some material is producing nearly zero etch tunnelling.

Most quartz growing is carried out on seeds cut from synthetic quartz, which may be many generations from the original natural seed. These seeds have high levels of dislocations from which further dislocations are generated during the growth process. See Figure 2.

First generation seed material is also capable of producing good material and Figure 4 shows the dislocation in such a block. It will be noticed that a fracture in the seed plate produces severe local dislocations.

All of the materials discussed in this paper have been grown in a well conditioned steel autoclave without special preparation or cleaning.

Sweeping

However, the production of quartz from natural seed relies on the availability of a limited supply of suitable material. It has been noticed by us and others* that swept quartz also produces low etch tunnel rates and investigations into sweeping techniques have been carried out.

CH2186-0/85/0000-0301$1.00©1985IEEE

Quartz with a high dislocation density has been swept using both vacuum and nitrogen atmospheres. Both methods will produce tunnel free elements and results have been achieved equal to that of quartz grown on natural seed. Further investigations have shown that quartz grown on swept multigeneration seed also shows improvement in etch tunnel counts and it is now possible to generate an adequate supply of good synthetic seeds from existing stocks of material.

Table 1 shows the purity analysis of quartz of various pedigrees. It can be seen that significant changes in impurity levels can be produced by good seed material and sweeping.

A typical specification for quartz, which will give good results when etch processing quartz blanks, is set out below.

a) Growth rate less than 0.35mm/day.

b) Q greater than 2.2×10^6.

c) Natural or swept seed.

d) Etch tunnel count in an AT-cut blank less than 10/sq cm.

The specification for etch tunnelling is set at the level of 10/sq cm as this gives a maximum of one tunnel per blank for a 5mm dia element with a 2mm electrode.

Preparation of the Quartz Blanks Prior to the Etching Process

The basis of high frequency fundamental bulk wave resonators are 5mm diameter AT-cut blanks prepared by conventional lapping processes up to 18 MHz and conventional crystal polishing up to 23.5 MHz.

All blanks are, however, hand-edged and stored individually between subsequent processes in order to avoid any surface damage arising from blank-to-blank contact. Likewise, any handling after the hand-edging stage is carried out using vacuum tweezers.

As surface finish and degrees of flatness and parallelism are critical in high frequency devices, much attention has been paid to the final polishing stage prior to the etching process. The criteria for VHF crystal blanks are a scratch-free central area and a thickness variation of less than 0.05μm, or 1 light fringe. A Haidinger fringe interferometer, employing a He/Ne laser, has been devised for assessing thickness variation and at the same time has proved a reliable method of assessing surface scratches. Several blank polishing rouges have been evaluated both for polishing performance and particle size distribution. It has been found that rouges with particle sizes less than 5 μm give optimum results.

After polishing, blanks are given a 10 minute etch in saturated ammonium bifluoride solution at 62.5°C to reveal clearly the degree of scratching.

Figure 6 (a) shows a polished blank before etching. Figure 6 (b) shows a blank which has been polished with a satisfactory rouge and then etched. Note that there is negligible loss of flatness on etching and the scratch-free central area. Some edge scratches from chipping are unavoidable. Figure 6 (c) demonstrates the scratching that can result from a rouge with a proportion of particles (10%) in the 10 m region. The lack of parallelism in this blank would also make it unacceptable for VHF processing.

Final Etching Process to Obtain the Finished Blank

Having passed the scratch and parallelism inspection, blanks may be liquid etched either to any frequency up to 75 MHz that is required. The liquid etching set-up is the same as that used for the 10 minute scratch reveal etch. Blanks are supported vertically in individual pockets of purpose designed teflon jigs, which may be double stacked to provide an adequate batch capacity.

The etchant is saturated ammonium bifluoride solution at 62.5°C and the jigs are gently raised and lowered through 10mm in the etchant during the cycle to ensure continuous replenishment of etchant at blank surfaces. The blanks are etched to near the required frequency, then frequency sorted and trimmed into frequency specification, again using the same etch conditions.

This process has proved very successful up to 60 MHz; above 60 MHz the thinness of the blanks has caused some handling problems, making the production of recessed blanks highly desirable.

Final Process Methods to obtain Package Resonators

The final processing of the etch thinned blank follows conventional technology. Special care is taken to ensure a high level of cleanliness, particularly with pre-metalisation cleaning which was 50/50 hydrogen peroxide/deionised water and the usual range of ultrasonic rinses in deionished water and isopropyl alcohol.

Electrodes of silver or aluminium are deposited by evaporation. Good film adhesion is required, therefore, a clean vacuum system is used with substrate heating to +100°C.

The plated element is usually mounted on either an HC45 or TO5 header and finally adjusted again in the conventional way using evaporators.

Blank handling requires care, but advantages of blank strength have been achieved by the use of high quality material, quartz strength is high when using low dislocation material.

The present practical limitation for wet etched blanks is in the region of 80 MHz.

Device Parameters

The results which have been achieved are set out below. The parameters of C1 (motional capacitance) and Q are analysed for batches of units produced, process consistency and control can be measured from these results. All these results have been achieved on 5mm diameter blanks. Figure 7 shows the relationship between C1 and electrode area. Control of C1 within 2% at 61 MHz, provided blank thickness variation is better than 50nm over 2mm, can be achieved. The plating is approximately 60nm of silver, ie. plateback of 1.5%. The offset represents edge effects.

Figure 6 illustrates that thickness variations become difficult to control above 80 MHz and Figure 9 gives an indication of effect of wedging in the blank, as determined by the Haidinger fringer using a He/Ne laser. As the device diverges from parallelism the C1 control worsens.

Figure 10 summarises the Q limitations and the results so far achieved.

Figure 11, Table 2 gives typical Q and C1 results achieved on fundamental, third overtone and fifth overtone mode.

Figure 12 shows some early results on a monolithic element at 45 MHz. The bandwidth and unwanted separation is not possible with an overtone design.

Conclusion

It has been shown that bulk wave devices operating in the fundamental mode up to 150 MHz can be manufactured by etch thinning techniques. Good parameter control can be achieved if good quality material is used. This control becomes more difficult above 70 MHz but provided blank features and wedginess can be controlled frequencies up to 150 MHz are possible.

PART 2

Plasma Processsing of SAW Components

Introduction

Surface acoustic wave technology is used to create frequency control and filtering devices over the range 30 MHz to 1.2 GHz. The lower frequency limit is determined by increasing pattern size and the availability of alternative bulk acoustic wave devices.

The upper frequency limit is determined by the minimum resolution at which the aluminium patterns can be defined. Where wet etching is employed this upper limit is about 1 GHz for devices which employ lambda/4 structure.

This frequency limit corresponds to about 0.78µM gaps for the finger width and calls for a high level of skill in the wet etch process.

Plasma etching provides a process which can be operated under precise control and has a capability of etching patterns with 0.5µM features in 1000 Å aluminium.

Figure 13 shows the fabrication process path and the three stages at which a plasma process can replace a wet process.

SAW Fabrication

In general, a SAW device consists of blocks of interdigitated transducers. These transducers are fabricated using standard photomechanical techniques in thin film aluminium on the surface of a polished slice of crystalline quartz. the slice is orientated very precisely with respect to the quartz atomic lattice to achieve a minimum frequency change over the required temperature band.

Transducer patterns are defined by a standard photomechanical process.

Figure 14 - The deposition of the aluminium film onto the slice is done by an electron beam evaporator in a vacuum in the range 10^{-6} to 10^{-7} TORR. Great care is applied to the cleaning of the slice prior to evaporation and also to minimising the variation in aluminium thickness across the slice and from slice to slice.

A thin layer of photoresist is spun onto the slice, baked and UV hardened. The transfer of the pattern from mask to slice is done by UV illumination in a direct contact mask aligner. Good contact over the whole area is essential to maintain finger to gap ratios within a narrow range, and both the use of relatively thin aluminium and uniform resist thickness contribute to faithful reproduction of the SAW pattern. The pattern is developed in AZ1350 developer and after hardening and UV exposure is ready for aluminium removal by plasma etching with a chlorinated gas mixture.

For good control the RF excitation power is adjusted during the process. Careful layout of the reactor is necessary to achieve good uniform gas distribution.

Plasma Etching

The plasma is used to etch away the unwanted regions of aluminium, the plasma chemistry and RF excitation are adjusted during the process to control etch rate and minimise quartz removal.

Gas dynamics and electrode configuration are arranged for maximum uniformity of etch.

Figures 15 and 16 show the comparison in definition between wet etched and plasma etch electrode structures.

Plasma Cleaning

An oxygenated plasma is used to remove the unwanted resist layer. The resist, which is composed largely of carbon and hydrogen, oxidises to form CO_2 and H_2O. Any other hydrocarbon contaminants will be removed at the same time.

Frequency Adjustment

The tolerance that can be achieved in a batch of SAW filters or resonators is typically ±100 ppm.

The mean frequency of a batch is affected by run-to-run variations in the evaporated aluminium thickness and by variations in the pattern mark-to-space ratio. The latter are due to changes in the photoresist exposure and development processes.

Precision frequency adjustment can be carried out after resist removal by using a fluorine species plasma to etch away quartz between the aluminium regions. This can be carried out at the slice level (open loop) or on assembled devices in a closed loop system.

Figure 17 shows the degree of adjustment available by this technique.

Figures 18, 19 and 20 show typical 1 GHz SAW resonator responses which are being achieved using the techniques described.

Conclusion

This report has described the application of etch processing to SAW and BAW devices. These evolving techniques, drawn from the semiconductor industry, contribute to the upward trend in maximum frequency over volume manufacture of both technologies. Their application to the processing of BAW structures has highlighted the limitations of commonly available synthetic quartz, and the improved material that is currently prepared for our BAW structures could, in turn, serve to increase the performance of SAW devices.

Acknowledgements

The authors would like to express their thanks to the following people for their help in carrying out this work.

L H Feasey, Dr A F B Wood and R W T Rabbetts for material development, D Bower and I Llewelyn for SAW process development, D Cox for BAW processing and R J Williamson for BAW parameter measurement.

References

* J C Gualtieri and J R Vig, "Sweeping and Irradiation Studies in Quartz", Proc. 38th Annual Symposium on Frequency Control, 42 (1984).

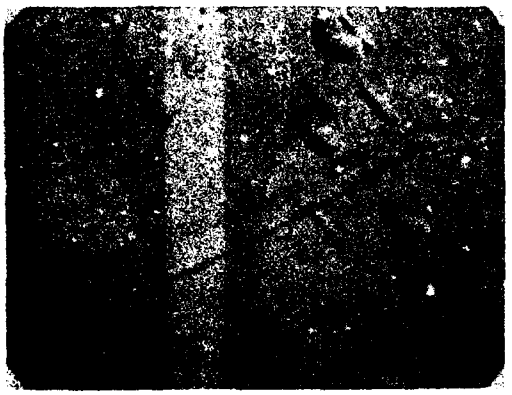


FIG.1 Surface of Polished Etched Quartz showing Tunnel Pits

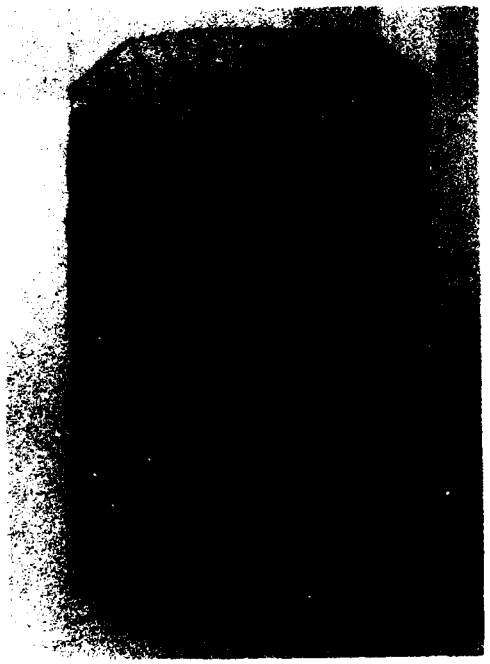


FIGURE 2

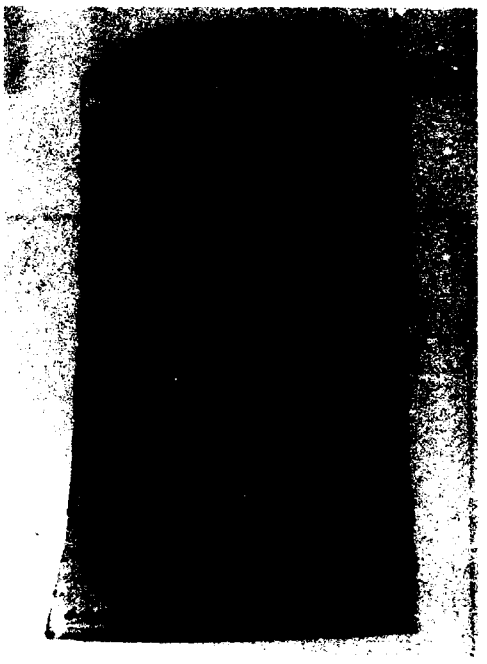


FIGURE 3

FIGURE 4

FIGURE 5

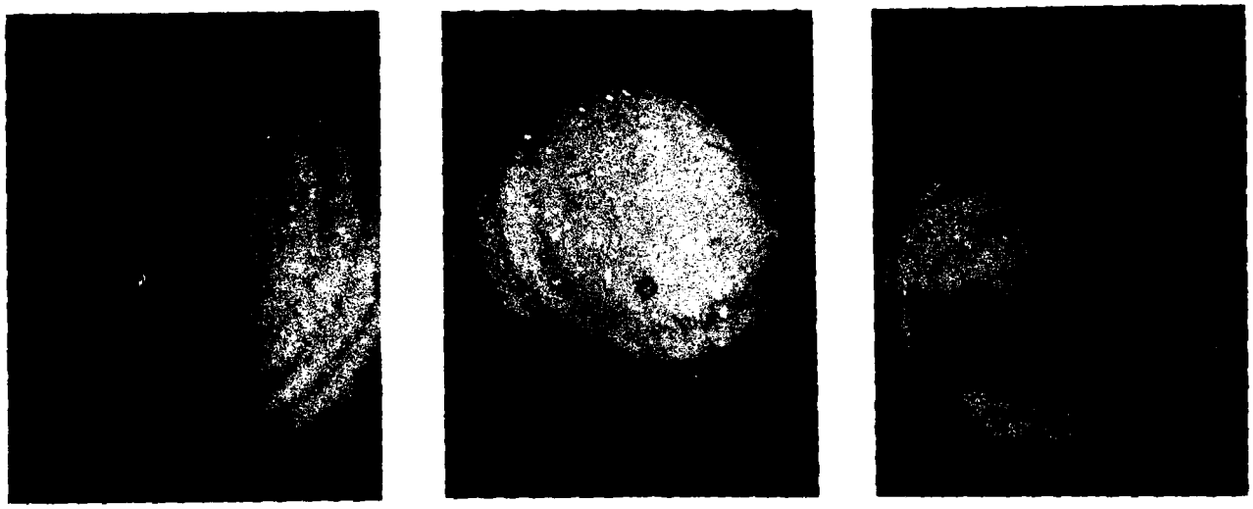
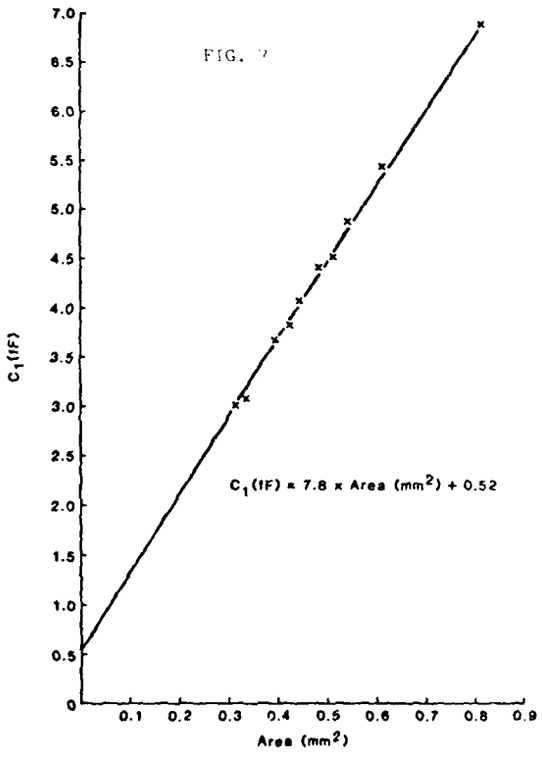


Fig. 6 (a) Polished Blank b) Polished, Etched Blank c) Polished, Etched Blank

IMPURITY	STC PREMIUM UNSWEPT MULTIGENERATION SEED	STC PREMIUM SWEPT IN NITROGEN FOR 2 WEEKS MULTIGENERATION SEED	STC PREMIUM VACUUM SWEPT MULTIGENERATION SEED	STC PREMIUM UNSWEPT NATURAL SEED
Li	4.86	[illegible]	[illegible]	[illegible]
Cs	9.19	[illegible]	[illegible]	[illegible]
Na	10.85	9.12	[illegible]	[illegible]
K	24.93	11.85	[illegible]	[illegible]
Ca	54.53	[illegible]	[illegible]	[illegible]
Fe	[illegible]	[illegible]	[illegible]	[illegible]
Zn	[illegible]	[illegible]	[illegible]	[illegible]
Y	[illegible]	[illegible]	[illegible]	[illegible]
Cu	[illegible]	[illegible]	[illegible]	[illegible]
Mn	[illegible]	[illegible]	[illegible]	[illegible]

TABLE 1

FIG. 7



Dynamic capacitance versus resonator area at 61 MHz

FREQUENCY MHz		DIAMETER mm	C1 fF	Q 130°	MATERIAL & CUT
40	F	0.85	5.15	60	AT
50	F	0.85	4.2	55	AT
65	F	0.85	4.4	40	AT
75	F	0.6	2.8	50	AT
100	F	0.65	2.5	35	AT
150	3OT	0.85	2.42	55	AT
250	5OT	0.85	3.14	36	AT

STC PARAMETERS OF ETCHED FUNDAMENTAL UNITS (PLATED ABOUT 6m ANGSTROMS)

TABLE 2

Electrode area 0.85 μm diameter
Thickness 2 x 7nm silver

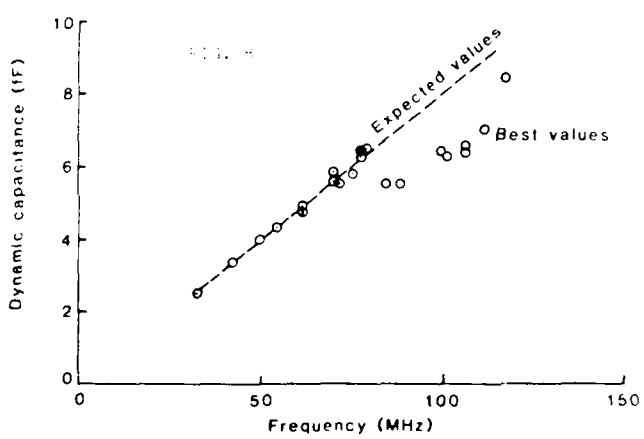
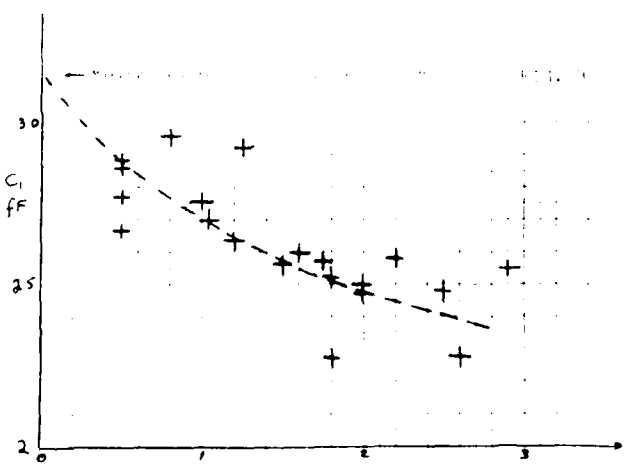
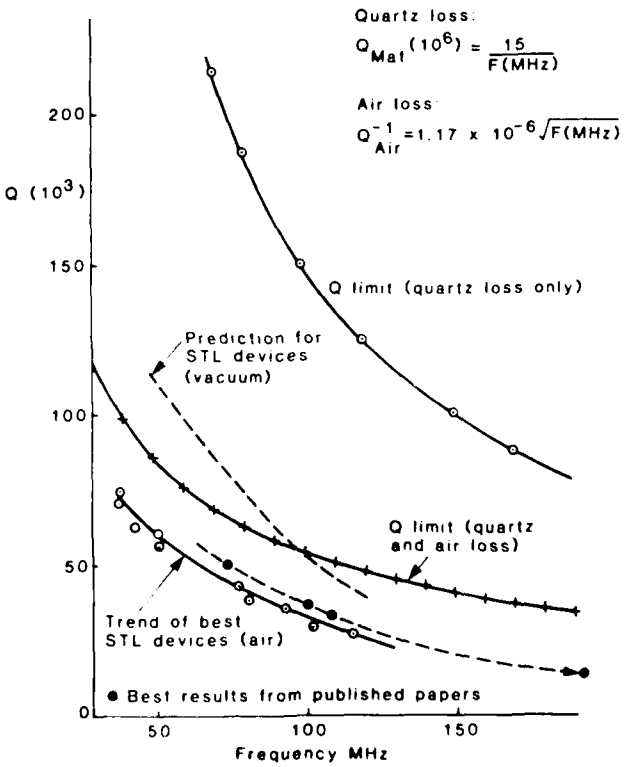


FIG. 14





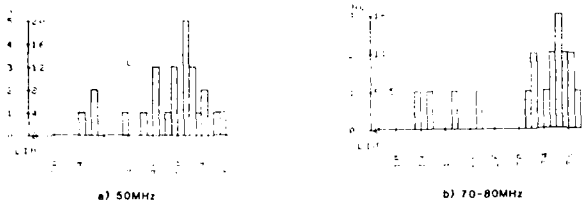
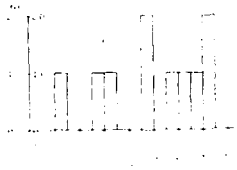


FIG. 11



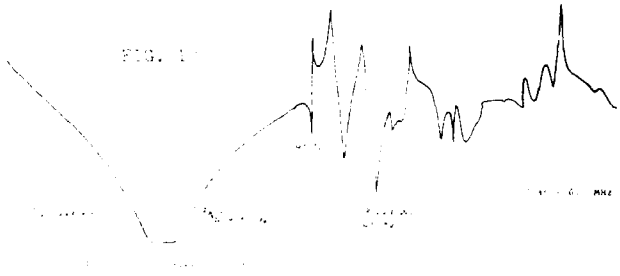
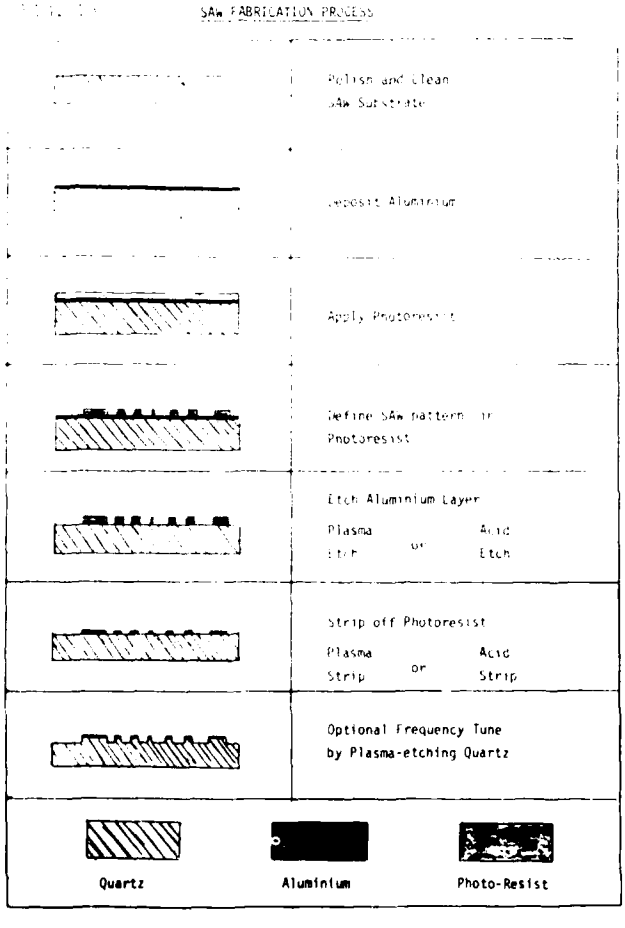


FIG. 12



Plasma Etching — Schematic

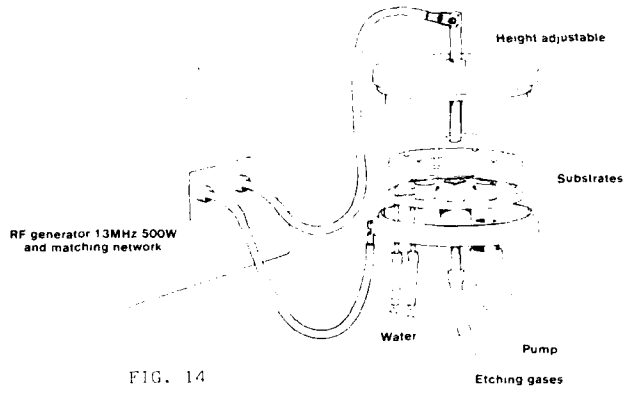
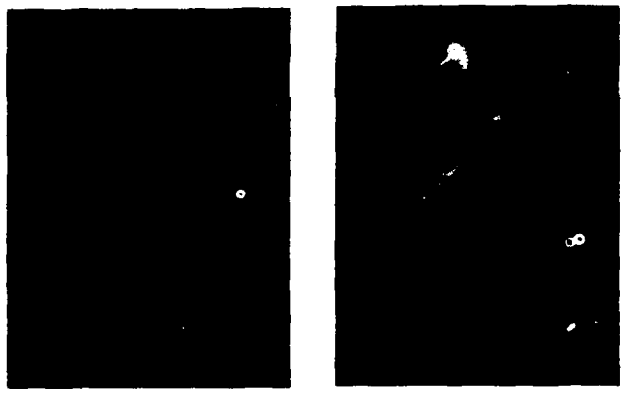


FIG. 14

Fig. 15 Wet Etched



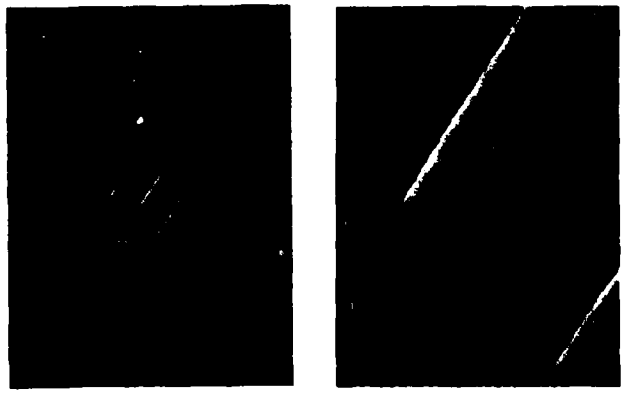
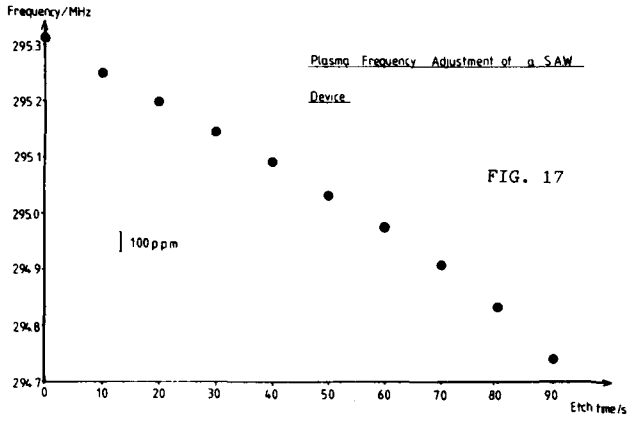


Fig. 16 Plasma Etched



Plasma Frequency Adjustment of a SAW Device

FIG. 17

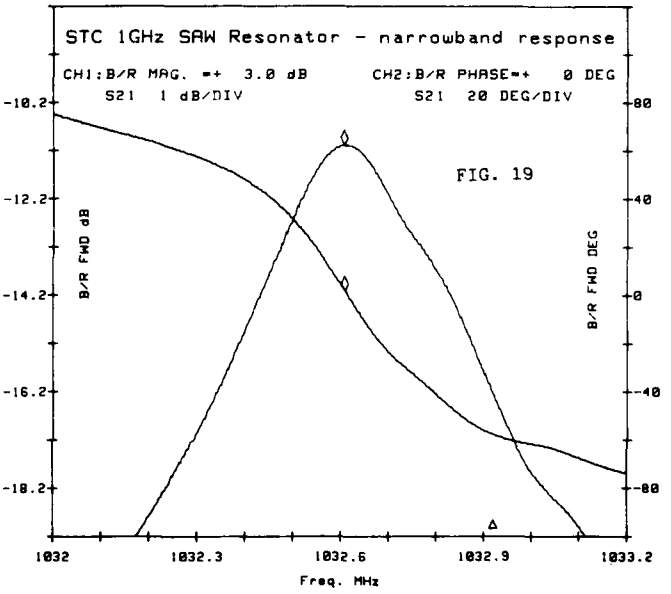


FIG. 19

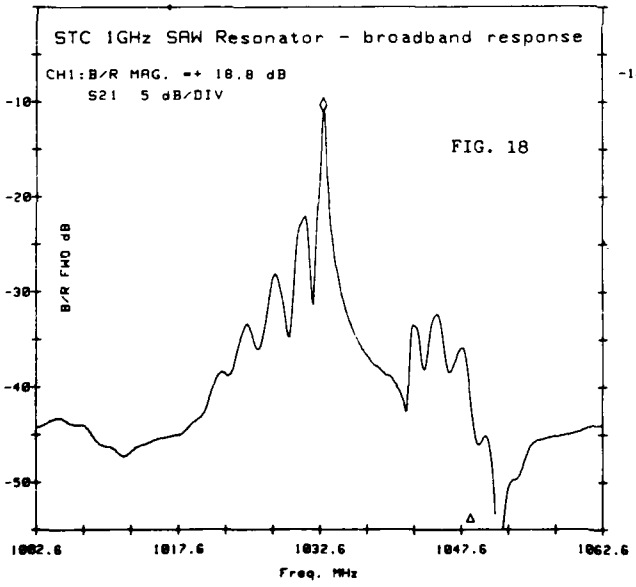


FIG. 18

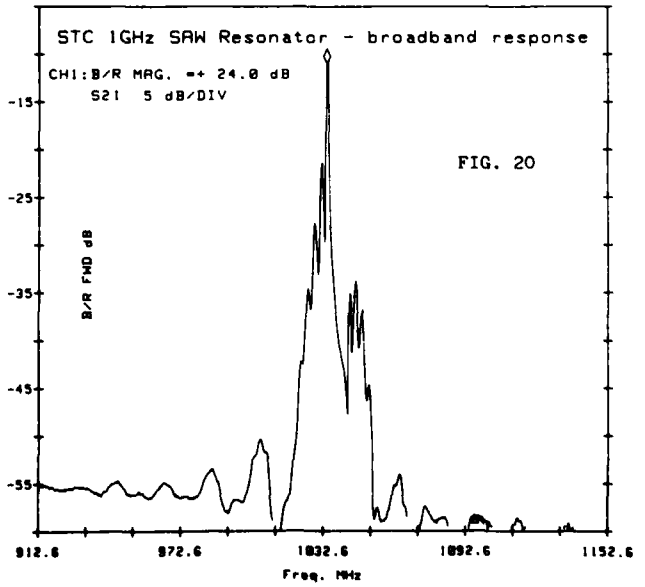


FIG. 20

THE AMPLITUDE-FREQUENCY EFFECT IN SC-CUT RESONATORS

Raymond L. Filler

US Army Electronics Technology and Devices Laboratory (ERADCOM)
Fort Monmouth, New Jersey 07703-5302

ABSTRACT

The effect of excitation level on the performance of quartz crystal resonators has been of interest for many years. A consequence of the nonlinearity of quartz is a distortion of the impedance vs. frequency function. The resonant frequency is, therefore, a function of the drive level. Above a certain drive level there are instabilities which cause discontinuities in the observed resonance curves.

The subject of amplitude-frequency behavior is of current interest due to the increasing importance of low phase noise oscillators. In this paper experimental results on the amplitude-frequency effect for SC-cut resonators of various contours, frequencies, and overtones will be reported. The results will be compared with those obtained from AT-cut resonators of the same overtones and frequencies. The measurements were performed on high-precision ultrahigh vacuum and high temperature processed ceramic flatpack enclosed devices.

As a review and starting point, the basic form of nonlinear resonance will be derived using a polynomial expansion of the motional capacitance in the equivalent circuit of the resonator.

Key words: quartz, quartz crystal resonator, SC-cut, AT-cut, nonlinearity.

INTRODUCTION

The subject of amplitude-frequency behavior is of current interest[1-3] due to the increasing importance of low phase noise oscillators. There is a direct relationship between resonator drive level and oscillator phase noise.[4] An oscillator designer must make a trade-off between the improved signal to noise ratio gained by increasing the drive level and the detrimental effects of high drive. Besides affecting the frequency, drive level can worsen an activity dip problem.[5,6] SC-cut resonators show a marked absence of activity dips, which make them attractive for high drive applications. In this paper, measurements on the amplitude frequency effect in SC-cut contoured quartz crystal resonators are reported.

NONLINEARITY

The amplitude-frequency effect is primarily due to the nonlinearity of quartz.[2] The phrase "nonlinearity" refers to the relationship between the amplitude of vibration and the applied force. In the simple case of a mass at the end of a massless spring, we have

$$F = -kx, \tag{1}$$

where F is the applied force, x is the displacement, and k is the spring constant. In this case, F and x are linearly related by the spring constant k. The natural frequency w_o of this system is

$$w_o = (k/m)^{1/2}, \tag{2}$$

where m is the mass on the end of the spring. The natural frequency is independent of the amplitude of the displacement.

In a real spring, the relationship between force and displacement is nonlinear. The nonlinear mass-spring system can be described by

$$F = -kx(1 + ax^2) = k'x, \tag{3}$$

where k' is the "effective spring constant." If the coefficient a is positive, the stiffness of the spring increases with increasing amplitude. This case is referred to as the "hard spring." If the coefficient a is negative, the stiffness of the spring decreases with amplitude, i.e., the spring is "soft." Since the stiffness of the nonlinear system is amplitude-dependent, it follows that the natural frequency is dependent on the amplitude of the vibration, i.e., the drive level.

We can write the equation of motion for the nonlinear system, resulting from a sinusoidal driving force, as

$$mx'' + kx + kax^3 = F_d\cos(wt+\phi). \tag{4}$$

This is Duffing's equation.

To obtain the electrical analog of this equation, one replaces the mass with an inductance, the reciprocal of the spring constant with a capacitance, the displacement with the charge, and the driving force with a voltage. The capacitance, which is analogous to the effective spring constant in equation (3), is then given by

$$1/C' = (1/C)(1 + aq^2). \tag{5}$$

The resulting equation for the sinusoidally driven LC network is

$$q'' + w_o^2 q + w_o^2 aq^3 = V\cos(wt+\phi), \tag{6}$$

where $w_o=(LC)^{-1/2}$ is the natural frequency without the nonlinear term. If we assume a solution of the form

$$q = Q\cos(wt), \tag{7}$$

and substitute equation (7) into equation (6), we arrive at an algebraic equation with $\cos^3(wt)$ terms. It is possible to transform these terms into cos(3wt) terms, indicating that harmonics are generated by the nonlinearity. We will ignore the harmonics for the present discussion, although harmonics generated through the nonlinearities have important consequences with respect to activity dips and intermodulation.

 "US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

The frequency of resonance is that frequency at which the driving voltage is in phase with the current. The resulting solution[7] for the resonant frequency w_r, assuming that the frequency shift is very small, is

$$(w_r - w_o) = (3/8)(a/w_o)I^2, \quad (8)$$

where I is the peak current. It can be seen from equation (8) that the resonant frequency of a nonlinear LC circuit, such as a quartz crystal, is a function of the square of the peak current. A drive sensitivity coefficient D can then be defined such that

$$\Delta f/f = DI^2. \quad (9)$$

Based on this analysis, Figure 1 is a plot of the current vs. frequency, in the region of resonance, through a 10 MHz resonator, for several driving voltages. A series resistance of 100 ohms and a drive sensitivity of 50 ppb/ma^2 were assumed.

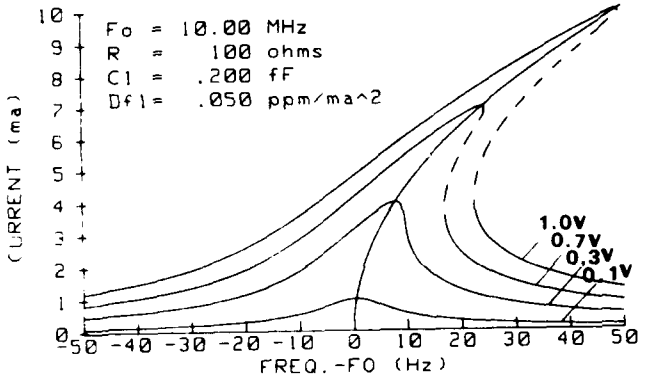


Figure 1. Nonlinear Resonance Curves - Current vs. Frequency for Several Voltages.

The peak of each curve occurs at the resonant frequency. There is a driving voltage above which the resonance curves are triple-valued. The highest and lowest of the three values are accessible experimentally, while the middle value is not. The current would exhibit discontinuities as either the frequency or the driving voltage was varied. The phase versus frequency for that resonator is shown in Figure 2.

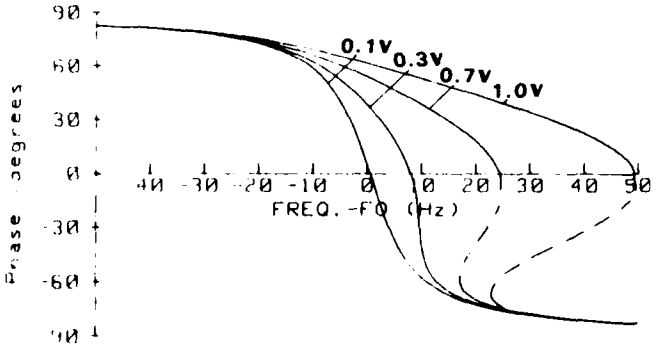


Figure 2. Nonlinear Resonance Curves - Phase vs. Frequency for Several Voltages.

EXPERIMENTAL CONFIGURATION

The measurements were made using the circuit shown schematically in Figure 3.[8]

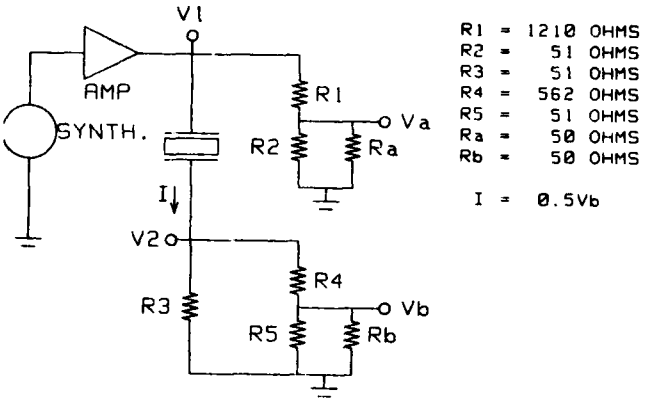


Figure 3. Experimental Circuit

Voltmeters with 50 ohm input impedances were connected to the Va and Vb ports. The R1-R2 and R4-R5 networks are included to avoid overloading the input to the voltmeter. Several configurations of the synthesizer and voltmeter were employed. In one configuration, the synthesizer was computer controlled, and Va and Vb were measured using a ZPV vector voltmeter or an HP-4192 LF Impedance Analyzer in its vector voltmeter configuration. The easiest and most successful method used the HP-3577 Network Analyzer as both the source and detector. The HP-3577 was configured to display current versus frequency directly. The frequency of maximum current and the value of that current were output to the controller. A typical output plot is shown in Figure 4. The ordinate is the difference between the

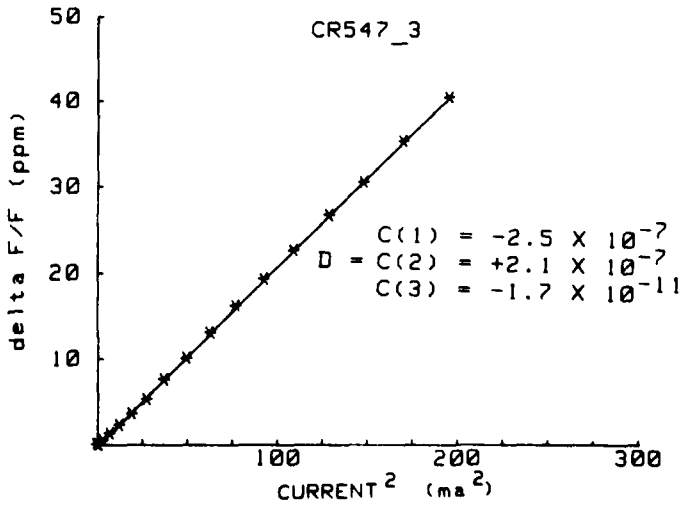


Figure 4. Typical Experimental Data and Curve Fit.

resonant frequency at each drive level and the resonant frequency at the lowest drive current used. The abcissa is the square of the drive current at resonance.

The frequency-versus-current data were fit to the function

$$\Delta f/f = C_1 + C_2 I^2 + C_3 I^4. \qquad (10)$$

The coefficients C_n for the sample data are printed on Figure 4. The coefficient C_2 is equivalent to the quantity D from eq. (9).

A fourth order polynomial was chosen to more closely fit the data. The deviation from the square law is most prominent on units which exhibit coupled modes that distort the resonance curve at high currents.

RESULTS

The measurements were performed on 14 mm diameter resonators with 5 mm diameter gold electrodes. The thickness of each resonator is indicated by its nominal frequency and overtone. For each nominal thickness, measurements were made on several units with different contours. Each resonator was operated on as many of its overtones as the design permitted. All of the resonators used in this experiment were high-precision, ultra clean processed ceramic flatpack enclosed devices with plano-convex quartz blanks. [9-16]

5 MHz FUNDAMENTAL MODE SC-CUT

The drive sensitivity coefficient D was measured for 5.115 MHz fundamental mode SC-cut resonators with contours ranging from 1.00 diopter to 2.37 diopter plano-convex. Each resonator in this group was measured on its fundamental, third, and fifth overtone. The data are plotted in Figure 5 and listed in Table I.

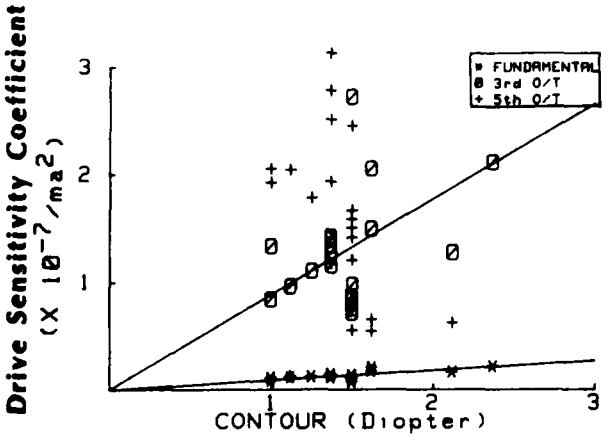


Figure 5. Drive Sensitivity of 5.115 MHz, fundamental mode, SC-cut Resonators as a Function of Contour.

The values of D for the fundamental mode and the third overtone, are proportional to the contour (inversely proportional to the radius of curvature). The scatter in the data for the fifth overtone is probably due to interfering modes.

TABLE I
Drive Sensitivity Coefficient
5.115 MHz Fundamental Mode SC-cut

Resonator	Diopter	fund	3rd	5th
		ppb/ma^2/diopter		
1	1.00	7.6	84	205
2	1.00	10.4	133	192
3	1.12	10.7	96	204
4	1.25	11.6	110	178
5	1.37	13.3	115	251
6	1.37	12.9	142	313
7	1.37	10.6	139	117
8	1.37	13.8	134	193
9	1.37	13.2	127	278
10	1.50	12.7	272	55
11	1.50	10.6	71	158
12	1.50	9.5	85	245
13	1.50	11.7	83	166
14	1.50	6.7	97	150
15	1.50	11.1	86	141
16	1.50	12.2	75	120
17	1.62	16.8	149	54
18	1.62	19.9	205	65
19	2.12	15.7	127	62
20	2.37	21.0	211	394

5 MHz 3rd OVERTONE SC-CUT

The same measurements were performed on 5.115 MHz, third overtone SC-cut units. The drive coefficient was determined for the third overtone and the fifth overtone. The data are plotted in Figure 6 and listed in Table II.

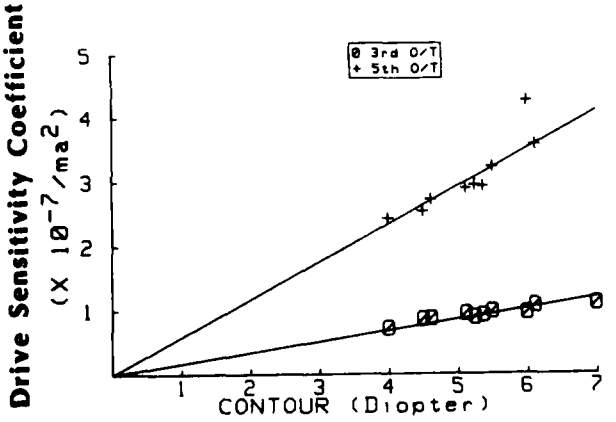


Figure 6. Drive Sensitivity of 5.115 MHz, third O/T, SC-cut Resonators as a function of Contour.

TABLE II
Drive Sensitivity Coefficient
5.115 MHz Third Overtone SC-cut

Resonator	Diopter	3rd	5th
		ppb/ma^2/diopter	
1	4.00	70.2	241
2	4.50	84.1	252
3	4.62	85.3	270
4	5.12	93.4	288
5	5.25	87.4	293
6	5.37	90.5	291
7	5.50	96.4	321
8	6.00	94.0	424
9	6.12	105.	355
10	7.00	108.	-

10 MHz 3rd OVERTONE SC-CUT

The drive coefficient was measured for 10MHz, third overtone, SC-cut units on the fundamental, third and fifth overtones. The fifth overtone could be measured for 1.00 diopter only. The results are plotted in Figure 7 and listed in Table III. The data are fit to a linear function, but the third overtone appears to be quadratic.

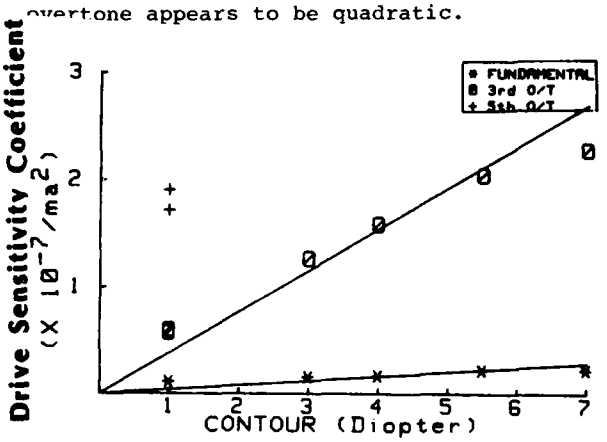


Figure 7. Drive Sensitivity of 10.00 MHz, third O/T SC-cut Resonators as a Function of Contour.

TABLE III
Drive Sensitivity Coefficient
10.00 MHz Third Overtone SC-cut

Resonator	Diopter	Fund	3rd	5th
		ppb/ma^2/diopter		
1	1.00	10.4	56.8	171
2	1.00	-	59.6	190
3	3.00	15.0	126.	-
4	4.00	16.0	158.	-
5	5.50	21.3	204.	-
6	7.00	21.7	228.	-

5.115 MHz 3rd OVERTONE AT-CUT

The drive coefficient was also measured for 5.115 MHz, third overtone, AT-cut devices. The results are plotted in Figure 8 and listed in Table IV. The drive coefficient is a function of contour, but the function does not appear to go linearly to zero at zero contour.

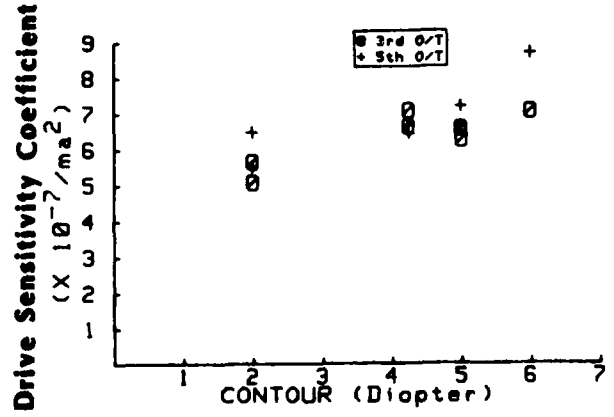


Figure 8. Drive Sensitivity of 5.115 MHz, third O/T AT-cut Resonators as a Function of Contour.

TABLE IV
Drive Sensitivity Coefficient
5.115MHz Third Overtone AT-cut

Resonator	Diopter	3rd	5th
		ppb/ma^2/diopter	
1	2.00	562	646
2	2.00	504	550
3	4.25	704	664
4	4.25	659	642
5	5.00	623	717
6	6.00	702	865

10 MHz 3rd OVERTONE AT-CUT

The behavior of the 10MHz, third overtone, AT-cut units was anomalous. The frequency decreased when the drive was increased. In effect, whereas all the other resonator types behaved like a hard spring, these devices behaved like a soft spring. Normally, all AT-cut devices are hard springs, while BT-cut devices are soft springs. The results for the drive coefficient as a function of contour for 10MHz, 3rd O/T AT-cut resonators are shown in Figure 9. It can be seen that the scatter in the data is very large. These devices were much more susceptible to interfering modes.

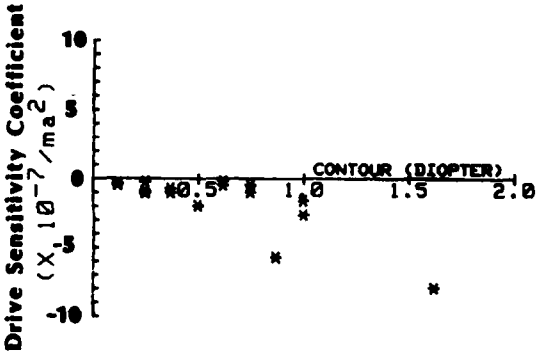


Figure 9. Drive Sensitivity of 10.00 MHz, third O/T AT-cut Resonators as a Function of Contour.

These interfering modes can be seen quite clearly on the HP-3577. An example is shown in Figure 10.

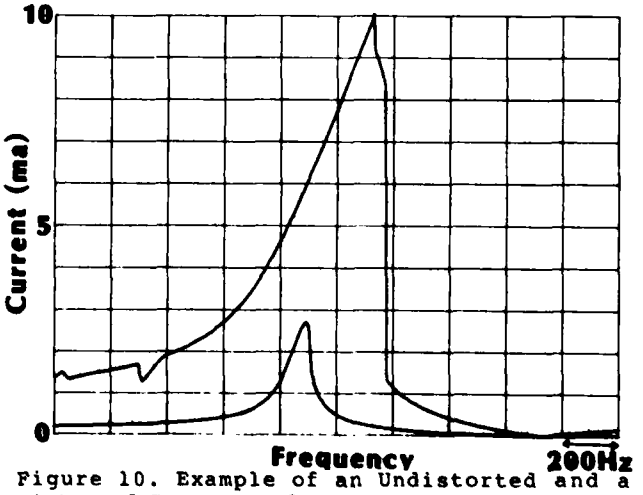


Figure 10. Example of an Undistorted and a Distorted Resonance Curve.

The lower curve is an undistorted resonance curve. The upper curve displays two unwanted modes to the left of the desired resonance and a severe distortion right at the resonance. In addition, the steep drop to the right of the resonance frequency in the upper curve is due to the inaccessible region of the nonlinear resonance curve.

Lateral Field Resonators

Amplitude-frequency measurements were performed on some 5 MHz, 4.25 diopter, plano-convex, 3rd O/T, SC-cut lateral field resonators (LFR).[17] The results for AT-cut thickness field resonators (TFR) and SC-cut LFR and TFR are plotted in Figure 11. It can be seen that the SC-cut LFR is more sensitive to drive level than a TFR of the same contour. The controlling factor is the particle displacement vs. drive current relationship which is dependent on geometry.[18]

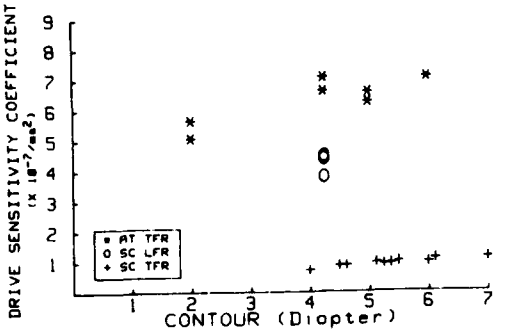


Figure 11. Drive Sensitivity of 5 MHz Lateral Field and Thickness Field Resonators.

SUMMARY

The magnitude of the amplitude frequency effect was measured for several types of SC and AT-cut resonators. The magnitude was found to be proportional to the contour in diopters for SC-cut resonators. The linear dependence on contour agrees with that calculated by Tiersten.[3] The magnitude for 5 MHz SC-cut is significantly lower than for AT-cut units of similar design. Table V gives a summary of the results for SC-cut devices.

TABLE V
Slope of the Drive Sensitivity Coefficient vs. Contour

Nominal	Thickness	fund	3rd	5th
		ppb/ma^2/diopter		
5.115 fund	0.38mm	9.0	89	-
10.000 3rd	0.56	4.1	39	-
5.115 3rd	1.08	-	17	58

CONCLUSIONS

The oscillator designer desiring to use a high drive current must consider the effects of current fluctuations. The slow drift of drive current is a contributor to aging, and current noise is a contributor to short-term stability. These effects increase as the square of the drive current. The data suggests that, for a given drive level, to minimize drive related frequency fluctuations, the lowest practical contour and overtone should be used. (Of course, other considerations suggest the use of higher overtones for maximum stability.)

ACKNOWLEDGEMENTS

The author thanks Mr. R.C. Smythe and Prof. H.F. Tiersten for many useful discussions, Mr. E. Simon for performing some of the measurements, and Mr. D. Boyce for fabricating the test fixtures.

REFERENCES

1. J.J. Gagnepain and R. Besson, Physical Acoustics , edited by W.P. Mason and R.N. Thurston,Academic, New York, 1975, Vol XI.

2. G. Theobald and J.J. Gagnepain, "Frequency Variations in Quartz Crystal Resonators Due to Internal Dissipation," J. Applied Phys., 50 , 10 , 1979, pp 6309-6315.

3. H.F. Tiersten, "An Analysis of Nonlinear Resonance in Electroded Contoured AT- and SC-cut Quartz Crystal Resonators," Proc. 38th Ann. Symp. Freq. Cont., 1984, pp 132-140. Copies available from IEEE, 445 Hoes Lane, Piscataway, NJ 08854, document number: 84CH2062-8.

4. B. Parzen, Design of Crystal and Other Harmonic Oscillators , John Wiley & Sons, New York, 1983.

5. A.F.B. Wood and A. Seed, "Activity Dips in AT-cut Crystals," Proc. 21st Ann. Symp. Freq. Cont., 1967, pp 420-435. Copies available from National Technical Information Service, Sills Bldg, 5285 Port Royal Road, Springfield, VA 22161, document number: AD659792.

6. C. Franx, "On Activity Dips of AT Crystals at High Levels of Drive," Proc. 21st Ann. Symp. Freq. Cont., 1967, pp436-454. Copies available from NTIS (see ref. 5).

7. W.J. Cunningham, Introduction to Nonlinear Analysis , McGraw-Hill, New York, 1958.

8. R.C. Smythe, private communication.

9. R.L Filler and J.R. Vig,"The Acceleration and Warm-up Characteristics of Four-Point Mounted SC- and AT-Cut Resonators," Proc. 35th Ann. Symp. Freq. Cont., pp 110-116, 1981; available from Electronics Industries Association, 2001 Eye St., Washington, DC 20006.

10. J.R. Vig and E. Hafner, "Packaging Precision Quartz Crystal Resonators," Technical Report ECOM-4143, US Army Electronics Command, Fort Monmouth, NJ, (July 1973). Copies available from NTIS (see ref. 5), document number: AD 763215.

11. P.D. Wilcox, G.S. Snow, E. Hafner, and J.R. Vig, "A New Ceramic Flatpack for Quartz Resonators," Proc. 29th Ann. Symp. Freq. Cont.,1975, pp 202-219, 1975. Copies available from NTIS (see ref. 5), document number: AD A017466.

12. R.D. Peters, "Ceramic Flatpack Enclosures for Precision Quartz Crystal Units," Proc. 30th Ann. Symp. Freq. Cont.,1976, pp 224-231. Copies available from NTIS (see ref. 5), document number: AD A046089.

13. J.R. Vig, J.W. LeBus, and R.L. Filler, "Chemically Polished Quartz," Proc. 31st Ann. Symp. Freq. Cont., 1977, pp 131-143. Copies available from NTIS (see ref. 5), document number: AD A088221.

14. J.R. Vig, R.J. Brandmayr, and R.L. Filler, "Etching Studies on Singly and Doubly Rotated Quartz Plates," Proc. 33rd Ann. Symp. Freq. Cont., 1979, pp 351-358; available from EIA (see ref. 9).

15. J.R. Vig, "UV/Ozone Cleaning of Surfaces: A Review." In Surface Contamination: Genesis, Detection and Control , K.L. Mittal, ed., Vol 1, Plenum Press, New York, 1979, pp 235-254.

16. R.L. Filler, J.M. Frank, R.D. Peters, and J.R. Vig, "Polyimide Bonded Resonators," Proc. 32nd Ann. Symp Freq. Cont, 1978, pp 290-298. Copies available from EIA (see ref. 9).

17. A.W. Warner and B. Goldfrank, "LF Resonators," Proc. 39th Ann. Symp. Freq. Cont., 1985. Copies available from IEEE (see ref. 3), document number: 85CH2186-5.

18. H.F. Tiersten, private communication.

EXPERIMENTAL EVALUATION OF THE EFFECTIVE NON-LINEAR ELASTIC CONSTANT FOR TRAPPED ENERGY AND CONTOURED RESONATORS

R.C. Smythe & P.E. Morley
Piezo Technology, Inc.
Orlando, Florida

Abstract

Elastic non-linearity in quartz gives rise to well-known effects in quartz resonators and filters. Analyses by Tiersten relate non-linear resonance and intermodulation in singly- and doubly-rotated contoured and trapped-energy resonators to an effective non-linear elastic constant, γ_e, which in turn is a combination of third- and fourth-order elastic constants.

Because the fourth-order elastic constants of quartz have not been determined, γ_e must be measured for each cut of interest. A number of resonators have been carefully fabricated to obtain an estimate of γ_e both from non-linear resonance measurements and, with the resonator incorporated into a filter network, from intermodulation measurements. Measurements of AT-, BT-, and SC-cut resonators, of both plane-parallel and contoured design are reported. For the BT- and SC-cuts, γ_e has not been previously determined, and published non-linear resonance data is inadequate to allow its calculation. Measured values of γ_e for the AT-cut are compared with previously reported measurements.

Introduction

Non-linear effects in quartz crystal resonators include non-linear resonance, motional resistance variation with drive level, and, in filter applications, intermodulation. Each of these effects results from a deviation from the classical linear model for the crystal resonator.

The mechanisms giving rise to these non-linearities can be considered to be in four distinct categories:

a) Bulk elastic non-linearity related solely to the intrinsic material properties of quartz.

b) Surface effects, usually associated with particulate contamination on or under the crystal electrodes, or with poor electrode adhesion [1].

c) Static & dynamic thermal gradient effects [2] resulting from spatially non-uniform dissipation in the crystal plate [3].

d) Non-linear coupling to resonances at harmonically related frequencies [4, 5], including other thickness modes as well as contour modes. This may include parametric excitation of two or more such coupled modes, and may be associated with any of the three mechanisms listed above.

For the purposes of this paper, only bulk anelasticity is discussed, in relation to its effects on resonance frequency and in-band intermodulation. The modes discussed will be restricted to overtones of thickness shear, particularly to the AT-, SC- and BT-cuts of quartz.

Non-Linear Resonance Theory

The theoretical relationship between resonance frequency and crystal current has been obtained by Tiersten for trapped-energy resonators [6, 7] and for contoured resonators [8, 9].[1] In either case, the relationship is quadratic, thus

$$\Delta f/f = A\gamma_e I I^* \qquad (1)$$

where A is a function of the linear parameters of the resonator, f is the change in resonance frequency due to current I, and we introduce an effective non-linear elastic constant, γ_e, which can be written as:

$$\gamma_e = \gamma + \gamma_n - \gamma_c \qquad (2)$$

In (2) γ_n and γ_c are complex functions containing the β-dependent terms of equation 2.4, [9]. The value of $(\gamma_n - \gamma_c)$ has been calculated [7] for the AT-, BT- and SC-cuts, and is shown in Table 1 for various overtones.

TABLE 1. VALUES OF $(\gamma_n - \gamma_c)$

CUT	$(\gamma_n - \gamma_c) \times 10^{-12}$ N/m^2		
	n=1	n=3	n=5
AT	-.29	-.15	-.20
SC	-.15	-.18	-.13
BT	-.02	-.06	-.03

The coefficient [1] depends only on the cut, and can be expressed as a combination of rotated elastic constants:

$$\gamma = C_{22}/2 + C_{266} + C_{6666}/6 \qquad (3)$$

Planar Trapped Energy Resonator

The electrode geometry of a trapped energy resonator is shown in fig. 1. First considering the thickness solution, where the mode shape is considered

[1] In this and the following section, much of the notation is Tiersten's, to whose work, as well as that of Stevens, the reader is referred; e.g., [8-12].

CH2186-0/85/0000-0317$1.00©1985IEEE

to be uniform over the electrode area, it has been shown [6] that

$$\Delta f/f = \gamma_e K_1 (n/fA_e)^2 \; I \; I^* \qquad (4)$$

where Δf is the change in the motional resonance frequency due to current I, A_e is the electrode area and K_1 is given by the expression

$$K_1 = 9/512 \; e_{26}^2 \; \bar{c}^{(1)} \qquad (5)$$

The rotated piezoelectric and stiffened elastic constants e_{26} and $\bar{c}^{(1)}$ are defined in [10]. For brevity the superscript of $\bar{c}^{(1)}$ will be omitted in subsequent equations.

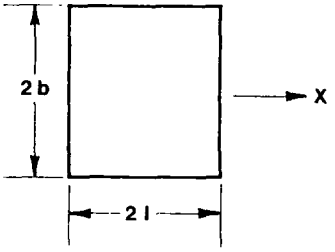


FIG. 1. ELECTRODE GEOMETRY, PLANAR TRAPPED-ENERGY RESONATOR. WAVE NUMBERS FOR X- AND Z-PROPAGATION ARE, RESPECTIVELY, $\bar{\xi}$ AND $\bar{\upsilon}$ IN ELECTRODED REGION AND ξ AND υ ELSEWHERE.

The motional capacitance of the hypothetical thickness solution is given by:

$$C_1 = 4 \; e_{26}^2 \; A_e/(n^2 \pi^2 \; \bar{c} \; h) \qquad (6)$$

where 2h is the plate thickness.

Substituting (6) into (4) yields the expression

$$\Delta f/f = \gamma_e \; K_2 (1/n^4 C_1^2) \; I \; I^* \qquad (7)$$

where:

$$K_2 = 9 \rho \, e_{26}^2/(2 \; \bar{c} \; \pi^4) \qquad (8)$$

When the mode shape of the trapped energy resonators is included, the equations become slightly more complex; but can be simplified to:

$$\Delta f/f = \gamma_e \; K_t \; K_1 \; (n/fA_e)^2 \; I \; I^* \qquad (9)$$

or

$$\Delta f/f = \gamma_e \; K_t' \; K_2 \; (1/n^4 \; C_1^2) \; I \; I^* \qquad (10)$$

The quantities K_t and K_t', which depend upon the mode shape, have been calculated from [2] and are illustrated graphically in fig. 2, assuming equal trapping for the two propagation directions ($\bar{\xi} \, l = \bar{\upsilon} \, b$).

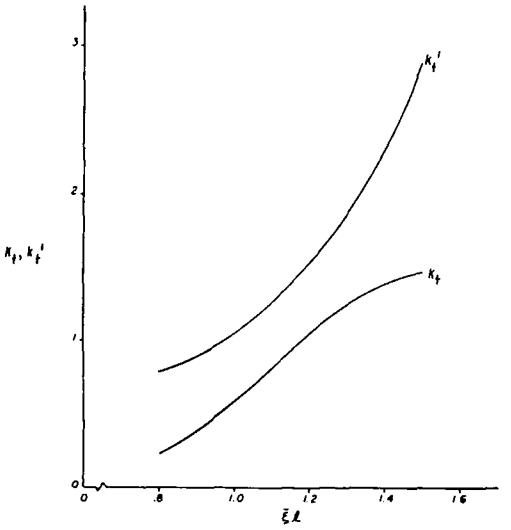


FIG. 2. PARAMETERS K_t AND K_t' AS A FUNCTION OF TRAPPING PARAMETER $\bar{\xi} \, l$, FOR THE CASE $\bar{\xi} \, l = \bar{\upsilon} \, b$.

Contoured Resonator

Adopting a similar procedure of substituting the motional capacitance expression into the non-linear resonance equation [8, 9], and assuming that for a contoured resonator, the electrodes cover the whole active area, an almost identical expression is obtained:

$$\Delta f/f = 2 \; \gamma_e \; K_2 \; (1/n^4 \; C_1^2) \; I \; I^* \qquad (11)$$

where K_2 is defined as before in equation (8) and C_1 is given by Eq. (4.19) of [11].

Intermodulation Theory

An analysis of third-order intermodulation (IM_3) due to anelasticity has been carried out by Tiersten for a single AT-cut trapped-energy resonator having strip electrodes [13, 14]. If I_1 and I_2 are the currents in the motional arm of the resonator equivalent circuit at circular frequencies ω_1 and ω_2, and I_Ω is the IM_3 current at the intermodulation frequency

$$\Omega = 2\omega_1 - \omega_2 \qquad (12)$$

Tiersten's result can be written:

$$I_\Omega = [Y_0(\Omega) + Y_1(\Omega)] \; V_\Omega - j \; \gamma_e \; K \; Y_1(\Omega) I_1^2 \; I_2^* \qquad (13)$$

where Y_0 and Y_1 are the static and dynamic admittance terms and V_Ω is the voltage across the resonator at Ω. K depends on the material constants of the cut, frequency, overtone, electrode dimensions or motional capacitance, and mode shape [15, 16]. By interchanging subscripts 1 and 2 the IM_3 current at $(2\omega_2 - \omega_1)$ is obtained.

Non-Linear Equivalent Circuit

Both non-linear resonance and intermodulation may be represented by the equivalent circuit of fig. 3, in which a controlled voltage source having instantaneous value, $\tilde{v}(i)$, can represent any current-dependent non-linearity.

For the anelastic analyses of the preceding two sections, the controlled source is of the form

$$\tilde{v}(i) = j\frac{2}{3}\,\alpha\, i^3 \qquad (14)$$

and

$$i = i(t)$$

For the case of non-linear resonance

$$i(t) = \sqrt{2}\, I \cos \omega t \qquad (15)$$

and v contains terms in ω and 3ω. Retaining only the ω term and using phasor notation

$$\tilde{V}_{\omega} = j\alpha I^2 I^* \qquad (16)$$

and it can readily be shown that

$$\Delta f/f = (\omega_s C_1/2) I\, I^* \qquad (17)$$

where ω_s is the motional arm (linear) resonance frequency. The coefficient α can be expressed in terms of the constants of Eq. (7), (9), (10), or (11) to obtain a non-linear circuit representation.

For the case of third-order intermodulation

$$i = \sqrt{2}\,[I_1 \cos\omega_1 t + I_2 \cos\omega_2 t]$$

For the Ω component of $\tilde{v}$ we have, in phasor form,

$$\tilde{V}_{\Omega} = j\,\alpha\, I_1^2 I_2^* \qquad (18)$$

Equation (13) can now be written [16]

$$I_{\Omega} = [Y_0(\Omega) + Y_1(\Omega)]\, V_{\Omega} - Y_1(\Omega)\, \tilde{V}_{\Omega} \qquad (19)$$

and

$$\alpha = \kappa \gamma_e \qquad (20)$$

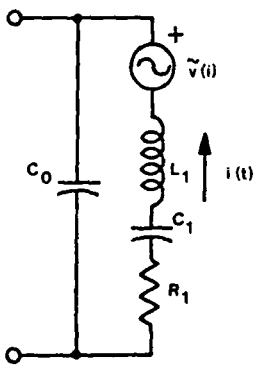


FIG. 3. RESONATOR NON-LINEAR EQUIVALENT CIRCUIT

Experimental Results

Several batches of resonators were carefully fabricated at various frequencies. Most units were made with square electrodes to closely match the model used for theoretical calculations. The crystals were cleaned thoroughly and were not overplated or otherwise adjusted in frequency.

For non-linear resonance measurements, an HP 3577A Network Analyzer was used and the crystal was measured in transmission between 50 ohm terminationsto measure crystal current and resistance. For resonance measurement at high level, the system was used in single sweep mode to reduce static thermal effects. The sweep speed was maximized for the same reason, taking care that the measurement was not affected by the rate of sweep.

Also, great care was taken to ensure that coupled modes were not affecting the result. Fig. 4 illustrates a "well-behaved" AT cut resonator, and fig. 5 a typical BT cut. Figs. 6 and 7 illustrate anomalous behavior which we believe is due to coupled modes, although it is difficult to identify these modes, this requiring further work.

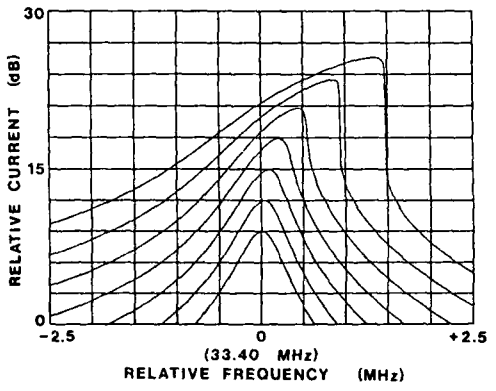


FIG. 4. NORMAL NON-LINEAR RESONANCE CURVES, AT-CUT

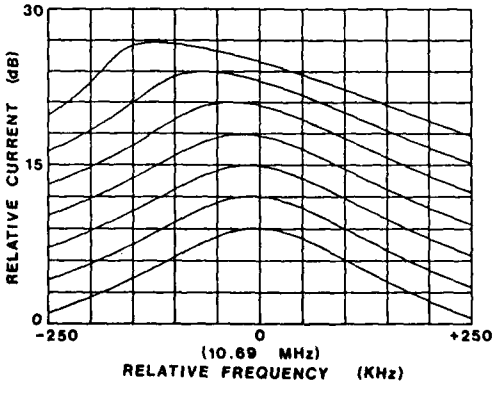


FIG. 5. NORMAL NON-LINEAR RESONANCE CURVES, BT-CUT

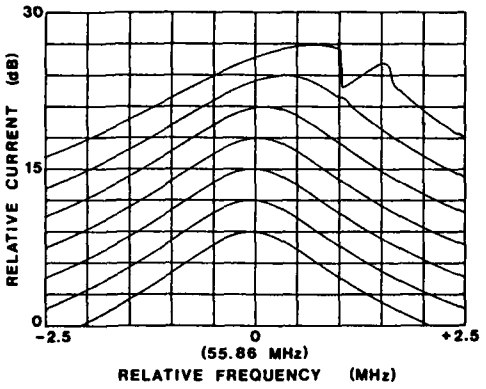


FIG. 6. ANOMALOUS NON-LINEAR RESONANCE CURVES, AT-CUT, SHOWING COUPLED MODE BEHAVIOR

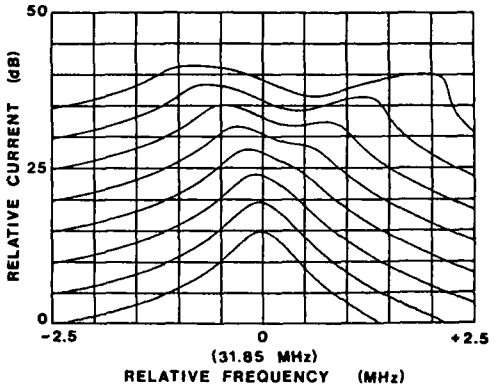


FIG. 7. ANOMALOUS NON-LINEAR RESONANCE CURVES, AT-CUT

The value of γ_e was determined from equation (10) for all of the crystal batches. The results are tabulated in Table 2. A small number of AT- and SC-cut contoured resonators have also been measured using the same technique, obtaining very similar results, Table 3. It can be seen that planar and contoured AT-cut resonators exhibit similarly inconsistent non-linear behavior.

Finally for a number of the AT- and SC-cut resonators, γ_e was calculated from third-order intermodulation measurements, Table 4.

TABLE 2. NON-LINEAR RESONANCE DATA FOR PLANAR TRAPPED-ENERGY RESONATORS

Cut	Ident	f (MHz)	n	$\bar{\xi}$1(Rad)	$\Delta f/I^2 f$ (A^{-2})	C_1(fF)	$\gamma_e(10^{12}N/m^2)$	$\gamma(10^{12}N/m^2)$
AT	A136	10.8	1	1.13	.011	13.9	1.4	1.7
	A136	32.3	3	1.40	.028	1.35	2.0	2.2
	A136	53.8	5	1.45	.065	0.42	3.8	4.0
	A50	33.4	3	1.25	.87	0.27	2.9	3.1
	A50	55.6	5	1.34	1.2	0.09	3.0	3.2
	A50/2	32.4	3	1.20	.87	0.27	3.2	3.4
	A80	10.8	1	1.07	.032	6.2	1.1	1.4
	A110	72.1	3	1.48	.029	2.0	4.5	4.7
	A150	33.5	3	1.45	.014	2.2	2.5	2.7
	A150	55.8	5	1.46	.040	0.76	6.2	6.4
	A42	104.3	3	1.25	0.41	0.40	3.1	3.3
	Res A	11.7	1	1.16	0.073	15.8	1.3	1.6
SC	S136	10.8	1	1.11	.0063	4.6	0.47	0.62
	S136	32.3	3	1.42	.014	0.42	0.49	0.67
	S136	53.8	5	1.47	.018	0.14	0.52	0.65
	S50	11.2	1	.82	.100	0.55	0.55	0.60
	S50	33.3	3	1.29	.272	0.090	0.49	0.67
	S50	55.4	5	1.36	.478	0.025	0.47	0.60
	S80	10.7	1	.95	.022	2.2	0.55	0.60
	S80	31.8	3	1.35	.064	0.22	0.64	0.82
BT	B136	10.8	1	1.09/1.29	-.0036	3.6	-0.97	-0.95
	B50	11.2	1	0.78/1.07	-0.036	0.46	-0.42	-0.40
	B80	10.8	1	0.91/1.17	-0.014	1.7	-1.4	-1.4

TABLE 3. NON-LINEAR RESONANCE DATA FOR CONTOURED RESONATORS

Cut	Ident	f (MHz)	n	Contour (diop)	$\Delta f/I^{2}f$ (A^{-2})	C_1(fF)	$\gamma_e(10^{12}N/m^2)$	$\gamma(10^{12}N/m^2)$
AT	A10	10.0	3	1	0.26	0.75	3.6	3.8
	A2.5	2.50	1	9	0.11	7.0	2.0	2.3
	A7.5	7.50	3	9	1.4	0.39	5.8	6.0
	A12.5	12.5	5	9	1.4	0.088	2.3	2.5
SC	S10	10.0	3	1	.045	0.27	0.43	0.61

TABLE 4. INTERMODULATION DATA FOR PLANAR TRAPPED-ENERGY RESONATORS

Cut	Ident	f (MHz)	n	BW_3(kHz)	P_{TT} (dBm)	IM_3(dB)	$\gamma_e(10^{12}N/m^2)$	$\gamma(10^{12}N/m^2)$
AT	A50	11.2	1	1.3	10	25.5	0.77	1.1
	A50	33.4	3	2.0	0	29.0	2.1	2.3
	A50	53.7	5	1.9	0	27.8	1.3	1.5
	A80	10.7	1	[illegible].3	10	54.8	0.88	1.2
	A80	53.1	5	4.0	10	34.8	0.84	1.0
	A136	10.8	1	10.8	10	75.0	0.97	1.1
	A136	32.2	3	8.3	0	55.0	1.6	1.8
	A136	53.7	5	6.0	0	40.5	3.3	3.5
	A	11.7	1	12.0	20	54.7	0.82	1.1
SC	S50	11.2	1	0.61	10	26.5	0.35	0.50
	S50	33.3	3	1.15	10	22.5	0.52	0.70
	S50	55.5	5	1.15	0	33.0	1.15	1.28
	S80	10.7	1	2.0	10	43.8	0.78	0.93
	S80	31.8	3	1.6	10	28.1	0.84	1.02
	S80	53.0	5	2.3	10	32.5	0.96	1.09
	S136	10.8	1	3.8	10	62.3	0.62	0.75
	S136	32.2	3	3.0	10	47.4	0.67	0.85
	S136	53.7	5	3.0	10	39.0	1.36	1.5

The measurements were made with the resonators in a one-pole filter network. Air core coils were used to avoid extraneous non-linearity. The measurements were typically made with an input power of 0dBm or +10dBm depending on the filter bandwidth. The value of γ_e was determined from equation (13).

Interpretation of Results

From the data shown it can be seen that for SC-cut resonators γ is in the range 0.6 to 0.8 x $10^{12}N/m^2$ based on non-linear resonance measurements while γ_e is between .43 and .64 x $10^{12}N/m^2$. When γ is calculated from IM measurements it ranges from .50 to 1.5 x $10^{12}N/m^2$ for the SC-cut. For AT-cut resonators, on the other hand, values of γ_e from non-linear resonance measurements vary by about a factor of 6 and by a factor of 4 using IM data. Limited measurements of BT-cuts indicated similar variability and were terminated pending further study of the AT-cut.

It should be noted that the IM analysis, having been carried out much earlier than the non-linear resonance analysis, differs from it very significantly. First, the IM analysis was carried out for a strip electrode model. Moreover, the treatment of non-linearity has been refined in the later work. Still further, we have made ad hoc modifications to the IM analysis to accommodate γ_e and to extend it to SC-cuts. Hence the values of γ obtained from IM measurements might be expected to differ substantially from non-linear resonance values.

A valid comparison can be made, however, by means of the non-linear equivalent circuit, fig. 3. This has been done in Table 5, assuming a cubic non-linearity, as in Eq. (14), and noting that α is proportional to γ_e. It can be seen that for the AT-cut, agreement is generally poor, while for the SC-cut it is somewhat better.

TABLE 5. COMPARISON OF INTERMODULATION & NON-LINEAR RESONANCE MEASUREMENTS

Cut	Ident	f (MHz)	n	α(IM)/α(NLR)
AT	A50	33.4	3	.86
	A50	53.7	5	.66
	A80	10.7	1	.93
	A136	10.8	1	.49
	A136	32.2	3	.46
	A136	53.7	5	.39
	A	11.7	1	.56
SC	S50	11.2	1	1.4
	S50	33.3	3	1.0
	S50	55.5	5	1.6
	S80	10.7	1	2.0
	S80	31.8	3	1.2
	S136	10.8	1	1.3
	S136	32.2	3	0.9
	S136	53.7	5	2.0

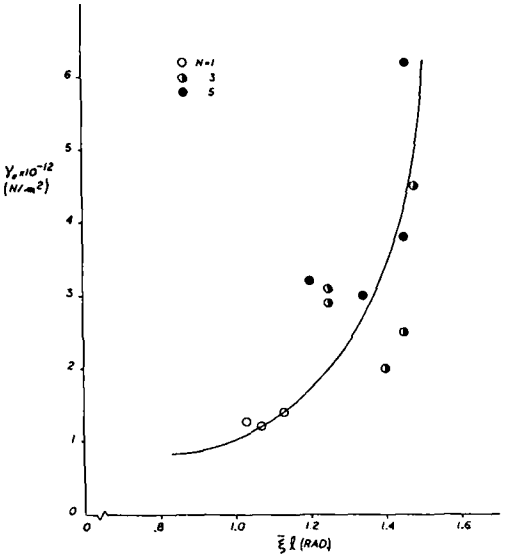


FIG. 8. γ_e FOR AT-CUT RESONATORS CALCULATED FROM NON-LINEAR RESONANCE MEASUREMENTS

Since, for each cut, γ is a constant, the AT-cut measurements and, to a lesser extent, the SC-cut IM measurements reflect mechanisms in addition to third-order anelasticity. (Because of the consistency in the SC non-linear measurements, the analytical description of third-order anelasticity is believed to be adequate.) Possible additional mechanisms include:

1) Surface effects. Every attempt was made to elimate these through careful processing.

2) Thermal gradient effects. Because the AT- and BT-cuts are highly sensitive to thermal gradients, these may be responsible at least in part, for discrepancies in measurements of AT's.

3) Non-linearly coupled modes. While resonators having obvious coupled modes were eliminated from the reported results this mechanism cannot be ruled out.

To inveetigate 2) further, measured values of γ_e for the AT-cut resonators were plotted versus the x-axis trapping parameter $\bar{\xi}$ l, figs. 8 & 9. For non-linear resonance measurements, fig. 8, γ_e tends to increase with the degree of trapping, qualitatively supporting 2). This data can also be interpreted as increasing with overtone, however. For γ_e based on IM measurements, fig. 9, no consistent variation with trapping is seen. Since the current levels for the IM measurements were lower than for the non-linear resonance measurements, this is not inconsistent with 2), which may be important primarily for non-linear resonance. The scatter in the IM data may be related to 3). By contrast, the cosistency of the SC-cut non-linear resonance data is shown in fig. 10.

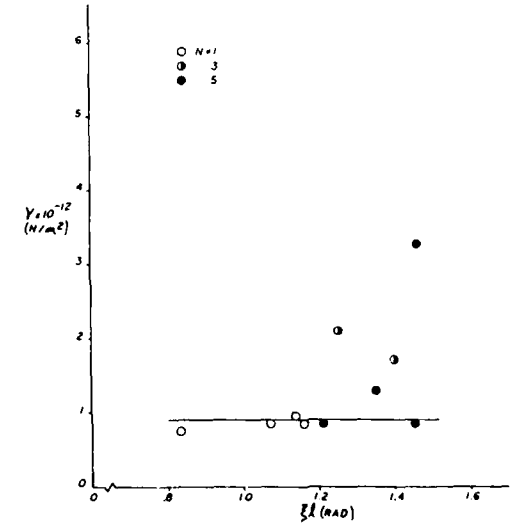


FIG. 9. γ_e FOR AT-CUT RESONATORS CALCULATED FROM THIRD-ORDER INTERMODULATION MEASUREMENTS

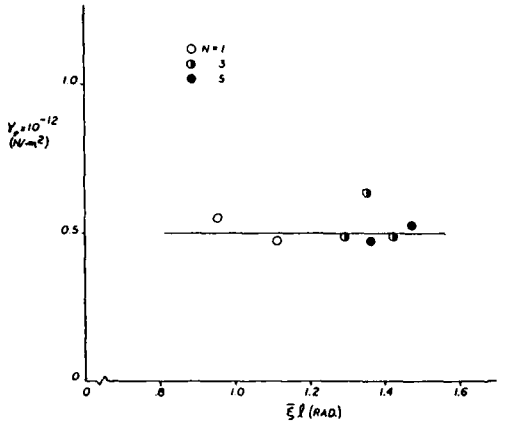


FIG. 10. γ_e FOR SC-CUT RESONATORS CALCULATED FROM NON-LINEAR RESONANCE MEASUREMENTS

In summary, the AT measurements indicate that γ_e is in the range 0.8 to 1.2 x 10^{12}N/m^2, but that for many resonators additional non-linear mechanisms, yet to be positively identified, are of importance. For BT-cuts based on limited data, an upper bound on γ_e is probably -1.0 x 10^{12}N/m^2. Again, other non-linear mechanisms must be considered.

Previous AT-cut values of γ_e (reported as γ) have been given by Smythe [16] who obtained values of .81 and .83 x 10^{12}N/m^2 using IM measurements and 1.3 x 10^{12}N/m^2 using non-linear resonance measurements, and by Planat, et al [18], who obtained values of .79 and .66 x 10^{12} using IM measurements.

Despite the uncertainty in γ and γ_e for AT- and BT-cuts it is useful to compare the non-linear performance of the three cuts using approximate "best guess" values of γ_e as follows:

AT-cut, $\gamma_e = 1.0 \cdot 10^{12}$ N/m^2

SC-cut, $\gamma_e = 0.5 \cdot 10^{12}$ N/m^2

BT-cut, $\gamma_e = -1.0 \cdot 10^{12}$ N/m^2

Assuming the values above, a comparison can be made between the three crystal types for various cases. The simplest, if somewhat impractical, case is to assume equal motional capacitance. Table 6 lists comparative frequency change for the fundamental, 3rd and 5th overtone.

TABLE 6. COMPARATIVE $\Delta f/f$, EQUAL C_1, f, I

CUT	OVERTONE		
	1	3	5
AT	100	1.2	0.16
SC	9.9	0.12	0.016
BT	-3.3	-0.04	-0.005

A more practical case, particularly for the design of crystal filters, is to assume equal electrode areas. Here, the crystal impedance is a function of crystal cut and overtone so the terminating impedance of a filter network incorporating the resonator must be adjusted accordingly. The comparison is made assuming equal input power to the filter network, Table 7.

TABLE 7. COMPARATIVE $\Delta f/f$, EQUAL ELECTRODE AREA, f, INPUT POWER

CUT	OVERTONE		
	1	3	5
AT	100	100	100
SC	34	34	34
BT	-12	-12	-12

Another case useful for comparison of filter resonators assumes that the anharmonic mode spectra are identical for the various cyrstal types, Table 8. Here the electrode dimensions and mass loading are varied to obtain identical propagation and trapping along the two major axes. Again, the terminating impedance is appropriately scaled.

TABLE 8. COMPARATIVE $\Delta f/f$, EQUAL MODE SPECTRA, f, INPUT POWER

CUT	OVERTONE		
	1	3	5
AT	100	40	22
SC	39	17	7.8
BT	-18	-3.8	-1.2

Conclusion

The non-linear analysis of Tiersten provides an analytical tool for determination of non-linear resonance and intermodulation in quartz crystals. We have interpreted these analyses in such a way as to relate these properties to known linear resonator parameters and to quantify the effects for AT-, SC- and BT-cut crystals.

Unfortunately, the results obtained from this work have shown a very high degree of inconsistency, particularly for AT-cut. This is considered to be due either to non-linear coupling to resonance modes at higher frequencies or to thermal effects, or both.

However, a comparison has been made between these crystal cuts on the basis of these results, and it is found that a considerable improvement in non-linear resonance or intermodulation performance can be

obtained by using SC- or BT-cut over AT-cut. This has been corroborated by measurements on monolithic crystal filters in a companion paper [17]. Also the apparent absence of anomolous behavior in the SC-cut may facilitate practical designs of low intermodulation SC crystal filters.

Acknowledgment

The authors would like to thank H.F. Tiersten for providing an analysis and for many valuable discussions, D.S. Stevens for providing a computer program and for helpful advice, and R.L. Filler for useful discussions and for preliminary experimental results.

References

[1] Horton, W.H., & Smythe, R.C., "Experimental Investigations of Intermodulation in Monolithic Cyrstal Filters," Proc. 27th Annual Frequency Control Symposium (AFCS), pp. 243-245; 1973.

[2] Holland, Richard, "Non-Uniformly Heated Anisotropic Plates: II. Frequency Transients in AT and BT Quartz Plates," Proc. IEEE Ultrasonics Symp., pp. 592-598; 1974.

[3] Theobald, G., & Gagnepain, J.J., "Frequency Variations in Quartz Crystal Resonators Due to Internal Dissipation," J. Appl. Phys., v. 50, no. 10, pp. 6309-6315; Oct. 1979.

[4] Wood, A.F.B., & Seed, A., "Activity Dips in AT-Cut Crystals," Proc. 21st AFCS, pp. 420-435; 1967.

[5] Franx, C., "On Activity Dips of AT-Cut Crystals At High Levels of Drive," Proc. 21st AFCS, pp. 436-454; 1967.

[6] Tiersten, H.F., "Analysis of Non-linear Resonance in Thickness-Shear and Trapped-Energy Resonators," J. Acoust. Soc. Am., vol. 59, no. 4; April 1976.

[7] Tiersten, H.F., Private communication

[8] Tiersten, H.F., "The Evaluation of the Coefficient of Non-linear Resonance for SC-cut Quartz Resonators," these Proceedings.

[9] Tiersten, H.F., "An Analysis of Non-linear Resonance in Electroded Contoured AT- and SC-cut Quartz Crystal Resonators," Proc. 38th AFCS; 1984.

[10] Stevens, D.S., & Tiersten, H.F., "An Analysis of SC-cut Quartz Trapped Energy Resonators with Rectangular Electrodes," Proc. 35th AFCS; May 1981.

[11] Tiersten, H.F. and Stevens, D.S., "An Analysis of Contoured SC-Cut Quartz Crystal Resonators," Proc. 37th AFCS, pp. 37-45; 1982.

[12] Stevens, D.S., "An Analysis of Arbitrarily Anisotropic Contoured Quartz Crystal Resonators," Ph. D. Thesis, Rensselaer Polytechnic Institute, Troy, N.Y.; Aug. 1984.

[13] Tiersten, H.F., "Analysis of Intermodulation in Rotated Y-cut Quartz Thickness-Shear Resonators," Proc. 28th AFCS; 1974.

[14] Tiersten, H.F., "An Analysis of Intermodulation in Thickness-Shear and Trapped-Energy Resonators," J. Acoust Soc. Am., vol. 57, no. 3; March 1975.

[15] Smythe, R.C., "Intermodulation in Thickness-Shear Resonators," Proc. 28th AFCS; 1974.

[16] Smythe, R.C., "Bulk Acoustic Wave Filters," in Precision Frequency Control, (E.A. Gerber & A.D. Ballato, edis), pp. 188-228; Academic Press, New York; 1985.

[17] Howard, M.D., Smythe, R.C. and Morley, P.E., "Monolithic Crystal Filters having Improved Intermodulation and Power Handling Capability," these Proceedings.

[18] Planat, M., Theobald, G., & Gagnepain, J.J., "Intermodulation in X-Cut Lithium Tantalate Resonators," El. Lett., v. 16, no. 5, pp. 174-175; 28, Feb., 1980.

THE EVALUATION OF THE COEFFICIENT OF NONLINEAR RESONANCE FOR SC-CUT QUARTZ RESONATORS

H.F. Tiersten and D.S. Stevens
Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, New York 12180-3590

Abstract

In the recent treatment of nonlinear resonance in contoured quartz resonators the influence of the quadratic nonlinearities was included in addition to the cubic nonlinearity. In that work the solutions resulting from the quadratic nonlinearities were taken in the form of infinite series, from each of which one dominant term was selected and the others were assumed to be negligible. However, since some of the other terms that were ignored are not actually negligible, the procedure is deemed to be not quite adequate. Consequently, in this work forms of solutions resulting from the quadratic nonlinearities are found, each of which consists of a sum of only two terms. Since these new forms contain only a relatively small number of terms, the entire effect of the quadratic nonlinearities may readily be included in the nonlinear resonance analysis. Hence, in this work the solution for nonlinear resonance in contoured SC-cut quartz resonators containing the complete influence of the quadratic nonlinearities is obtained. Furthermore, the expressions for the coefficients of the quadratic nonlinearities for the SC-cut, which were not obtained in the earlier work, are obtained here. The more complete solution for nonlinear resonance in contoured quartz resonators obtained in this work is employed in the evaluation of the coefficient of nonlinear resonance for the SC-cut from measurements on the fundamental and third and fifth harmonic, and the data is shown to be consistent with the analytical description.

1. Introduction

In the recent analysis[1] of nonlinear resonance in contoured quartz crystal resonators the influence of the quadratic nonlinearities was included as well as the cubic nonlinearity. The solutions due to the quadratic nonlinearities were taken in the form of infinite series, from each of which one dominant term was retained and the remaining ones were neglected. In this work, new forms of solutions resulting from the quadratic nonlinearities are found, each of which contains only two terms and, hence, enables the entire effect of the quadratic nonlinearities to be included. Consequently, in this work the solution for nonlinear resonance in doubly-rotated contoured quartz resonators containing the complete influence of the quadratic nonlinearities is obtained. In addition, the expressions for the coefficients of the quadratic nonlinearities for doubly-rotated orientations and, in particular, for the SC-cut, which were not presented in the earlier work[1], are presented here. In all other respects the analysis follows the procedure employed in the earlier work[1]. In particular we again observe that since the modes are essentially thickness modes with slow transverse variation, only the thickness dependence of the elastic nonlinearities need be retained.

First the case of pure thickness resonance is considered and then the treatment is extended to include the transverse shape of the harmonic modes in the contoured resonator. The linear portions of the equation in the dominant thickness eigendisplacement of interest, which is used in the description of the transverse behavior of the modes in the contoured resonator, are the same as those in the new equation[2] describing the transverse modal behavior of doubly-rotated contoured quartz resonators, but with a change in the thickness-wavenumber caused by the quadratic nonlinearities considered here. Since the analysis is to be applicable to the doubly-rotated SC-cut, in the case of both the pure thickness resonator and that of the contoured resonator the mechanical displacement field is decomposed along the eigenvector triad of the pure thickness solution, exactly as in the linear case[2]. Of course, all conditions imposed on that work are applicable here and since the nonlinearity is small also, only the thickness dependence of all electrical variables is included in the treatment, as in all the other work in this area[3,4]. Naturally, the nonlinear portions of the equation in the dominant thickness eigendisplacement are the same as those that occur in the pure thickness case, but with a slow transverse variation, since it has been shown[4] that this equation reduces to the pure thickness equation when the transverse variation is suppressed.

The steady-state solutions to the nonlinear forced vibration problems are obtained by means of an asymptotic iterative procedure and an expansion in the eigensolutions while retaining the nonlinear correction to the eigensolution that has a resonant frequency in the vicinity of the driving frequency. The slow transverse variation in the mode is included in the nonlinear correction by appropriately averaging over the plate. An equation relating the change in resonant frequency resulting from the nonlinearity to the current through the crystal, independent of any external circuitry, is obtained. Both this latter equation and earlier calculations[1] of nonlinear resonance curves indicate a significant dependence of the change in resonant frequency resulting from the nonlinearity on the order of the harmonic overtone of the contoured resonator. Finally, the analysis is employed in the determination of the cubic nonlinearity for the SC-cut from measurements[5] of nonlinear resonance on contoured SC-cuts.

2. Thickness Equations

A schematic diagram of a plano-convex quartz crystal resonator is shown in Fig.1 along with the associated coordinate system. Since the modes of interest in contoured crystal resonators are essentially thickness-modes varying slowly along the plate and the amplitude of the motion itself is small, it is appropriate to consider only the thickness (X_2) dependence of the elastic nonlinearities, by virtue of the small piezoelectric coupling of quartz. Accordingly, we write the pure thickness equations for the general anisotropic case in the form

$$\bar{c}^{(1)}u_{1,22} - \rho\ddot{u}_1 + \rho\omega^2 K X_2 = -\gamma[(u_{1,2})^3]_{,2} - \beta_1[(u_{1,2})^2]_{,2} - 2\beta_2[u_{1,2}u_{2,2}]_{,2} - 2\beta_3[u_{1,2}u_{3,2}]_{,2}, \quad (2.1)$$

$$\bar{c}^{(2)}u_{2,22} - \rho\ddot{u}_2 = -\beta_2[(u_{1,2})^2]_{,2}, \quad (2.2)$$

$$\bar{c}^{(3)}u_{3,22} - \rho\ddot{u}_3 = -\beta_3[(u_{1,2})^2]_{,2}, \quad (2.3)$$

since u_1 is taken to be the "large" thickness-driven eigendisplacement in this work. We note that the mechanical displacement field u_1 is referred to the

CH2186-0/85/0000-0325$1.00©1985IEEE

eigenvector triad of the linear piezoelectric thickness solution, in accordance with Sec. 2 of Ref. 2, and the $\bar{c}^{(n)}$ denote the piezoelectrically stiffened eigenvalues. As a result of decomposing u in the diagonal system, the linear portions of Eqs. (2.1) - (2.3) are uncoupled in this general anisotropic case. The expressions for the effective nonlinear elastic constants on the right-hand sides of (2.1) - (2.3) are obtained in terms of the fundamental elastic constants of second, third and fourth order by transforming the mechanical displacement components $\hat{u}_r$ in the original coordinate system to the eigenvector components u_j in the nonlinear thickness (X_2)-dependent terms with the aid of existing relations[6], and are given by[7]

$$\gamma = \frac{1}{2} c_{2222} + Q_{1M}Q_{1R}c_{2MR222} + \frac{1}{6} Q_{1M}Q_{1R}Q_{1N}Q_{1K}c_{4MR2N2K2},$$

$$\beta_1 = \frac{3}{2} Q_{1M}c_{2M22} + \frac{1}{2} Q_{1M}Q_{1K}Q_{1R}c_{3MK2R2},$$

$$\beta_2 = \frac{1}{2} Q_{2M}[c_{2M22} + Q_{1K}Q_{1R}c_{3MK2R2}],$$

$$\beta_3 = \frac{1}{2} Q_{3M}[c_{2M22} + Q_{1K}Q_{1R}c_{3MK2R2}], \qquad (2.4)$$

in which the Q_{LM} are the components of the orthogonal transformation from the original coordinate system to the thickness-eigendisplacement coordinate system, and are discussed in Sec. 2 of Ref. 2.

Also referred to the eigenvector triad, the corresponding components of the Piola-Kirchhoff double vector take the form

$$K_{21} = \bar{c}^{(1)}u_{1,2} + \left(e_{26}C + c^{(1)}K + e_{26}\frac{V}{2h}\right)e^{i\omega t} + \gamma(u_{1,2})^3 + \beta_1(u_{1,2})^2 + 2\beta_2 u_{1,2}u_{2,2} + 2\beta_3 u_{1,2}u_{3,2}, \qquad (2.5$$

$$K_{22} = \bar{c}^{(2)}u_{2,2} + \beta_2(u_{1,2})^2,$$

$$K_{23} = \bar{c}^{(3)}u_{3,2} + \beta_2(u_{1,2})^3, \qquad (2.6)$$

where, since u_1 is the "large" driven eigendisplacement, we have included[7] the linear (in X_2) eigenpotential, which causes the term $e_{26}C$ and the inhomogeneous forcing term $\bar{c}^{(1)}K + e_{26}V/2h$ in (2.5) where

$$c^{(1)} = \bar{c}^{(1)}(1 - k_1^2), \quad k_1^2 = e_{26}^2/\epsilon_{22}\bar{c}^{(1)}, \qquad (2.7)$$

and ω is the driving frequency. Note that the first index 2 in K_{2j} refers to the normal to the plate and the second index j refers to the respective coordinates of the eigenvector triad. Note further that we have omitted the linear (in X_2) eigenpotential from K_{22} and K_{23} in (2.6) which would have caused terms $e_{22j}C$ ($j = 2, 3$) in K_{22} and K_{23}, respectively, that couple the $\bar{c}^{(j)}$ eigensolutions for shorted electroded[8-10]. Since the piezoelectric coupling is small in quartz and the $\bar{c}^{(j)}$ are sufficiently widely separated in the cut of interest, the resulting coupled transcendental frequency equation[8-10] very accurately uncouples and we effectively have three independent transcendental frequency equations[9], one for each thickness eigendisplacement u_j. By virtue of the forms taken for K_{2j} in (2.5) and (2.6), which indicate that the u_1 eigendisplacement is driven linearly by the voltage while the u_2 and u_3 eigendisplacements are not, the aforementioned uncoupling is already built into the description. Since both the thickness of the electrodes $2h'$ and the amplitude of the dominant eigendisplacement u_1 are small, the boundary conditions take the form

$$K_{21} = \mp 2\rho' h' \ddot{u}_1, \bar{\varphi} = \pm \frac{V}{2} e^{i\omega t}, \text{ at } X_2 = \pm h, \qquad (2.8)$$

$$K_{22} = 0, \quad K_{23} = 0, \quad \text{at } X_2 = \pm h. \qquad (2.9)$$

The expression for the X_2 component of the electric displacement vector, which determines the current, takes the form[2]

$$D_2 = e_{221}u_{1,2} - \epsilon_{22}\varphi_{,2}. \qquad (2.10)$$

As already alluded to verbally, the components of e_{22j}, K_{2j} and u_j occurring in this work are respectively related to the $\hat{e}_{22r}$, $\hat{K}_{2r}$ and $\hat{u}_r$, which are referred to the conventional plate axes, by the transformations[2]

$$e_{22j} = Q_{jr}\hat{e}_{22r}, \quad K_{2j} = Q_{jr}\hat{K}_{2r}, \quad u_j = Q_{jr}\hat{u}_r, \qquad (2.11)$$

in which Q_{jr} denotes the same orthogonal matrix as the Q_{LM}, which is obtained from the thickness eigenvalue problem[2].

3. Nonlinear Thickness Vibrations

As noted in the Introduction, although this problem was treated in Ref. 1, the solutions resulting from the quadratic nonlinearities were taken as infinite series, from each of which one dominant term was retained. Since not all the terms that were ignored were actually negligible, new forms of solutions due to the quadratic nonlinearities are obtained here, each of which contains only two terms and enables the entire effect of the quadratic nonlinearities to be included, thereby increasing the accuracy. In addition, solution of the nonlinear pure thickness vibration problem serves as a convenient, if not essential, model for the solution of the nonlinear resonance problem for the contoured resonator, since the mode in the contoured resonator is very nearly a thickness mode. Consequently, the relevant part of the solution of the nonlinear pure thickness vibration problem, namely the nonlinear eigensolution, is presented here.

Since the linear solution forms the essential starting point for the nonlinear resonance solution, it is briefly reproduced[11] here. In order to satisfy the inhomogeneous boundary condition $(2.8)_2$, we have taken $\bar{u}_1$ and $\bar{\varphi}$ in the form

$$\bar{u}_1 = u_1 + KX_2, \quad \bar{\varphi} = \frac{VX_2}{2h} + \varphi, \qquad (3.1)$$

in which $e^{i\omega t}$ has been suppressed, and we note that K and $V/2h$ already appear in the equations in Sec. 2 since they are on the u_1 and φ variables. Since $2h'$ is small, from (2.5) and (2.8), we have

$$K = -\frac{e_{26}V}{2hc^{(1)}}, \qquad (3.2)$$

and we note that the linear portion of (2.1) takes the form

$$\bar{c}^{(1)} u_{1,22} + \rho\omega^2 u_1 = -\rho\omega^2 K X_2, \qquad (3.3)$$

and

$$\varphi = \frac{e_{26}}{\epsilon_{22}} u_1 + C X_2, \quad \varphi = 0 \text{ at } X_2 = \pm h, \qquad (3.4)$$

and we note that C already appears in (2.5). The linear eigensolution (V = 0) takes the form

$$u_{1n} = A_n \sin \eta_n X_2, \quad \varphi_n = \frac{e_{26}}{\epsilon_{22}} u_{1n} + C_n X_2, \qquad (3.5)$$

$$\omega_n = \sqrt{\frac{\bar{c}^{(1)}}{\rho}}\, \eta_n, \quad C_n = -\frac{e_{26} A_n}{\epsilon_{22} h} \sin \eta_n h, \qquad (3.6)$$

$$\tan \eta_n h = \frac{\eta_n h}{k_1^2 + \hat{R}\eta_n^2 h^2}, \quad \hat{R} = \frac{2\rho' h'}{\rho h}, \qquad (3.7)$$

and since $k_1 \ll 1$ and $\hat{R} \ll 1$, from $(3.7)_1$, we obtain

$$\eta_n h = \frac{n\pi}{2}\left(1 - \frac{4k_1^2}{n^2\pi^2} - \hat{R}\right). \qquad (3.8)$$

In the vicinity of a resonance, say the Nth, one term dominates the series solution of the forced vibration problem, and we are interested in finding the influence of the nonlinearities on the Nth eigensolution at the driving frequency ω, which is near ω_N. To this end we obtain the solution to the nonlinear pure thickness vibration problem by systematically iterating from the linear solution to successive orders in the dominant but small amplitude A_N. From $(3.5)_1$ we see that the zeroth iterate, which is the linear solution, for the Nth eigensolution may be written using complex notation in the form

$${}_0u_1 = \sin \eta_N X_2 \frac{1}{2}\left[A_N e^{i\omega t} + A_N^* e^{-i\omega t}\right], \qquad (3.9)$$

the substitution of which in (2.2) and $(2.9)_1$ yields

$$\bar{c}^{(2)}\, {}_1u_{2,22} - \rho_1 \ddot{u}_2 = \frac{\beta_2}{2}\eta_N^3 \sin 2\eta_N X_2 \left[A_N A_N^* + A_N^2 e^{i2\omega t}\right] \qquad (3.10)$$

$$\bar{c}^{(2)}\, {}_1u_{2,2} + \beta_2 \eta_N^2 \cos^2 \eta_N h \left[A_N A_N^* + A_N^2 e^{i2\omega t}\right] = 0 \quad \text{at } X_2 = \pm h, \qquad (3.11)$$

where the real part of the complex bracket is understood in the usual way. Since the piezoelectric coupling and amplitude A_N of the dominant displacement are both small, the influence of the external circuitry on the nonlinear eigensolution can be shown[7] to be negligible and can be ignored. As a solution of (3.10) and (3.11) consider

$${}_1u_2 = B_{NO} \sin 2\eta_N X_2 + B_{N2} \sin 2\eta_N X_2 e^{i2\omega t} + H_{NO} X_2 + H_{N2} \sin \eta_{NO} X_2 e^{i2\omega t}, \qquad (3.12)$$

which satisfies (3.10) provided

$$\eta_{NO} = 2\omega/\sqrt{\bar{c}^{(2)}/\rho}, \qquad (3.13)$$

$$B_{NO} = -\frac{\beta_2 \eta_N A_N A_N^*}{8\bar{c}^{(2)}}, \quad B_{N2} = \frac{\beta_2 \eta_N A_N A_N^*}{8(\bar{c}^{(1)} - \bar{c}^{(2)})}, \qquad (3.14)$$

and satisfies (3.11) provided

$$H_{NO} = -\frac{\beta_2 \eta_N^2 A_N A_N^*}{4\bar{c}^{(2)}}, \quad H_{N2} = \frac{\beta_2 \eta_N^2 A_N^2}{8(\bar{c}^{(1)} - \bar{c}^{(2)})\eta_{NO} \cos \eta_{NO} h}, \qquad (3.15)$$

and we note for further use that for ω in the vicinity of ω_N, we have approximately but very accurately

$$\eta_{NO} = 2\kappa\eta_N, \qquad (3.16)$$

where

$$\kappa = \sqrt{\bar{c}^{(1)}/\bar{c}^{(2)}}. \qquad (3.17)$$

A similar first iterate solution can be found for ${}_1u_3$ but not for ${}_1u_1$ because the 2ω terms for ${}_1u_1$ result in a secular case[7]. For the $\omega = 0$ terms the ${}_1u_1$ solution is of the same form as the ${}_1u_2$ solution in (3.12), but for the 2ω terms the ${}_1u_1$ solution for the secular case takes the form[7]

$${}_1u_1^{2\omega} = (\hat{H}_{N2} \sin 2\eta_N X_2 + \hat{B}_{N2} X_2 \cos 2\eta_N X_2) e^{i2\omega t}, \qquad (3.18)$$

where

$$\hat{B}_{N2} = -\beta_1 \eta_N^2 A_N^2/8\bar{c}^{(1)}, \quad \hat{H}_{N2} = \beta_1 \eta_N A_N^2/16\bar{c}^{(1)}. \qquad (3.19)$$

We will provide further details here for the ${}_1u_2$ solution only[7], but the final results will be written for the full solution. Substituting from (3.9) and (3.12) in both the quadratic and cubic terms in (2.1) and $(2.8)_1$, with (2.5) in the absence of β_1 and β_3, while retaining all terms of order $A_N^2 A_N^*$ and containing the time dependence $e^{i\omega t}$ only and employing some trigonometric identities, we obtain the inhomogeneous differential equation

$$\bar{c}^{(1)}\, {}_2u_{1,22} - \rho\, {}_2\ddot{u}_1 = -\rho\omega^2 K X_2 e^{i\omega t} + \frac{9}{16}\gamma A_N^2 A_N^* \eta_N^4 (\sin \eta_N X_2 + \sin 3\eta_N X_2) e^{i\omega t} + \frac{\beta_2^2 A_N^2 A_N^*}{8}\left[\frac{(7\bar{c}^{(2)} - 6\bar{c}^{(1)})}{\bar{c}^{(2)}(\bar{c}^{(1)} - \bar{c}^{(2)})}\eta_N^4 \sin \eta_N X_2 + \frac{3(3\bar{c}^{(2)} - 2\bar{c}^{(1)})}{\bar{c}^{(2)}(\bar{c}^{(1)} - \bar{c}^{(2)})}\eta_N^4 \sin 3\eta_N X_2 + \frac{\eta_N^3}{2(\bar{c}^{(1)} - \bar{c}^{(2)})\cos \eta_{No} h} \cdot \left(\eta_N^+ \sin \eta_N^+ X_2 + \eta_N^- \sin \eta_N^- X_2\right)\right] e^{i\omega t}, \qquad (3.20)$$

where

$$\eta_N^+ = \eta_{NO} + \eta_N, \quad \eta_N^- = \eta_{NO} - \eta_N, \qquad (3.21)$$

and homogeneous boundary conditions

$$\bar{c}^{(2)}\, {}_2u_{1,2} - \frac{e_{26}^2}{\epsilon_{22} h} A_N \sin \eta_N h \pm 2\rho' h' \ddot{u}_1 = 0 \quad \text{at } X_2 = \pm h, \qquad (3.22)$$

since each nonlinear term in $(2.8)_1$ with (2.5) contains $\cos \eta_N h$ which vanishes to the order being considered by virtue of (3.8) and the fact that k_1, $\hat{R}$ and $A_N \eta_N$ are small. Equations (3.20) and (3.22) are for the second iterate ${}_2u_1$ solution of the nonlinear thickness problem. Since we regard A_N and η_N appearing in Eqs. (3.20 and (3.22) as presently unknown and the $\sin \eta_N X_2$ term will arise as a solution of the linear differential operator appearing in Eq. (3.20), the

sin $\eta_N X_2$ is regarded as homogeneous and we say the iterate is corrected[12]. As the Nth eigensolution of the linear differential equation (3.20), we take

$$ {}_2u_1 = u_N e^{i\omega t}, \tag{3.23} $$

where

$$ u_N = A_N \sin \eta_N X_2 + L_N \sin 3\eta_N X_2 + S_N \sin \eta_N^+ X_2 + D_N \sin \eta_N^- X_2, \tag{3.24} $$

the substitution of which in (3.20) yields

$$ \rho\tilde{\omega}_N = \bar{c}^{(1)}\eta_N^2 + \frac{9}{16}\gamma\sigma\eta_N^4 - \frac{\beta_2^2(7\bar{c}^{(2)} - 6\bar{c}^{(1)}}{8\bar{c}^{(2)}(\bar{c}^{(1)} - \bar{c}^{(2)})}\sigma\eta_N^4, \tag{3.25} $$

$$ L_N = -\frac{9}{128}\frac{\gamma}{\bar{c}^{(1)}}\eta_N^2\sigma A_N - \frac{3\beta_2^2(3\bar{c}^{(2)} - 2\bar{c}^{(1)})}{64\bar{c}^{(1)}\bar{c}^{(2)}(\bar{c}^{(1)} - \bar{c}^{(2)})}\eta_N^2\sigma A_N, $$

$$ S_N = \frac{\beta_2^2\eta_N^3\eta_N^+\sigma A_N}{16\bar{c}^{(1)}(\bar{c}^{(1)} - \bar{c}^{(2)})(\eta_N^2 - \eta_N^{+2})\cos \eta_{NO}h}, $$

$$ D_N = \frac{\beta_2^2\eta_N^3\eta_N^-\sigma A_N}{16\bar{c}^{(1)}(\bar{c}^{(1)} - \bar{c}^{(2)})(\eta_N^2 - \eta_N^{-2})\cos \eta_{NO}h}, \tag{3.26} $$

where

$$ \sigma = A_N A_N^*. \tag{3.27} $$

Substituting from (3.23), (3.24), (3.26) and (3.27) into (3.22) and observing that since $k_1^2 \ll 1$, $\hat{R} \ll 1$ and $\sigma \ll 1$, the roots $\eta_N h$ of (3.22) must differ from $N\pi/2$ by small quantities, say $\tilde{\Delta}_N$, so that we may write

$$ \eta_N h = (N\pi/2) - \tilde{\Delta}_N, \quad N \text{ odd}, \tag{3.28} $$

and expanding and retaining terms linear in $\tilde{\Delta}_N$, we obtain

$$ \tilde{\Delta}_N = \frac{2}{N\pi}k_1^2 + \frac{N\pi\hat{R}}{2} - (-1)^{\frac{N-1}{2}}\frac{\beta_2^2\sigma\eta_N^2}{16\bar{c}^{(1)}(\bar{c}^{(1)} - \bar{c}^{(2)})} \cdot \cdot \left[\frac{\eta_N^{+2}\cos \eta_N^+ h}{(\eta_N^2 - \eta_N^{+2})\cos \eta_{NO}h} + \frac{\eta_N^{-2}\cos \eta_N^- h}{(\eta_N^2 - \eta_N^{-2})\cos \eta_{NO}h}\right]. \tag{3.29} $$

Now, the substitution of (3.28) and (3.29) into (3.25) yields the nonlinear eigenfrequency $\tilde{\omega}_N$ for pure thickness vibrations in the form

$$ \tilde{\omega}_N^2 = \frac{\bar{c}^{(1)}}{\rho}\frac{N^2\pi^2}{4h^2}\left[1 - \frac{8k_1^2}{N^2\pi^2} - 2\hat{R} + \left(\frac{9}{16}\frac{\gamma}{\bar{c}^{(1)}} - \beta_2^2\nu_2\right)\frac{N^2\pi^2}{4h^2}\sigma + \frac{\beta_2^2\zeta_{N2}\sigma}{h\bar{c}^{(1)}(\bar{c}^{(1)} - \bar{c}^{(2)})}\right], \tag{3.30} $$

where

$$ \nu_2 = \frac{7\bar{c}^{(2)} - 6\bar{c}^{(1)}}{8\bar{c}^{(1)}\bar{c}^{(2)}(\bar{c}^{(2)} - \bar{c}^{(1)})}, \quad \zeta_{N2} = \frac{(-1)}{8}^{\frac{N-1}{2}} \cdot \cdot \eta_N\left[\frac{\eta_N^{+2}\cos \eta_N^+ h}{(\eta_N^2 - \eta_N^{+2})\cos \eta_{NO}h} + \frac{\eta_N^{-2}\cos \eta_N^- h}{(\eta_N^2 - \eta_N^{-2})\cos \eta_{NO}h}\right], \tag{3.31} $$

and we recall that the influence of the nonlinear coefficients β_1 and β_3 is not contained in (3.30) because they were omitted in the treatment[7]. Since we are interested in the driven nonlinear solution for the contoured resonator, we carry the nonlinear solution for pure thickness vibrations no further, i.e., we do not obtain the driven solution for pure thickness vibrations or present the dependence on β_1 and β_3. As we shall see, the amount of the solution to the nonlinear pure thickness vibration problem presented in this section is crucial for obtaining the solution to the problem of the nonlinear vibrations of the contoured resonator, which is presented in the next section.

4. Nonlinear Vibrations of Contoured Resonators

In the linear case it has been shown that the inhomogeneous differential equation for the forced vibrations of contoured SC-cut quartz resonators driven by the application of a voltage across the surface electrodes may be written in a form given in Eq. (3.24) of Ref. 2. In view of that equation and the reasoning leading to it and the solution to the nonlinear pure thickness problem presented in the last section, it can be shown[13,7] that, for a driving frequency ω in the vicinity of the linear resonant frequency ω_{NMP} of one of the families of modes associated with the Nth harmonic, the inhomogeneous differential equation for the nonlinear forced vibrations of a contoured SC-cut quartz resonator driven by a voltage across the surface electrodes, i.e., the equation for the second iterate, may be written in the form

$$ \sum_{n\neq N}\left[M_n\frac{\partial^2 u_1^n}{\partial X_1^2} + Q_n\frac{\partial^2 u_1^n}{\partial X_1\partial X_3} + P_n\frac{\partial^2 u_1^n}{\partial X_3^2} - \frac{n^2\pi^2\hat{c}_n^{(1)}}{4h^2}u_1^n - \rho\ddot{u}_1^n\right] + M_N\frac{\partial^2 u_1^N}{\partial X_1^2} + Q_N\frac{\partial^2 u_1^N}{\partial X_1\partial X_3} + P_N\frac{\partial^2 u_1^N}{\partial X_3^2} + \left(\rho\omega^2\hat{u}_N + \bar{c}^{(1)}\frac{\partial^2\hat{u}_N}{\partial X_2^2}\right)\tilde{u}_1^N e^{i\omega t} $$
$$ = \frac{\rho\omega^2 e_{26}VX_2 e^{i\omega t}}{c^{(1)}2h} + \frac{9}{16}\gamma\tilde{u}_1^{N3}\hat{A}_N^2\hat{A}_N^*\eta_N^4(\sin \eta_N X_2 + \sin 3\eta_N X_2)e^{i\omega t} $$
$$ + \frac{1}{8}\beta_2^2\tilde{u}_1^{N3}\hat{A}_N^2\hat{A}_N^*\left[\frac{7\bar{c}^{(2)} - 6\bar{c}^{(1)})}{\bar{c}^{(2)}(\bar{c}^{(1)} - \bar{c}^{(2)})}\eta_N^4 \sin \eta_N X_2 + \frac{3(\bar{c}^{(2)} - 2\bar{c}^{(1)})}{\bar{c}^{(2)}(\bar{c}^{(1)} - \bar{c}^{(2)})}\eta_N^4 \sin 3\eta_N X_2 + \frac{1}{2(\bar{c}^{(1)} - \bar{c}^{(2)})\cos \eta_{NO}h}\eta_N^3(\eta_N^+ \sin \eta_N^+ X_2 + \eta_N^- \sin \eta_N^- X_2)\right] e^{i\omega t}, \tag{4.1} $$

where M_n, P_n and Q_n are given in Eqs. (3.18) of Ref. 2 and

$$ \tilde{u}_1^N = \tilde{u}_1^N(X_1, X_3), \quad \hat{c}_n^{(1)} = \bar{c}^{(1)}\left(1 - \frac{8k_1^2}{n^2\pi^2} - 2\hat{R}\right), $$
$$ \hat{u}_N = \hat{A}_N \sin \eta_N X_2 + \hat{L}_N \sin 3\eta_N X_2 + \hat{S}_N \sin \eta_N^+ X_2 + \hat{D}_N \sin \eta_N^- X_2, $$
$$ 2h = 2h_o[1 - (X_1^2 + X_3^2)/4Rh_o], \tag{4.2} $$

and we note that $\tilde{u}_1^N$ is a slowly varying function of X_1 and X_3. As noted in Sec. 3 of Ref. 2, for each driven harmonic family, or equivalently, for each harmonic family of eigensolutions, the mixed derivative term may be eliminated by referring the differential equation to

the planar orthogonal coordinate system X_1', X_3', which is rotated from X_1, X_3 by the angle $\hat{\beta}_n$ defined in Eq.(3.29) of Ref.2. When referred to the X_1', X_3' coordinate system, the driven linear equation for the Nth harmonic family takes the form given in Eq.(3.30) of Ref.2, which is

$$M_n' \frac{\partial^2 \tilde{u}_1^n}{\partial X_1'^2} + P_n' \frac{\partial^2 \tilde{u}_1^n}{\partial X_3'^2} - \frac{n^2\pi^2\hat{c}_n^{(1)}}{4h^2}\tilde{u}_1^n - \rho\ddot{\tilde{u}}_1^n = \rho\omega^2(-1)^{\frac{n-1}{2}} \frac{e_{26}4Ve^{i\omega t}}{n^2\pi^2}, \tag{4.3}$$

where M_n' and P_n' are given in terms of M_n, P_n, Q_n and $\hat{\beta}_n$ in Eqs.(3.31) of Ref.2. Note that the transformation to the primed coordinate system does not affect the nonlinear terms on the right-hand side of (4.1).

It has been shown that the linear eigensolutions of the associated homogeneous problem, i.e., of (4.1) without the nonlinear terms, i.e., (4.3), and with $V = 0$, can be written in the form[1,4]

$$u_{1nmp} = e^{-\alpha_n \frac{X_1'^2}{2}} H_m\left(\sqrt{\alpha_N}X_1'\right) e^{-\beta_n \frac{X_3'^2}{2}} \cdot H_p\left(\sqrt{\beta_n}X_3'\right) \sin\frac{n\pi X_2}{2h} e^{i\omega_{nmp}t}, \tag{4.4}$$

where H_m and H_p are Hermite polynomials and

$$\alpha_n^2 = n^2\pi^2\hat{c}_n^{(1)}/8Rh_o^3M_n', \quad \beta_n^2 = n^2\pi^2\hat{c}_n^{(1)}/8Rh_o^3P_n'. \tag{4.5}$$

The linear eigenfrequencies ω_{nmp} may be found from the relation

$$\rho\omega_{nmp}^2 = (\hat{c}_n^{(1)}n^2\pi^2/4h_o^2) + M_n'\gamma_{mn}^2 + P_n'\mu_{pn}^2, \tag{4.6}$$

where

$$\gamma_{mn}^2 = \alpha_n(1+2m), \quad \mu_{pn}^2 = \beta_n(1+2p), m,p = 0,2,4,\ldots, \tag{4.7}$$

which are determined from the condition that the series for H_m and H_p terminate and they be polynomials. Since only the harmonics are of interest in this work, we have

$$\tilde{u}_1^N = e^{-\alpha_N \frac{X_1'^2}{2}} e^{-\beta_N \frac{X_3'^2}{2}}, \quad m = p = 0, \quad \omega = \tilde{\omega}_{N00}. \tag{4.8}$$

Consider the Nth homogeneous equation, i.e., the nonlinear one consisting of all terms not included in the sum in (4.1) but with $V = 0$ and rotated to the primed coordinate system through the angle β_N, multiply by $\tilde{u}_1^N$ and integrate over the entire surface to obtain the weighted average with respect to X_1 and X_3 since the solution function in (4.4) is essentially the slowly varying thickness solution. The weighted average turns out to be most convenient because of the use of orthogonality in the forced vibration analysis. Then from the resulting X_2-dependent problem, by following the procedure used in the pure thickness problem treated in Sec.3, we obtain

$$\hat{L}_N = \frac{-9}{256}\frac{\gamma}{\bar{c}^{(1)}}\eta_N^2\hat{\sigma}\hat{A}_N - \frac{3\beta_2^2(3\bar{c}^{(2)} - 2\bar{c}^{(1)})}{128\bar{c}^{(1)}\bar{c}^{(2)}(\bar{c}^{(1)} - \bar{c}^{(2)})}\eta_N^2\hat{\sigma}\hat{A}_N,$$

$$\hat{S}_N = \frac{\beta_2^2\eta_N^3\eta_N^{+}\hat{\sigma}\hat{A}_N}{32\bar{c}^{(1)}(\bar{c}^{(1)} - \bar{c}^{(2)})(\eta_N^2 - \eta_N^{+2})\cos\eta_{NO}h}$$

$$\hat{D}_N = \frac{\beta_2^2\eta_N^3\eta_N^{-}\hat{\sigma}\hat{A}_N}{32\bar{c}^{(1)}(\bar{c}^{(1)} - \bar{c}^{(2)})(\eta_N^2 - \eta_N^{-2})\cos\eta_{NO}h}, \tag{4.9}$$

$$-M_N'\alpha_N - P_N'\beta_N - \frac{N^2\pi^2}{4h_o^2}\left(1 - \frac{8k_1^2}{N^2\pi^2} - 2\hat{R} - \frac{\beta_2^2\nu_2N^2\pi^2\hat{\sigma}}{2\ 4h_o} + \frac{\beta_2^2\zeta_{N2}\hat{\sigma}}{2h_o\bar{c}^{(1)}(\bar{c}^{(1)} - \bar{c}^{(2)})}\right)\bar{c}^{(1)} + \rho\tilde{\omega}_{N00}^2 = \frac{\bar{c}^{(1)}N^2\pi^2}{4h_o^2}\frac{9}{32}\frac{\gamma N^2\pi^2\hat{\sigma}}{4h_o^2}, \tag{4.10}$$

where

$$\hat{\sigma} = \hat{A}_N\hat{A}_N^*. \tag{4.11}$$

In obtaining (4.10) we have used the condition

$$\hat{\gamma}_N h = (N\pi/2) - \hat{\Delta}_N, \tag{4.12}$$

which is analogous to (3.28) and from which we obtain

$$\hat{\Delta}_N = \frac{2}{N\pi}k_1^2 + \frac{N\pi\hat{R}}{2} - \frac{\beta_2^2\hat{\sigma}\eta_N\zeta_{N2}}{4\bar{c}^{(1)}(\bar{c}^{(1)} - \bar{c}^{(2)})}, \tag{4.13}$$

in the same way that (3.29) was found. Substituting from (4.6) and $(4.2)_2$ into (4.10), we may write

$$\tilde{\omega}_{N00}^2 = \omega_{N00}^2 + \frac{N^2\pi^2\bar{c}^{(1)}}{4h_o^2\rho}\hat{\alpha}\hat{\sigma}, \tag{4.14}$$

where

$$\omega_{N00}^2 = \frac{N^2\pi^2}{4h_o^2}\frac{\hat{c}_N^{(1)}}{\rho}\left[1 + \frac{1}{N\pi}\sqrt{\frac{2h_o}{R}}\left(\sqrt{\frac{M_N'}{\hat{c}_N^{(1)}}} + \sqrt{\frac{P_N'}{\hat{c}_N^{(1)}}}\right)\right], \tag{4.15}$$

and we have employed (4.5) in writing (4.15). In writing (4.14) we have included the influence of β_1 and β_3, which have been omitted from the derivation[7], as we said we would, in the expression for $\hat{\alpha}$, which is of the form

$$\hat{\alpha} = \left(\frac{9}{32}\frac{\gamma}{\bar{c}^{(1)}} - \frac{\beta_2^2\nu_2}{2} - \frac{\beta_3^2\nu_3}{2} - \frac{53}{256}\frac{\beta_1^2}{\bar{c}^{(1)2}}\right)\frac{N^2\pi^2}{4h_o^2} + \frac{1}{2h_o}\left(\frac{\beta_2^2\zeta_{N2}}{\bar{c}^{(1)}(\bar{c}^{(1)} - \bar{c}^{(2)})} + \frac{\beta_3^2\zeta_{N3}}{\bar{c}^{(1)}(\bar{c}^{(1)} - \bar{c}^{(3)})}\right), \tag{4.16}$$

in which ν_3 and ζ_{N3} may be obtained[7] from the respective expressions for ν_2 and ζ_{N2} in (3.31)with (3.16), (3.17).

We now write the steady state solution of the inhomogeneous equation (4.1) in the form

$$u_1 = \tilde{u}_{NOO}(\hat{A}_N \sin \eta_N X_2 + \hat{L}_N \sin 3\eta_N X_2 + \hat{S}_N \sin \eta_N^+ X_2 +$$

$$\hat{D}_N \sin \eta_N^- X_2)e^{i\omega t} + \sum_{\substack{n\neq N \\ \text{together}}} \sum_{m\neq 0} \sum_{p\neq 0} H^{nmp}\tilde{u}_{nmp} \sin \eta_n X_2 e^{i\omega t}, \tag{4.17}$$

and we note that we are interested only in the vicinity of nonlinear resonance, i.e., when the terms containing $\tilde{u}_{NOO}$ are dominant. Substituting from (4.17) into (4.1) with (4.4), employing (4.6) in the linear terms and (4.10) in the nonlinear terms along with (4.9), using the orthogonality of the eigensolutions and, as usual, replacing the circular electrode by the circumscribed square, we obtain

$$(\omega^2 - \tilde{\omega}_{NOO}^2)\hat{A}_N = (-1)^{\frac{N-1}{2}} \frac{4e_{26}\mathcal{I}_{NOO}V\omega^2}{c^{(1)}N^2\pi^2 L_{NOO}}, \tag{4.18}$$

and we do not bother to write any of the H^{nmp} because, as already noted, we are interested in the solution only when ω is in the vicinity of ω_{NOO} and the NOOth eigenmode is dominant and all other eigenmodes are negligible and where[15]

$$\mathcal{I}_{NOO} = \frac{2\pi}{\sqrt{\alpha_N}\sqrt{\beta_N}} \operatorname{erf}\sqrt{\frac{\alpha_N}{2}}\,\ell \operatorname{erf}\sqrt{\frac{\beta_N}{2}}\,\ell, \quad L_{NOO} = \frac{\pi}{\sqrt{\alpha_N}\sqrt{\beta_N}}. \tag{4.19}$$

From Eqs. (4.11) and (4.14), we may write

$$\tilde{\omega}_{NOO}^2 = \omega_{NOO}^2 + \omega_N^2\hat{\alpha}\hat{A}_N\hat{A}_N^*, \quad \omega_N^2 = (\bar{c}^{(1)}/\rho)(N^2\pi^2/4h_o^2). \tag{4.20}$$

Thus, for an ω in the vicinity of ω_{NOO}, the solution can very accurately be written

$$u_1 = \hat{A}_N \sin\frac{N\pi X_2}{2h_o}\tilde{u}_{NOO}e^{i\omega t} - \frac{e_{26}VX_2}{c^{(1)}2h}e^{i\omega t},$$

$$\varphi = \frac{VX_2}{2h}e^{i\omega t} + \frac{e_{26}}{\epsilon_{22}}\hat{A}_N\tilde{u}_{NOO}\left(\sin\frac{N\pi X_2}{2h} - (-1)^{\frac{N-1}{2}}\frac{X_2}{h}\right)e^{i\omega t}, \tag{4.21}$$

where $\hat{A}_N$ must satisfy (4.18) with (4.20).

In order to find the required relation between the amplitude $\hat{A}_N$, the voltage across the crystal V and the current through the crystal I_c, we substitute from (4.21) into (2.10), which is then substituted into

$$I_c = -\int_{A_e} \dot{D}_2\, dX_1 dX_3, \tag{4.22}$$

where A_e is the area of the electrode, to obtain

$$I_c = \frac{i\omega\epsilon_{22}}{2h}(1+\hat{k}_1^2)\hat{A}_e V - i\omega e_{26}\frac{\tilde{A}_N}{h}\mathcal{I}_{NOO}, \tag{4.23}$$

in which, for convenience, we have again replaced the circular electrode by the circumscribed square and

$$\hat{A}_e = A_e\left(1 + \frac{\ell^2}{8Rh_o}\right), \quad \tilde{A}_N = (-1)^{\frac{N-1}{2}}\hat{A}_N, \quad \hat{k}_1^2 = \frac{k_1^2}{1-k_1^2}. \tag{4.24}$$

A very interesting and valuable relation giving the change Δ_N from the linear resonant frequency at maximum current due to the nonlinearity can be obtained from $(4.20)_1$ and (4.23), independent of any external circuitry, by noting that at resonance V can be neglected in (4.23) to obtain

$$I_c \approx -i\omega\epsilon_{26}\frac{\tilde{A}_N}{h}\mathcal{I}_{NOO}. \tag{4.25}$$

Then substituting $\tilde{\omega}_{NOO} = \omega_{NOO} + \Delta_N$ in $(4.20)_1$ and employing (4.25), we obtain

$$\Delta_N = \frac{\hat{\alpha}h^2 I_c I_c^*}{2\omega_{NOO}e_{26}^2\mathcal{I}_{NOO}^2}, \tag{4.26}$$

which is the complete equation relating the change in resonant frequency to the square of current.

5. The Coefficient of Nonlinear Resonance

Since the second[16] and third[17] order elastic constants are known for quartz, β_1, β_2 and β_3 can readily be calculated from (2.4). However, since the fourth order elastic constants are not known, γ cannot be calculated. Hence, all quantities appearing in the expression for $\hat{\alpha}$ in (4.16) can be calculated except γ. An examination of Eqs. (4.26) and (4.16) reveals that the coefficient γ can be evaluated from nonlinear resonance measurements, i.e., the change in resonant frequency with current level for any odd harmonic of a contoured resonator with the given orientation. In view of the foregoing, the consistency of the experimental data with the analytical description presented here can conveniently be checked by evaluating γ from data on successive overtones of the same resonator and different resonators with the same orientation to see to what extent it remains constant, as it should. Unfortunately, this is not exactly the case.

The coefficient γ has been evaluated from data provided by Filler[6] on the first, third and fifth harmonics of a number of 5.115 MHz fundamental contoured SC-cut quartz resonators, each of which has an electrode diameter of 5.52 mm, and the results are shown in Table I. It can be seen from the table that the agreement (or consistency) of the value of γ obtained from the different harmonics of any one resonator is quite decent, but that it differs a bit more from one resonator to another. We do not presently understand the reasons for this. From the table about the most we can estimate is that γ is somewhere between 5 and 6×10^{11} N/m^2 for the SC-cut. It should also be pointed out that other data provided by Filler[5] on only two harmonics of either 5 or 10 MHz third overtone contoured SC-cut resonators results in derivations in γ of up to twice the amount shown in the table. It should also be pointed out that data provided by Filler[5] on contoured AT-cut resonators results in considerably greater deviations in γ than for the SC-cut.

Acknowledgements

We wish to thank R.L. Filler of the U.S. Army Electronics Technology and Devices Laboratory for kindly providing the data on drive level sensitivity used in the evaluation of γ and R.C. Smythe of Piezo Technology, Inc. for many valuable discussions. This work was supported in part by the Army Research Office under Contract No. DAAG-29-82-K-0130.

References

1. H.F. Tiersten, "An Analysis of Nonlinear Resonance in Electroded Contoured AT- and SC-Cut Quartz Crystal Resonators," Proceedings of the 38th Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey and Institute of Electrical and Electronics Engineers, New York, IEEE Cat. No. 84CH2062-8, 132 (1984).
2. D.S. Stevens and H.F. Tiersten, "An Analysis of Doubly-Rotated Contoured Quartz Crystal Resonators," Proceedings of the 39th Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey and Institute of Electrical and Electronics Engineers, New York, these proceedings (1985).
3. 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).
4. 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).
5. R.L. Filler, "The Amplitude Frequency Effect in SC-Cut Resonators," Proceedings of the 39th Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey and Institute of Electrical and Electronics Engineers, New York, these proceedings (1985).
6. H.F. Tiersten, "Nonlinear Electroelastic Equations Cubic in the Small Field Variables," J. Acoust. Soc. Am., 57, 660 (1975).
7. For more detail see H.F. Tiersten and D.S. Stevens, "An Analysis of Nonlinear Resonance in Electroded Contoured Quartz Crystal Resonators," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.
8. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.9, Sec.2.
9. H.F. Tiersten, "Electromechanical Coupling Factors and Fundamental Material Constants of Thickness Vibrating Piezoelectric Plates," Ultrasonics, 8, 19 (1970).
10. A. Ballato, "Doubly-Rotated Thickness Mode Plate Vibrators," in Physical Acoustics, edited by W.P. Mason and R.N. Thurston (Academic, New York, 1977), Vol.XIII, Sec.III.
11. H.F. Tiersten, "Analysis of Intermodulation in Thickness-Shear and Trapped Energy Resonators," J. Acoust. Soc. Am., 57, 667 (1975).
12. The procedure is a straightforward iterative procedure except when the functional form resulting from the nonlinear terms is identical with that satisfying the basic linear differential equation, at which point the coefficients associated with the functional form due to the nonlinearity are treated as unknown and identical with the coefficients of the linear solution. This is the reason the sin $\gamma_N X_2$ term in (3.20) is homogeneous whereas (3.10) is an inhomogeneous equation. This procedure is appropriate because the term on the right-hand side of (3.20) actually is a homogeneous term and appears to be inhomogeneous (known) only because of the nature of the formal procedure employed which tends to be misleading whenever this type of coincidence of functions occurs.
13. Equation (4.1) has been obtained by expansion in the plate waves for the contoured resonator. The terms for $n \neq N$ are just the usual linear terms and only one of the terms with subscript N is actually nonlinear, i.e., the one for $m = M$ and $p = P$. The terms containing M_N, P_N, Q_N are linear terms because of the slow variation of $\tilde{u}_1^N$ and the fact that $\hat{A}_N$ is small. The remaining terms containing N are due to either thickness (X_2) or time derivatives and simply contain $\tilde{u}_1^N(X_1, X_3)$. These latter terms are the same as those occurring in the nonlinear pure thickness problem treated in Sec.3, but now contain $\tilde{u}_1^N$, which is never differentiated in the nonlinear terms. This occurs essentially because the linear and nonlinear behavior remain separate in the equations since $\tilde{u}_1^N$ varies slowly and $\hat{A}_N$ is small, and, as a consequence, the linear terms that do not appear in a pure thickness problem result in exactly the same solution functions at the second iterate stage as at the zeroth iterate stage.
14. C.J. Wilson, "Vibration Modes of AT-Cut Convex Quartz Resonators," J. Phys., D7, 2449 (1974).
15. P.M. Morse and H. Feshbach, Methods of Theoretical Physics (McGraw-Hill, New York, 1953), p.786.
16. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 110, 1060 (1958).
17. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

TABLE I

COEFFICIENT OF NONLINEAR RESONANCE FOR THE SC-CUT

5.115 MHz Fundamental

Unit #/N	Diopters	γ (N/m^2)
502 - 1	1.0	4.8×10^{11}
3		5.84
5		5.40
545 - 1	1.25	5.07
- 3		5.78
- 5		4.74
792 - 1	1.37	5.47
- 3		6.67
- 5		6.04
361 - 1	1.37	5.82
- 3		6.57
- 5		5.51
703 - 1	1.5	4.06
- 3		4.11
- 5		3.79

Electrode Diameter $2\ell = 5.52$ mm

Data due to R.L. Filler, Ft. Monmouth.

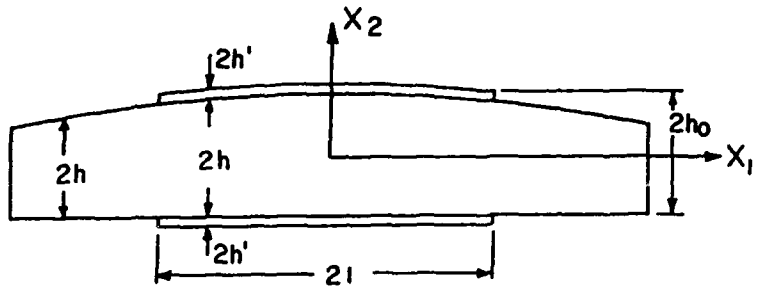


Figure I Schematic Diagram Showing a Cross-Section of the Plano-Convex Resonator

LASER PROCESSED MINIATURE $LiTaO_3$ RESONATORS AND MONOLITHICS FILTERS

R. LEFEVRE, L. JENSELME and D. SERVAJEAN

Centre National d'Etudes des Télécommunications
Laboratoire de Bagneux
196 Rue de Paris - 92220 BAGNEUX (FRANCE)

Summary

Because of it's high Q and large electro-mechanical coupling coefficient, Lithium Tantalate ($LiTaO_3$) is a very attractive material for piezoelectric resonators and monolithic filters which cannot be realized by Quartz or Ceramics. Theses properties and good frequency temperature characteristics [1] allow realization of large relative bandwidth filters and large resonance frequency shift resonators. On the other hand, miniature resonators and monolithic filters are expanding rapidly : They are used especially in mobile equipments such as portable measuring instruments, cellular Radio, conventional mobile radio, electronic clocks... [2].

For this kind of devices there are, at present time, two ways to manufacture them : Photolithography and mechanical sawing.

Mechanical sawing is very simple but does not allow all possible shapes and sizes. Typically only plates (rectangular and circular), bars and tuning forks can be cut up by this method [3].

Photolithography is a very powerfull process but it's possible only when etching-speed is sufficient. This is not the case for Lithium Tantalate that has an etching-speed of about 2 μm per hour in a boiling hydrofluoric-nitric solution. (In the same conditions, etching-speed of quartz is about 74 μm per hour).

So this paper reports a new machining process to cut up Lithium Tantalate plates by using a Q-switched YAG laser without no more damage induced in the crystal than in the case of mechanical sawing. Devices processed by this powerfull method are $LiTaO_3$ Y-45° cut resonators and monolithic filters vibrating in the length extension mode in the 0.15 to 4 MHz frequency range.

Electrical characteristics are measured and results are discussed.

Basic Principle of process

Laser machining is a well-known process already used to cup up, well and drill material. So, originality of our method is in the way by which cutting of Lithium Tantalate is performed.

In fact, Lithium Tantalate is transparent to the 1.06 μm laser radiation. Then to cut it up, it needs a high electrical field given by a giant laser pulse which induces a local breakdown in the crystal [4]. So at this point, absorption of laser radiation grows up quickly and crystal is burnt and cut up by the end of laser pulse.

This cutting mechanism is common for all crystal that presents a low threshold of optical damage (for instance this is also the case for $LiNbO_3$).

Cutting is very clean and crystal damage is limited to the laser spot size as in the case of mechanical sawing. Loss of material is negligible due to small size of laser beam (about 15 μm). To avoid thermal shock, we must adjust laser parameters (especially pulse rate). To achieve a fast cutting, displacement speed of substrate must be modulated. Typically, with a peak power of 1 KW and a 200 ns pulse duration it needs a mean time of about 4 minutes to cut up a two poles monolithic filter with a resonators size of about 6.3 x 0.6 mm.

Laser apparatus is fully automatized and driven by a desk computer so many shapes of devices and integrated suspensions can be achieved.

In our apparatus, laser beam is fixed and substrate displacement made by driven stepping motor tables (accuracy ± 1 μm).

Electrical characteristics of devices

Resonators

If some damage is induced in the crystal, electrical parameters such as electro-mechanical coupling coefficient K, Q-factor and aging of resonators will be different from thoses of same devices cut up by mechanical sawing. Experimental results on rectangular bar resonators cut up by the two methods and mounted with brass-nail do not show differences between their coupling coefficient which value is about 25.5 % (theoretical value is 26 % for this $LiTaO_3$ Y-cut). Some differences appear with the type of integrated suspension. We have designed two kinds of suspension named type I and type II (stirrup) [5] as shown in fig 1. Measures on these resonators are listed in the table I. With type II suspension, K values lie near 25.5 %. But with type I suspension additional static capacitance reduces the value of K. In the case of the 4.12 MHz resonator, parasitic static capacitance to resonator static capacitance ratio is about 5.7. This reduces K value down to 9 % ; this value is in good agreement with experimental result as shown in the Table I.

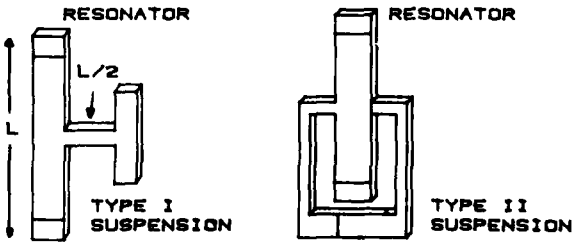


Fig. 1 : Y-45° cut $LiTaO_3$ resonators with integrated suspension

CH2186-0/85/0000-0333$1.00©1985IEEE

Measures on these resonators are listed in the table I. With type II suspension, K values lie near 25.5 %. But with type I suspension additional static capacitance reduces the value of K. In the case of the 4.12 MHz resonator, parasitic static capacitance to resonator static capacitance ratio is about 5.7. This reduces K value down to 9 % ; this value is in good agreement with experimental result as shown in the Table I.

TABLE I : Characteristics of some Y-45° cut $LiTaO_3$ resonators and variation of electro-mechanical coupling coefficient with the type of suspension.

RESON. FREQUENCY	SUSPENSION	Q-FACTOR	K(%)	RESONATOR SIZE L*W*T(mm)
.455 KHz	TYPE I	50,000	21	6.32*.63*.15
	TYPE II	50,000	25.6	
1 MHz	TYPE I	80,000	19.8	3.2*.32*.15
	TYPE II	80,000	25.5	
2 MHz	TYPE I	90,000	14.4	1.6*.16*.15
4.12 MHz	TYPE I	37,000	8.5	.75*.07*.15

Fig 2 shows the "Family" of resonators with type I suspension and 2 MHz two pôles monolithic filters while, on fig 3, resonators with type II suspension can be seen mounted in their can.

Fig 2 : "Family" of resonators with type I suspension and 2 MHz two pôles monolithic filters. Resonance frequencies are : 455 KHz ; 1 MHz ; 2 MHz and 4.12 MHz.

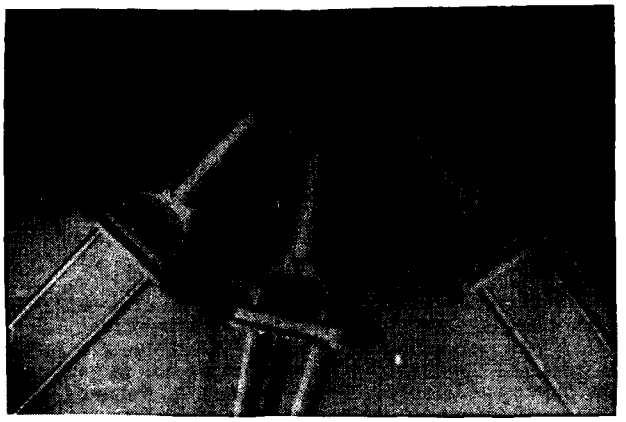


Fig. 3 : Resonators with type II suspension (strirrup) mounted in their can. Resonance frequency is : 455 KHz

The resonance mode of theses devices is free from spurious responses in the practical frequency range as it can be seen on fig 4.

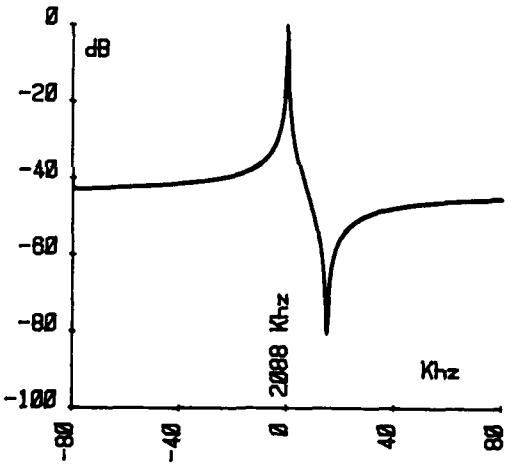


Fig. 4 : Frequency response of a 2 MHz resonators with a type I suspension.

However, theses devices are not yet optimized and new kinds of integrated suspension are under investigation, especially to achieve high Q values.

Monolithic filters

Filters are, two times two pôles monolithic filters (fig 5) in the .15 to 2 MHz frequency range. Relative bandwidth up to 2 % (with type I suspension) have been achieved with an inband ripple of about 0.2 dB and insertion loss as low as 0.5 dB (fig 6 and fig 7). But this limit can be overshot by using a double stirrup (fig 8) which reduces parasitic static capacitance. On the other hand, this kind of suspension allows welding of leads without disturb the electrical behaviour. No trimming has been performed on each cell.

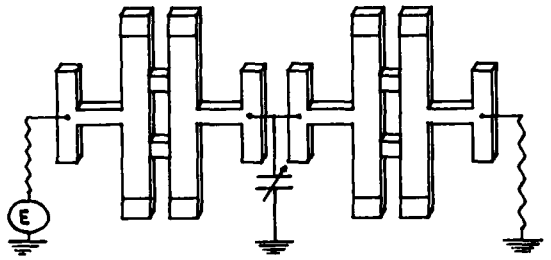


Fig. 5 : Two times two pôles monolithic filter with type I suspension.

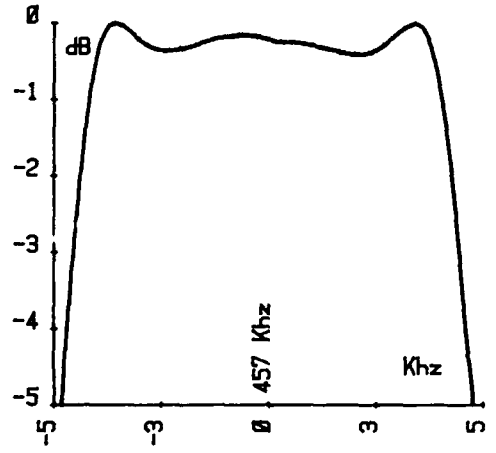


Fig. 6 : Pass-band of a two times two pôles monolithic filter at 457 KHz center frequency.

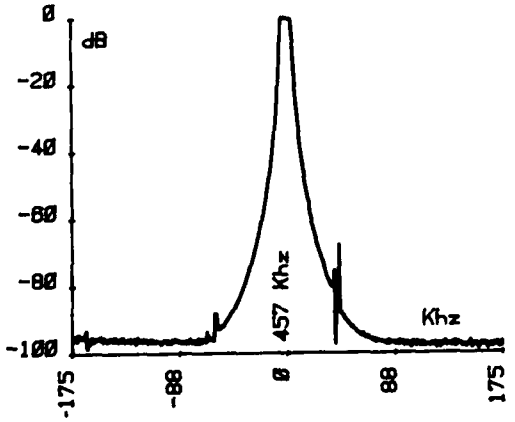


Fig 7 : Stopband of the same filter

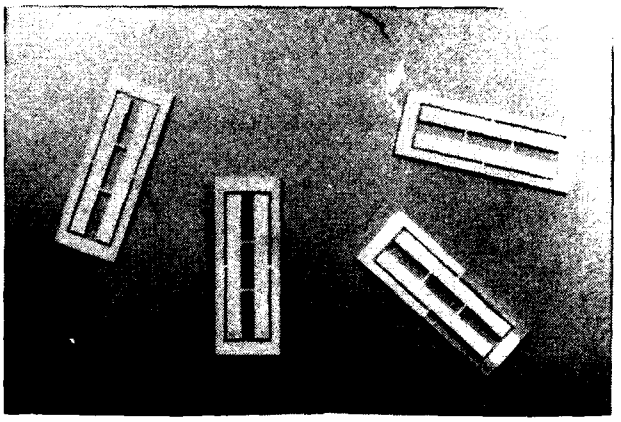


Fig. 8 : Two pôles monolithic filter at a 455 KHz center frequency with a double strirrup to reduce parasitic static capacitance and allow better mounting conditions.

Aging

We report some aging curves of resonators at various frequencies (455 KHz ; 1 MHz ; 2 MHz) and over 900 days. In fig 9, fig 10 and fig 11, we see that devices cut up by laser present a classical aging and there is no difference with same devices processed by mechanical sawing. From this curve we have computed a mean aging value of 3×10^{-7} per month after two years at room temperature.

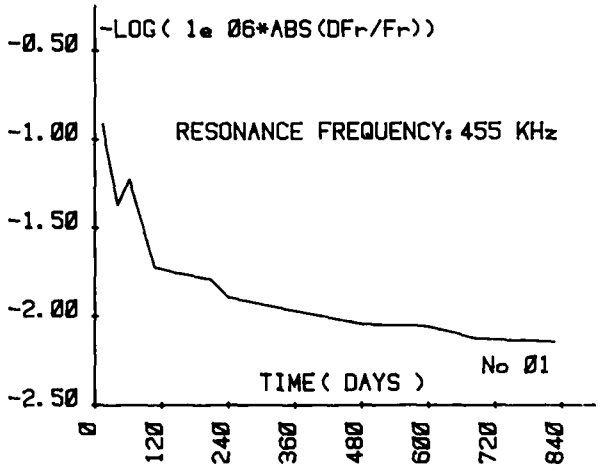


Fig. 9 : Aging characteristics of $LiTaO_3$ resonators with type I suspension measured at room temperature.

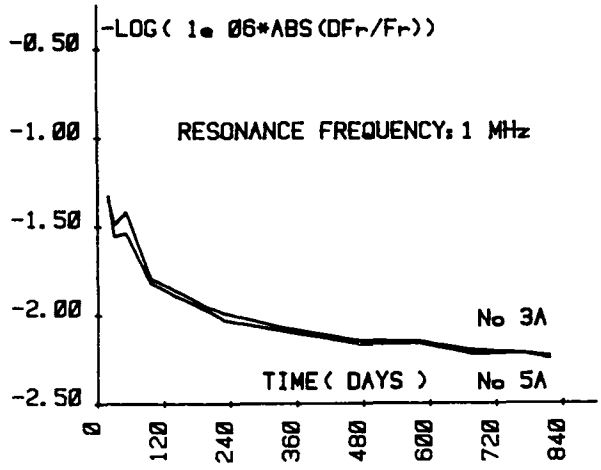


Fig. 10 : Aging characteristics of two resonators with type I suspension at room temperature.

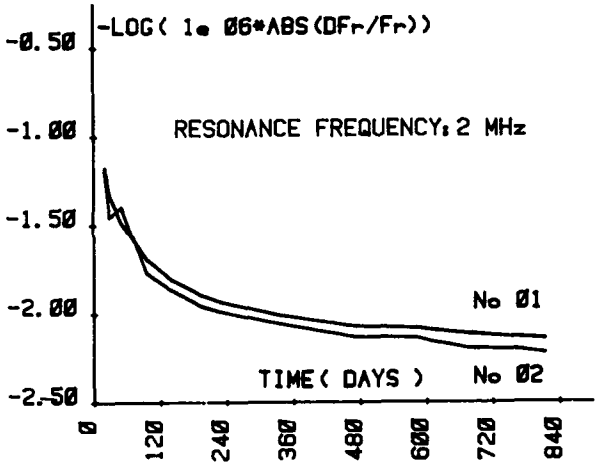


Fig. 11 : Aging characteristics of two resonators with type I suspension at room temperature.

Typical steps of the laser process

1) We start with rectangular plates of 20 mm length, 16 mm width and 0.15 mm thick.
2) Cut up is made by leaving little bridge of material to prevent fall of devices in the further operations (fig 12).
3) Plates are then cleaned during about 10 minutes in an hydrofluoric-nitric solution which removes also cut-up pieces.
4) Devices receive their thin film coating through copper masks processed by the same laser (fig 13).
5) Plate is put again in the laser apparatus for eventual frequency trimming. All parts that must be trimmed (resonator ends and couplers) are without coating to prevent further aging.
6) Finally devices are separated and mounted in their can.

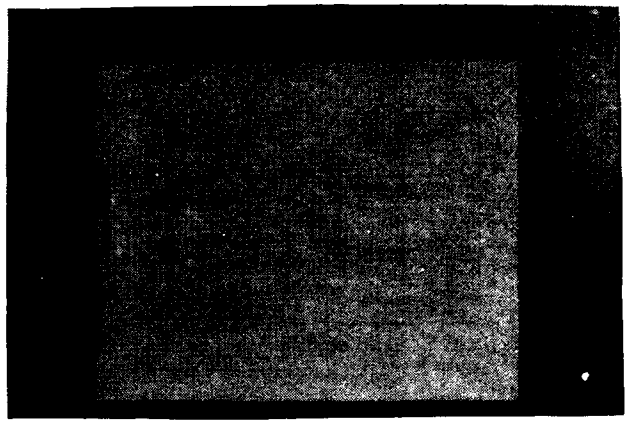


Fig. 12 : Aspect of plate after cutting. We can see the two pole monolithic filters with double stirrup suspension.

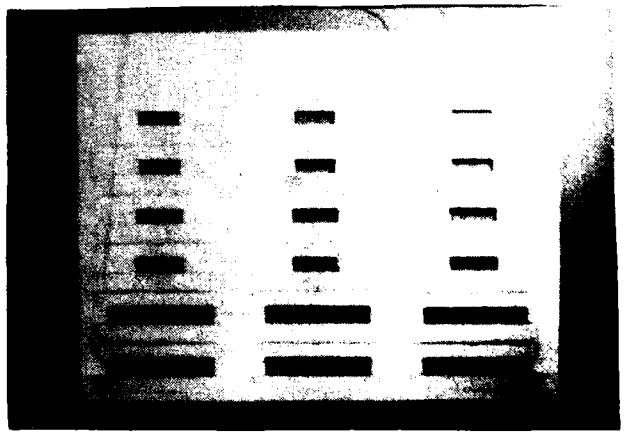


Fig. 13 : Aspect of the same plate after cleaning and metallization. With this kind of suspension trimming can be performed without separate devices.

There is low dispersion in the size of devices cut-up by this method. This allows to be close of final electrical characteristics and reduce trimming. After cutting, on a whole plate which may contain up to 36 resonators, maximum resonance frequency dispersion is no more than 0.11 percent for a 455 KHz center frequency. Now if we compute the mean value of all resonance frequencies of resonators on a plate and compare it to that of an other plate taken from the same crystal, the relative deviation is no more than 0,032 percent.

Conclusion

We have shown that cutting of Lithium Tantalate by a YAG laser is possible without damage of devices characteristics (compared with devices processed by mechanical sawing). All shapes and suspension types are possible and very miniature resonators and monolithic filters have been designed. Frequency trimming can be performed by the same laser before separation of devices. Of course the frequency range investigated

in this paper is not limited and this process can be applied to high frequency thickness mode resonators. In this case, laser can finely adjust size of devices to remove spurious responses. Other materials such as lithium tetraborate, can be processed by the same method.

Further developments include a study of integrated suspension that optimize both electro-mechanical coupling coefficient and Q-factor. This process is protected by a patent.

Acknowledgements

The authors wish to thank Mr JOLY for his usefull experimental work.

Reference

[1] J. DETAINT and R. LANCON, "temperature characteristics of high frequency lithium Tantalate plates" in Proc. 30th Ann. freq. cont. Symp., 1976, pp 132-140.

[2] R.C. SMYTHE and M.D. HOWARD, "Current trends in crystal filters", in Proc. 37th Ann. freq. cont. Symp., 1983, pp 349-353.

[3] Yoshiro FUJIWARA, Sumio YAMADA, and Noboru WAKATSUKI, "Miniaturized $LiTaO_3$ strip Resonator", in Proc. 37th Ann. Freq. Cont. Symp., 1983, pp 343-348.

[4] M. BASS, "Nd : YAG laser-irradiation-induced damage to $LiNbO_3$ and KDP", IEEE J. of Quantum Electronics, 1971, pp 350-359.

[5] R.J. DINGER, "A miniature quartz resonator vibrating at 1 MHz", in Proc. 35th Ann. Freq. Cont. Symp., 1981, pp 144-148.

A PROGRESS REPORT ON MANUFACTURING METHODS AND TECHNOLOGY FOR PRODUCTION OF HIGH-STABILITY, VIBRATION-RESISTANT QUARTZ CRYSTAL UNITS

Jack C. Korman

Frequency Electronics, Inc.
Mitchel Field, New York 11553

Summary

This report describes the progress to date in the development of equipment, techniques, and procedures required to produce quartz crystals enclosed in ceramic flatpacks.

The equipment consists of modular stations which are capable of processing batch quantities of as many as 48 ceramic flatpack resonators containing SC-cut crystals in the 5 MHz, 5.115 MHz, 10 MHz and 20 MHz ranges.

The functions performed independently in the system modules include ultraviolet/ozone cleaning and ultrahigh vacuum chambers with high temperature capability for deposition of gold electrodes to specified resonant frequency, followed by vacuum bake and thermocompression sealing of the ceramic enclosure.

System fixtures include crystal blank carriers for electrode plating, cover carriers for sealing, and interchangeable plating masks.

The system is presently operational, and test quantities of 5 MHz flatpacks have exhibited good stability plus high Q and low resistance values.

Work is still in progress to refine methods in order to establish set operational procedures.

Introduction

The design goals of this manufacturing technology are based upon the importance and need, in present and future electronic systems, to have the availability of precise, vibration resistant oscillator/clocks. Therefore, quantity production of high-stability, vibration resistant ceramic flatpack resonators is essential to meet this need.

System concepts are based upon mobile, modular processing stations which incorporate ultrahigh vacuum systems combined with temperature control capability. The modular concept enables each operation to be performed independently, so that several batch quantities of flatpacks can be moved through each process as independent units.

The mobility feature of the equipment enables rapid relocation of individual stations for service without interrupting other elements of the system. The entire system is known as the Modular Quartz Crystal Fabrication Facility (MQXFF).

Flatpack Processes

The ceramic resonator is shown in Figure 1. Its components consist of the crystal-holding frame plus two identical ceramic sealing covers (Figure 2). All mating sealing surfaces are metallized gold. Table 1 contains a sequence of operations in the manufacturing processes for ceramic resonators.

Off-Line Preparation

Prior to off-line operations, the ceramic components are thoroughly cleaned and baked out under vacuum and a high temperature environment. They are then subjected to ultraviolet cleaning during all phases of the preparatory work. Best results have been achieved with a solvent pre-cleaning followed by irradiation with short wave ultraviolet light. This procedure has proved highly effective in removing a variety of contaminants from the surfaces of the resonator components.[1]

Following X-ray and angle correction procedures, resonator fabrication commences, which produces a finished SC-cut quartz crystal blank. The blank is then bonded to four gold plated molybdenum ribbon clips which are welded onto metallized gold areas on the ceramic frame. Bonding to the crystal blank is accomplished at four points where gold had previously

Table 1

Flatpack Processes

OFF-LINE

1. Prepare quartz blank - plano convex.
2. Gold plate 4 gold pads on outer edge of blank.
3. Clean ceramic parts and vacuum bake and UV clean.
4. Weld support clips to ceramic frame.
5. Assemble quartz blank to support clips and UV clean.
6. Assemble crystal/frame unit into holding fixture and UV clean.
7. Assemble and press gold gasket onto covers and UV clean.
8. Assemble gasket/cover units into sealing fixture and UV clean.

MACHINE OPERATIONS

PLATING SYSTEM

1. Place crystal/holding fixtures into slots on carousel - can process as many as 48 units per run.
2. Seal plating chamber and pump down to vacuum in 10^{-8} torr range.
3. Set temperature in chamber at 100°C and soak overnight in vacuum.
4. Plate electrodes onto each quartz blank individually - coarse plate @ 250°C.
5. After coarse plating all blanks, reset temperature control to turnover temperature and fine tune each crystal to desired frequency.
6. Cool to room temperature, break vacuum and remove plated crystals.

SEALING SYSTEM

1. Assemble gasket/cover fixtures on both sides of crystal holder.
2. Place assembled sealing fixtures into slots on sealing carousel. As many as 48 can be run.
3. Seal chamber and pump down to vacuum in 10^{-8} torr range.
4. Set temperature in chamber at 100°C and soak overnight in vacuum.
5. To seal, increase temperature to 320°C and activate air powered rams to thermocompression bond both covers onto each side of crystal frame.
6. Cool to room temperature, break vacuum and remove flatpacks.

CH2186-0/85/0000-0338$1.00©1985IEEE

been sputtered onto the crystal perimeter to enable clip attachment.

Various methods of bonding the crystal blank to the clips include: cementing with polyimide, welding, or brazing techniques. Each of these is currently under investigation to determine the optimum bonding method.

Resonator Fabrication

Resonator fabrication is then performed in the Modular Quartz Crystal Fabrication Facility (MQXFF). The fabrication processes consist of ultraviolet cleaning of all components, base plating of the electrodes and final frequency tuning followed by vacuum baking and sealing of the crystal within the ceramic flatpack.

The MQXFF consists of a number of modular systems, each performing a sequential process in flatpack production. Each module is independent of others and all are designed to be moved about and transported with relative ease of setup and maintenance. Figure 3 shows the electrode plating and resonator sealing systems.

In order to develop suitable means for processing the flatpack components during resonator fabrication, a special holder was designed. This device, which is made of machinable ceramic, holds the crystal properly oriented for electrode plating, and provides for electrical contact with the electrode terminals so that crystal frequency can be monitored during plating. Two gold plated tungsten wires (shown in Figure 4) hold the crystal/frame assembly in its holder by spring pressure against the gold electrode terminals on the ceramic frame. These gold plated wires also act as electrical contacts for use in monitoring frequency during electrode plating of the crystal blank.

Electrode Plating

The plating chamber contains provision for processing 48 flatpacks in holders. These are inserted into the chamber as a group onto a transport carousel which is indexed external of the chamber via a rotary feedthrough mechanism. A linear motion feedthrough selects and positions a single flatpack holder between two nozzle beam gold sources[2] for simultaneous plating, through masks, of both sides of the crystal. Figure 5 is an internal view of the plating chamber.

Crystal frequency is monitored during these operations via contacts made between a pair of terminals on the holder and a mating pair connected externally, by RF feedthroughs, to a monitoring network and instrumentation.

Following overnight bakeout at 100°C under vacuum, base plating of all crystals is completed under high vacuum combined with a temperature of 250°C. Final frequency adjustment is then completed by plating all crystals after reducing the temperature to the turnover 1 point of the crystals.

After plating, each flatpack holder is returned to its position in the carousel which is then indexed to the next holder, and plating is repeated similarly upon each subsequent crystal.

After tuning operations are complete, the chamber is allowed to cool to room temperature, vacuum is broken, and the plated crystal/frame assemblies are removed for sealing operations.

Flatpack Sealing

Upon completion of the plating operations, the crystals in their holders are removed to an off-line operation. Each crystal holder is then sandwiched between two stainless-steel cover holders, each of which contains a precleaned and preassembled ceramic cover. Each cover contains a pure gold gasket which has been pressed onto the sealing surfaces. One of the cover holders is fixed and the other is gimbaled, to allow for even presure distribution during sealing.

Following the sandwich assembly, each is then placed into the sealing chamber, onto another carousel. This carousel is similar in design but larger than the plating carousel, in order to accommodate the wider cover sealing asemblies. This is shown in Figures 6 and 7.

The sealing vacuum chamber is then pumped down into the 10^{-8} torr range, with a preliminary bakeout temperature of 100°C for several hours. The chamber temperature is then raised to 320°C for one hour.

Each cover sealing assembly is then drawn out of the carousel by a linear feedthrough and into position between two sealing rams. One of the rams is fixed into a stationary position; the other is caused to move against the gimbaled side of the fixture. This compresses both covers against each side of the flatpack frame, with a pressure of about 60 psi air.

After maintaining this condition for 5 minutes, ram pressure is removed and the completed assembly is replaced into its position in the carousel by the linear feedthrough.

The sealing operation is then repeated until all flatpacks in the chamber have been sealed. The chamber is then allowed to cool to room temperature and is opened for removal of the completed resonators.

Test Results

In preliminary tests a total of 25 flatpacks were plated and sealed successfully. 5 MHz, 3rd overtone, SC-cut quartz blanks were used in these tests.

Electrical measurements and temperature slews were performed before and after temperature cycling (-55°C to +95°C) twelve times, with encouraging results:

Resistance measurements averaged 70 ohms, with deviations of less than 2 ohms in the temperature range of -55°C to +95°C.

The average value of Q was 2.44×10^6.

Long-term stability results over a 90-day period showed frequency deviations which were running better than 2pp 10^{-11} per day.

The crystals have exhibited a "g" sensitivity of better than $5 \times 10^{-10}/g$.

Figures 8 and 9 are typical temperature slews of the same flatpack. Figure 9 is a slew run made after temperature cycling twelve (12) times from -55°C to +95°C. No change in frequency or resistance is apparent.

Conclusions

The MQXFF system, in its present form, is operational. Some refinements in areas of temperature control and mechanism functions will be completed, but these are not expected to seriously impede flatpack production currently in progress.

Testing of flatpack parameters will continue, to further establish their superior performance characteristics.

Acknowledgements

This work was supported by the U.S. Army Electronics Research and Development Command, Fort Monmouth, N. J.

The author wishes to thank the staff of Frequency Electronics' Crystal Department for their support.

References

1. J. R. Vig, J. M. LeBus and R. L. Filler, "Further Results on UV Cleaning and Ni Electrobonding," Proceedings of the 29th Annual Symposium on Frequency Control, 1975.

2. Robert J. Ney, "Quartz Crystal Fabrication Facility," General Electric Neutron Division Report No. HH610964CPW3/02A dated May 1980, for U.S. Army ERADCOM, Ft. Monmouth, N. J.

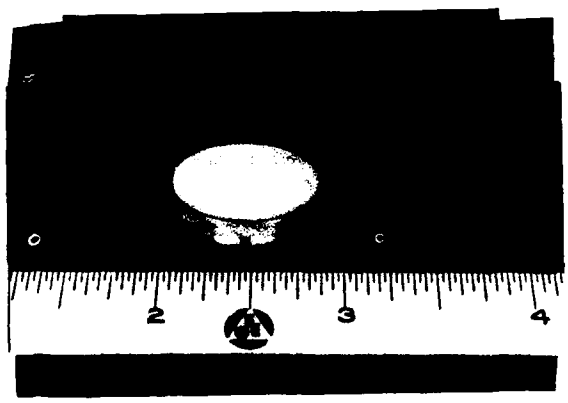


Figure 1. Ceramic Flatpack Resonator

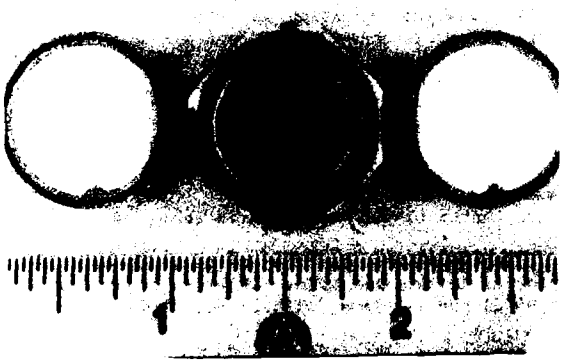


Figure 2. Resonator Components

Figure 3. Resonator Plating and Sealing System

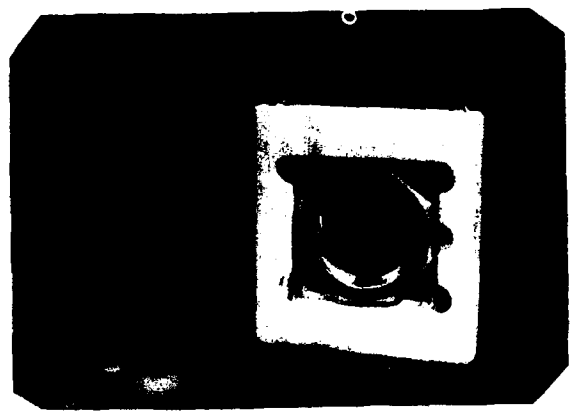


Figure 4. Crystal Holder with Electrical Contact Terminals

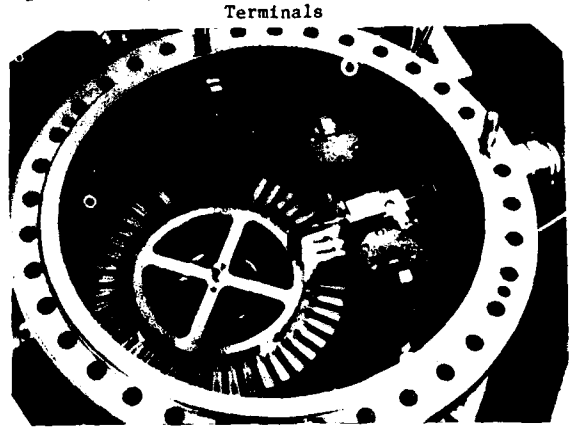


Figure 5. Internal View of Plating Chamber

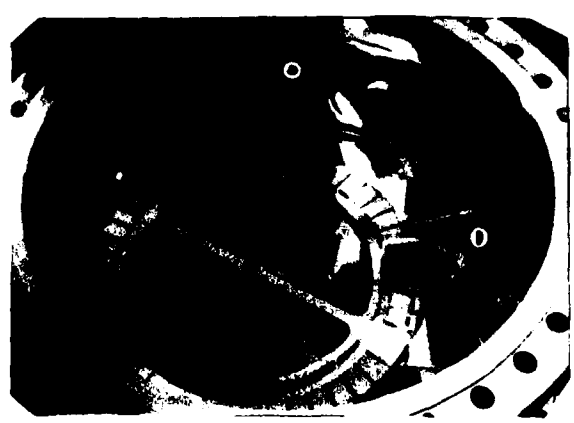


Figure 6. Internal View of Flatpack Sealing Chamber

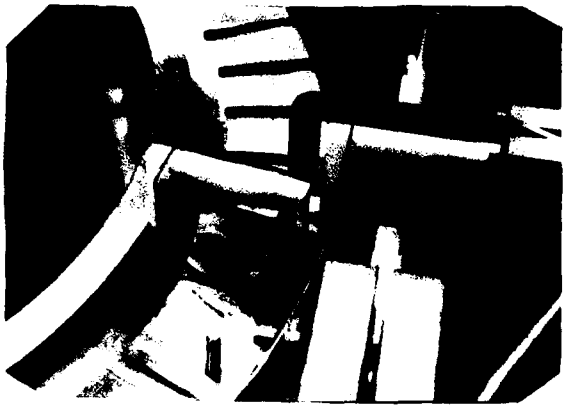


Figure 7. Close-up of Sealing Station Showing Fixture Assembly

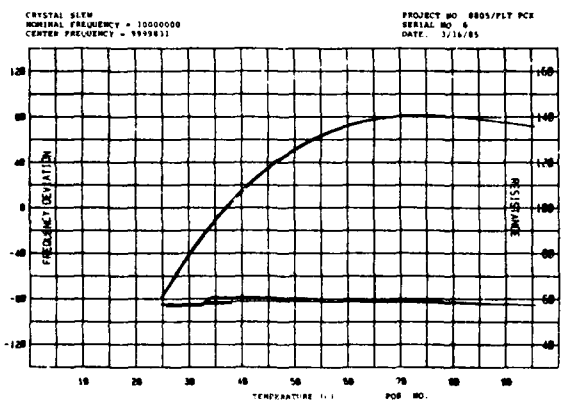


Figure 8. Typical Temperature Slew Before Cycling

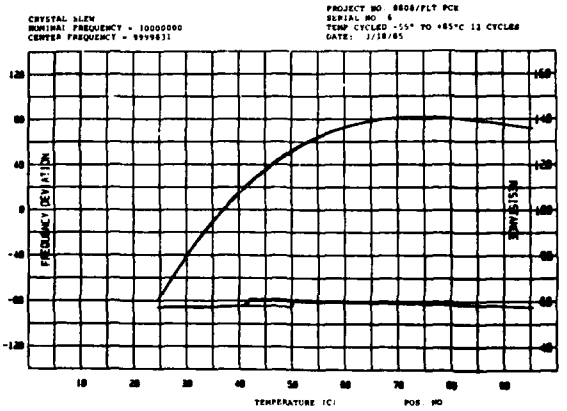


Figure 9. Typical Temperature Slew After Cycling

CUT AND GRIND

Dr. A. W. Warner, Jr., J. Tsaclas and J. Korman

Frequency Electronics, Inc.
Mitchel Field, New York 11553

With the advent of the SC-cut crystal, which has an inflection temperature at 95°C, orientation angles have become tighter in tolerance because the operating temperature is usually near the inflection temperature, and the angles are difficult to measure for two reasons. First, the measurement necessitates the use of weak X-ray planes requiring a very sensitive detector and some calculations. Second, the small tolerance requires a very precise fixture to hold the quartz plate and thus defines the surface to be measured. Often, a laser beam must be used to determine the position of the central surface of the blank.

Crystal units for oven operation above 75°C, that is, with turnover temperatures between 75°C and 95°C, will require θ angles which vary no more than ±0.3 minutes of arc from some predetermined angle. This is only ±20 seconds of arc tolerance, assuming everything else is stable, i.e., quartz blanks must not only be cut and corrected, but also measured to this tolerance. This requirement is depicted in Figure 1.

Cut and grind refers to a process in which a quartz bar is properly oriented as in a saw, and an initial saw cut made. However, the arbor on which the saw blade is mounted also has a rigid face grinding diamond wheel on it. The bar surface, after being cut, can now be faced off by the grinding wheel; the process is then repeated. The result is a group of blanks with one accurately ground surface. The other side of the blank may be finished in a contour generator and polishing is accomplished by a chemical etch, a process which does not disturb or change the orientation.

This surface grinder, shown in Figure 2, is of the usual design except for a crossfeed screw accurate to 50 millionths of an inch. The little box on top is an added readout for both vertical and horizontal motion, accurate to a tenth of a mil. In Figure 3, the 8 inch saw blade and the 6 inch face grinding diamond wheel can be seen with the quartz bar in grinding position. Figure 4 is a rear view showing the sine bar and gauge blocks for the θ angle, with the bar in the cutting position.

The principles behind the cut and grind process are as follows:

1. It is easier to measure and correct the orientation of a bar than it is to measure and correct a quartz blank. The larger surfaces of a bar are more easily defined, i.e., the large flat surface can be vacuum chucked or clamped against a flat reference surface, either for X-raying or for grinding. The X and Z surfaces can be directly measured by X-ray, since these X-ray planes are parallel to the surface of the bar. For example, an air-cooled X-ray tube and an ionization chamber detector are sufficient. A measurement is made by reversing the bar 180° and noting the difference in the micrometer screw reading of the X-ray goniometer.

2. A grinder-sine bar combination can produce surfaces that are accurate and reproducible to seconds of arc. It is an old and well established art. For example, a variation of only 1 second of arc in a 5 inch sine bar at 30° is accomplished by a 20 millionth of an inch change in gauge blocks--within the specifications for gauge blocks and sine bars. Five seconds of arc corresponds to 0.1 mil., which is within the accuracy of a micrometer screw.

3. Each blank is corrected right on the bar without disturbing the bar setting. One simple X-ray correction in one direction on one side is all that is needed for twenty 5 MHz blanks. Only one angle, ψ, needs to be correct, since Ø and θ can be set by the sine bars on the grinder. Individual grindings take about 1 minute extra of time. The whole process could be automated once the bar is set. There is no need to X-ray the cut blanks, as the error in X-ray measurement at this stage is greater than the variation in blanks. Verification can be obtained by measuring the turnover temperature of the first blank cut.

Results

The flat surface generated by the diamond wheel is not, strictly speaking, flat, but is slightly elliptical. By choosing the location for grinding, for example, at the 3 o'clock position, the axis of the ellipse can be located along the Z' axis so that rotation about X' (θ angle) is known and measurable. In the X' direction the curvature varies from 1 to 3 microns, and is not really important since the opposite side will be heavily contoured to about 4 diopters.

The diamond wheel used is resinoid bonded 400 mesh and the quartz blank is actually semipolished directly from the grinding wheel. The shape of the surface can readily be seen by use of an optical flat, as depicted in Figure 5. This semipolished surface lends itself well to chemical polishing. About 15 f^2 of etching will produce the typical polish obtained by etching a 5 micron lapped surface.

Initial results from quartz plates processed from one bar are shown in Figure 6. This was a test of reproducibility, and the bar was not prepared. The range of turnover temperatures is from 69°C to 73°C and the frequency vs. temperature curve shows the care taken in measurement, i.e., frequency to 0.01 Hz and temperature to 0.1°C. The indicated range in angle θ is 20 seconds of arc.

A later group of quartz plates more closely approached our goal, which is a slope of +0 to -0.04 ppm/°C at 95°C. Figure 7 gives the turnover temperature. It can be seen that the X-ray measurements are not very meaningful except that the lowest angle corresponds to the highest turnover temperature. The most recent bar cut---performed by a cut and try method---gave even better results, both in numbers and accuracy. In this case, the bar was not prepared but set in position and one blank cut. This blank was processed quickly and measured for turnover and inflection temperature. The results were used to reset the sine bars for further cutting. This, of course, is an alternative method to that of correcting the bar itself. While this method sounds attractive, it can require more time because the Ø and θ angles are not separately corrected, which is the case if the bar is exact in the ψ direction. Of the 19 pieces cut from

CH2186-0/85/0000-0342$1.00©1985IEEE

this bar, the first three were test cuts. The results from the third test cut are shown in Figure 8, and since it was on target, the rest of the bar was cut. All units, #3 through #19, made from this bar have turnover temperatures between 80°C and 83°C. The spread in slope at 95°C is 0.012 ppm/°C, which translates to an angular tolerance of ±6 seconds of arc.

Cost Factors

The sawing operation is the same as for our regular saw. The additional time for correction after sawing is one minute per blank compared to the much longer time for an average of three stages of X-ray measurement, correction, and relapping. Time for rounding is the same as before. The time for contouring the opposite surface is one minute per blank when using the contour grinder, and considerably more time if individual bowl lapping to frequency is employed. Chemical polishing is used in all cases.

One detailed estimate for 100 very close tolerance blanks by conventional means came to 5 bars of quartz and 140 hours of time. Cut and grind would eliminate at least 36 hours of work and parts of the quartz bar would be left over for conventional processing.

The Mitsui grinder, with fine crossfeed, costs about 15 thousand dollars, and it can eliminate the need for a 50 thousand dollar X-ray goniometer. A grinder with automation would cost considerably more, as would a more sophisticated, temperature controlled grinder. On the other hand, this method appears to be the only practical way to automate the production of very close tolerance blanks. The quartz bar could be horizontal instead of tilted, of course, using orientation angles worked out by Walter Bond[1] and presented at the 31st symposium proceedings, which would simplify automation. A rotated D bar could also be used horizontally.

An added advantage to the cut and grind method is that any crystal orientation, even those with no convenient X-ray reference planes, can be produced just as easily as any other cut,e.g., FC's, NT's, AK's,[2] etc.

Conclusion

When it is considered that this method of producing quartz blanks eliminates an X-ray machine, eliminates all laps, eliminates mechanical polishing, permits a high degree of automation, and is capable of a very high order of precision, isn't it time to give it a try?

This work was supported by the U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey, under Contract DAAB07-82-C-J052.

References

1. Bond, Walter, L., "Making Doubly Rotated Quartz Plates," presented at the 31st proceedings of the AFCS, 1977.

2. Kahan, A. and Euler, F., "Harmonic and Anharmonic Modes of AK-Cut Crystal Resonators," presented at the 38th proceedings of the AFSC, 1984.

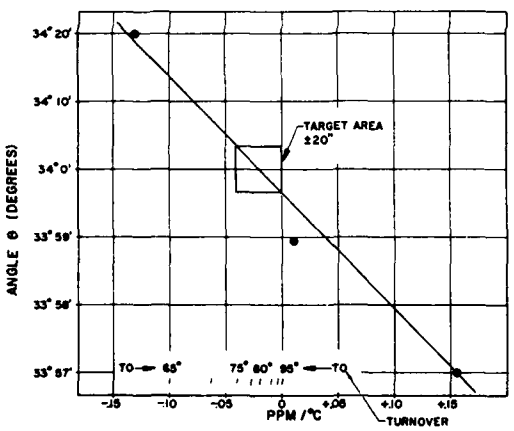


Figure 1. Slope at Inflection Vs. Angle θ

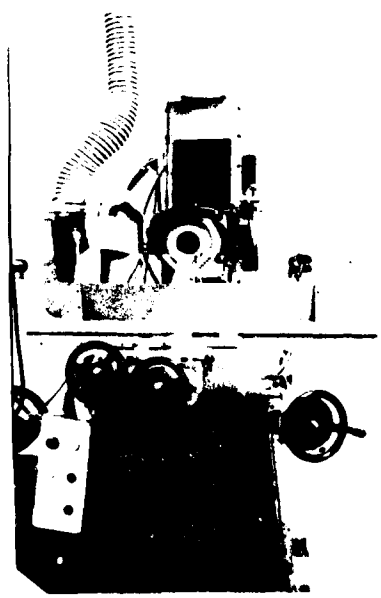


Figure 2. Front View of Mitsui Grinder

Figure 3. Close-Up Oblique View of Blade/Grinding Wheel

Figure 4. Close-Up Side View of Grinder Showing Sine Bar, Gauge Blocks and Quartz Bar

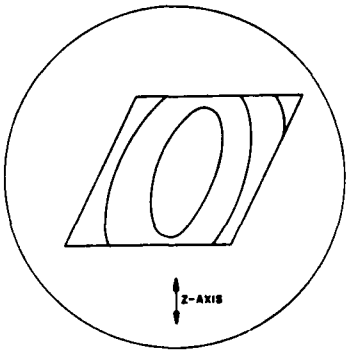


Figure 5. Optical Flat Inspection of Cut and Ground Quartz Blank

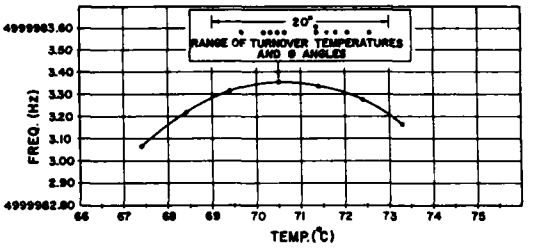


Figure 6. Frequency Vs. Temperature 5 MHz, 3rd Overtone Crystal

GRIND-AND-CUT CRYSTAL MEASUREMENTS

CRYSTAL NO.	ANGLE θ"	SLOPE AT 95°C	TURNOVER
1	48°11'	0.008 PPM/°C	87°C
2	48°11'	0.040	82°C
3	48°11'	0.016	86°C
4	48°11'	0.004	88°C
5	48°11'	0.038	83°C
6	48°11'	0.004	87°C
7	48°10½'	0	95°C

Figure 7. Grind and Cut Crystal Measurements

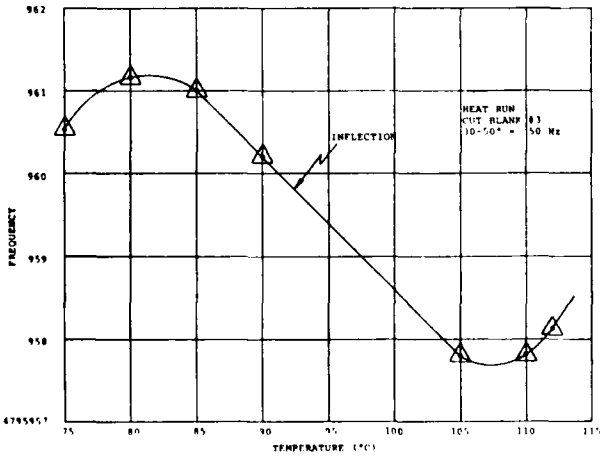


Figure 8. Heat Run of Blank No. 3

LOW PROFILE GLASS PACKAGED CRYSTAL UNIT

M. Sato, H. Eguchi, S. Ishigami
Tokyo Denpa Co., Ltd.
Tokyo, Japan

K. Yamamoto
Science University of Tokyo
Tokyo, Japan

Abstract

The manufacturing processes of the usual metal packaged crystal units are complicated and low efficient for automation of each process. To solve these problems and to make cost down without degradation of performance, the DIP and the surface mounted-type glass packaged crystal units were newly developed.

For materials of the package of crystal units, soda lime glass were used for its high productivity and low cost. And for the sealing of package, frit glass for high temperature was applied. To prevent the performance of the packaged crystal units from deterioration, improved electrode materials and newly developed conductive adhesives were applied.

At the same time, automatic assembly machines were designed to make cost down and to have high efficient mass production.

Electrical and sealing performances and structural rigidity, etc. of the above crystal units were tested and the experimental results are explained in the following sections.

Introduction

In the case of production of crystal units, the selection of the composing material, supporting method of quartz blank and sealing method etc. are generally considered. For materials, metals, glass, ceramics, and plastics etc. are usually used, and for supporting method of quartz blanks, 2, 3 and/or 4 points support are applied. For sealing method, soldering, cold welding and resin adhesion etc. are commonly considered.

The improvement of the specification of crystal units have been performed in combination of above materials and by supporting methods and welding. For example, the studies of ceramic materials used for the package of the units and sealing method with Au-diffusion are reported [1], [2]. The studies of polyimide adhesive which has a high temperature resistance, and the improvements of supporting methods are also reported [3]. But all these studies are for the purpose of having the high performance of the units without considering to make the production cost down. Then in order to have low production cost, the units, for which package the molded resin materials are used, are developed in the similar way to the IC production processes. But, compared with the usual metal packaged units, the resin or plastic packaged units are in disadvantage for humidity resistance.

Then, in this paper, the developments of units, which have the same order airtight and the elastic characteristics as that of the usual metal packaged units, are described. Moreover, the automatic assembly machines newly developed to make the cost down are also described.

The Development of Package

Selection of Materials

For the package materials, ceramic and glass are generally used. Among them, soda lime glass plate was selected for its high productivity, formative facility and moreover high mechanical resistance. For the electrical terminal, 426 alloy was used for its near thermal expansion coefficient to that ot soda lime glass. For sealing material, Frit glass of low melting point was used.

The physical properties of above mentioned materials are listed in table 1. The thermal expansion coefficients are 72×10^{-7} to 100×10^{-7}/°C in the temperature range of 0°C to +300°C. In this range, these materials are resistant for thermal shock in production process and after-treatment. The thermal expansion coefficients are depicted in Fig.1.

Structure of the Crystal Unit

In order to have the unit structure convenient for its automatic assembly, quartz blanks were set on and attached to the four tabs which are located properly through conductive adhesive.

External appearance of the package is given in photo 1 and internal structure in Fig.2. As shown in these figures, terminals are four, two of which are connected to the same electrode so as to have the electrical stability. In the sealing part of the terminal, it has a "narrowing" to have high airtight by its sealing path. Furthermore, to decrease the induced deformation of the terminal through the printed circuit board(PCB) by torsion and bending after putting on the PCB, the terminals are designed to absorb shock.

Mechanical Test

To ascertain the reliability of this structure, the following experiments should be carried out;

(A) Adhesive intensity test between quartz blank and its support part. This adhesive intensity corelates seriously to mechanical vibration and shock resistance.

(B) Adhesive intensity test between leads and frit glass, and package glass. These intensities corelate to airtight.

(C) The shape effect of the shock absorber on the adhesive intensity.

These tests are described in detail in the following section.

Adhesive Intensity Test: The adhesive intensity between quartz blank and its supporting part was measured by increasing the weight suspended by thin metal wire. The method is illustrated in Fig.3 and the obtained results are listed in table 2. In the case of thick blank for low frequency, the blank was separated from the adhesive plane. In the case of thin blank for high frequency, blank was broken. This adhesive intensity is corelated to the resistance to

CH2186-0/85/0000-0345$1.00©1985IEEE

shock and/or vibration which is applied to the unit.

So, these properties were tested and the results are shown in Fig.4 and Fig.10. From these results, it was confirmed that the units with the newly developed glass package have considerably high mechanical resonant frequencies and enough resistance for mechanical shock.

Adhesive intensity test between leads and frit glass, and/or package glass: This intensity was measured by using tension gauge. The measuring method was illustrated in Fig.5 and obtained results are listed in table 3 and shown in histogram 1. From these figures, it is clear that the adhesive intensity considerably depends on the kinds of frit glass, glazing thickness, area and homogeneity of the frit glass. And this adhesive intensity is considered to be corelated to airtight.

The variations of the adhesive intensity according to the kind of frit glass are shown in table 3. Then, it was concluded that the adhesive intensity in this part is enough to resist the three times as large as Ca. 600 gf, which is the critical stress designed to cause deformation when the terminal with a shock absorber was impacted as described in the next.

The shape effect of the shock absorber on the adhesive intensity: In the case of units attached to the flexible PCB such as epoxylene glass, the units have cracks near the foot of the terminals come to break down due to the induced stresses in the longitudinal or transversal directions through torsion or bending of the PCB. To avoid these stresses, the terminal has a rectangular hole as the shock absorber as shown in Fig.2. This absorber begins to deform at the load of about 600 gf.

Improvement of Electroconductive Adhesive

For electroconductive adhesion, some resins as epoxylene and polyimide, ceramics and Au-Ge alloy etc. are generally used. Among these adhesives, polyimide resin was used because of its high temperature resistance when sealing and easy handling in adhesion. This adhesive was exposed in the air at 450°C following heating by a given temperature schedule in the electric furnace. In this case, degradation of electric conductivity, gas emission at high temperature and deterioration of adhesive properties due to thermal dissociation of the resin should be researched.

To obtain the superior properties of the adhesive, various experiments on its temperature characteristics were performed. Components of the adhesive, that is, silver metal powder, polyimide and N-methyl-pyrrolidone etc. were changed in its amount and/or kind. Then curring conditions and thickness of the adhesive were variously changed and the electrical conductivity and adhesive intensity were examined.

From the thermal experiment, the obtained results showed that degradation of electric conductivity depends considerably on the contents of silver metal powder in the adhesive resin. The adhesive intensity decreases and electric conductivity increases with increase of the contents of silver metal powder.

Then the adhesive was newly improved in which the content of silver metal powder was determined so as to match both adhesive intensity and electric conductivity. With use of this newly developed adhesive, the degradation of electric conductivity was remarkably suppressed.

The components of emitted gases from resin at high temperature were mainly N-methyl-2-pyrrolidone, CO_2 and CO gas. These emitted gases from the small package results in pushing out the frit glass used for sealing and in breaking down airtight as shown in Fig.6. This phenomenon is named as "blow off". By changing curring conditions of the adhesive in order to avoid the blow off, the complete airtight of the package was obtained.

Variation in Frequency of Packaged Units by Sealing

Before and after sealing process of the package, the variation of frequency was measured and the obtained frequency distributions were depicted in Fig.7. From these results, variation in frequency of packaged unit after sealing was about twice with comparison to the variation before sealing. These facts are considered to be caused by the surface oxidation of the electrodes, change of the surface state of the blank and difference in partial evaporation amount of material when sealing at high temperature.

The Automatic Assemblies

To decrease the assembly processes and assure the uniformity of the units, the automatic assembly machines are newly designed and applied in all production process. The mainly developed assembly machines were introduced.

Automatic Bonding Machine

This machine is suitable to automatically supply lead frames and quartz blanks, and to bond blanks set on frames. The appearance and details of this machine are shown in Photo 2.

In the case of adhesion, the improved adhesives of the thixotropy were applied to avoid "threading" phenomenon.

Automatic Lead Cutting and Automatic Marking Machine

By using this machine, the leads were automatically cut in a given length and packaged units were automatically marked. After these processes, the completed units were automatically carried into the magazine. This machine is shown in Photo 3.

The Automatic Inspection Machine

After the units carried with the magazine were automatically pull out, the insulating resistance, characteristics of low drive levels, CI values and frequencies are checked automatically and then the passed units through these checks are carried automatically into the magazine. The appearance and details are shown in Photo 4.

Results of Various Tests

Various kinds of test were performed on the new crystal unit to confirm its reliability. Among these tests, thermal shock test, leak test and vibration test were performed with special interests in glass materials applied to the unit. The results are represented as follows. According to these results, the reliability of the new crystal unit are confirmed.

Leak Test

The units were held in helium gas at 3.5 kg/cm^2 during 8 hours. The histogram of the results are shown in Fig.8.

Frequency-Temperature Characteristics

The characteristics in the temperature range of -40°C to +85°C are shown in Fig.9.

Vibration Test

Vibration tests were performed on three orthogonal directions. Testing conditions were as follows.

frequency:	10 to 500Hz
acceleration:	2.5G
sweep speed:	1 octave/min.
duration time:	2 hours

The results are shown in Fig.10.

Thermal Shock

The units were subjected to the temperature change between -40°C and +85°C up to 1000 cycles. The testing conditions and the results are shown in Fig.11.

Long Term Stability (Aging)

The units were exposed to high temperature environment(+105°C) and checked the frequencies periodically. The results are shown in Fig.12.

Conclusion

A new crystal unit with glass materials for its package and sealing has been developed. The shape of the package has IC-type appearance suitable for automatic insertion in printed circuit board. To realize low cost production, some automatic assembly machines have been developed for the unit. High reliability of the unit is confirmed through various kinds of test.

References

[1] R. D. Peters, "Ceramic Flat Pack Enclosures for Quartz Crystal Units", Proceedings 30th Annual Symposium on Frequency Control, US Army Electronics Command, Fort Monmouth, N.J. pp. 224-231, (1976)

[2] R. L. Filler and J. R. Vig, "The Acceleration and Warmup Characteristics of Four-Point-Mounted SC and AT-Cut Resonators", Proceedings 35th Annual Symposium on Frequency Control, US Army Electronics Command, Fort Monmouth, N.J. pp 110-116, (1981).

[3] R. L. Filler and J. M. Frank and J. R. Vig, "Polyimide Bonded Resonators", Proceedings 32nd Annual Symposium on Frequency Control, US Army Electronics Command, Fort Monmouth, N.J. pp 290-297, (1978).

Table 1. Various Properties of Materials for Package

426 Alloy	
Tensile Strength	50 kg/mm^2
Yield Strength	35 kg/mm^2
Elongation	35 to 40%
Erickson Value	approx. 8.5
Rockwell Hardness	15T 80
Thermal Expansion Coefficient	75 to 110 ($\times 10^{-7}$), 30°C to 500°C
Soda Lime Glass	
Thermal Expansion Coefficient	92×10^{-7}/°C, 0°C to 300°C
Specific Gravity	2.40 g/cm^3
Volume Resitivity	6.4 log(ohm-cm) at 50°C
Dielectric Constant	7.2 1MHz at 25°C
Frit Glass	
Thermal Expansion Coefficient	72.5×10^{-7}/°C, 30°C to 250°C
Specific Gravity	7.05 g/cm^3
Softning Point	390°C
Sealing Temperature	430°C
Dielectric Constant	34.5 1MHz at 25°C

Table 2. Adhesive Intensity between Blank and Supporting Part

Thick Blank		
weight(gw) F=kX	displacement(mm) X	coefficient(g/mm) k
184.2	0.05	3684
283.0	0.075	3773
481.4	0.125	3851
600	separated	----
Thin Blank		
weight(gw) F=kX	displacement(mm) X	coefficient(g/mm) k
84.6	0.05	1692
198.4	0.125	1587.2
283.0	0.20	1415
400	broken	----

Table 3. Adhesive Intensity of Package According to the Kind of Frit Glass

frit #	x (kg)	s	min.(kg)	max.(kg)
1	3.14	0.58	2.2	4.2
2	4.3	0.82	3.2	5.6
3	2.92	0.47	2.2	3.9
4	3.36	0.61	2.2	4.0
5	2.79	0.42	2.0	3.8

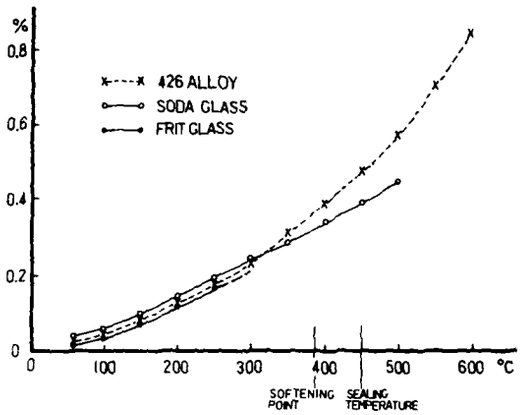


Fig. 1 Thermal Expansion

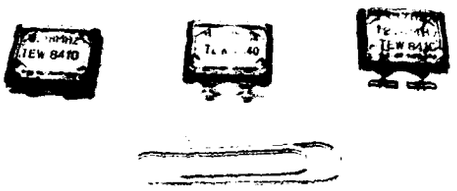


Photo 1 Appearance of the Crystal Unit

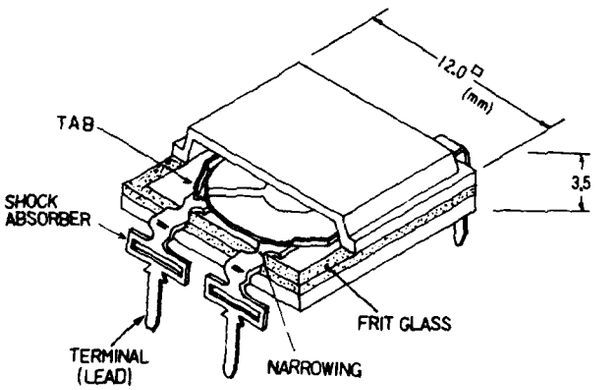


Fig. 2 Internal Structure of New Crystal Unit

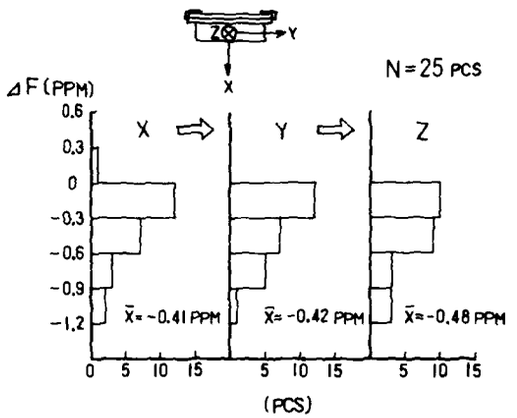


Fig. 4 Results of Shock Test

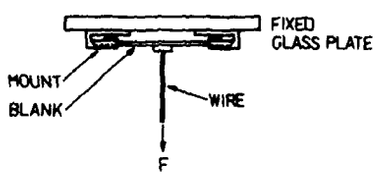


Fig. 3 Test Method of Adhesive Intensity between Blank and Supporting Part

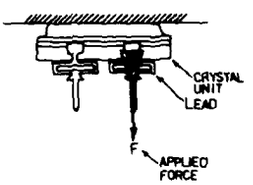
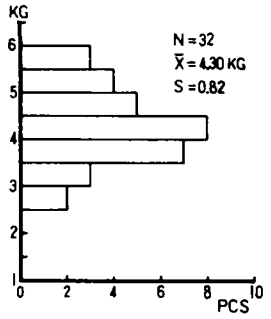


Fig. 5 Test method of Adhesive Intensity between Leads and Frit Glass and/or Package Glass



Histogram 1 Adhesive Intensity of the Package

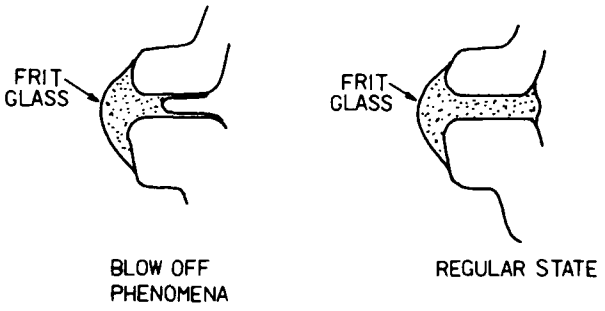


Fig. 6 Illustration of "Blow Off" Phenomena

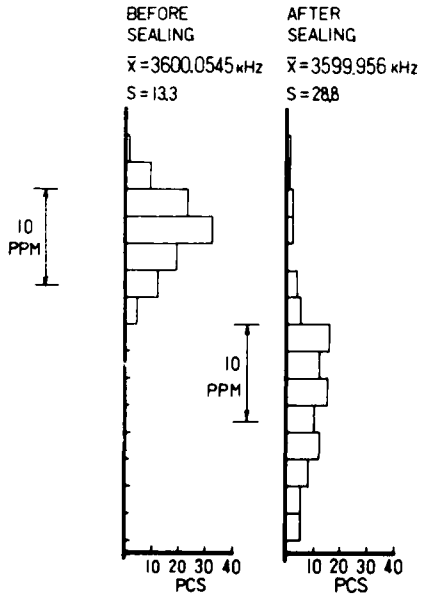


Fig. 7 Variation of Frequency of Packaged Units by Sealing

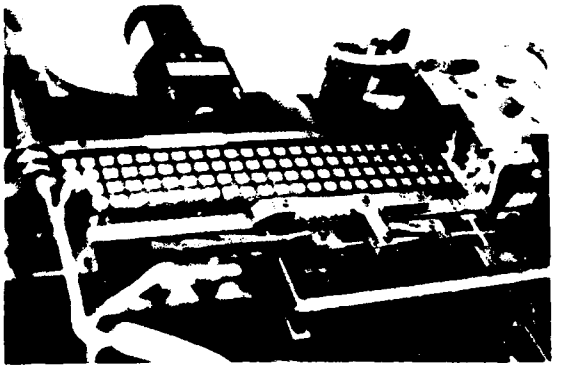


Photo 2 Appearance of Automatic Bonding Machine

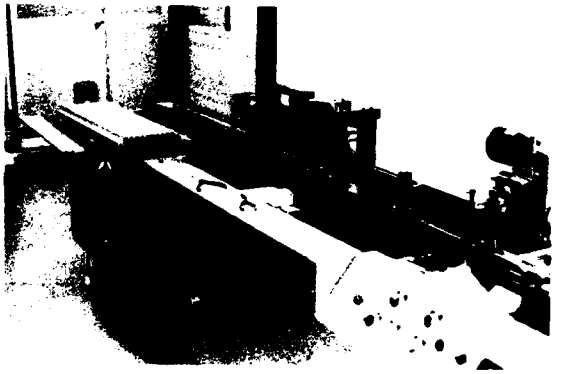


Photo 3 Appearance of Automatic Lead Cutting and Code Marking Machine

Photo 4 Appearance of Automatic Inspection Machine

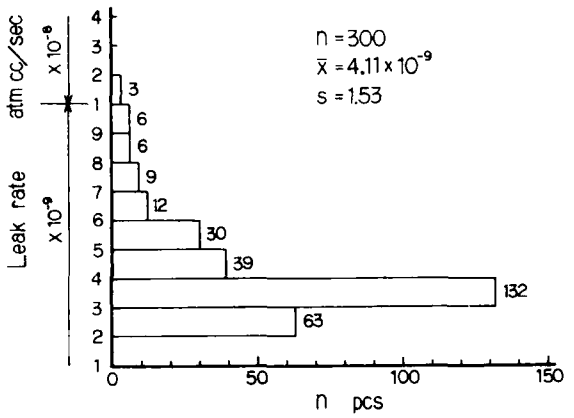


Fig. 8 Results of Leak Test

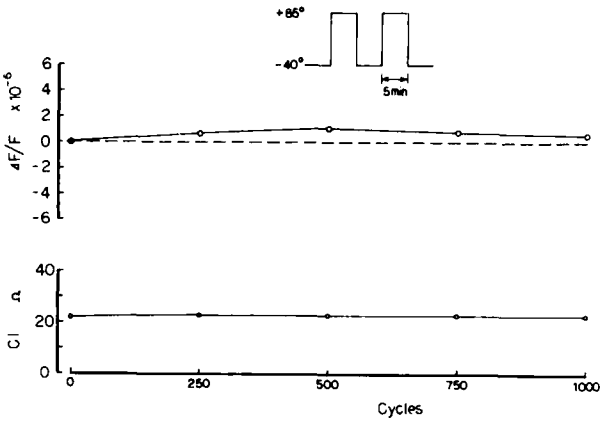


Fig. 11 Results of Thermal Shock Test

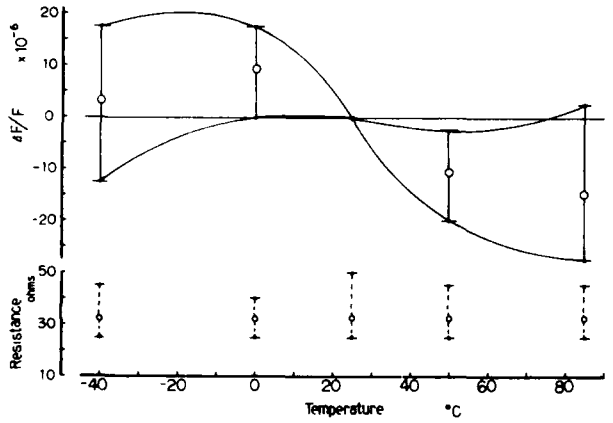


Fig. 9 Frequency-Temperature Characteristics

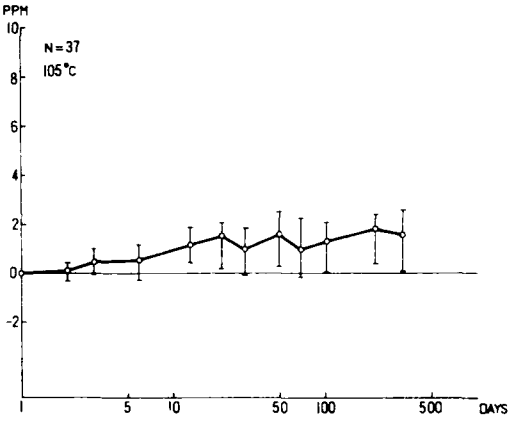


Fig. 12 Long Term Stability (Aging)

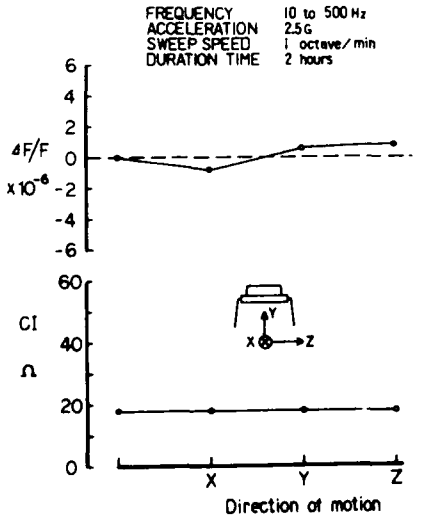


Fig. 10 Results of Vibration Test

STRIP TYPE RESONATOR OF LITHIUM TETRABORATE

Yoshiro FUJIWARA, Masaaki ONO, Masayuki SAKAI and Noboru WAKATSUKI

FUJITSU LIMITED

1015 Kamikodanaka Nakahara-ku Kawasaki 211, JAPAN

ABSTRACT

The thickness shear slow mode of 51° rotated Y-cut lithium tetraborate has zero temperature coefficient on first order, a high coupling coefficient of 26% and velocity of 3240 m/sec. The coupling coefficient of thickness shear first mode is zero, that of the thickness extension mode is 20.5% and the velocity in thickness extension mode is 6880 m/sec.

We have studied the strip type resonator of 51° rotated Y-cut plate. With the longitudinal direction set up parallel to the displacement direction (parallel to Z'-axis), the strip type resonator operates in single mode.

The width of the strip plate is designed to separate the main mode from spurious modes which are generated by the shear mode traveling laterally and reflected at the edge of strip. The optimum ratio of element width to thickness is 4.0.

The length of the strip plate is designed to realize small size but to keep a high Q and to be free from edge reflection. The optimum ratio of element length to thickness is about 30.

The mechanical Q factor is more than 10,000 and capacitance ratio is 20. The resonance mode is free from spurious response in the practical frequency range. The temperature characteristic curve is a parabola with turnaround point at room temperature and the second order temperature coefficient is -0.26 ppm/°C^2.

1. Introduction

Electronic devices are becoming more and more digitized and miniaturized, greatly increasing the demand for stable chip type resonators. Therefore, mechanical resonators using piezoelectric material are attracting attention due to their high Q, high stability, and small size.

Many piezoelectric materials, quartz, $LiTaO_3$, $LiNbO_3$, $Li_2B_4O_7$, and piezoelectric ceramics, have been studied for use in bulk acoustic wave devices, used in the frequency range from several MHz to several tens of MHz.

Quartz is the most useful material because of its low temperature coefficient and high stability, but it has a low electro-mechanical coupling factor, so that it is difficult to make a miniature resonator of quartz. Piezoelectric ceramic is suitable material for miniaturized resonators because of its high coupling factors and low material cost but its temperature stability is not so good. $LiTaO_3$ crystal has a parabolic temperature characteristic curve with turnaround point at room temperature and 20 times higher coupling coefficient than AT-quartz. Table 1 lists the relevant characteristics of thickness mode resonators using various materials. The newly developed lithium tetraborate is an interesting material. $Li_2B_4O_7$ has the first order zero temperature coefficient in rotated Y-cut plate [1][2][3].

On the other hand, the miniaturization of resonators using thickness mode has progressed using the piezoelectric strip technique [4]. When the displacement direction of the vibration mode is parallel to the longitudinal direction of the strip and the driving electrodes reach throughout the lateral direction, the spurious response caused by the twist overtone cannot be generated. Many papers reported on strip type resonators [5][6][7][8].

This paper reports the temperature characteristics of thickness shear slow mode of $Li_2B_4O_7$ resonator using double rotation plate. We applied the piezoelectric strip to $Li_2B_4O_7$ resonator and studied the dimensions of the strip to separate the spurious responses for enough from the main mode and to miniaturize the strip.

Table 1 Relevant characteristics of thickness mode resonator using various materials.

MATERIALS	Coupling factor (k)	Temperature stability (−20-70 C)
AT-Quartz	0.086	· 30 ppm
163 -Y $LiNbO_3$	0.61	150 ppm
X $LiTaO_3$	0.47	· 3000 ppm

2. Temperature Coefficient

To analyze the electro-mechanical coupling coefficient and temperature coefficient, we used the classical infinite, one-dimensional resonator model adding the electric field along the normal direction of the piezoelectric plate [9]. The plate orientation of double rotation plate is defined by two angles ϕ and θ as shown in Fig. 1.

CH2186-0/85/0000-0351$1 00© 1985IEEE

The results are presented in altitude charts in polar diagrams showing only a quarter part because $Li_2B_4O_7$ is class 4mm crystal so that X and Y axes are symmetrical. Electro-mechanical coupling factor is shown in Fig. 2, first order resonance frequency temperature coefficient in Fig. 3, and first order antiresonance frequency temperature coefficient in Fig. 4.

The frequency response of thickness shear slow mode on 51° rotated Y-cut $Li_2B_4O_7$ resonator was single mode with high Q, as shown in Fig. 5. The resonator was a disk 8 mm in diameter and 0.25 mm thick with gold electrodes 3 mm in diameter. This resonator did not generate spurious responses without thickness overtone, as shown in Fig. 6.

We confirmed, experimentally, the second order temperature coefficient of rotated Y-cut plate using the above disk. The resonator had zero first order temperature coefficient at 51° rotated Y-cut plate, and the resonance frequency temperature characteristic indicated the parabolic curve as shown in Fig. 7. The second order temperature coefficient was -0.26 ppm/°C^2 which is the same order value as for $LiTaO_3$.

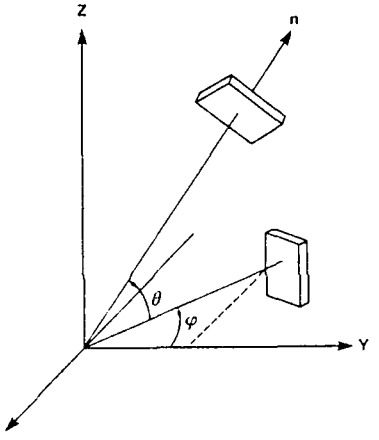


Fig. 1 Plate orientation of resonator with respect to crystal axes XYZ. (a) Single rotation (b) Double rotation electric field normal to the plate

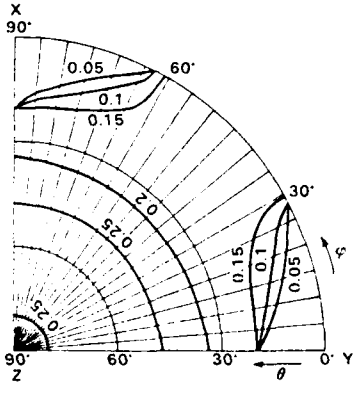


Fig. 2 Coupling coefficient of resonator as a function of rotation angle
(Thickness shear slow mode)

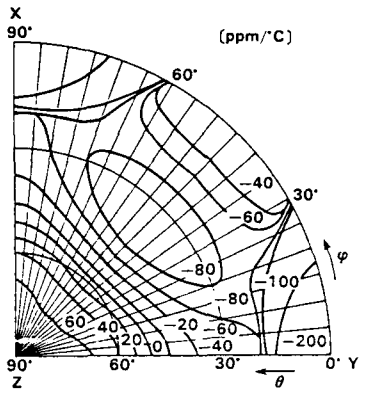


Fig. 3 First order temperature coefficient of resonance frequency as a function of rotation angle
(Thickness shear slow mode)

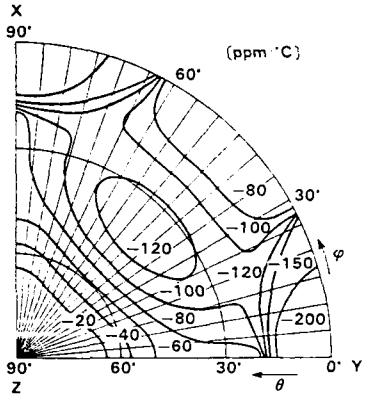


Fig. 4 First order temperature coefficient of antiresonance frequency as a function of rotation angle
(Thickness shear slow mode)

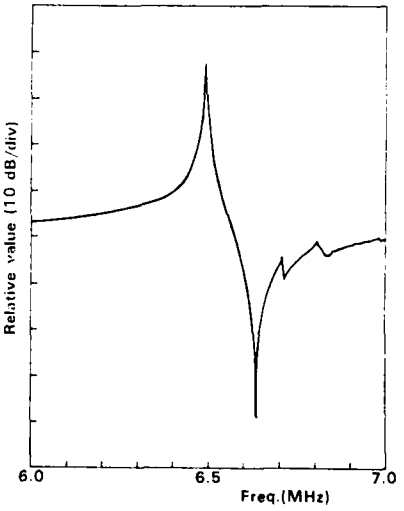


Fig. 5 Narrow band frequency response of 51° rotated Y-cut $Li_2B_4O_7$ resonator

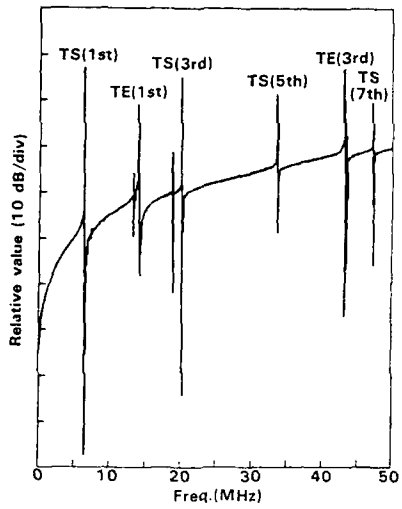


Fig. 6 Wide band frequency response of 51° rotated Y-cut $Li_2B_4O_7$ resonator

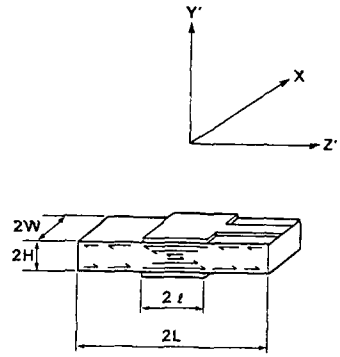


Fig. 8 Construction of piezoelectric strip

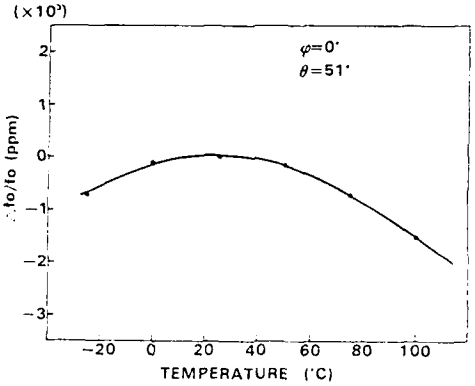


Fig. 7 Temperature characteristics of thickness shear slow mode on 51° rotated Y-cut plate

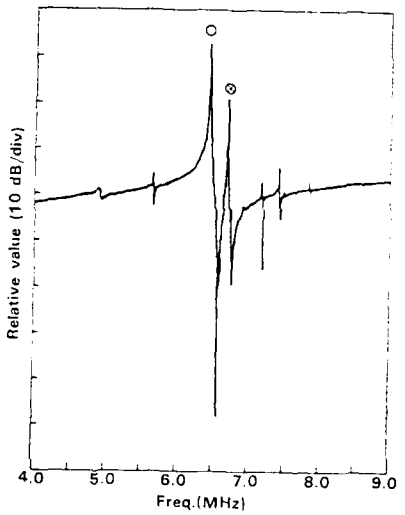


Fig. 9 Frequency response of strip type resonator (○; Main mode, ⊗; Spurious mode)

3. Miniaturization of Strip

The construction of the piezoelectric strip is defined in Fig. 8, where 2W is the width of the strip, 2H is the thickness, and 2L is the length of the strip. The driving electrodes are first confined to the center of the strip on the top and bottom surfaces and then reach throughout the lateral direction, so that the thickness twist overtone cannot be generated.

Figure 9 shows the frequency response of the 51° rotated Y-cut strip type resonator. The resonator was a rectangular plate 8 mm long, 2 mm wide and 0.25 mm thick. The longitudinal direction was at a right angle to the X axis which was the calculated direction. In this strip, the main mode responses are free from twist overtone spurious response, but the spurious responses due to the width of the strip could not be avoided. In Fig. 9, the response at 6.6 MHz was the main mode and the response at 6.7 MHz was the spurious mode due to the width of the strip. The frequency constants of spurious responses were affected by the ratio of width to thickness of the strip (W/H) as shown in Fig. 10. These spurious responses were at higher frequencies if the ratio of width to thickness (W/H) was small. When W/H was the most suitable ratio (W/H=4.0), the spurious response frequencies were far from the main mode, and the level of the nearest spurious response was sufficiently low.

The displacement amplitude along the longitudinal direction is the highest at the center of the strip and decreases on either side. The support on the edge of the strip therefore does not influence the resonance energy if the strip is long enough. To miniaturize the strip, we studied the length of the strip. When the ratio of length to thickness (L/H) was 28, Q was 10,000, and it was not increased by further increasing the ratio (L/H), as shown in Fig. 11. The minimum length of the strip was 28 times the thickness.

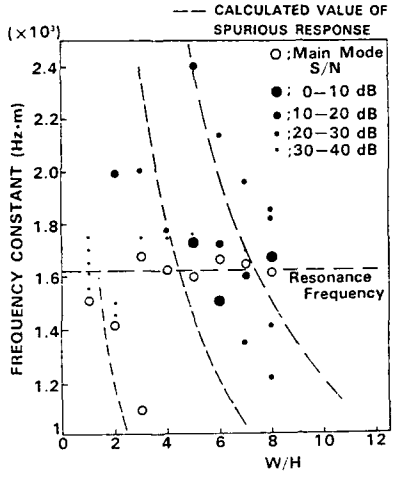


Fig. 10 Frequency spectrum of spurious response

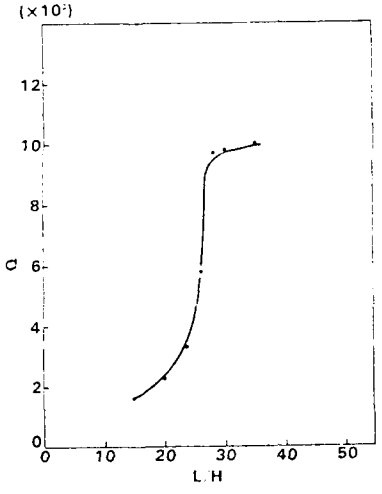


Fig. 11 Q factor of strip type resonator as a function of L/H ratio

4. Characteristics of Strip Type Resonator

Figure 12 shows frequency response and Table 2 lists the equivalent circuit constants of the 6.5 MHz strip type resonator with 51° rotated Y-cut plate. The resonator was a simple rectangular plate 7 mm long, 1 mm wide and 0.25 mm thick with gold electrodes 3 mm long as shown in Fig. 13. The capacitance ratio is 21 and Q is 12,643. The temperature characteristics curve is parabolic with the turnaround point at room temperature, and the second order coefficient is -0.26 ppm/°C².

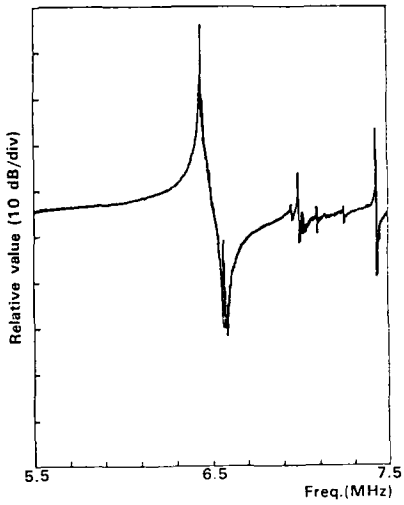


Fig. 12 Frequency response of strip type $Li_2B_4O_7$ resonator

Table 2 Equivalent-circuit constants of strip type $Li_2B_4O_7$ resonator

Item	Value
fr(Resonance frequency)	6.5071 MHz
Q(Resonance sharpness)	12,643
Co(Dumped capacitance)	3.11 pF
L_1(Series inductance)	4.02 mH
R_1(Resonance resistance)	13.0 Ω
r(Capacitance ratio)	20.9
Second order temperature coefficient	−0.26 ppm/°C²

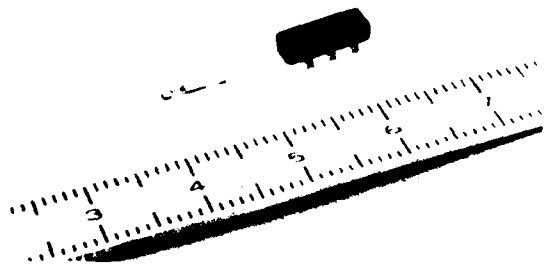


Fig. 13 Photo of $Li_2B_4O_7$ strip type resonator

5. Conclusion

This paper reported on a $Li_2B_4O_7$ strip type resonator with stable temperature characteristics. The 51° rotated Y-cut plate has first order zero temperature coefficient and the turnaround point is at room temperature. We applied the piezoelectric strip to the $Li_2B_4O_7$ resonator and studied the dimensions of the strip to separate the spurious responses sufficiently from the main mode and to miniaturize the strip.

The strip is small. The capacitance ratio is small (20) and the Q is high (12,000). The turnaround point is at room temperature and the second order temperature coefficient is -0.26 ppm/$^{\circ}C^2$. These characteristics are suitable for a highly stable voltage controlled oscillator with a wide variable frequency range.

Acknowledgements

The authors wish to thank Dr. Hara, Mr. Aritomo and Mr. Tominaga for their useful guidance and suggestions during this study.

Reference

[1] M.N. Shorrocks, R.W. Whatmore and F.W. Ainger, "Lithium Tetraborate -- a new piezoelectric substrate material for surface acoustic wave devices", Procs. IEEE ultrasonics Symposium, p337-340, 1981

[2] C.D.J. Emin and J.F. Werner, "The Bulk acoustic wave properties of Lithium Tetraborate", Procs. 37th Annual Freq. Control Symposium, p136-143, 1983

[3] R.C. Peach, C.D.J. Emin, J.F. Werner and S.P. Doherty, "High Coupling Piezoelectric resonators using Lithium Tetraborate", Procs. IEEE Ultrasonics Symposium, p521-526, 1983

[4] K. Nakamura and H. Shimizu, "On the Equivalent Circuits for Thickness Twist Modes and Thickness Shear Modes in Piezoelectric Plates", Paper of technical group U.S., IECE Japan, 69-21, October 1969

[5] Y. Fujiwara, S. Yamada, and N. Wakatsuki, "Strip type $LiTaO_3$ chip resonator with zero temperature coefficient", Paper of technical group US, IECE Japan 82-46, November 1982

[6] M. Okazaki and S. Watanabe, "Miniature $LiTaO_3$ X-cut Strip Resonator", Procs. 37th Annual Freq. control Symposium, p337-342, 1983

[7] Y. Kojima, Y. Fujiwara, S. Yamada, and N. Wakatsuki, "Chip Type Resonator with Load Capacitors", Procs. 38th Annual Freq. Control Symposium, p114-118, 1984

[8] M. Okazaki and N. Manabe, "AT-cut Strip Resonators Enclosed in Cylindrical Package", Procs. 38th Annual Freq. Control Symposium, p119-125, 1983

[9] M. Onoe and H. Jumonji, "Analysis of piezoelectric resonators vibrating trapped energy modes", National Convention Record of IECE of Japan, Vol.48, No.9, p.1574, 1965

RECENT DEVELOPMENTS ON MEMBRANE BULK-ACOUSTIC-WAVE RESONATORS

J.S. Wang, A. Kong, K.F. Lau and K.H. Yen
TRW Electronic Systems Group
One Space Park, Redondo Beach, CA 90278

Abstract

Fundamental-mode filters and oscillators in the VHF and UHF range have been realized utilizing quartz membrane and thin film bulk-acoustic-wave resonators. A high speed etching technique has been developed to fabricate quartz membrane resonators. An etch rate as high as 30 μm/hr has been obtained, and a high Q quartz resonator has been achieved. A temperature-compensated thin film resonator on Si has also been made using shear wave AlN film. This paper will report our recent developments on both membrane and bulk-acoustic-wave resonators, including fabrication techniques, device evaluation and applications.

Introduction

Crystal resonators are widely used in communication systems as a frequency source for filters and oscillators. These devices have applications well into the UHF range, but limitations on crystal size and thickness have restricted the range of fundamental resonant frequencies to less than 50 MHz. Recently, a new class of miniature bulk-acoustic-wave resonators has been investigated. One is the so-called inverted mesa bulk-acoustic-resonator (IMBAR)[1] which consists of a quartz crystal membrane in an inverted mesa structure. The other is a semiconductor bulk-acoustic-resonator (SBAR)[2,3] composed of a thin piezoelectric film membrane on a semiconductor substrate. These resonators exhibit several attractive features: They operate at their fundamental mode in the VHF/UHF frequency range, the wafers readily dice into a plane chip form for circuit hybridization, or they can be integrated with semiconductor devices on the same substrate. A high Q IMBAR device has been made on quartz. This demonstrates that IMBAR devices with great temperature and frequency stability are obtainable using the advantages of AT and SC quartz as a substrate. The SBAR device on GaAs substrates also shows great promise for the future integration of millimeter wave integrated circuit devices. This paper will summarize our research efforts and recent results of experiments on the fabrication and evaluation of these devices in TRW.

Inverted Mesa Bulk-Acoustic-Wave Resonators (IMBAR)

The quartz membrane resonator employs a special supported geometry. Its thin membrane is difficult to fabricate by conventional mechanical fabrication techniques due to the fragile nature of its structure. Since the quality of this resonator is critically dependent on its surface parallelism, a technique to obtain a smooth etched surface is very desirable. Chemical polishing[4] and ion milling[5] were two methods that had been used during the last decade. More recently, progress has been made using reactive ion beam etching.[6] In this work, a high speed etching technique for fabricating thin membrane crystal resonators has been developed. The technique, which combines the advantage of reactive ion beam etching and microelectronic processing, has proven superior in this application and will be of importance in the crystal resonator industry.

The system used was a modified version of an ion milling machine which allowed the use of reactive freon gases, such as CF_4, C_2F_6 or C_3F_8. The quartz substrate was mounted on a rotating water-cooled stage at ground potential, as shown in Fig. 1. The stage was tilted 35 degrees with respect to the incident ion beam to provide the highest etching rate while simultaneously creating a flat and smooth surface. A Si mask with 35 degree angled window sidewalls was used during etching in order to prevent the mask edges from shadowing portions of the substrate and producing a roughly V-shaped surface. The sidewalls accommodate the 35 degree incident ion beam and reduce the bombardment by ions reflected from the walls of the etched substrate. The Si mask, which is composed of an array of square openings, was prefabricated by a sequence of microelectronic semiconductor processing steps: oxidation of the <100> Si wafer; creation of an array of square windows in the SiO_2 layer for exposure of the Si during subsequent chemical etching; and selective chemical etching of the Si in an EPW solution[7] to produce the square openings. The window array of the Si mask enables the production of many membrane resonators from a single substrate, increasing the manufacturing yield from each run.

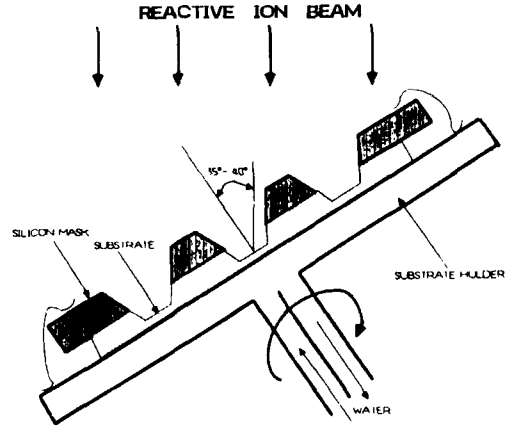


Figure 1. Schematic of substrate mounting assembly for reactive ion beam etching.

A mixture of freon and Ar gases was used in the etching, as previously reported. The Ar serves to physically sputter the accumulated carbon layer which reduced the etch rate and caused an irregular surface. Figure 2 shows a mass spectrum of ions observed during etching when using C_2F_6 and Ar gases. The fragments, primarily CF_3^+, CF^+, Ar^+ are evidence that extensive fragmentation of C_2F_6 in the ion gun is occurring and Ar fragments are exiting the ion gun and impinging on the substrate. Etch rates as high as 30 μm/hr were achieved with a 40% Ar and 60% C_2F_6 mixture. The surface etched in this mixture were extremely smooth and flat. Surface profiles reveal etched surfaces which are slightly convex

CH2186-0/85/0000-0356$1.00©1985IEEE

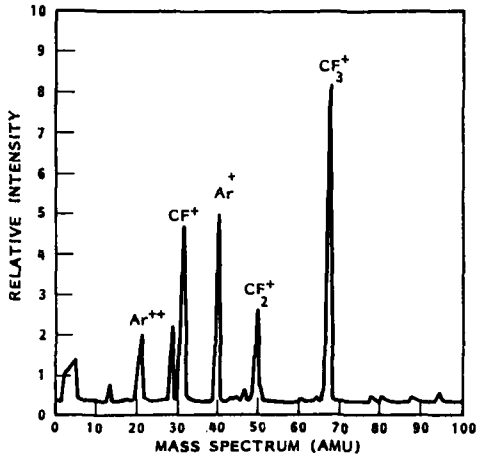


Figure 2. Mass spectrum of C_2F_6 and Ar during etching.

and are suitable for resonator fabrication. This high rate etching technique has been applied to AT, BT and SC-cut quartz and the same surface finishes have been obtained.

Quartz resonators have been fabricated using reactive ion beam etching. Figure 3 shows an array of resonators fabricated on a 3/8 inch in diameter quartz plate. The substrate was then diced into a chip form and mounted in a TO-5 package, as shown in Fig. 4. In order to obtain two floating electrodes, a piece of alumina or glass substrate was mounted on the package below the resonator. Consequently, the device and package were electrically isolated. A schematic of the device mounting is shown in Fig. 5. The packaged devices had undergone mechanical tests to insure that the devices can be securely fastened to the TO-5 packages. The tests included 500 g mechanical shock test at six directions (MIL-STD-883C, Method 2002) and 20 g vibration fatigue test along three different axes of vibration for 96 hours (MIL-STD-883C, Method 2005). No degradation of device performance due to these tests was observed. The main features of these miniature devices--features which could be of importance for manufacturing a hybridized frequency source on a substrate which can be mounted in a small package--are small volume, light weight, and planar mounting.

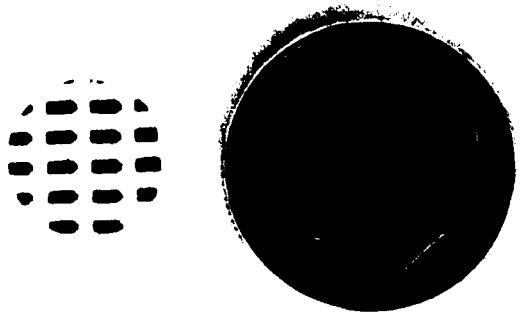


Figure 3. IMBAR devices fabricated on a quartz wafer. The diameter of the quartz wafer is 3/8 inch.

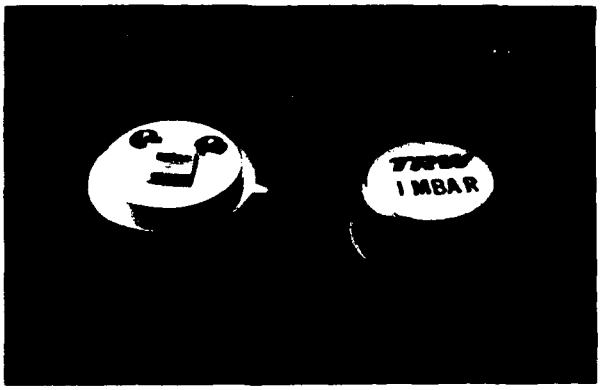


Figure 4. Photograph of a miniature IMBAR chip mounted in a TO-5 package.

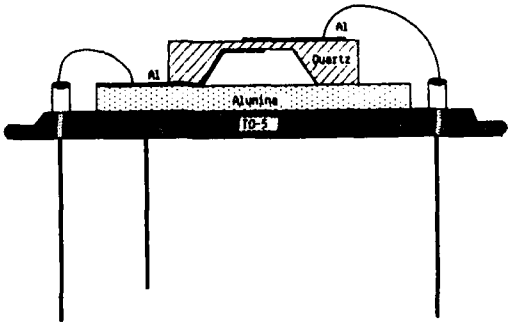


Figure 5. Schematic of a planar mounted IMBAR device.

IMBAR devices having a thickness of less than 10 μm have been fabricated. A 7 μm thick quartz resonator exhibits 225.404 MHz fundamental series resonant frequency with 65 ohm series resistance and 2410 ohms parallel resistance. The unloaded Q was evaluated to be 34,000 by computer-aided measurements on the phase slope of the impedance with respect to frequency at series resonance. The high Q of this resonator demonstrates that extremely smooth and finished surfaces are obtainabale with this etching technique. An IMBAR device with thickness as thin as 3.5 μm has also been fabricated. Its phase and the absolute value of the impedance is plotted as a function of frequency in Fig. 6. The fundamental series resonant frequency was measured to be 435.735 MHz and the Q was 14,000. The Smith chart plot of the device impedance around the resonant frequency is also shown in Fig. 7. Spurious responses were observed on the high frequency side of resonance. A method to suppress these spurs is under investigation.

A feedback oscillator using IMBAR as a frequency source has been made. The Pierce oscillator is of interest because of its simplicity and superiority in the VHF/UHF range. Figure 8 shows an IMBAR oscillator mounted in a TO-8 package. This hyridized circuit consists of a resonator, an amplifier, and two shunt capacitors for phase trimming. An output power level of -1.6 dBm and phase noise of -125 dBc/Hz (1 kHz carrier offset frequency) at 145.3 MHz oscillator frequency has been measured. The phase noise sideband spectrum of the oscillator is plotted in Fig. 9. The short term stability was measured to be 3×10^{-9} in time domain for a 1 second average.

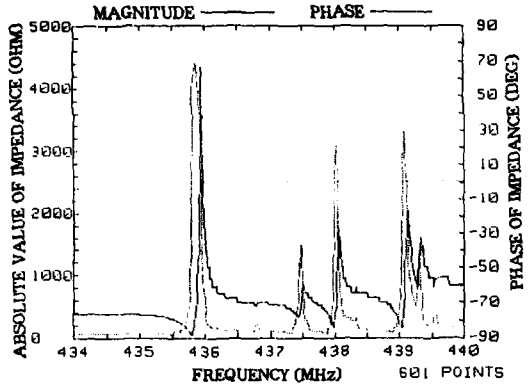


Figure 6. Input impedance plot of an IMBAR around the fundamental resonant frequency of 436 MHz.

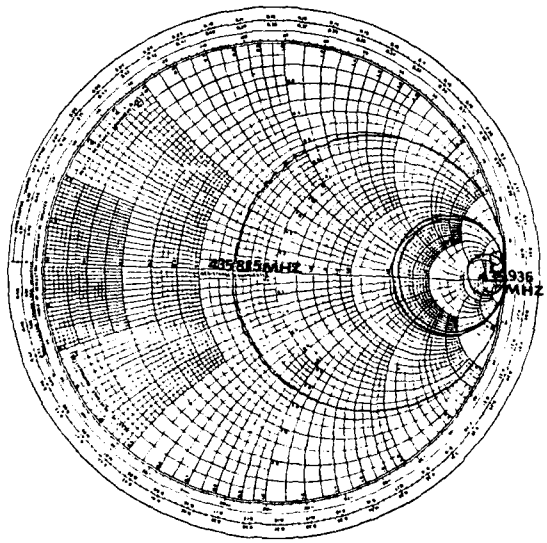


Figure 7. A Smith chart plot of the IMBAR in Figure 6.

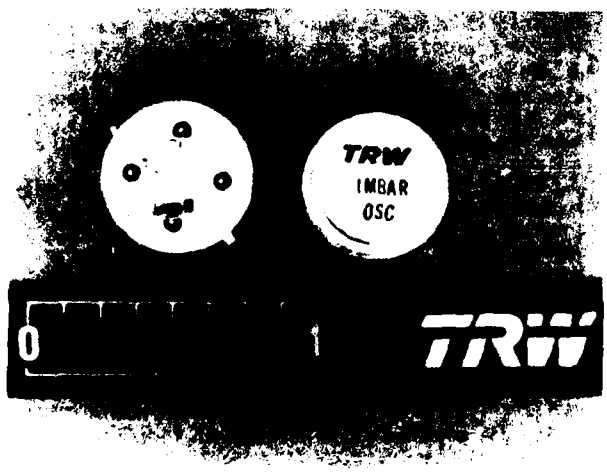


Figure 8. An IMBAR oscillator mounted in a TO-8 package.

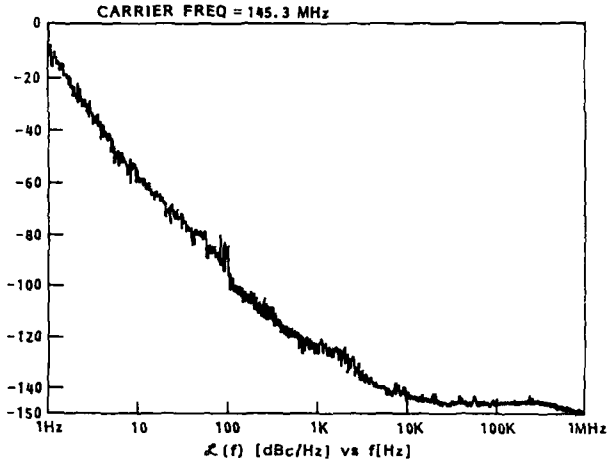


Figure 9. A phase noise spectrum of an IMBAR oscillator.

Semiconductor Bulk-Acoustic-Wave Resonators (SBAR)

The thin film membrane resonator consists of a thin piezoelectric film, such as ZnO or AlN, on a semiconductor membrane that is supported by the bulk of the host semiconductor substrate. Configuration of the composite structure can be learned from the previous reports.[2,3] The piezoelectric film is grown by the thin film sputtering technique, and the semiconductor membrane is fabricated using microelectronic etching methods. Both ZnO/Si and AlN/Si structures that support longitudinal or shear modes have been fabricated. The quasi-shear acoustic wave was excited using C-axis inclined piezoelectric films.[6] The AlN films are of more interest because of their higher acoustic wave velocity, resulting in higher resonant frequency. They also exhibit better chemical stability, allowing more flexibility in the fabrication processes and lower sputtering temperature for semiconductor integration consideration. A frequency-temperature characteristic due to the temperature variation for the AlN/Si structure using a shear wave mode is plotted in Fig. 10. A temperatuare coefficient of 400 ppm was measured in the 0 to 100 degree C temperature range for a 2.6 Si/AlN thickness ratio of the resonator. This result is very close to the calculated value,[9] which indicated that temperature compensated resonators can be fabricated.

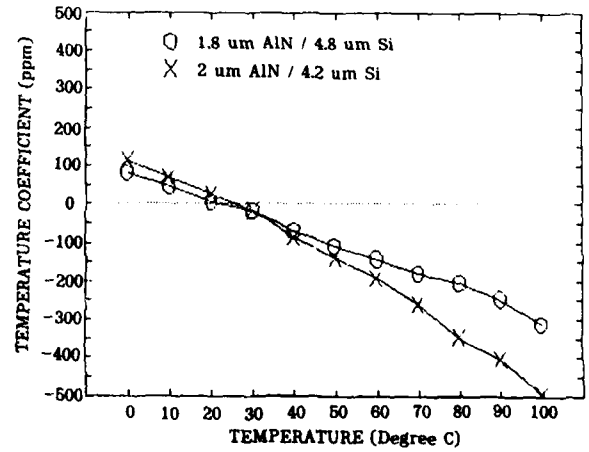


Figure 10. Temperature behavior of AlN/Si SBAR devices.

The dimensions of a SBAR device are in the range of one hundred microns, and the total resonant thickness is about 10 microns. Figure 11 shows monolithic SBAR filters which have been diced in a chip form. The size of these miniature filters is 1 mm x 2 mm. Like IMBAR, the SBAR devices have been mounted in a TO-8 package and have passed the shock and vibration tests. Typically, SBAR devices exhibit a fundamental resonant frequency in a range from 300 MHz to 500 MHz, and have Q's of about 4000. The phase and absolute value of the impedance for a 5 μm AlN on 6 μm Si SBAR device is plotted vs. freqeuncy in Fig. 12. Due to additional parallel capacitance which is intrinsic to the structure, the resonant peak was not much above the background response at series resonant frequency. In order to use the resonator in the oscillator circuit, an inductor across the resonator was used to compensate for this capacitance. A 500 MHz SBAR oscillator in a flat pack is shown in Fig. 13, in which elements of the spiral inductor and shunt capacitors are visible. Figure 14 shows the frequency spectrum of the single-stage oscillator at 507 MHz. The output power level is 8.80 dBm.

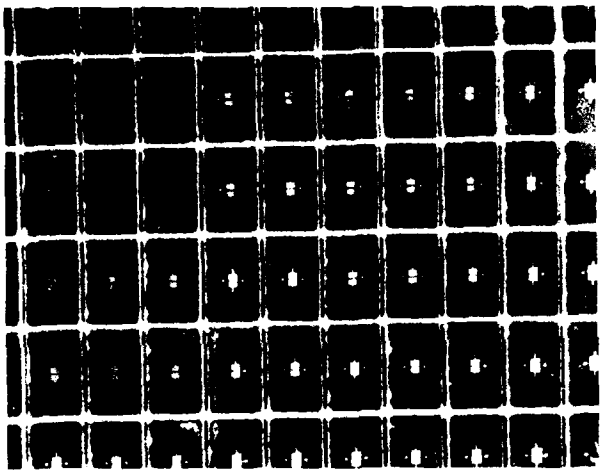


Figure 11. Photograph of diced monolithic SBAR filters fabricated on Si.

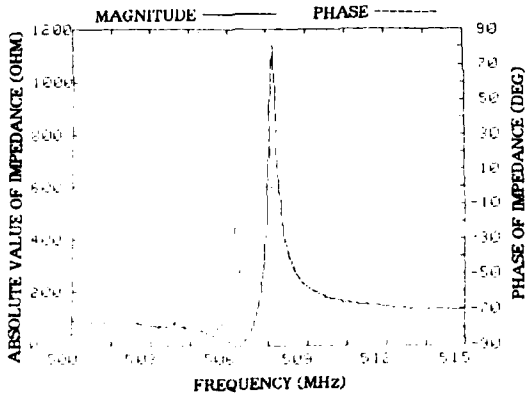


Figure 12. Input impedance plot of a typical AlN/Si SBAR.

A SBAR oscillator using discrete chip elements has also been fabricated on an alumina substrate. Figure 15 shows a prototype oscillator mounted on a test fixture. The components on the circuit can be easily mounted and

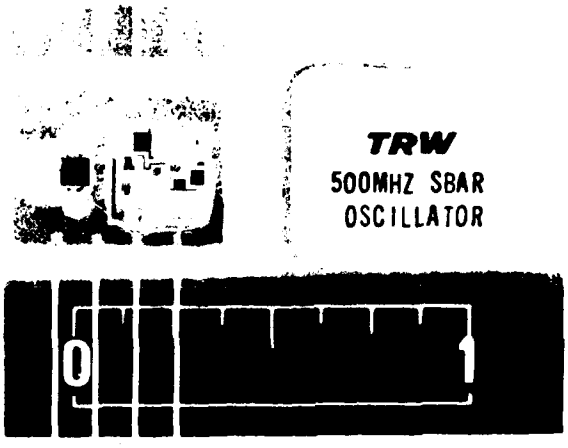


Figure 13. Photograph of a 500 MHz SBAR oscillator mounted in a flat package.

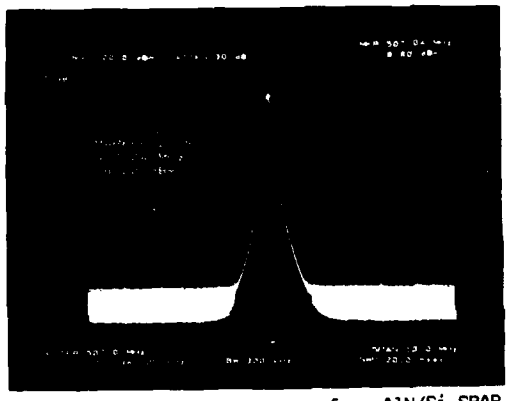


Figure 14. A frequency spectrum of an AlN/Si SBAR oscillator at 507 MHz.

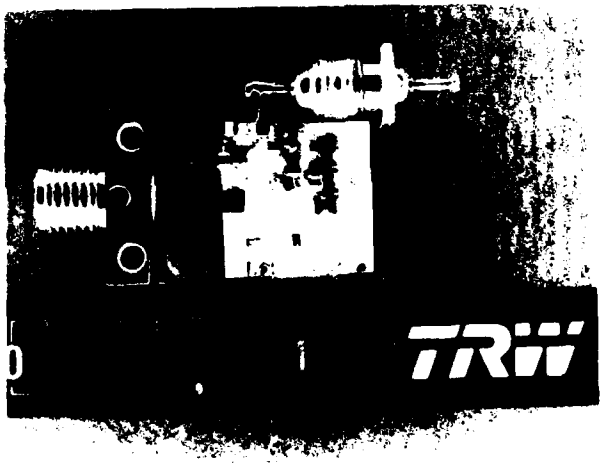


Figure 15. A prototype SBAR oscillator fabricated on an alumina substrate using discrete chip components.

replaced. This method is very useful in the design and analysis of a UHF oscillator circuit. The schematic diagram of the single-stage circuit as a result of this effort is shown in Fig. 16. Since all the elements can be fabricated on a semiconductor substrate, the SBAR device shows great promise for the future integration of oscillators.

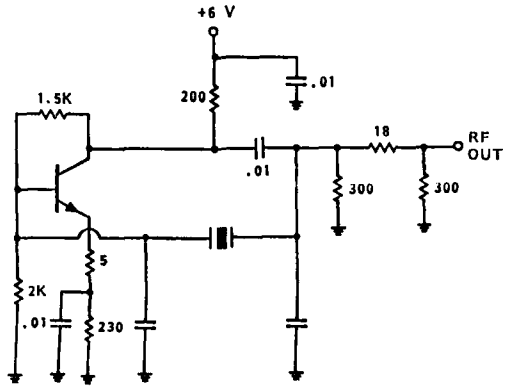


Figure 16. A schematic diagram of a single-stage SBAR oscillator.

Summary

This paper has described our recent work on IMBAR and SBAR devices, including the fabrication technique and device evaluation. A high speed etching technique has been developed to fabricate a thin membrane quartz crystal resonator. Etch rates as high as 30 μm/hr have been achieved, and mirror-like surface finishes have been produced. An IMBAR with a Q of 34,000 at 225 MHz fundamental resonant frequency has been obtained. The Q of the device is, however, less than the expected Q of the crystal, especially in the UHF range. In order to achieve a frequency stability in the 10^{-12} range, improvements on resonator fabrication processes and UHF oscillator circuitry need further investigation. So far, a SBAR oscillator has been fabricated on an alumina substrate using discrete chip components. When combined with a well-developed IC processing tehcnique, these elements could be implemented on the Si or GaAs substrate. Additional work is currently underway in the areas of final frequency trimming, phase noise and stabilization of the oscillators utilizing IMBAR and SBAR resonators. It is expected that stable IMBAR oscillators at desired UHF frequency and an integrated SBAR device can be made in the near future using state-of-the-art technology.

Acknowlegements

The authors wish to thank Phillip Ng, Quang Vo, Jay Crawford, Van Tran, Dong Ho and Kathy Lau for the technical assistance in the course of this research and development.

References

1. G. K. Guttwein, A. D. Ballato and T. J. Lukaszek, United States Patent 3.694.677-26, September 1972.

2. K. M. Lakin and J. S. Wang "Acoustic Bulk Wave Composite Resonators," Appl. Phys. Lett., 38, 125 (1981).

3. T. W. Grukowski, J. F. Black, T. M. Reeder, D. E. Cullen and R. A. Wagner, "Fundamental Mode UHF/VHF Miniature Acoustic Resonators and Filters on Silicon," Appl. Phys. Lett., 37 (22), 993 (1980).

4. J. R. Vig, J. W. Lebus and R. W. Filler, "Chemically Polished Quartz," Proc. 31st Annual Symp. on Freq. Control, p. 131, 1977.

5. M. Berte, "Acoustic Bulk Wave Resonators and Filters Operating in the Fundamental Mode at Frequencies Greater than 100 MHz," Elec. Lett., Vol. 14, p. 248, 1977.

6. J. S. Wang, S. K. Watson, and K. F. Lau, "Reactive Ion Beam Etching for VHF Crystal REsonators," Proc. 38th Annual Symp. on Freq. Control, p. 101, 1984.

7. A. Bohg, "Ethylenediamine-Pynocatechol-Water Mixture Shows Etching Anomaly in Boron-Doped Silicon," J. Electrochem. Soc., Vol. 118, p. 401, Feb. 1971.

8. J. S. Wang and K. M. Lakin, "C-Axis Inclined ZnO Piezoelectric Shear Wave Films," Appl. Phys. Lett., 42, 352 (1983).

9. J. S. Wang, A. R. Landin and K. M. Lakin, "Low Temperature Coefficient Shear Wave Thin Films for Composite Resonators and Filters," Proc. 1983 Ultrasonics Symposium, 491 (1983).

AN AIR-GAP TYPE PIEZOELECTRIC COMPOSITE THIN FILM RESONATOR

Hiroaki Satoh, Yasuo Ebata, Hitoshi Suzuki and Choji Narahara

Toshiba Research & Development Center, Toshiba Corporation,
Kawasaki, 210, Japan

Abstract

A self-supported thin film resonator with a very thin air-gap between the semiconductor surface and the diaphragm bottom surface has been newly developed. The features, fabrication process, experimental results, theoretical studies using a distributed constant equivalent circuit for realization of non-inharmonic overtone response and a VCO application are described in this paper.

The miniature resonator made on trial with a $SiO_2/ZnO/SiO_2$ diaphragm was 300μm x 900μm in size including the bonding area. This means that the area necessary for installation in ICs was 300μm square in the case of full integration. A typical response was measured as follows: 386MHz for fundamental series resonant frequency(f_o), 1100 for series Q-factor and 53 for the ratio (γ) of static capacitance to motional capacitance. As a result of the study on energy trapping condition, non-inharmonic overtone response has been realized with 418MHz for f_o, 970 for Q and 63 for γ.

The practicality of the resonator has also been verified by the voltage controlled oscillator application.

1. Introduction

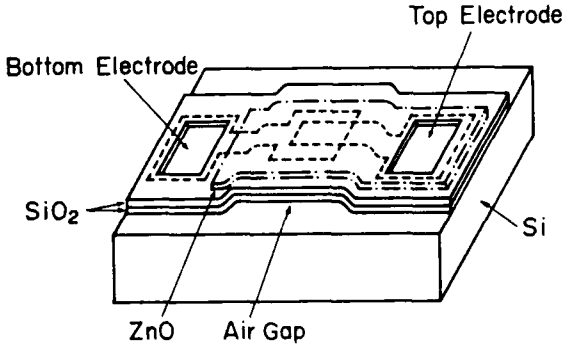
The process technology for making electronic active devices such as integrated circuits has progressed rapidly and has brought about dramatic size reductions in electronic systems in recent years. As a result, electronic circuitry has been greatly used during these years and also the amount will continue to increase rapidly. However, many passive component parts, such as resonators and filters, have not made so much progress and are behind the advance of process technology. One example of this is the presently unfulfilled need for a miniature resonator used in VHF or UHF communication systems.

Recently, much effort has been made to overcome this problem. As an example, a thin film bulk acoustic wave composite resonator with a diaphragm consisting of a piezoelectric zinc oxide (ZnO) film on a silicon (Si) or a silicon dioxide (SiO_2) film has been reported in several papers.[1-4] The resonator has received considerable attention, because of its low temperature coefficient of resonant frequency at its fundamental thickness mode in the VHF/UHF frequency range, its compatibility with active semiconductor circuits in the fabrication process, and the device size.

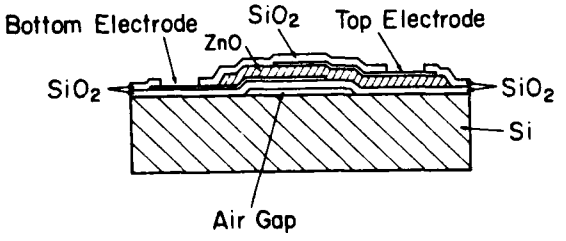
The diaphragm is made by etching anisotropically the (1 0 0) Si substrate, covered on the top side by a back-stop layer such as a heavily boron doped Si layer or SiO_2 film, from the bottom surface through the mask window, to form a pyramid-shaped hole in the substrate. However, this pyramid-shaped area is very large in comparison with the actual resonance region, because the size is proportional to the thickness of the substrate. This area is wasteful for practical use, and the etching process seems to increase the risk of damaging such outer devices as transistors in the case of integrating the resonator into an active circuit. In addition, it is necessary for actual installation in an IC to align the hole pattern by observing from both sides of the substrate.

In this paper, a new type structure with a very thin air-gap between the semiconductor surface and the diaphragm bottom surface has been developed and studied as a solution of these problems and as the most practical method for realization of full integration with ICs.

2. Air-gap Type Resonator



(a) Perspective



(b) Cross Section

Fig. 1 Configuration of an Air-gap Type Thin Film Resonator with a $SiO_2/ZnO/SiO_2$ Diaphragm

CH2186-0/85/0000-0361$1.00©1985IEEE

2.1 Structure

A newly developed resonator structure is shown schematically in Fig. 1(a), (b). The thickness extensional mode resonator has a very thin air-gap between the upper surface of the silicon (Si) substrate and the bottom surface of the diaphragm. A typical composite diaphragm consists of SiO_2 films and a well C-axis oriented piezoelectric ZnO film. The resonator structure takes the shape of a bridge and the diaphragm is self-supported at both ends of the air-gap as seen in Fig. 1(b). The air-gap distance was experimentally determined, as discussed in detail in a later section. The features of this structure are as follows:

1) It becomes possible to fabricate a resonator easily on a flat dielectric passivation film in an IC.
2) A diaphragm can be fabricated from just one side.
3) The device size can be minimized because of its self-supported structure.

2.2 Fabrication Procedure

Device fabrication is summarized as shown in Fig. 2. The ZnO and SiO_2 films are deposited by an RF planer magnetron sputtering method using a ZnO and a SiO_2 target, respectively. The fabrication generally begins with a uniform deposition of a thin non-piezoelectric ZnO film. During the deposition, the Si substrate is kept at a constant temperature of 300°C and a self-biasing voltage of about -25V. The film is photolithographically processed to create a partial layer for the air-gap position with a thin acetic acid etchant.

In the 2nd step, the layer is covered with a sputtered SiO_2 film at the substrate temperature of 20°C. The SiO_2 film becomes the bottom layer of the diaphragm. Next, a titanium-gold electrode (Au/Ti) is deposited on it and formed by using the conventional lift-off method.

In the 3rd step, a well C-axis oriented polycrystalline ZnO film is grown under the conditions of the substrate temperature of 190°C and substrate self-biasing voltage of about -2V. The ZnO film is etched away except for the resonator region. A Au/Ti top electrode is created on it by using the vapor deposition method and the lift-off method.

In the 4th step, these films are overlaid with a SiO_2 film in the case of a $SiO_2/ZnO/SiO_2$ composite diaphragm. Next, the film and the bottom SiO_2 film are etched away with a buffered HF solution to form the contact hole and the side wall. Finally, the bottom non-piezoelectric ZnO film is etched away with a HCl solution through the etching window in the side wall.

The Si substrate was unaffected chemically through the whole process. Since the air-gap layer is made in the final step, the thermal distribution within the substrate can be maintained more uniform during the deposition of each film than that for the conventional thin film resonator. This air-gap structure, as is obvious from this procedure, should enable thin film resonators to be fabricated easily not only on a Si substrate but also on a SOS substrate, on a flat dielectric passivation film in an IC and on other substrate such as glasses and ceramics.

Two key techniques are necessary for the realization of the air-gap type resonator. One is a perfect step coverage technique to prevent the etching damage of the piezoelectric ZnO film in the final etching process for the air-gap making. The best experimental result was obtained by forming small angle slopes as shown in Fig. 3. Another is to make the air-gap very thin in order to prevent the deformation and crack caused by residual stress in the sputtered film. It was found by our experiments that the gap thickness of 1000 Å was appropriate for this purpose.

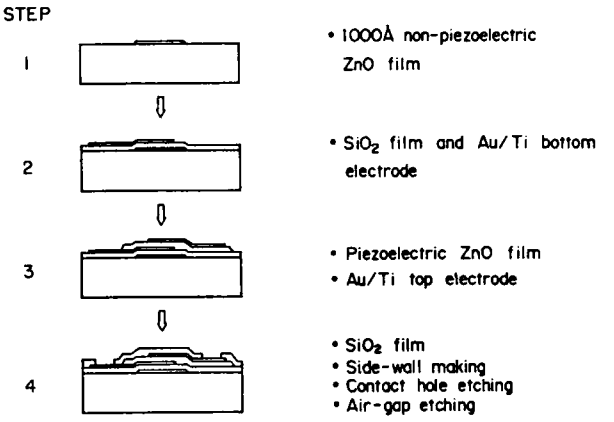


Fig. 2 Fabrication Procedure for a Resonator with a $SiO_2/ZnO/SiO_2$ Diaphragm

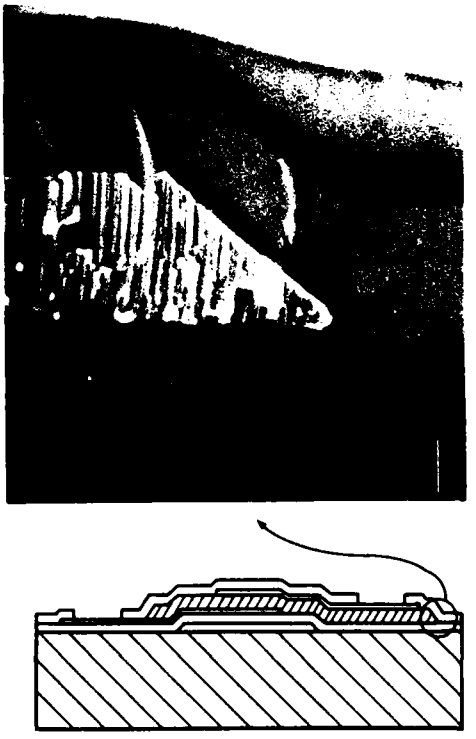


Fig. 3 Cross Sectional SEM Photomicrograph of a Perfectly Covered Step Region

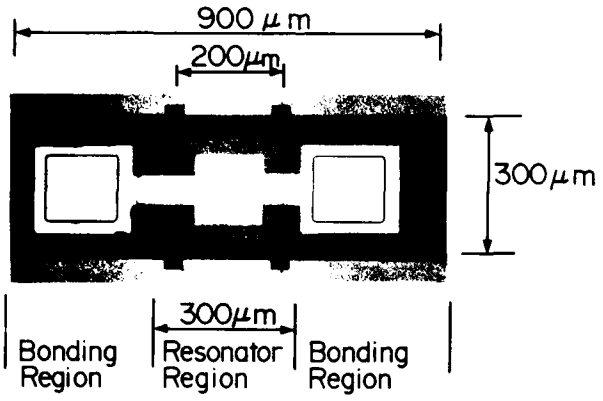


Fig. 4 Top View of a Typical Fabricated Resonator

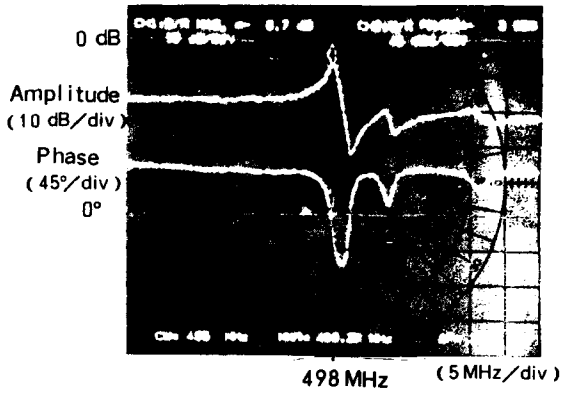


Fig. 5 Frequency Response of the Resonator with a SiO_2/ZnO Diaphragm

2.3 Experimental Results

The experiments were carried out by using two kinds of composite films. One used ZnO/SiO_2 and the other used SiO_2/ZnO/SiO_2 as the diaphragm. A top view of the fabricated resonator is shown in Fig. 4 as a typical example. The resonator region was about 300μm square. And the air-gap size was 200μm in width. The total size including the bonding area was about 300μm in width and 900μm in length. As seen in this figure, the area necessary for installation is 300μm square in the case of full integration. It can be said that this resonator is very practical in terms of the size.

In the case of a ZnO/SiO_2 diaphragm, the air-gap and the contact hole were fabricated by using a photo-resist film instead of the upper SiO_2 film in the above-mentioned process. The thickness of the ZnO film and SiO_2 film were 4.3μm and 1.8μm, respectively. The measured frequency characteristic of a resonator inserted in series into a 50Ω transmission line is shown in Fig. 5. Using the characteristic near the main resonance, the following values were obtained: 498 MHz for fundamental series resonant frequency (f_0), 500 for series Q-factor, 60 for the ratio (γ) of static capacitance to motional capacitance.

In the case of SiO_2/ZnO/SiO_2 diaphragm, the thickness of the upper SiO_2 film, ZnO film and bottom SiO_2 film were 2.6μm, 4.5μm and 2.6μm, respectively. Fig. 6 shows the frequency transmission characteristics measured by a network analyzer with 50Ω input/output impedance under the condition of being series into the line. In this case, the size of the top electrode was 60μm in width and 150μm in length.

From the characteristic near the main resonance, the values of equivalent circuit parameters were obtained as listed in Table 1. The unloaded Q-factor was 1130 which was twice as much as that for the ZnO/SiO_2 structure and the capacitance ratio was 50. It seems that the increase in the Q-factor depends not on the layer structure difference, but on the surface cleanness of the diaphragms.

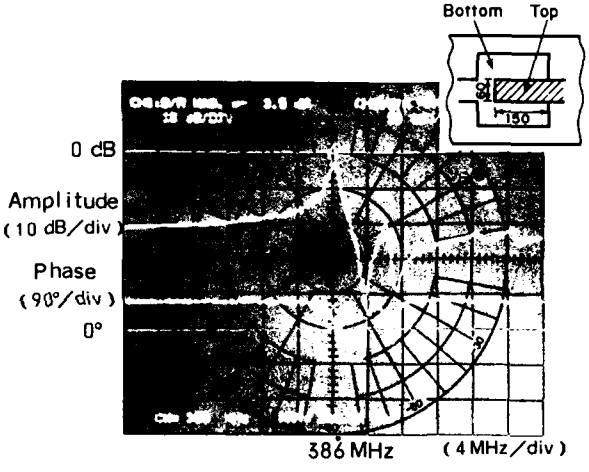


Fig. 6 Frequency Response of the Resonator with a SiO_2/ZnO/SiO_2 Diaphragm and with Top Electrode Size of 60μmx150μm

Table 1 Equivalent Circuit Constants (Top Electrode: 60μmx150μm)

Fr (MHz)	386
L_1 (μH)	23.8
C_1 (pF)	0.007
R_1 (Ω)	51
C_0 (PF)	0.381
γ	53
Q	1130

3. Theoretical and Experimental Study

3.1 Realization of Non-inharmonic Overtone Response

The previous resonators have inharmonic spurious responses as seen in Figs. 5 and 6. In order to obtain a single resonance condition, the frequency characteristic was analyzed theoretically by using a distributed constant equivalent network for an energy trapping

resonator. Fig. 7 shows typical dispersion curves for the Lamb wave propagated in a composite film with a symmetrical structure, in the case of the ratio of ZnO film to total thickness being 0.464. The solid and dashed curves respectively show the dispersion for both free and metalized surfaces. Frequency Ω is normalized by the ratio of the ZnO film thickness t_p to the longitudinal propagation velocity ($\sqrt{C_{33}^E/\rho}$) in the ZnO medium.

$$\Omega = \frac{\omega t_p}{V\ell} = \omega t_p / \sqrt{\frac{C_{33}^E}{\rho}},$$

where ω is the angular frequency, and C_{33}^E and ρ are the elastic stiffness constant and density of the ZnO crystal, respectively. These curves were obtained by using the conventional two dimensional field analysis technique for a composite film membrane structure.[5] The following approximations, however, were made in this analysis.

1. The titanium-gold electrode layer is of infinitesimal thickness and has a finite mass per unit area which is considered as a mass loading effect in the boundary condition between the SiO_2 film and ZnO film.
2. The electrodes have infinitesimal resistance and are short-circuited electrically at zero potential.
3. Acoustic losses in the silicon dioxide and zinc oxide are infinitesimal.
4. The polycrystal ZnO film has C-axis orientation normal to the diaphragm surface and also has the properties of a single crystal.
5. The diaphragm is perfectly flat and uniform infinitely along the propagation direction.

The conventional energy-trapping technique is applicable to the thickness extensional mode (TE 1) of this diaphragm because the resonant frequency (Ω_0) in the metalized region is lower than the resonant frequency (Ω_0') in the free region and, among these frequencies, the wave number (k') in the free region is imaginal and the wave number (k) in the metalized region is real as seen in Fig. 7.

The dispersion curves near the resonance region can be approximately expressed by two quadratic curves as follows.

$$\Omega^2 = \beta' k'^2 + \Omega_0'^2 \quad \text{and} \quad \Omega^2 = \beta k^2 + \Omega_0^2.$$

A distributed-constant 3-port equivalent network can be expressed by a form including the energy trapping effect as shown in Fig. 8 within the frequency range in which this approximation is valid.[6,7] A driving point admittance in the equivalent circuit can be obtained as the following expression,

$$Y = j\omega C_0 + \frac{N^2}{2\theta Z_0}\left\{1 - \frac{1}{\theta(\cot\theta + j\frac{Z_0}{Z_0'})}\right\},$$

where $\theta = kW$, $Z_0/Z_0' = k/k'$ and C_0 is an electromechanical clamped capacitance. The value (N^2/Z_0) can be calculated by equalizing the following two equivalent inductances. One is the equivalent inductance obtained by expression (1) under the condition of $Z_0' = 0$, and the other is calculated using Mason's one-dimensional equivalent circuit.

Fig 9 shows the theoretical frequency amplitude response using the original piezoelectric e-constant. The parasitic capacitances between bonding electrodes are also included. However, the calculated value of the equivalent inductance (L) was obviously lower than the measured value. In order to obtain an agreement between these values, we have made calculations by

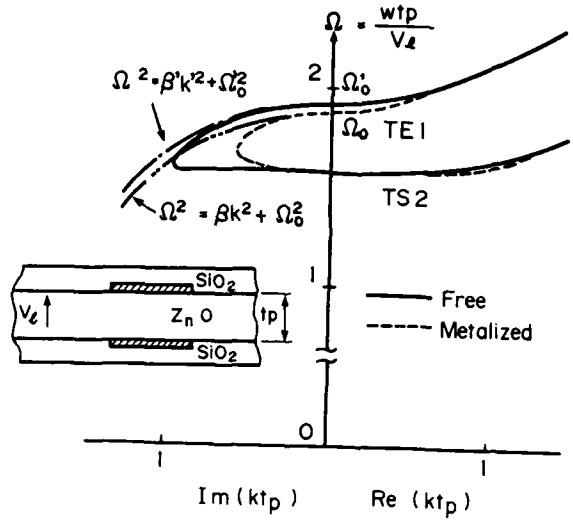


Fig. 7 Dispersion Curves for a Symmetric Composite Plate

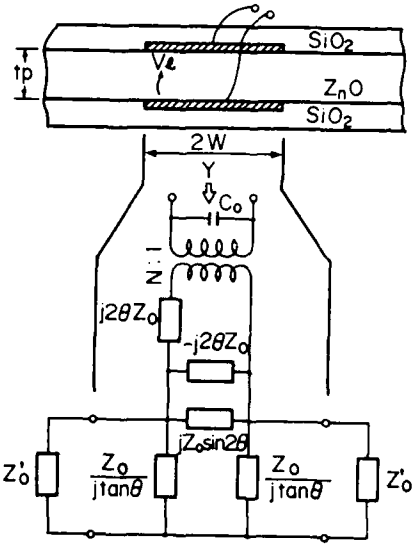


Fig. 8 Distributed-constant Equivalent Circuit Representing a Thickness Expansion Wave Resonator

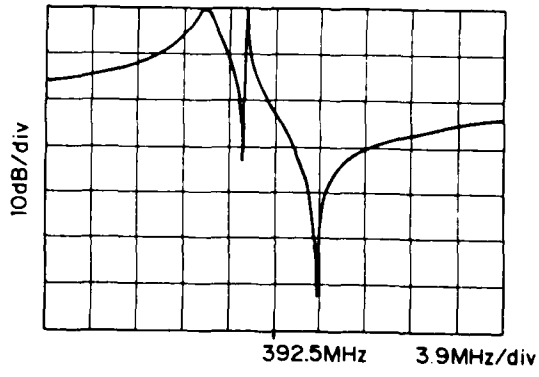


Fig. 9 Theoretical Response of the Resonator with a Top Electrode Size of 60μmx150μm in the case of e_{33}=1.14C/m²

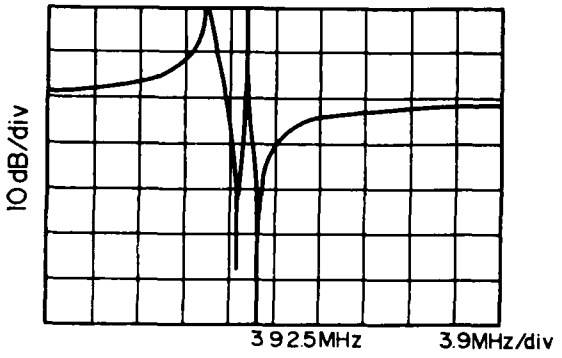


Fig. 10 Theoretical Response of the Resonator with a Top Electrode Size of 60μmx150μm in the case of e_{33}=0.7C/m^2

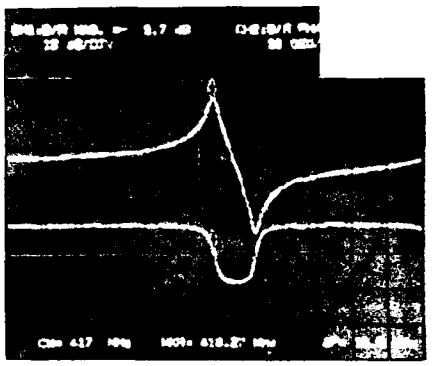


Fig. 13 Measured Non-Inharmonic Overtone Response for Top Electrode Size of 60μmx60μm

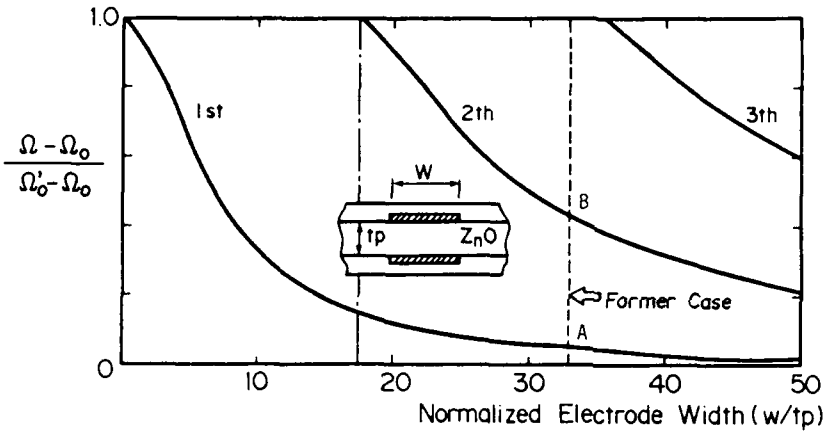


Fig. 11 Resonant Frequency Spectrum in the Energy Trapping Region for the Thickness Ratio of ZnO Film to Entire Film being 0.464

Table 2 Equivalent Circuit Constants (Top Electrode: 60μmx60μm)

Fr (MHz)	418
L_1 (μH)	42.0
C_1 (pF)	0.003
R_1 (Ω)	114
C_0 (PF)	0.216
γ	63
Q	970

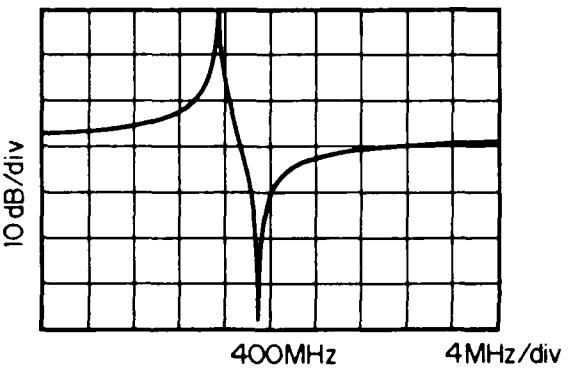


Fig. 12 Theoretical Non-Inharmonic Overtone Response

changing the value of the e-constant. An agreement was obtained under the condition of the e-constant being about 0.7C/m^2. Fig. 10 shows the calculated frequency response. The profile of the main resonance curve has also a close resemblance to the measured one. This indicates that the effective value of the e-constant is about 0.7C/m^2 in our case.

Fig. 11 shows the theoretical resonant frequency spectrum in the energy trapping condition, subject to the e-constant of 0.7C/m^2. In the former case shown in Fig. 6, the electrode width normalized by ZnO film thickness, (W/t_p), was about 33, and two resonances appear as indicated by "A" and "B" points in Fig. 11. As can be understood from this figure, the single resonance condition can be obtained in the region of $W/t_p \leq 18$. A typical frequency response calculated under the condition of W/t_p=14 does not have any spurious inharmonic responses, as shown in Fig. 12. A successful realization of a non-spurious resonator was made by reducing the electrode overlap length in accordance with our theory. Fig. 13 shows the measured frequency response of the resonator and the equivalent circuit constants are listed in Table 2. The capacitance ratio was 63 and the unloaded Q-factor was 980 which were slightly inferior to the value listed in Table 1 because the influence of parasitic capacitance and resistance is a little greater than that of low impedance resonator.

3.2 Voltage Controlled Oscillator Application

A Colpitts-type voltage condtrolled oscillator (VCO) was taken up as a typical application example. Fig. 14 shows the performances of the VCO with a resonator having a top electrode of 60μmx150μm in size, operating under the condition of the source voltage of about 3V and controlled voltage between 0 and 5V. The oscillation frequency could be controlled in the range of 450kHz. The carrier to noise ratio C/N (C; carrier, N; 12.5kHz offset from the carrier and 8 kHz band width) was more than 85dB. Fig. 15 shows the relation between the effective driving current in the series resonant circuit of the resonator and the measured carrier to noise ratio. In this case, at least 1 mA driving current is necessary for obtaining a good side band noise characteristic.

It has become clear from the experimental result that the sideband noise characteristics are very excellent and the measured carrier to noise ratio C/N satisfies the specification of C/N≧75dB required in 12.5kHz channel separation mobile communication systems. It should especially be noticed that an unloaded Q-factor of several hundreds is within practical range in terms of spectrum purity required for a VCO.

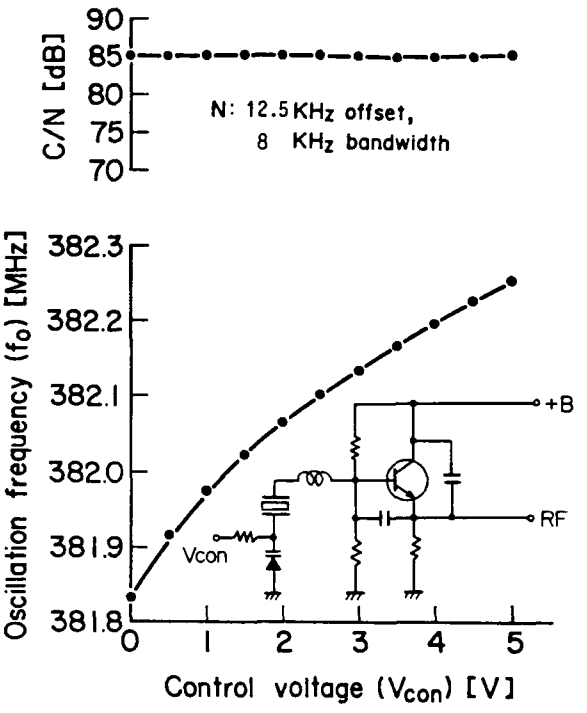


Fig. 14 Perforamnces of a Colpitts-type VCO

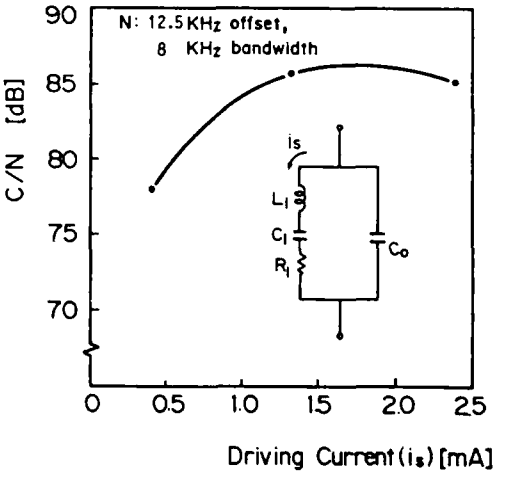


Fig. 15 Driving Current Dependency of C/N in the VCO

Summary

An air-gap type thin film resonator with the features of easy fabrication on ICs, small device size and single side processing has been proposed and developed. Concerning fabrication, it is very important to make the step coverage perfect and to make the air-gap very thin. A non-inharmonic over tone resonator has been realized as a result of the discussion for the energy trapping condition. And a VCO application has clearly indicated that the resonator is within practical range in terms of its spectrum purity.

The successful development of the resonator should be able to promise that VHF/UHF signal processing circuits will be fabricated in a fully monolithic form in the near future.

Acknowledgments

The authors wish to thank Prof. H. Shimizu, Prof. K. Yamanouchi and Assist. Prof. K. Nakamura of Tohoku Univ. for their valuable advice to our theoretical analyses. They also wish to thank Dr. N. Shimomura, T. Fukuda and Y. Mikuni of Toshiba Corporation for their encouragement and support of this work, and to T. Kodama and S. Morishita for valuable discussions and technical contribution during the course of this work.

References

[1] K.M. Lakin and J.S. Wang: Proc. IEEE 1980 Ultrasonic Symposium, pp. 834-837 (1980)
[2] T.W. Grudkowski, J.F. Black, T.M. Reeder, D.E. Cullen and R.A. Wagner: Proc. IEEE 1980 Ultrasonic Symposium, pp. 829-833 (1980)
[3] K. Nakamura, H. Sasaki and H. Shimizu: Electron. Lett., vol. 17, p. 507(1981)
[4] M. Kitayama, T. Fukuchi, T. Shiosaki and A. Kawabata: Jpn. J. Appl. Phys., vol. 22. Appl. 22-3, pp. 139-141 (1983)
[5] R.F. Milsom: Proc. IEEE 1982 Ultrasonic Symposium, pp. 484-489 (1982)
[6] K. Yamada and H. Shimizu: Proc. IEEE 1984 Ultrasonic Symposium, pp. 383-387 (1984)
[7] T. Obara, K. Nakamura and H. Shimizu: Reports of 1984 Spring Meeting, Acoust. Soc. Japan, No. 3-5-4, pp. 677-678 (1984) (In Japanese)

PROCESSING OF A FIVE RESONATOR VHF CRYSTAL DEVICE

By: C.W. Shanley, L.N. Dworsky,
J.A. Whalin, G.C. Clifford,
and M.N. Scansaroli

Motorola, Inc.
8000 W. Sunrise Blvd.
Plantation, FL 33322

Abstract

A simple yet elegant receiver can be constructed by placing a narrowband filter at the operating frequency and mixing the output of that filter down to the audio region. Such a receiver has been implemented by integrating a four pole filter and separate resonator on a single quartz blank. The design of this crystal is detailed in a companion paper.[1]

Due to the tight tolerances of the filter specifications, the large number of pinouts required, and the high frequency of operation, many new processing and packaging techniques had to be developed. Requirements for accurate and precise electrode dimensions dictated a photolithographic process for resonator definition. A technique was developed which used a dry film photoresist laminated to the blank.

In order to permit electrical connection to the five electrodes as well as present a low profile, a TO-8 package is used. Inside the package is a ceramic header with four pedestals near the center onto which the blank is mounted. Wirebonds are used to make the connections from the ceramic to the crystal.

The crystals are tuned in an improved version of the DC Plasma Anodization system described in an earlier Symposium.[2]

This process has been used to fabricate fifth overtone four pole filters and resonators on a single blank at frequencies from 35 MHz to as high as 174 MHz.

Introduction

There has been a great deal of interest in recent years in the fabrication of crystal filters at high frequencies. Much of this work has been with crystal blanks which have been locally thinned by chemical etching [3] or ion milling techniques.[4] Such devices are capable of operation at frequencies of several hundred MHz or higher but are often difficult or expensive to manufacture.

Surface acoustic wave (SAW) resonator filters are also capable of operation in the VHF range or higher.[5] The bandwidth of these devices is typically on the order of .1%, or more than 100 KHz at VHF frequencies. This bandwidth implies that additional filtering at some intermediate frequency is necessary to separate out a single 25 KHz wide radio channel.

More conventional bulk mode monolithic crystal filters have also been demonstrated at frequencies up to 162 MHz.[6] A four pole device such as this operating in the VHF band may serve as an almost ideal front end filter for a single channel radio, or as a very high IF filter for a much higher frequency radio. Widespread use of such devices has been hampered by the problems associated with shadow mask electrode definition, pinout limitations of conventional HC series packages, and the difficulty in tuning several small electrodes on a single crystal blank.

These issues have been addressed with the development of a process to manufacture a device containing a four pole filter and a separate resonator on a single crystal blank. The process involves dry film photolithography, an unusual packaging arrangement, and plasma anodization tuning of the resonators.

Metallization

The metallization process consists of cleaning the crystal blank, and coating them with a uniform layer of aluminum on each side.

Polished AT-cut blanks are received pre-sorted into narrow frequency windows. The blanks have been sorted at the operating overtone with an air gap technique described at a previous symposium.[7] Third overtone crystals are used for devices up to 90 MHz, while fifth overtone crystals are used from 90 to 174 MHz. The presorting is necessary because of the lack of a bandwidth adjust capability and the relatively narrow frequency adjustment range of the tuning technique.

The blanks are cleaned prior to metallization with the following process:

1. Ten minutes in a 30% Sodium Hydroxide bath.

CH2186-0/85/0000-0367$1.00©1985IEEE

2. Hot distilled water rinse.
3. Ten minutes in a Chromerge solution (Manostat, Inc.).
4. Hot distilled water rinse.
5. Monitored (resistivity greater than 19 megohm-cm) distilled water rinse.

After cleaning, the blanks are mounted in a planetary evaporation system. Each planet consists of an outer ring and a central disk, and holds thirty-four crystals. Figure 1 shows one of the planets. The central disk is spring loaded to flip 180 degrees upon command, thereby permitting the coating of both sides of the blank in a single pumpdown. Up to eight planets may be used during each evaporation.

The evaporator itself is a custom design, fixtured for planetary evaporation. It features a ten inch cryogenic pump (Cryotorr 10, CTI Cryogenics Corp.) to ensure clean, rapid pumping to a high vacuum. The aluminum deposition is done by conventional electron beam evaporator techniques.

After evaporation, the crystals are removed from the system for photolithography.

Photolithography

Two factors indicate that a photolithographic process be used to fabricate the electrode pattern rather than traditional shadow mask techniques. First, radio requirements dictate that the filter parameters be unusually precise and accurate. Too narrow a bandwidth increases insertion loss and can decrease radio sensitivity over temperature. Conversely, too wide a bandwidth will reduce adjacent channel selectivity. Second, if the resonators were initially too far off frequency, the tuning process could not tune them down to the desired frequency.

Experiments with liquid resists were discouraging. A meniscus usually formed on the edge of the blank, handling was difficult, and the application of resist to both sides of the blank complicated the procedure. Attempts were made to laminate an aluminum metallized blank to Riston dry film photoresist (Dupont de Nemours, Inc.) which proved successful. Eventually a custom machine was fabricated which mechanized the operation.

In order to facilitate the exposure and development process, the crystals are placed in groups of eight in a stainless steel shim. The shim is patterned in a 35 mm film format because of the availability of sprockets and drive mechanisms. The group of eight crystals occupy the center strip of the 35 mm format shim, and are covered with photoresist on both sides. The drive holes in the shim above and below the crystals are kept clear of photoresist.

The lamination process proceeds as follows. First, the shim is laminated with Riston photoresist on one side. The shim is turned over and metallized blanks are carefully placed in the center of the eight holes of the shim. The resist is slightly tacky, which helps to hold the crystals in place. Next, a second strip of photoresist is laminated over the shim so as to capture the blanks between two layers of photoresist. Three groups of eight crystals are laminated into each shim. The laminated shims are put aside for 30 minutes to stabilize the resist prior to exposure. Figure 2 shows three of the shims. The first is empty, the second has photoresist laminated to one side, and the third shows eight crystals dropped in the center pockets of the shim. Figure 3 shows a closeup of the eight crystals in the pockets before the second lamination step.

The photo process is done in a custom double sided contact photolithography machine. The machine holds two iron oxide masks, one with the pattern for the front of the crystal and one with the pattern for the back. Each pattern has been step and repeated in a pattern of eight to match the clustering of crystals in the shim. The shim is advanced so that the eight crystals are between the two masks, and the masks are brought into contact with the photoresist covered shims. Exposure of the front and back patterns of all eight crystals are made simultaneously.

Because the electrodes are so small and the filter tolerances so narrow, careful front to back alignment of the pattern is critical. Since the iron oxide masks are semitransparent, this is easily checked. The masks are simply brought together with no shim between them, and the masks checked with a microscope. This may be done as often as necessary between exposures to guarantee alignment.

In addition to the electrode pattern, four tabs are exposed on each crystal. These tabs extend from the crystal across the gap to the shim, and well into the shim. Their purpose is to hold the crystals in the shims after development.

After exposure, the shims are put aside for 30 minutes for the photoresist to stabilize. The protective Mylar layer is peeled off and the shims placed in a custom spray develop/etch machine. This machine is again designed to handle the 35 mm format shims, and moves each shim through a sequence of three tanks. The first tank sprays developer (Dupont D-2000), the second a water rinse, and the third an aluminum etchant. The etchant is made of 80% phosphoric acid, 5% acetic acid, 5% nitric acid, and 10% water.

The purpose of the tabs connecting the crystal to the shim is now clear. After the photoresist is developed in the first tank, something is needed to hold the crystals in the shim for the rinse and etch process. Without the tabs the crystals would fall into the bottom of the developer tank. Figure 4 shows a grouping of eight crystals after development and etching, held in the shim by tabs of developed photoresist.

After development and etching, the shims are dipped in acetone to dissolve the photoresist. The crystals are then chemically cleaned and plasma cleaned prior to mounting. Figure 5 shows the resolution obtained on the separate resonator with this process. The resonator is approximately 1.47 x .66 mm (58 by 26 mils).

Mounting and Packaging

The crystals are mounted in an unusual manner. In order to permit electrical connection to the five electrodes as well as present a low profile, a TO-8 package is used. Inside the TO-8 header is a ceramic substrate with four small pedestals near the center. The pedestals are located along lines rotated approximately 60 degrees from the X-axis of the crystal, where the force sensitivity of resonance frequency is known to be minimized.[8] Around the periphery of the substrate are wire-bonding pads electrically connected to the package pins. The pedestals and the pads are made of electroplated copper approximately one mil thick, with a thin flash of gold to facilitate bonding. The crystal is mounted to the pedestals with small amounts of conductive epoxy (C110-1 adhesive, Amicon, Inc.).

Electrical connection to the electrodes on the top surface of the crystal are made with aluminum wirebonds from the bond pad of the ceramic substrate. Wirebonding is done with an ultrasonic wirebonder. During the bonding operation, the crystal blank is actually deflected down to the substrate surface. Connections to the bottom (ground) electrodes are made through the pedestals, which have thin film metalization patterns running to the package pins. Thus, the resonators may be located anywhere on an annulus between the pedestals and the edge of the blank. The mounting system has proven to be excellent in terms of shock and vibration resistance, with virtually zero crystal breakage problems reported from use in the very rugged portable pager environment.

Tuning

The crystals are tuned using an improved version of the DC Plasma Anodization technique described at a previous symposium. [2] In this implementation eight cassettes, each holding four crystals, are inserted in a vacuum system. Each crystal is placed in front of its own anodization "gun." All thirty-two crystals (160 resonators) are tuned simultaneously. During the tuning operation, the oxygen pressure is 800-1000 mtorr and the oxygen flow rate 35-50 cc/minute. The anodization gun voltage is 630 V, and the DC bias applied to the electrodes undergoing tuning is 30 volts.

A sophisticated distributed computer system handles the tuning operation. In addition to sequencing the vacuum system, the system measures the frequency of each resonator by monitoring the reflection coefficient as the frequency is swept. The initial anodization times are calculated based on each resonator's frequency and historical data. The computer notes initial rates and bases the subsequent anodization times on that information and the known functional form of the frequency versus anodization time curve. The algorithm is kept conservative to avoid overshooting the target frequency. The entire operation, from initial pumpdown to completion of the anodization process, requires no operator intervention.

After anodization the crystals are sealed in a conventional resistance welder. Finally, the crystals are tested to ensure that they meet all specifications. Figure 6 shows a finished 162 MHz crystal just prior to sealing.

Conclusion

A manufacturing process for multipole VHF crystal filters has been described. As many as five resonators, normally configured as a four pole filter and a separate resonator, can be fabricated on a single blank with current equipment. Unique aspects of the process include the use of a laminated dry film photoresist, an unusual and rugged mount, and the use of DC plasma anodization to tune the resonators. The process is well mechanized to permit the production of these devices in high volume.

References

1. L.N. Dworsky and C.W. Shanley, "The Motorola Multi-Pole Monolithic Crystal Filter Project," Proceedings of the 39th Annual Symposium on Frequency Control, 1985.

2. C.W. Shanley and L.N. Dworsky, "DC Plasma Anodization of Quartz Resonators," Proccedings of the 36th Annual Symposium on Frequency Control, 1982.

3. P.J. Kavolis and W.P. Hanson, "Filter Application of High Frequency Chemically Polished Fundamental Mode Bulk Wave Quartz Crystal Resonators," Proceedings of the 36th Annual Symposium on Frequency Control, 1982.

4. B. d'Albaret and P. Siffert, "Recent Advances in UHF Crystal Filters," Proceedings of the 36th Annual Symposium on Frequency Control, 1982.

5. W.J. Tanski, "Multipole SAW Resonator Filters," Proceedings of the 36th Annual Symposium on Frequency Control, 1982.

6. R.C. Smythe and M.D. Howard, "Current Trends in Crystal Filters," Proceedings of the 37th Annual Symposium on Frequency Control, 1983.

7. L.N. Dworsky and G. Kennedy, "Air-Gap Probe Evaluation of Thin Quartz Plates," Proceedings of the 35th Annual Symposium on Frequency Control, 1981.

8. J.M. Ratajski, "The Force Sensitivity of AT-Cut Crystals," 20th Annual Symposium on Frequency Control, 1966.

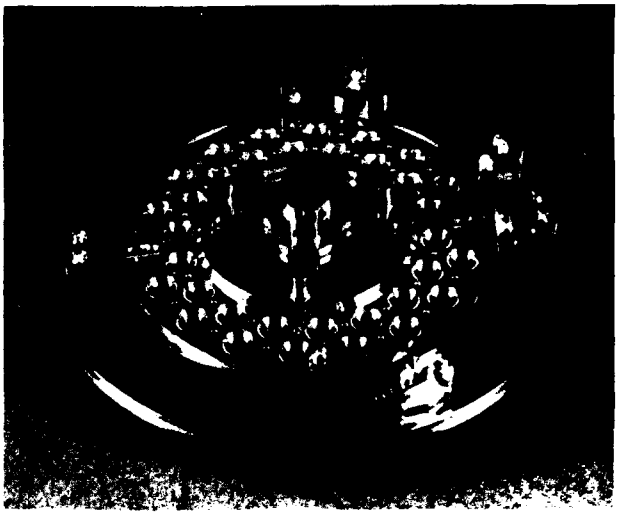


Figure 1: Crystal Holder for Planetary Evaporation System

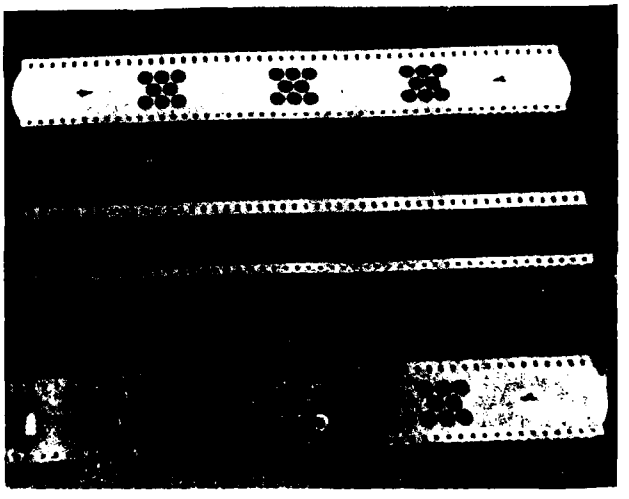


Figure 2: Lamination process. A stainless steel shim (top) is laminated with photoresist on one side (middle) and reversed. Crystals are dropped in the resulting pockets (bottom), and another strip of photoresist is applied.

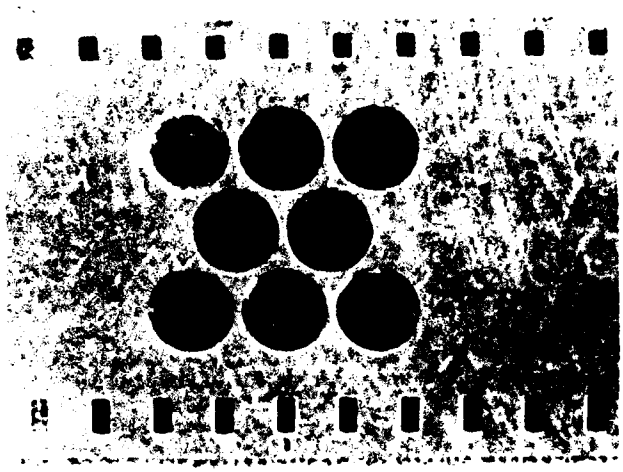


Figure 3: Closeup of Crystals in Pockets

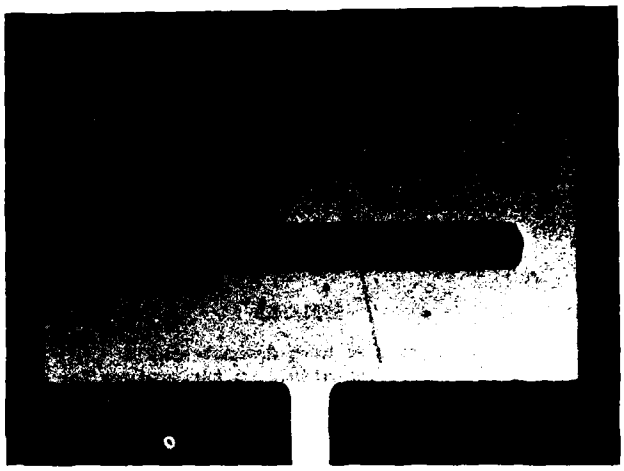


Figure 5: Closeup of the Separate Resonator Used in a Local Oscillator Circuit. Polarized light was used to show the top and bottom electrode in different shades of grey.

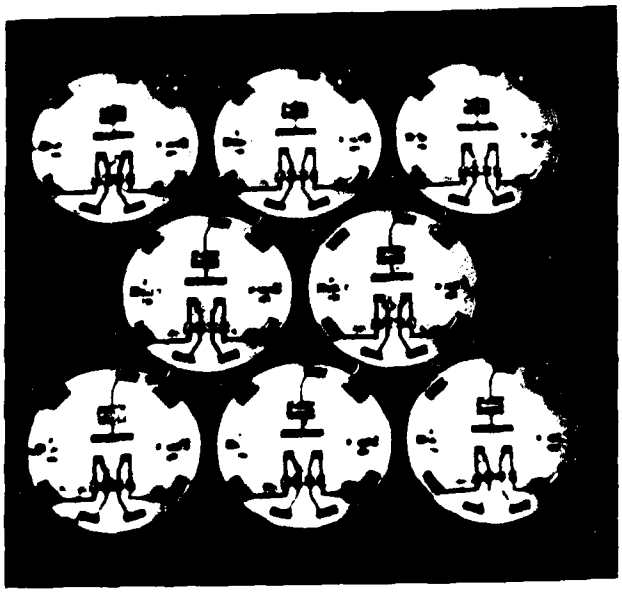


Figure 4: Crystals After Development and Etching. At this point they are held in the shim by tabs of developed photoresist.

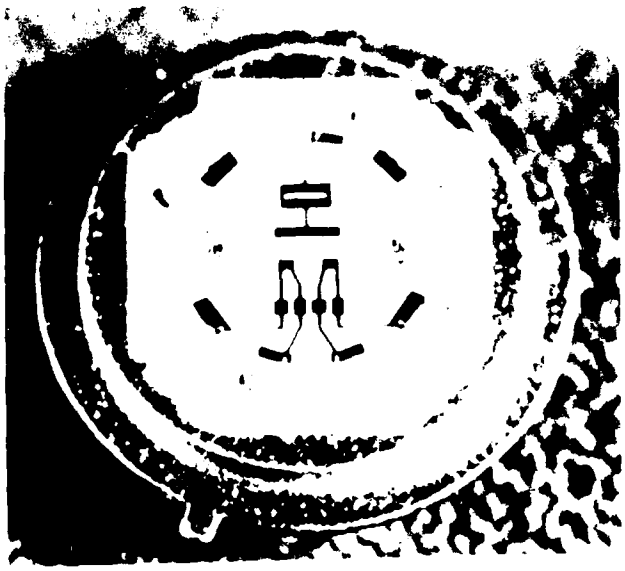


Figure 6: Tuned Crystal Prior to Sealing

HIGHLY STABLE, OVENIZED BULK SHEAR MODE RESONATORS*

B. R. McAvoy and S. V. Krisnaswamy
Westinghouse R&D Center
Pittsburgh, PA 15235

H. L. Salvo, Jr., and R. A. Moore
Westinghouse Defense and Electronics Center
Baltimore, MD 21203

Abstract

In previous work we have demonstrated a parabolic frequency dependence on temperature for specific cuts of lithium tantalate excited by shear mode ZnO transducers[1,2]. These cuts confirmed the validity of the computer model used to determine the first order zero temperature coefficient propagation directions[3,4]. This work also demonstrated that nonpiezoelectric microwave resonators exhibiting temperature stability of an overtone frequency could be produced.

Our current work is an extension of this effort with three goals:

(1) Find more useful cuts in terms of providing a higher temperature turnover appropriate for stabilization by means of a thermostated oven.

(2) Provide increased vibration resistance by identifying crystal cuts which have minimum internal piezoelectric coupling with the requisite temperature behavior as in (1).

(3) Permit the highest loaded Q with the requirements provided for in (1) and (2).

Toward these goals we have examined the temperature frequency characteristic of a new singly rotated cut of lithium tantalate. Several identical cuts were tested at a number of different microwave frequencies. The turnover region for this new cut is from +60°C to +80°C. The resonator plate orientation for which these results were obtained is a single rotation of the plate about x away from z to an angle near 10°. Loaded Q values in excess of 10^4 at 2 GHz have been measured. This value is believed to be limited by material defects distributed throughout the bulk of the crystal. These results show temperature stable high overtone bulk acoustic resonators (HBARs) with an application potential substantially greater than those shown before with a frequency stability exceeding parts per million with relatively simple temperature stabilization.

Introduction

The high overtone bulk acoustic resonator (HBAR) has been under investigation during the past few years for potential low noise oscillator applications in the 1 to 3 GHz frequency range[5,6]. This work has utilized high overtone longitudinal modes generated by ZnO transducers aligned on opposing faces of a nonpiezoelectric substrate. This configuration provides a transmission cavity resonator which takes advantage of the high Q, low propagation loss, properties in materials such as YAG and $MgAl_2O_4$. With the use of shear mode transducers in an identical configuration on $LiTaO_3$, the additional property of zero temperature coefficient of frequency may be achieved. Table 1 lists the temperature coefficient of frequency for various materials using an HBAR configuration for each measurement. Lithium tantalate is the only known low loss material with the potential for a zero temperature coefficient of frequency.

TABLE 1

TEMPERATURE COEFFICIENT OF FREQUENCY
USING HBAR CONFIGURATION

• SPINEL ($MgAl_2O_4$) [1,1,1] LONGITUDINAL MODE NEAR 1.6 GHz:	-28.8 ppm/°C
• YAG [1,1,1] CORED LASER STOCK LONGITUDINAL MODE NEAR 1.6 GHz:	-48 ppm/°C
• GaAs [1,00] HIGH RESISTIVITY LONGITUDINAL MODE NEAR 3 GHz:	-50 ppm/°C
• $LiTaO_3$ θ = -19 . ϕ =0 SHEAR MODE NEAR 1.6 GHz	+10 ppm/°C

These temperature stable propagation directions exist for the bulk shear modes as exemplified for one specific cut as shown in figure 1. In the upper right of the figure the temperature coefficient of delay is plotted as a function of angle in the z-x plane for the two shear modes (dashed lines) and for the longitudinal mode (solid line). Other plots include the slowness and power flow data in and out of the plane. The small circle on the x axis identifies the zero coordinate for both the temperature coefficient and power flow angles. This and similar plots as identified in reference 3 and 4 are being used in this study.

Experimental

In an earlier paper[7] we discussed the generation of shear bulk waves at microwave frequencies using transducers fabricated with RF

*This effort was sponsored in part by the Electro-magnetic Sciences Division, Rome Air Development Center, Air Force Systems Command under contract F19628-84-C0099.

CH2186-0/85/0000-0372$1.00©1985IEEE

diode sputtered ZnO films. Shear mode transducer films of ZnO also have been fabricated using low energy ion bombardment during growth.[8] Details of the system employed and the deposition parameters used are reported in reference 7. The transducers are fabricated using aluminum bottom contacts. The ZnO deposition is made on this bottom contact through an appropriate mask. The top contact is formed with a second aluminum deposition.

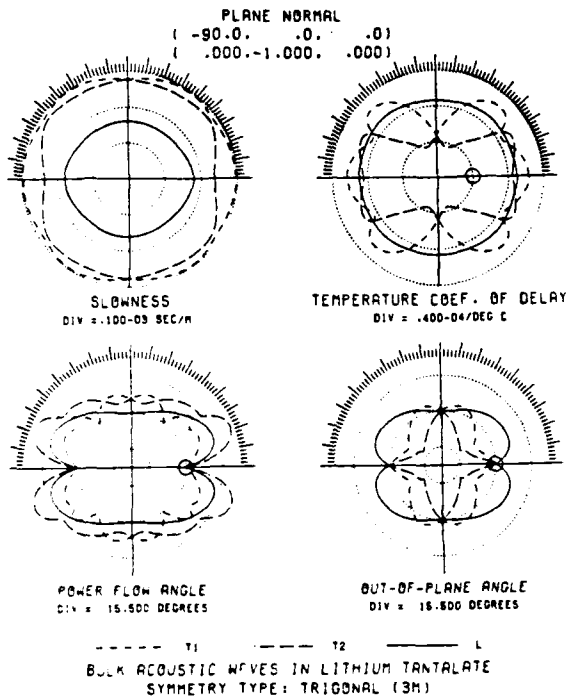


Figure 1. Polar Plots in the Z-X Plane Showing Zero Temperature Shear Mode (T_1, T_2) Directions.

Resonator response is measured using a HP 8410C network analyzer with a HP 8340A synthesized sweeper which provides a resolution of Hz at 2 GHz. A typical resonator mounting p dure is shown in figure 2. The microstripline fixture shown accomodates three mounted HBARs on a common heat sink. This arrangement facilitates the comparison of several resonators as a function of temperature. The HBARs are not sealed in their packages but are open to the air as shown in figure 2. This requires a modification to the environmental oven used for the temperature measurements. The interior of the oven is shown in figure 3. It is a Tenney, Jr. with a controlled temperature range of -80°C to +177°C. An RTD temperature detector attached to the resonator package is used together with a Doric 410A-P7 temperature monitor. This arrangement provides an overall temperature accuracy of ± 0.2°C. As shown in figure 3, to the rear of the oven, a heat exchanger temperature conditions dry nitrogen which is let into the chamber at a slightly positive pressure. The dry nitrogen at the oven ambient protects the open resonators. The microstrip with mounted HBARs is enclosed in the third aluminum box at the top of the stack in the middle of the oven.

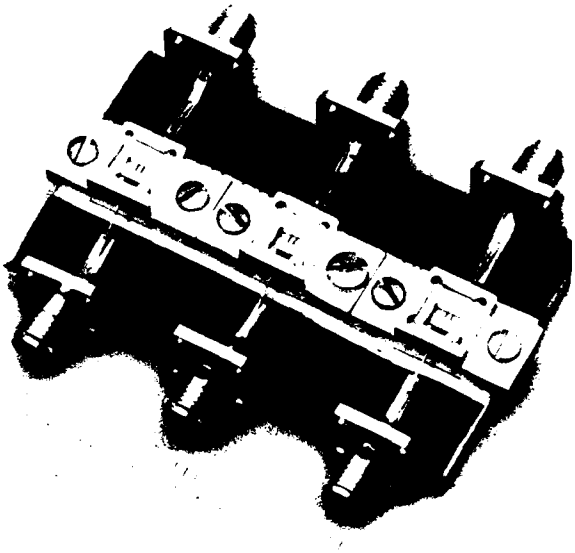


Figure 2. RF test mount for HBARs (High Overtone Bulk Acoustic Resonator) which provides a common heat sink.

Figure 3. Environmental oven for temperature stability tests. Heat exchanger for the dry nitrogen input is shown in the rear. Resonators are housed in small box at top of tier.

Results and Discussion

Figure 4 shows the frequency shift with temperature of a 2 GHz compensated lithium tantalate HBAR. As can be seen in the figure the turnover temperature is about 70°C and the frequency shift is small from 60°C to 80°C. This range is convenient in that it provides for temperature stabilization with simple temperature controls. Figure 5 shows the percent frequency shift for three HBARs over the same temperature range at three different frequencies: 1.0 GHz, 2.0 GHz and 2.2 GHz. The frequency resolution of these measurements is 0.001% and stability over about 10°C to within this resolution is indicated. Further tests of this cut will be undertaken to determine the optimum loaded Q and vibration sensitivity.

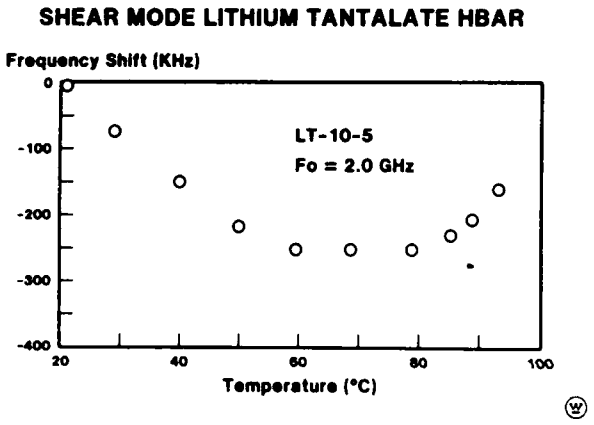


Figure 4. Frequency shift as a function of temperature for a shear mode $LiTaO_3$ HBAR (High Overtone Bulk Acoustic Resonators).

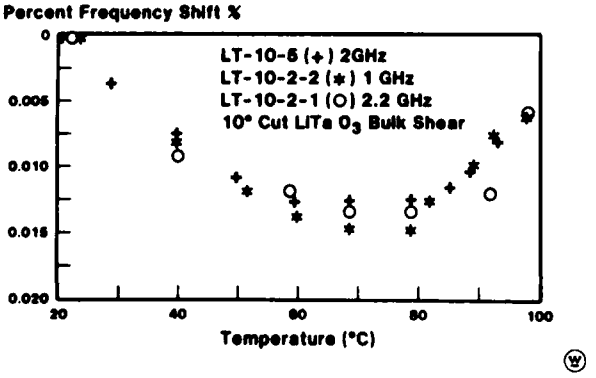


Figure 5. Percent of frequency shift vs. temperature for three shear mode HBARs (High Overtone Bulk Acoustic Resonators) at three different frequencies.

We are pleased to acknowledge G. A. Ferguson and D. H. Watt for their assistance in film preparation and resonator fabricaton. We would like to thank G. B. Draper for his help in device measurement.

References

1. S. V. Krishnaswamy and B. R. McAvoy, Appl. Phy. Lett 45(4), 15 August 1984.

2. S. V. Krishnaswamy, B. R. McAvoy and R. A. Moore, 38th Annual Frequency Control Symposium, Philadelphia, PA, May 30, 31, June 1, 1984.

3. J. Murphy and M. M. Gad, 1978 IEEE Ultrasonics Sympsoium Proceedings, Cherry Hill, NJ (IEEE, New York, 1978), p. 172.

4. J. Murphy, B. R. McAvoy and R. B. Feldman, 1979 IEEE Ultrasonics Symposium Proceedings, New Orleans, LA CIEEE, New York, 1979, p. 157.

5. J. T. Haynes, H. L. Salvo, Jr., R. A. Moore, and B. R. McAvoy in Proc. 37th Annual Symp. on Frequency Control, Philadelphia (IEEE, New York, 1983), p. 87.

6. H. Rossman and J. T. Haynes in Proc. 37th Annual Symp. on Frequency Control, Philadelphia (IEEE, New York, 1983), p. 272.

7. S. V. Krishnaswamy, B. R. McAvoy, W. J. Takei, and R. A. Moore in 1982 IEEE Ultrasonics Symposium Proc., San Diego (IEEE, New York, 1982), pg. 47G.

8. S. V. Krisnaswamy, B. R. McAvoy and R. A. Moore in 1982 IEEE Ultrasonics Symposium Proc., Atlanta (IEEE, New York, 1983), pg. 531.

A NOVEL MINIATURE ZT-CUT RESONATOR

J.W. Hermann

Swiss Center for Electronics and Microtechnology, Inc.
-Research and Development-
Neuchâtel, Switzerland

Abstract

A novel miniature ZT-cut resonator is described. It consists of a vibrational section, composed of two resonant plates connected by a resonant arm, and of a supporting section. The latter includes a clamping area where movement and stresses are evanescent, so that the crystal can be glued with epoxy onto a tiny pedestal.

Quartz wafers, comprising several tens of resonators, are batch-processed by using photolithography and chemical milling.

The preferred frequency range of these miniature resonators extends from 1 to 5 MHz. Their main features may be summarized as follows: very good cubic frequency-temperature characteristics with adjustable inflection point, small sensitivities to temperature transients and to acceleration, as well as high shock resistance.

Introduction

In 1979 a new quartz crystal cut, called the ZT-cut, was described[1]. This orientation, shown in Fig. 1, is suitable for width-extensional rectangular resonators with a unique combination of properties, namely:

- A cubic frequency-temperature curve, with a relatively low third-order coefficient (about $55 \times 10^{-12}/°C^3$ for an inflection point at room temperature).
- The possibility of shifting the inflection temperature within a large range, depending upon the application.
- A very small sensitivity of the first-order temperature coefficient to the width to length ratio, in sharp contrast to GT-cut resonators.

The drawback of contour mode rectangular resonators is the necessity of using supporting wires soldered onto the plate, which has an adverse effect on aging properties. Furthermore, the suspension system has to be very carefully designed to avoid any unwanted resonances, not an easy task at frequencies above 1 MHz.

The purpose of the novel design described in this paper is to obtain a resonator which, while retaining all the advantages of the ZT-cut, can be clamped in an area where movement and stresses are evanescent.

Novel resonator

The simplest example of the new design is shown in Fig. 2. The vibrational section consists of two resonant plates connected by a resonant arm. A clamping zone is attached to the latter by a suspension arm. This supporting section, in dashed line in the figure, produces only a slight perturbation of the main movement, which is a very pure extension along the Y' axis, as shown in Fig. 3. This design is only applicable because of a particularity of the ZT-cut, namely the fact that both planar elastic constants γ_{12} and γ_{26} are very nearly zero. Consequently, the extension along Y' produces only a negligible coupling with either the extension along X' or with the surface shear. The frequency, as well as the first-order temperature coefficient, are therefore almost independent upon the width to length ratio of the plates, in the same way as shown previously for the rectangular ZT-cut resonator[1]. As for the second-order temperature coefficient -or in other words the inflection point of the cubic frequency-temperature curve- it depends mainly upon the width to length ratio of the plates, and to a lesser degree on the location of the resonant arm.

It is to be noted that the perturbation produced by the suspension arm induces a very small movement of the clamping area in the X' direction. This can be compensated for by adjusting the length of the resonant arm, so as to produce a slight bending of the entire resonator. With proper conditions it is possible to minimize the movement of the clamping zone, and thereby obtain high Q values.

Another point worth mentioning is that, for a given location of the resonant arm, the sensitivity of the first-order temperature coefficient to the length of that arm can be annulled.

The design presented is not the only conceivable, as far as more than one resonant arm may be used. Experiments have also been done with resonators symmetrical about the Y' axis, with two or three resonant arms side by side. Nevertheless, an important advantage of the solution shown here is that complete electrodes may be used on both sides of the resonator, without coupling significantly any unwanted mode. In fact, with this design, all other modes lie at least 40 dB below the main mode.

Production techniques

It is fortunate that ZT-cut wafers can be suitably etched by chemical milling, so that intricate shapes are easily realized by using this process. The etchant most experimented with is a solution of ammonium bifluoride, saturated at room temperature and heated at 85°C. Etching rates of about 30 µm/hour were measured. Various mixtures of ammonium fluoride and hydrofluoric acid were also evaluated, but without any significant advantage.

The processing steps, including electrode deposition and photolithography, are essentially the same as those previously decribed for several miniature resonators[2][3].

Fig. 4 shows a one inch square ZT-cut wafer, which accomodates 28 resonators as well as a test zone in the center. A close-up of a single resonator is shown in Fig. 5.

CH2186-0/85/0000-0375$1.00©1985IEEE

Any convenient evacuated encapsulation, such as a ceramic package sealed by a glass lid, can be used to house the resonators. The clamping area has to be glued onto a tiny pedestal, and the electrical connections are then made by thermo-compression bonding of small wires.

No frequency adjustments have yet been performed but it is expected that, under the right conditions, this could be done without affecting the frequency-temperature curve.

Experimental results

Main Properties

The experimental work has been centered on 2.1 MHz watch crystals. The main features of these resonators may be summarized as follows (typical values):

overall dimensions	:	$4x2x0.15$ mm^3
frequency	:	2.1 MHz $\pm$ $1x10^{-3}$ (unadjusted)
quality factor	:	300 to $400x10^3$
static capacitance	:	2 pF
motional capacitance	:	1 fF
first-order temperature coefficient	:	$-0.03\pm0.08x10^{-6}$/°C (no adjustment needed)
inflection temperature	:	23 $\pm$ 4°C
third-order temperature coefficient	:	55 $\pm$ $5x10^{-12}$/°C^3
aging at 28°C	:	$0.5x10^{-6}$/first year (without preaging)
temperature transients sensitivity	:	0.1 to 0.2 µs/°C
static acceleration sensitivity	:	$3x10^{-10}$/g
maximum static acceleration	:	$15x10^3$ g

Amplitude-Frequency Effect

The series resonance frequency was measured for several resonators as a function of the energy stored in the crystal. All values fit very well the single curve shown in Fig. 6. This same figure also gives the corresponding maximum strain, determined under the assumption that the movement of the resonant section is a pure sine. The breakage strain is found to be of the order of 3 to 4×10^{-3}.

Short Term Stability

The short term stability was determined in the frequency domain, as well as in the time domain, at the LPMO (laboratoire de Physique et de Métrologie des Oscillateurs) in Besançon, France. Fig. 7 shows the spectral density of frequency fluctuations, where the f^{-1} dependence is clearly obtained.

Beyond the half bandwidth, the f^{-3} slope is due to the filtering effect of the quartz resonator. The synthesizer noise power is also represented by the dashed line in the same figure. The experiment was performed with an energy stored in the crystal of about 1 µJ, and the corresponding Allan variance is computed to be $8x10^{-12}$.

The influence of the drive level upon the variance is given in Fig. 8. It can be seen that the frequency stability remains approximately constant above 2 µJ, even at levels where the amplitude-frequency effect becomes important. However, it appears that the variance increases to 1.4×10^{-11} at very low levels. No explanation has yet been found for this behavior.

Measurements in the time domain essentially confirmed the results obtained in the frequency domain.

Future work

The design presented offers several degrees of freedom, which are currently being investigated. On the other end, as already mentioned, other designs with two or even three resonant arms were also tested. It is expected that some characteristics of this type of resonators may be further improved, such as the frequency-temperature behavior and the acceleration sensitivity. As for that last point, sensitivities lower than 1.5×10^{-10}/g were recently measured.

The frequency range of these resonators extends up to 5 MHz, but this limit could possibly be increased by adopting a somewhat different geometry, based on overtone modes.

Conclusions

The miniature quartz resonators described in this paper present several desirable properties, namely: very good cubic frequency-temperature characteristics with adjustable inflection temperature, small sensitivities to temperature transients and to acceleration, as well as a high shock resistance. They could be easily mass-produced by using photolithography and chemical etching.

Their preferred frequency range, from 1 to 5 MHz, lies in a gap where few good resonators are available.

Acknowledgments

The author wishes to acknowledge many helpful discussions with C. Bourgeois, who devised the program used to compute the movement of the resonator.

He also wishes to thank M. Planat, G. Marianneau and J.-J. Gagnepain, from the Laboratoire de Physique et de Métrologie des Oscillateurs du CNRS in Besançon, for the short term stability characterization.

Many thanks to F. Crevoisier for his assistance in resonator fabrication.

This work was supported in part by Rolex Watch Company in Geneva, whose encouragements are well appreciated.

References

1. J. Hermann and C. Bourgeois: "A New Quartz-Crystal Cut for Contour Mode Resonators", Proceedings 33rd Annual Frequency Control Symposium, May-June 1979, pp. 255-262.

2. J. Staudte: "Subminiature Quartz Tuning Fork Resonator", Proceedings 27th Annual Frequency Control Symposium, June 1973, pp. 50-54.

3. H. Kawashima and al.: "New Frequency Characteristics of Miniaturized GT-Cut Quartz Resonators", Proceedings 34th Annual Frequency Control Symposium, May 1980, pp. 131-139.

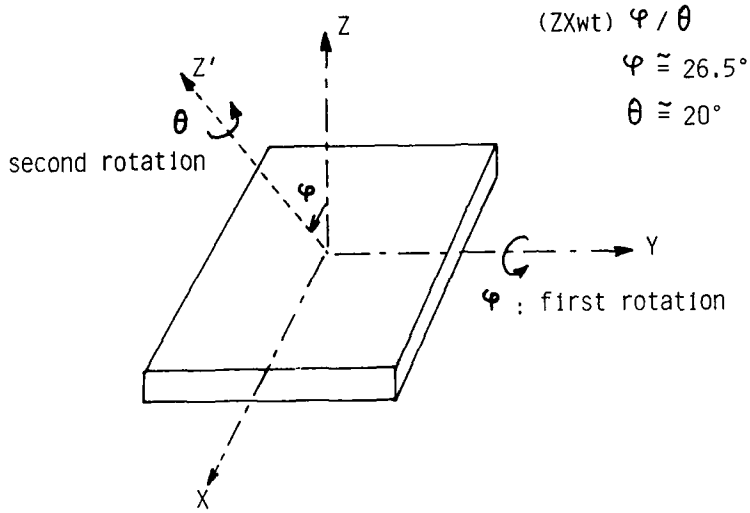


Fig. 1 ZT-Cut Quartz Plate

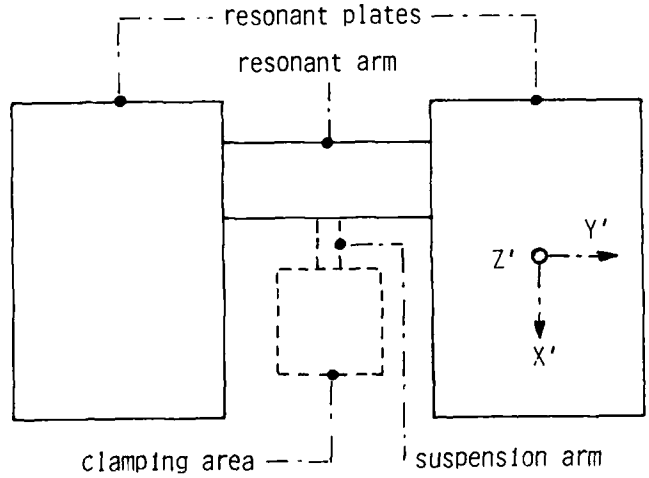


Fig. 2 New ZT-Cut Resonator

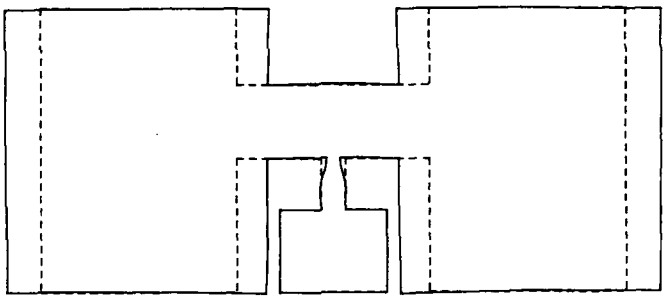


Fig. 3 Mode of Vibration

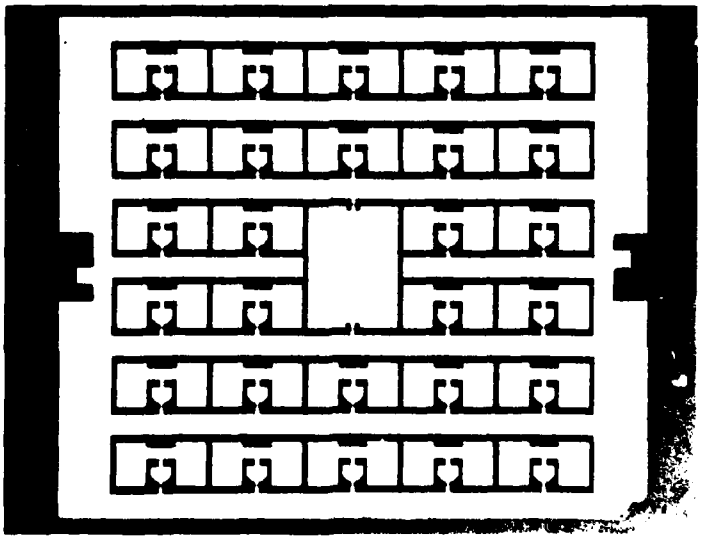


Fig. 4 Quartz wafer

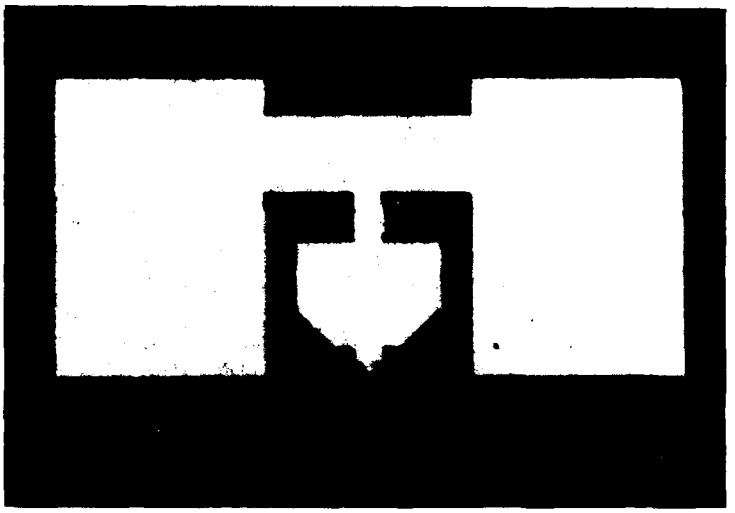


Fig. 5 Close-up of a Resonator

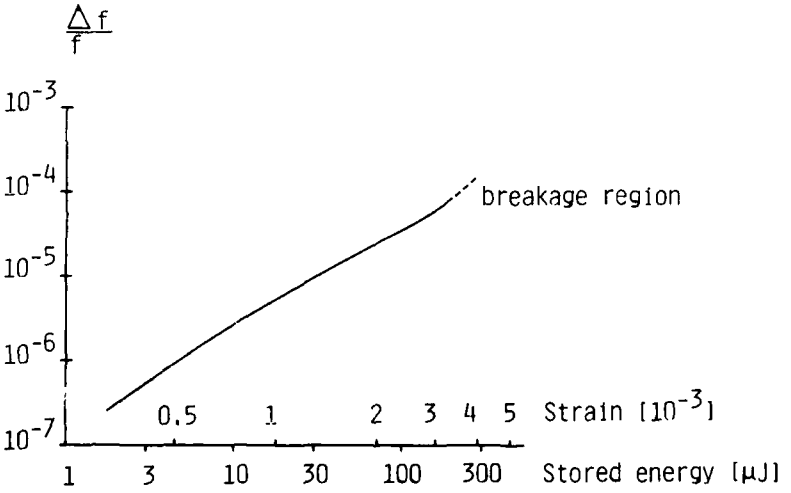


Fig. 6 Frequency-Amplitude Effect

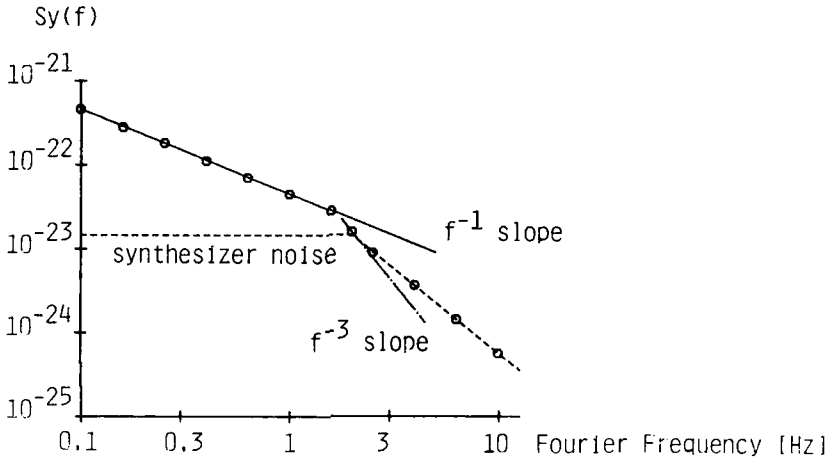


Fig. 7 Spectral Density of Frequency Fluctuations

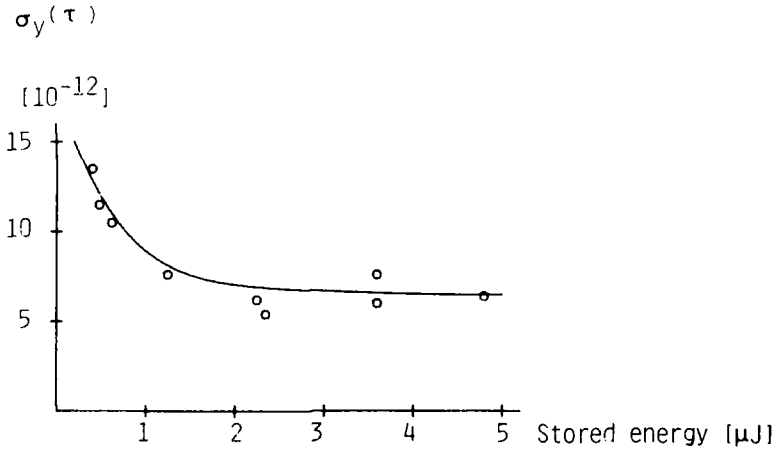


Fig. 8 Frequency Stability versus Drive Level

Characteristics of a Quartz Crystal Tuning Fork with shortened Arm Length for High Frequencies.

Masanobu OKAZAKI*, Hiroshi TOHMA*
and
Yoshiro TOMIKAWA**

* NDK (Nihon Dempa Kogyo Co., Ltd.)
1-21-2 Nishihara, Shibuyaku, Tokyo, Japan
** Department of Electrical Engineering, Yamagata University, Yonezawa, Yamagata, Japan

Summary

Theoretical and experimental characteristics of a quartz crystal tuning fork with shortened arm length for use in a high frequency range from about 50KHz to 500KHz are reported.

The first part of this paper describes the theoretical calculation by finite element analysis. According to the analysis, it was expected that if width of base-end is widened to minimize the displacement, good characteristics might be achieved.

The second part outlines experimental result. As expected, leakage of vibration energy at base-end is minimized by widening the base-end good electrical characteristics are achieved.

Introduction

Because of its outstanding capability against shock and vibration, a 32.768KHz quartz crystal tuning fork has been widely adopted for a wristwatch application as a stabel time base.(1)(2)Recent tendency, however, reports that many other applications than that of the wristwatch require such samll crystal units having excellent ant-environmental characteristics but for higher frequency than 32.768KHz.

To give an answer to such requirement, one of authers has previously analyzed and reported it with the range of about 50KHz to 500KHz of shortening arms length of tuning fork by the finite element method. (3) According to the analysis, it was found that displacement of base-end of tuning fork became large when arms of tuning fork was reduced for obtaining higher resonant frequency. It was also reported that when tuning fork having large displacement at base-end was mounted on a hermetic base, Q factor was worsened due to leakage of vibration energy to hermetic base through base-end of mounted portion together with suppression of vibration. Therefore, it was expected that if width of base-end is widen to minimize the displacement while arms of tuning fork is reduced to get higher target frequency, good characteristics might be achived. However, it was just theoretical examination but not the practicaly examined result.

This paper describes experimental result of most suitable condition of blank shape to offer high frequency tuning fork crystal unit. At first, electrical characteristics change is experimentaly confirmed when length of arms of tuning fork is gradually shortened to bring up resonant frequency from standard 32.768KHz using normal blank dimension at base-end. As expected, leakage of vibration energy at base-end is anavoidable as far as width of base-end is kept as the standard when resonant frequency is getting higher and higher. In addition to the above-mentiond, experiments to determine adequate width of base-end is performed at some fixed high frequency condition. And it is found that there is suitable width of base-end related to the arms length, which authers already have expected by the finite element method.

According to the experiments, it is confirmed that tuning fork widned at base-end offers higher Q value than that of not widened when compared with the same length of arms. For example, the Q value of widened base tuning fork at 110KHz shows upto 30×10^4, while normal base tuning fork is measured at 15×10^4 . This means that leakage of vibration energy through base-end can be minimized by widening width of base-end of tuning fork.

Theoretical analysis

To increase for resonant frequency of tuning fork crystal, the arms length of tuning fork must be shortened. Fig.1 shows the calculation result by the finite element method for the relative displacement at base-end against the top of the arm. In this figure each lines are obtained by shortening the length of arms from normal type tuning fork blank. According to the calculation, relative displacement at base-end is getting large when the length of arm is becoming short. Therefor it is expected that for high frequency type tuning fork, characteristics might be changed by supporting the base-end on the holder, if it is cared only reducting the length of arms.

The model of tuning fork blank used for calculation is shown in Fig.2. When calculating, following points were taken into consideration.

(1) The length of ℓ is chosen as assumption.

(2) Very small change of length at one arm are given.

(3) Resonant frequency for both case of free base-end and fixed are obtained.

Fig.3 shows the calculation result. The vertical axis means resonant frequency difference Δf between base-end free and base-end clamped. The horizontal axis means the resonant frequency normalized by the resonant frequency at minimum Δf in the case of base-end clamped. It is desirable to minimize the Δf and to increase the secondary coefficient of parabolic curve. The minimum value of Δf is obtained when resonant frequencies of each arms are agreed precisely and are vibrated ideally.

CH2186-0/85/0000-0381$1.00©1985IEEE

The minimum Δf at each arms are protted on Fig.4 (a). Fig.4 (b) shows the secondary coefficient of Δf curves at each length of arm. According to the Fig.4. (a), 2 to 3mm length of arm is not practical, because minimum value of Δf becomes too big. Further, in the case of less than 2mm length of arm, there are difficulties of frequency adjustment to resonant point because secondary coefficient of Δf curve becomes too small. Especially, in the case of 1.7mm arm length, it is very difficult to make a judgement of mechanical balance of arm. These were main reasons why normal type tuning fork blank could not go higher frequency than 50KHz.

Besides, one of authers expected that by widening base-end, the displacement of base-end may be reduced, using the finite element method. Fig.5 shows a model of tuning fork blank for calculation. Solid line shows the shape of widened base-end while dotted line shows reduced one.

Fig.6 shows the calculation result of relative displacement by changing width of base-end. According to the Fig.6, zero relative displacement at base-end may be possible in case the ℓw is determined as 0.11mm at first-mode. Besides, at third-mode, zero displacement will be obtained by reducing the width at base-end.

Experiment

To confirm the expectation, pilot production of high frequency tuning fork crystal driven by first mode are made and was compared with our calculation result. Fig.7 shows the one we produced for our study. Cut angle of crystal blank is +2°X and electrode material is gold. The blank is mounted on the base by epoxy adhesive.

Fig.8 shows the actual dimension of blank which has different width of base-end. Namely, such as length of arm, width of arm are kept as constant. ℓw of each specimen are 0, 0.09, 0.22 and 0.38.

Mean value of equivalent constants measured for each specimen are shown in Table 1. According to the table, motional inductance, motional capacitance and sunt capacitance are little changed by changing width of base-end. However, Q value changed very much.

Fig.9 shows measured Q value of each specimen. The sample No.2 and No.3 shows higher Q value and distribution of it is small. So these are called excellent crystal. Namely, it is considered that, because of 0.1 to 0.2mm of ℓw, displacement of base-end is small and does not effecte so much to the blank by mounting. These are resulted to acheive good characteristics. The most suitable value of ℓw obtained by experiment is very close to the value calculated.

In addition we found minimum value of Δf and it's secondary coefficient of widened type blank are better than normal type. Fig.10 shows measured value of Δf. No.1 is measured value of normal type. And it is not practical due to large Δf, since Δf of No.2 and No.3 are nearly equal to zero.

To compair with Fig.4, minimum Δf and secondary coefficient of Δf curve are plotted. Fig.11 (a) shows minimum values of Δf for each width. Fig.11 (b) shows the secondary coefficient of Δf curve for them. By widening of both edge of base-end as 0.1 to 0.2mm, Δf becomes nearly zero, and at the same time, secondary coefficient of Δf becomes large. In this case, it is easy to make a judgement of dimensional balance of arm. Therefore, after adjustment of balance, excelent characteristics can be realized.

Conclusion

Investigations on characteristics of a tuning fork with shortened arm length for use in high frequencies allow the following conclusions :

(1) Characteristics previously obtained by finite element analysis, that the shortened arm length tuning fork with a wide base portion shows a high quality factor under widening the base-end, are verified experimentally.

(2) It is also found that judgement of the balance can be easily made in this type of tuning fork.

(3) A high frequency tuning fork crystal with high quality factor is realized by chosing ℓw as 0.1 to 0.2mm.

References

(1) H. Yoda : Proc. 28th Ann. Frequency Control Symp. (1974)

(2) M. Konno et al : Proc. 30th Ann. Frequency Control Symp. (1976)

(3) K. Sato, Y. Tomikawa and M. Konno : Paper of Technical Group on SU78-26 IECE Japan (1978) (in Japanese)

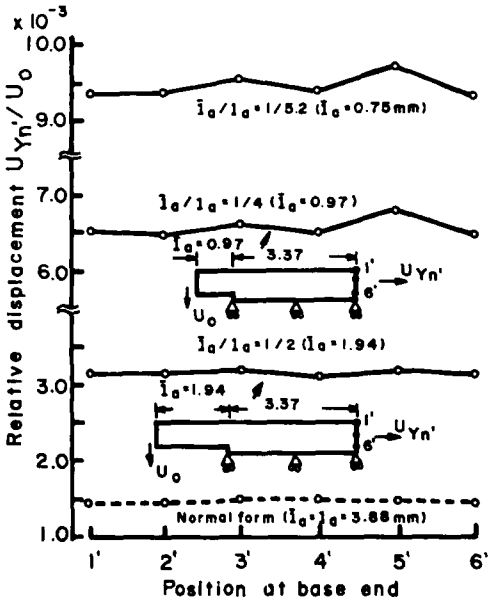


Fig.1 Relative displacement at base-end.

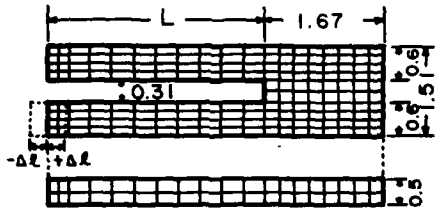


Fig.2 Model of normal tuning fork blank for calculation.

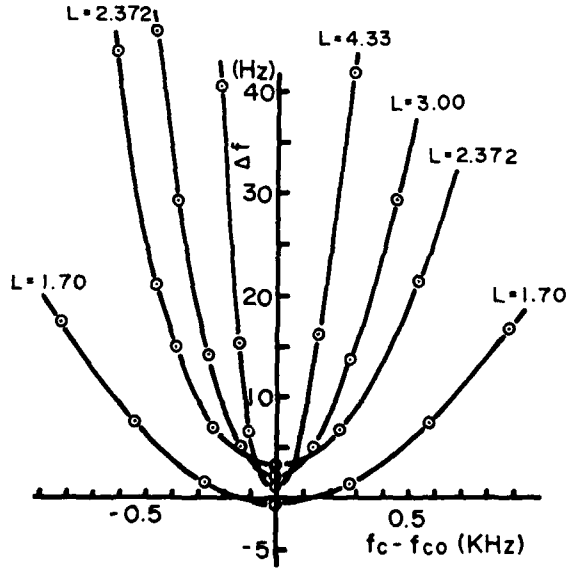
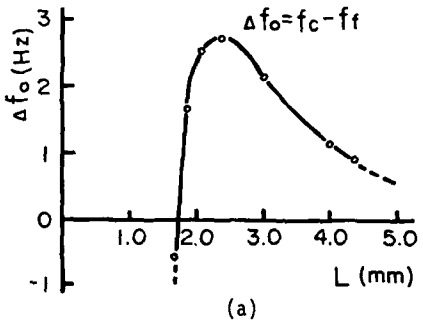
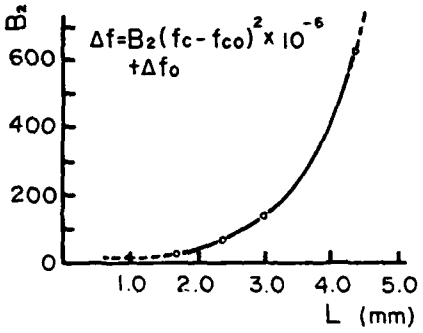


Fig.3 Δf curve characteristics (u-form) analyzed by finite element method.



(a)



(b)

Fig.4 (a) Minimum Δf at each length of arm.

(b) Secondary coefficient of Δf curves at each length of arm.

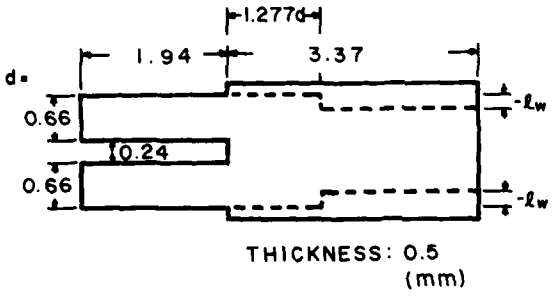


Fig.5 Actual dimensions of tuning fork blank for calculation.

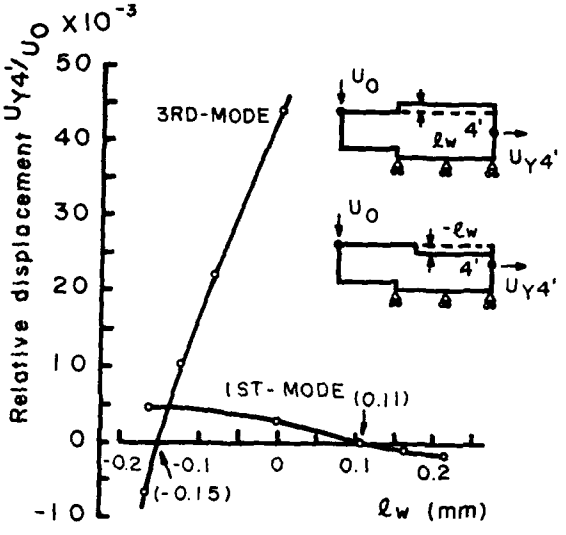
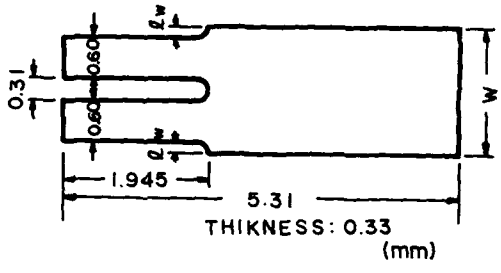


Fig.6 Relative displacement of the first and third modes.



SAMPLE NUMBER	ℓ_w (mm)
No. 1	0
No. 2	0.09
No. 3	0.22
No. 4	0.38

Fig.8 Dimensions of tuning fork for experiment

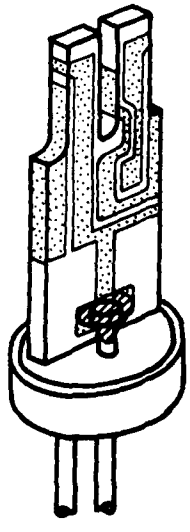


Fig.7 Quartz crystal tuning fork with widened base-end.

SAMPLE NUMBER	FREQ. (KHz)	Rs (Kohm)	Q x10^4	L_1 x10^3 (H)	C_1 x10^{-3} (pF)	Co (pF)	C_o/C_1
No. 1	101	4.9	19.6	1.46	1.68	1.38	891
No. 2	107	3.2	30.3	1.45	1.52	1.38	904
No. 3	110	3.3	30.6	1.48	1.41	1.40	989
No. 4	114	3.9	27.1	1.46	1.33	1.36	1186

Tabel 1 Mean value of equivalent constants.

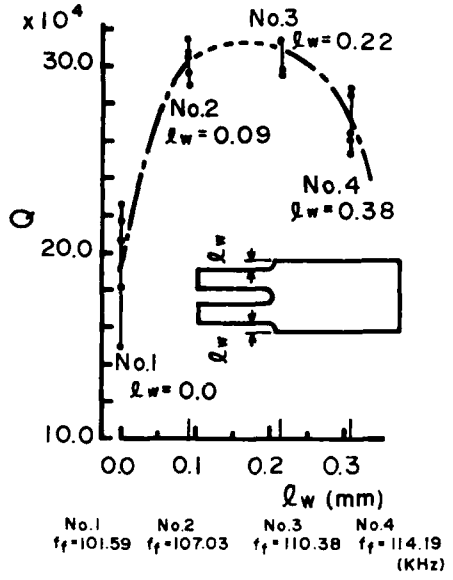


Fig.9 Quality factor for each specimen.

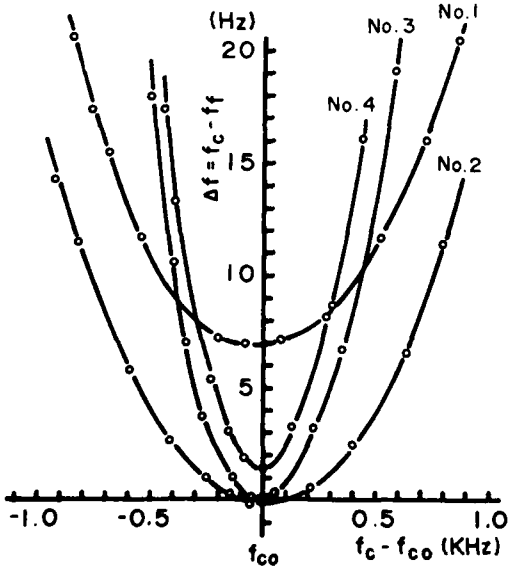


Fig.10 Measured curve of Δf for each specimen.

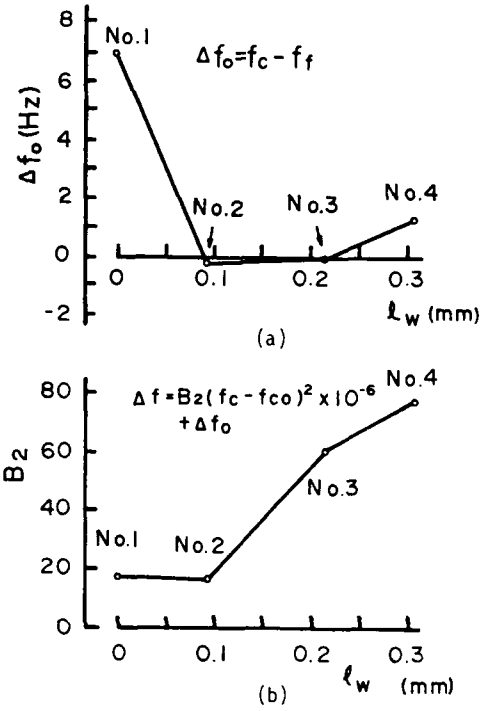


Fig.11 (a) Minimum Δf at each specimen.

(b) Secondary coefficient of Δf curves at each specimen.

Analysis of Trapped Energy Resonators with Tabs

Hitoshi Sekimoto
Tokyo Metropolitan University, Tokyo, Japan.

Hozumi Nakata and Masaaki Miura
Nihon Dempa Kogyo Co., Ltd., Tokyo, Japan.

Abstract

It is well known that tab electrodes brought out to plate edges for electrical connection have a significant effect on trapping characteristics of trapped energy resonators. The presence of tab can be interpreted as a perturbation of acoustical parameters in the unelectroded region surrounding central electrodes on a plate. Consequently, a sufficiently wide tab will produce a reduction in the resonant frequency and mode strength of a mode. In practice such a phenomenon has been observed in spurious modes of many overtone resonators where relatively wide tab electrodes are usually used. However, so far, theoretical papers on that problem have never been published.

This paper presents a simple theoretical model for analyzing effects of tab electrodes on trapped energy resonators with rectangular electrodes. A tab of strip type extending along the Z' (,or X) axis on an AT-cut rectangular plate is treated. The model is based on the mode-matching method, so that it can be applied to modes decaying exponentially and also modes with a sinusoidal variation along the tab. Equations which determine the resonant frequencies, mode shapes and mode strength are derived.

The calculated results of resonant frequencies and mode strength for AT-cut third overtone resonators are compared with measured values. It is shown from the comparison between experiment and theory that the model predicts well the strong acoustical effect of tab electrodes on anharmonic spurious modes with phase changes in the tab direction, and little effect on the main and other spurious modes.

I. Introduction

Trapped energy resonators of AT-cut quartz plates have been widely used in frequency selective circuits because of their high mechanical Q and high stability. It is well known that the configuration of partial electrodes on resonators has a significant effect on the resonant frequencies, mode shapes and mode strength. A number of theoretical studies of the effect of electrode thickness and lateral dimensions have been presented in the literature. The resonators with rectangular or circular electrodes were treated by Tiersten[1], Nakamura et al[2] and the authors [3],[4]. But their analyses are restricted to resonators without tab electrodes.

The resonators usually have two electroded regions: a region of central electrodes for exciting thickness-shear vibrations, and a region of tab electrodes brought out to the plate edges for electrical connection. The acoustical effect of the tab region, in which mass loading is one-half that in the central electroded region, was qualitatively discussed by Horton and Smythe [5]. A very narrow tab has little effect. If a tab is sufficiently wide, however, a portion of vibratory energy having been confined to the central electrode area can fringe out, or propagate into this tab region and ,therefore, the tab will produce a reduction in the resonant frequency and mode strength of a mode. In practice such a phenomenon has been observed in anharmonic spurious modes of many overtone resonators where relatively wide tab electrodes are usually used, and thus there is a need to theoretically predict the influence of tab on spurious characteristics for the purpose of designing overtone resonators. But, to this authors' knowledge, theoretical papers on that problem have never been published.

This paper presents a simple theoretical model for analyzing the effects of tab electrodes on trapped energy resonators with rectangular electrodes. A tab of strip type extending along the Z' (,or X) axis on an AT-cut rectangular plate is treated. The model, which is an extension of the previous one [3], is based on the mode-matching method, and so can be applied to modes decaying exponentially and also modes with a sinusoidal variation along the tab. Experimental measurements of resonant frequencies and mode strength for anharmonic modes on third overtone AT-cut quartz resonators are provided to substantiate the theory's validity. Calculated and measured results show that the tabs have a strong acoustical effect on anharmonic spurious modes with phase changes in the tab direction, and little effect on the main and other spurious modes.

II. Approximate Equations

A schematic diagram of a trapped energy resonator with central rectangular (length:$2l$, width:$2w$) electrodes and strip tab (length:l_t, width:$2w_t$) electrodes on a rectangular (length:$2l_q$, width:$2w_q$) AT-cut quartz plate is shown in Fig.1 where the x_1, x_2 and x_3 coordinates are chosen as the X(or Z'), Y' and Z'(or X) axes, respectively. This resonator has three separate cutoff frequencies for the central electroded, tab electroded and surrounding unelectroded regions, designated ω_o, ω_o and ω_o'. A basic assumption for the present analysis of trapped energy resonators is that the x_2 dependence of displacements is obtained from the pure thickness-shear solutions at cutoff frequencies. In addition, a two-dimensional scalar wave equation in each region is introduced in order to express displacement fields in the x_1x_3 plane near cutoff.

The thickness-shear displacement u_1 for the central electroded region in the vicinity of a cutoff frequency ω_o is approximated as

$$u_1 = u_o(x_2)\, U(x_1, x_3) \ , \quad u_o = \sin q_o x_2 \tag{1}$$

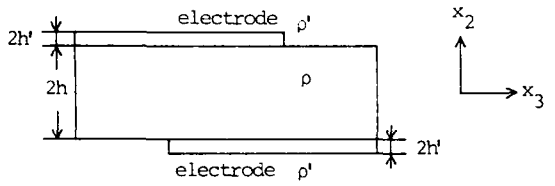
where q_o is the wavenumber in the x_2 direction at ω_o for pure thickness-shear vibrations of a fully electroded plate, and U represents the amplitude in the x_1x_3 plane.

The variation of U due to frequency must obey the dispersion relation. For straight-crested thickness waves propagating at an angle ψ to the x_1 axis, the dispersion relation near cutoff is given with sufficient accuracy as [6]

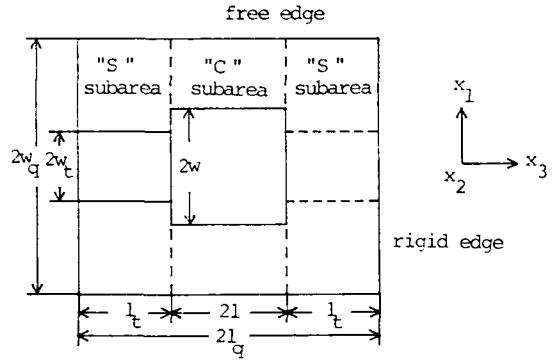
$$\omega^2/\omega_o^2 - 1 = (\alpha_1 \cos^2\psi + \alpha_3 \cos^2\psi)\, k_\psi^2 h^2 \tag{2}$$

where $2h$ is the plate thickness, k_ψ is the propagation wavenumber, and a_i(i=1 or 3) is the curvature of a dispersion curve at ω_o for thickness-shear waves propagating along the x_i axis, in which the influence of anisotropy, piezoelectric stiffening, electrode mass loading and electrical short-circuiting is included.

CH2186-0/85/0000-0386$1.00©1985IEEE



(a)



(b)

Fig.1. A rectangular trapped energy resonator with rectangular central and strip tab electrodes. (a) Side view. (b) Top view.

Accordingly, we assume that U obeys a wave equation:

$$G_{11}U_{,11} + G_{33}U_{,33} + H_o(\omega^2 - \omega_o^2)U = 0 \tag{3}$$

$$H_o = h^{-1}\left[\int_{-h}^{h} \rho\, u_o^2(x_2)\, dx_2 + 4\rho' h' u_o^2(h)\right] \tag{4}$$

$$G_{ii} = \alpha_i H_o \omega_o^2 h^2 \tag{5}$$

where ρ is the plate mass density, ρ' and $2h'$ are the mass density and thickness of the electrode plating, respectively. Here, H_o and G_{ii} may be considered to be the effective mass density and stiffness constant for the present approximation.

For the tab and surrounding regions, we also employ wave equations in the same form as Eq.(3) with the proper values of a_i and H_o. In this paper, a_i for the surrounding unelectroded region as well as the central electroded region is numerically calculated by using the published method [6], and one for the tab electroded region is approximated to be equal to that in the unelectroded region. H_o are taken to be

$$H_o = h^{-1}\left[\int_{-h}^{h} \rho\, u_o^{t\,2}(x_2)\, dx_2 + 2\rho' h' u_o^{t\,2}(\pm h)\right],$$

$$u_o^t = \cos q_o^t(x_2 \pm h) / \sin q_o^t h \quad , \quad \pm l \le x_3 \le \pm l_q \tag{6}$$

for the tab region, where q_o^t is the wavenumber in the x_2 direction at ω_o^t for pure thickness vibrations of a plate possessing an electrode on one surface only, and

$$H_o = \rho \tag{7}$$

for surrounding region.

It is further assumed that the mechanical boundary conditions to be satisfied at a discontinuity normal to x_j between adjacent regions are the continuity of

$$U \text{ and } G_{jj}U_{,j} \tag{8}$$

and the boundary conditions at plate edges are

$$U=0 \text{ at the rigidly mounted edge normal to } x_3 \tag{9a}$$

$$U_{,1}=0 \text{ at the free edge normal to } x_1 \tag{9b}$$

Here, we have neglected a difference in the x_2 dependence of displacements between adjacent regions.

III. Mode Matching

The trapped energy resonator with tab electrodes is analyzed using the mode-matching procedure. The authors have applied this procedure to an analysis of resonators without tabs, and proved its usefulness for a precise evaluation of resonant frequencies, motional inductances and mode fields [3]. In this paper, the previous model is extended.

To utilize the mode-matching procedure, the plane of a rectangular plate is divided into a cascade connection of three subareas along the x_3 axis, which are denoted as "C" and "S" subareas in Fig.1-(b). Each subarea contains a strip tab region or a central rectangular electroded region, sandwiched between two unelectroded regions. Two dimensional thickness-shear wave solution in the x_1x_3 plane for each subarea is obtained from Eqs.(3)-(9) as a sum of guided modes by the parallel plate edges with the x_3 axis. Both the transverse fields along the x_1 direction and the longitudinal propagation wavenumbers in the x_3 direction of guided modes can be determined by the well-known transverse-resonance technique. Since we are interested in the modes coupling to a driving voltage, only the even solutions in both x_1 and x_3 will be part of the analysis.

Accordingly, the solution U^c for the central subarea may be written in the form

$$U^c = \sum_{n=1}^{\infty} a_n^c u_n^c(x_1) \cos \gamma_n^c x_3 \qquad x_3 \le l \tag{10}$$

$$\int_{-w_q}^{w_q} u_n^c(x_1)\, G_{33}^c(x_1)\, u_m^c(x_1)\, dx_1 = \delta_{nm} \tag{11}$$

where

$$G_{33}^c(x_1) = \begin{cases} \bar{G}_{33} & |x_1| \le w \\ G'_{33} & w \le |x_1| \le w_q \end{cases} \tag{12}$$

in which $\bar{G}_{33}$ and G'_{33} are effective stiffness constants in the central electroded and unelectroded regions, respectively. a_n^c, γ_n^c and u_n^c are the arbitrary constant, propagation constant in the x_3 direction and symmetric transverse eigenfunction of the n-th guided mode, respectively.

Similarly, the solution U^s for each subarea containing a tab is

$$U^s = \sum_{n=1}^{\infty} \pm a_n^s u_n^s(x_1) \sin \gamma_n^s(x_3 \mp l_q) \qquad \pm l \le x_3 \le \pm l_q \tag{13}$$

$$\int_{-w_q}^{w_q} u_n^s(x_1)\, G_{33}^s(x_1)\, u_m^s(x_1)\, dx_1 = \delta_{nm} \tag{14}$$

where

$$G_{33}^{s}(x_1)=\begin{cases} G_{33}^{t} & |x_1|\leq w_t \\ G_{33}' & w_t\leq|x_1|\leq w_q \end{cases} \qquad (15)$$

in which G_{33}^{t} is effective stiffness constant in the tab electroded region. a_n^s, γ_n^s and u_n^s are the arbitrary constant, propagation constant in the x_3 direction and symmetric eigenfunction of the n-th guided mode, respectively.

Such a generalized Fourier series expansion is essential to an analysis of the effects of strip tab electrodes. There are innumerable guided modes: a finite number of lower-order modes which decay exponentially from an electrode edge in the x_1 direction, and a infinite number of higher-order modes which vary sinusoidally. From Eq.(8), the solutions in three subareas must satisfy the following continuity conditions at the discontinuities $x_3=\pm 1$ between adjacent subareas

$$U^c = U^s \quad \text{and} \quad G_{33}^{c}(x_1)\,U_{,3}^{c} = G_{33}^{s}(x_1)\,U_{,3}^{s} \qquad (16)$$

These conditions may be transformed, according to the mode-matching method, into

$$\int_{-w_q}^{w_q} G_{33}^{s}(x_1)\,[U^c - U^s]\,u_m^s(x_1)\,dx_1 = 0 \ ,$$

$$\int_{-w_q}^{w_q} [G_{33}^{c}(x_1)\,U_{,3}^{c} - G_{33}^{s}(x_1)\,U_{,3}^{s}]\,u_m^s(x_1)\,dx_1 = 0$$
$$x_3=\pm 1 \ , \qquad m=1,2,\ldots \qquad (17)$$

Substitution of Eqs.(10) and (13) with the orthogonal relation Eq.(14) of u_m^s into Eq.(17) yields

$$\sum_{n=1}^{\infty} a_n^c \left[\gamma_n^c \tan\gamma_n^c l \int_{-w_q}^{w_q} u_n^c(x_1)\,G_{33}^{c}(x_1)\,u_m^s(x_1)\,dx_1 \right.$$
$$\left. - \gamma_m^s \cot\gamma_m^s l_t \int_{-w_q}^{w_q} u_n^c(x_1)\,G_{33}^{c}(x_1)\,u_m^s(x_1)\,dx_1 \right] = 0$$
$$m=1,2,\ldots \qquad (18)$$

$$a_m^s = \sum_{n=1}^{\infty} a_n^c \cos\gamma_n^c l_t \int_{-w_q}^{w_q} u_n^c(x_1)\,G_{33}^{c}(x_1)\,u_m^c(x_1)\,dx_1$$
$$/ \sin\gamma_m^s l_t \qquad (19)$$

If N terms are kept in both series of a_n^c and a_m^s, N linear homogeneous equations will be obtained from Eq.(18). It is then possible to compute approximate resonant frequencies and displacement fields for the resonances.

When the vibrational field and the surface charge density on the electrode are known, the equivalent motional inductance that characterizes the relative strength of each resonant mode can be evaluated. Since it is reasonable to assume in accordance with the approximation in Eq.(1) that the charge distribution on the central electrode is proportional to U^c, the present expression for the equivalent inductance L is

$$L = \int_V \rho u_1^2\,dV \Big/ \left(\int_S Q_o U^c\,dS\right)^2 \qquad (20)$$

where V is the total volume including the electrodes, S is the area of central electrode, and Q_o represents the free surface charge density on the central electrode associated with the pure thickness-shear vibration at ω_o.

In the above model, the tab can be interpreted as a perturbation of acoustical parameters in the surrounding unelectroded region for a trapped energy resonator with no tab electrodes because a cutoff frequency in the tab region is lower than that in the unelectroded region. So, an anharmonic spurious mode with phase changes in the tab direction, of unperturbed trapped modes, may be strongly influenced by the perturbation. Therefore, the resonant frequency and strength will decrease greatly. Moreover, an additional untrapped resonance may occur at a frequency between the cutoff frequencies of the tab and surrounding regions. However, this new resonance caused by the existence of tabs may be very weak because most of the acoustical energy is confined in the tab region.

IV. Experimental Results

Measurements of the resonant frequencies, the motional inductance of main resonance and the relative amplitudes of spurious resonances on resonators with tabs of strip type extending along the Z' axis were carried out as a function of the central electrode width 2w for the fixed central electrode area 4wl, tab X-width $2w_t$ and plate back R ($=2\rho'h'/\rho h$). Two groups of resonators were fabricated on circular AT-cut quartz plates (5 mm Φ), provided that the difference between the rectangular and circular plate configurations did not cause a change of vibrational fields in the vicinity of electroded

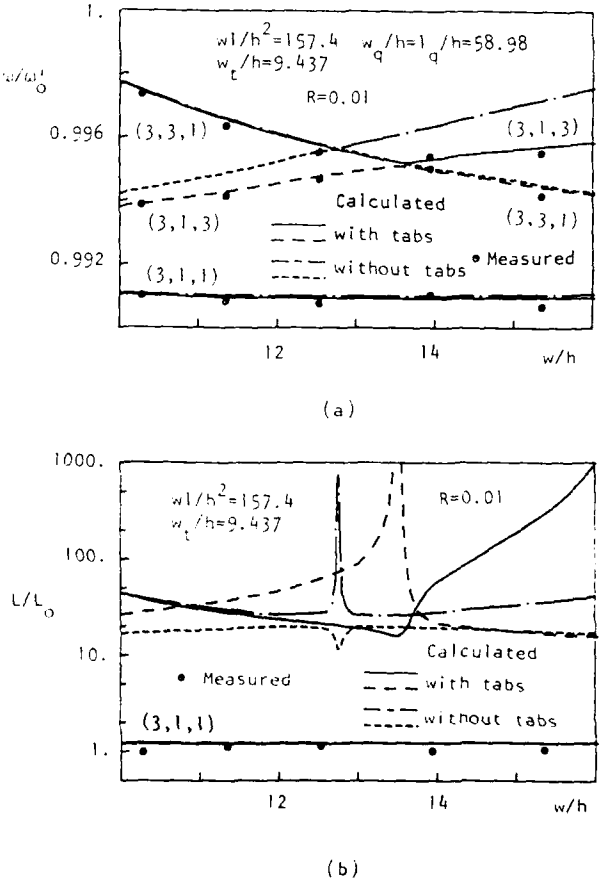


Fig.2. Measured and calculated results of (a)frequencies and (b)equivalent inductances as a function of central electrode width-to-plate thickness w/h for the first group, where $4wl=1.13mm^2$, $2w_t=0.8mm$, R=0.01 and the equivalent inductance of pure thickness-shear vibration $L_o=20mH$.

regions, and operated in the third overtone thickness-shear mode at about 59 MHz. Three modes could be distinguished among responses for each group, which were identified by the calculation of vibrational fields as a main trapped mode (3,1,1), and two anharmonic trapped spurious modes (3,3,1) and (3,1,3), where the n and m of mode parameter (3,n,m) correspond to the number of antinodes along the X and Z' axes, respectively.

The experimental results for the three modes were compared with the calculated results by both the model including and the model neglecting tabs. The measurements generally have agreed well with the prediction of the present theory. The comparison between measurement and theory showed that the tabs had a strong effect on the (3,1,3) mode but little effect on the (3,1,1) and (3,3,1) modes. The simple theoretical model accurately evaluated the large reduction of resonant frequencies for the anharmonic spurious mode (3,1,3) which is caused by tabs. Furthermore, the theory predicted well a change in the relative relation of frequencies and amplitudes between two anharmonic spurious modes with changes in the central electrode width. These are illustrated in Figs.2, 3 and 4 for the first group ($4wl=1.13mm^2$, $2w_t=0.8mm$, R=0.01) and Figs.5 and 6 for the second group ($4wl=1.1mm^2$, $2w_t=0.37mm$, R=0.0091) with the calculated results, in computation of which 15 to 20 terms in both series of a_n and a_m are employed.

Fig.2 shows the measured and calculated results of (a) resonant frequencies and (b) motional inductances for the first group, where $\omega_o/\omega_o'=.98976$, $\omega_o^t/\omega_o'=.99503$, and the equivalent inductance of pure thickness vibration $L_o=20mH$. The experimental points fit the theoretical curves including tabs quite well. The calculated results of motional inductance for two anharmonic spurious modes in Fig.2(b) were justified only qualitatively by the measurements of frequency-impedance characteristics. The resonators were inserted simply between the input and output of a network analyzer. The characteristics are plotted in Fig.3. We can see that the lower and higher spurious responses have interchanged their strengths at two values of w/h as predicted by the theory including tabs. Fig.4 shows the computed vibrational patterns along the $x_1=X$ and $x_3=Z'$ axes for anharmonic modes. It is apparent that these modes have been trapped in frequencies above ,as well as below, ω_o^t.

Figs.5 and 6 show the results for the second group where $\omega_o/\omega_o'=.99064$, $\omega_o^t/\omega_o'=.99547$, and $L_o=20.6mH$. Since the tab is narrower than that for the first group, its influence on the (3,1,3) mode is weaker. The good agreement between measurement and theory is observed again. In Fig.5, a dotted short line represents an calculated example of additional untrapped modes, in which the coupling with the (3,3,1) mode occurs near w/h=10. However, this resonance could not be measured. The analysis of vibrational field showed that most of the acoustical energy was confined in the tab electroded region.

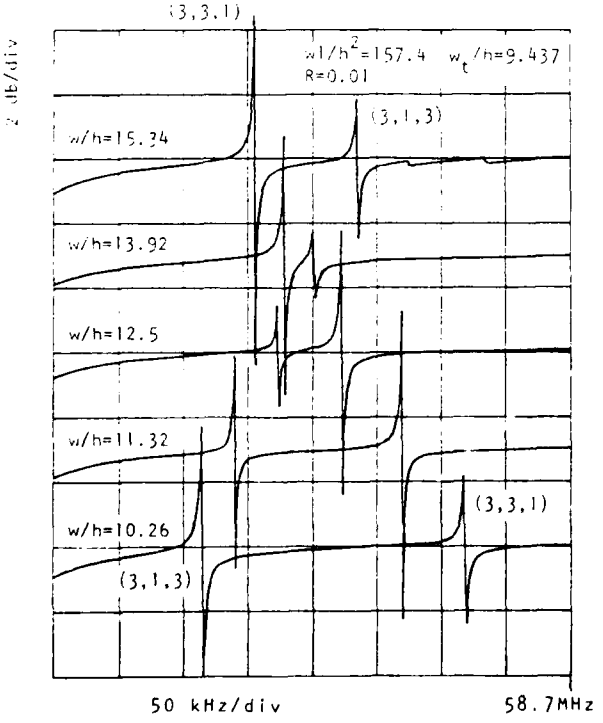


Fig.3. Measured frequency-impedance characteristics of spurious modes for the first group.

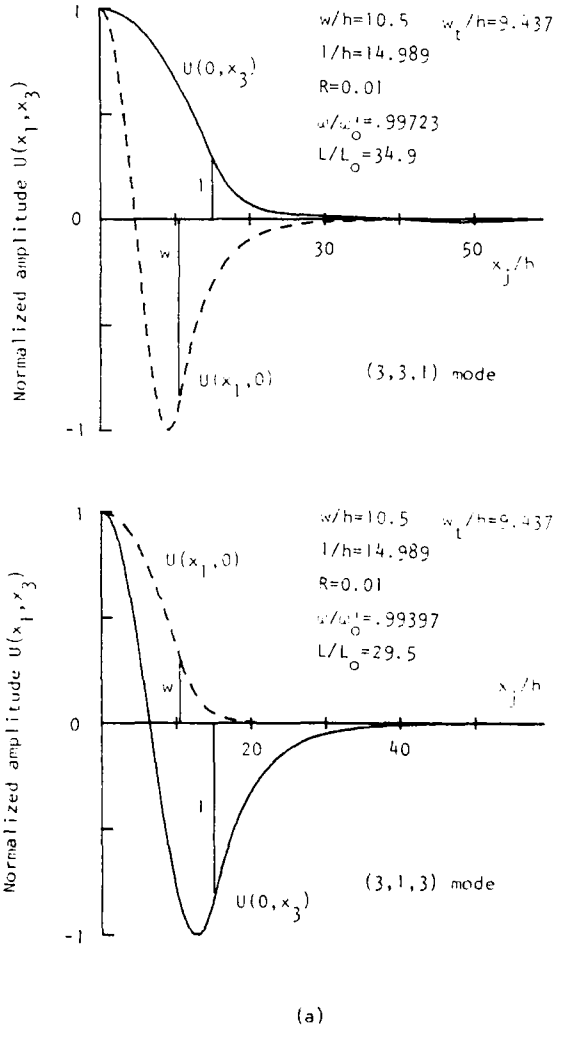


Fig.4 (a). Calculated vibrational patterns of spurious modes for the first group at w/h=10.5.

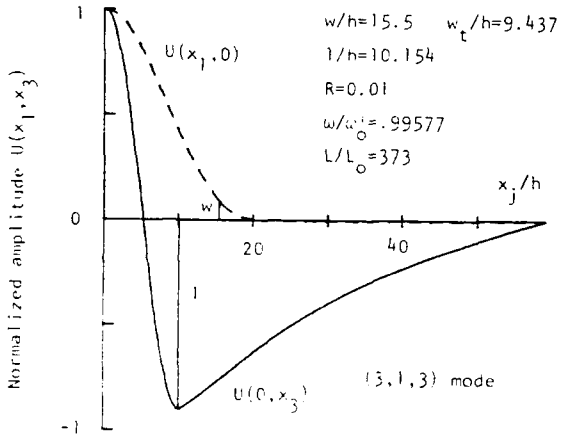
Also, an another analysis using the free edge condition $U_{,3}=0$ at $x_3=\pm l_q$ indicated that the resonant frequencies of untrapped modes in the range $\omega_o^t < \omega < \omega_o'$ were very sensitive to a change in the plate edge condition. Hence, we judged that such untrapped resonances would be effectively suppressed by the lossy mounting.

V. Concluding Remarks

The present simple theory using many assumptions and approximations describes well the contribution of tab electrodes to resonant frequencies and mode strength. Calculated and measured results indicate that the strong acoustical influence of tabs is observed only on the anharmonic spurious modes with phase changes in the tab direction. The results given here, of course, are not complete. The theory's validity for the equivalent inductance of spurious modes have been only qualitatively justified. Also, experimental studies on the additional untrapped modes have not been made. However, in spite of such limitations, we believe that our theoretical model is useful in the design of trapped energy resonators.

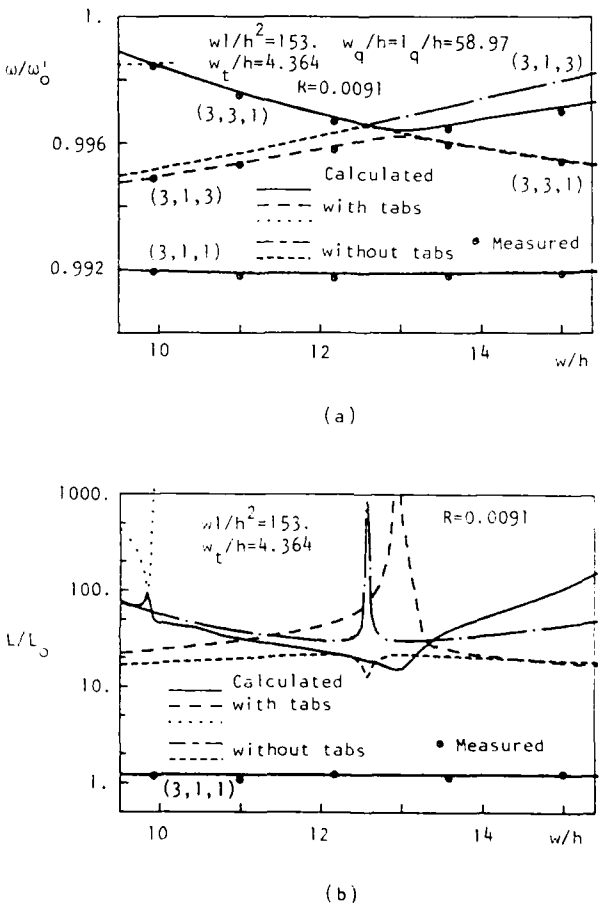
Acknowledgements

The authors would like to thank the staff of Nihon Dempa Kogyo Co. Ltd. for their help.



(b)

Fig.4 (b). Calculated vibrational patterns of spurious modes for the first group at w/h=15.5.



(a)

(b)

Fig.5. Measured and calculated results of (a)frequencies and (b)equivalent inductances as a function of central electrode width-to-plate thickness w/h for the second group, where $4wl=1.1mm^2$, $2w_t=0.37mm$, R=0.0091 and the equivalent inductance of pure thickness-shear vibration L_o=20.6mH.

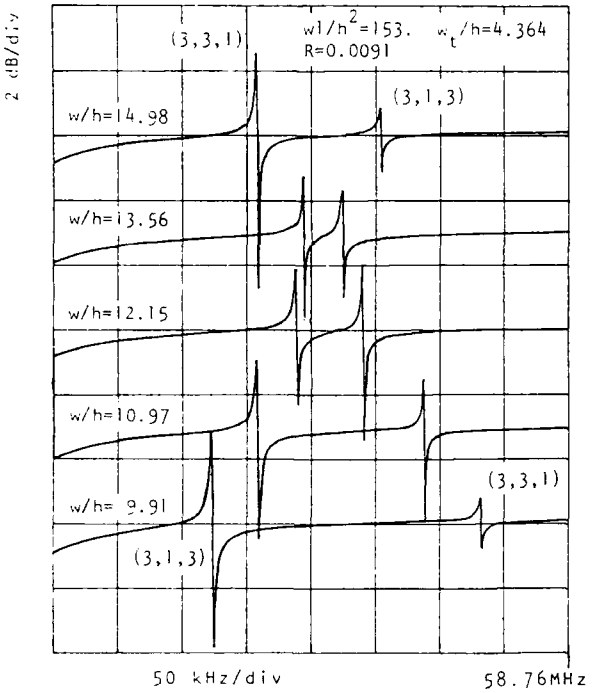


Fig.6. Measured frequency-impedance characteristics of spurious modes for the second group.

References

[1] H.F.Tiersten,"Analysis of trapped energy resonators operating in overtones of coupled thickness-shear and thickness-twist," J. Acoust. Soc. Am., vol.59, pp.879-888,1976

[2] K.Nakamura and H.Shimizu,"Analysis of two-dimensional energy trapping in monolithic crystal filters," Trans. IECE Japan, vol.62-A, pp.349-356,1979(in Japanese)

[3] H.Sekimoto and M.Ariga,"An analysis of trapped energy resonators with rectangular electrodes," Trans. IECE Japan, vol.63-A, pp.530-532,1980 (in Japanese)

[4] H.Sekimoto,"Analysis of trapped energy resonators with circular electrodes," IEEE Trans. Sonics Ultrason., vol.SU-31, pp.664-669,1984

[5] W.H.Horton and R.C.Smythe,"Theory of thickness-shear vibrations, with extensions and applications to VHF acoustically-coupled-resonator filters," *Proc. 21th Ann. Symp. on Freq. Cont.*, pp.160-178, 1967

[6] H.Sekimoto and M.Ariga,"Dispersion characteristics near cutoff frequencies for thickness waves in high coupling piezoelectric plates," Trans. IECE Japan, vol.61-A, pp.980-987,1978(in Japanese)

A VARIATIONAL METHOD FOR THE DESIGN OF TRAPPED ENERGY RESONATORS

R C Peach

GEC Research Laboratories Limited
Hirst Research Centre
East Lane
Wembley Middlesex HA9 7PP

The Q and motional parameters of a trapped energy acoustic resonator depend on several factors, including electrode size and shape, electrode plating thickness, and blank contour. In addition frequency temperature performance may be modified by resonator geometry. For these reasons a reliable simulation procedure is an extremely useful tool in developing new resonator designs.

The objective of the present work was to develop a system which could analyse resonators fabricated on any orientation of any piezoelectric material, with either rectangular or circular electrodes.

As with most semi-analytic treatments the procedure starts by eliminating the thickness coordinate of the plate, and generating approximate two dimensional equations of motion for an effective mode amplitude. The parameters of the two dimensional equations are deduced using an analytic procedure.

The approximate equations are then solved for the specified geometry by a variational method. The technique can be applied very efficiently to both rectangular and circular electrodes. Extensions to include the calculation of motional parameters have also been incorporated.

INTRODUCTION

The Q and the motional parameters of a trapped energy acoustic resonator depend on several factors, including electrode size and shape, electrode plating thickness, and blank contour. Furthermore, in certain cases, such as the SC cut, the frequency temperature behaviour may be noticeably modified by the resonator geometry. Developing a new resonator design entirely by empirical methods can therefore be a lengthy task, particularly if it differs significantly from previous designs, and for this reason a comprehensive simulation procedure is an extremely useful tool.

The present work was initiated and supported by Salford Electrical Instruments Ltd., and was intended to provide an extension of the earlier techniques developed by Werner and Dyer [1]. These were only applicable to rotated Y-cuts of quartz with rectangular electrode geometries, and it was decided that the extended system should be capable of analysing any mode in resonators fabricated on any orientation of any piezoelectric material, with either rectangular or circular electrodes.

The system employs a semi-analytic method of solution mostly adapted from previous work by the author [2]. The problem can be divided into two distinct parts; first, the construction of approximate equations of motion, and second, the solution of these equations for the specified geometry. In the earlier work parameters for the approximate equations of motion were calculated by a general acoustic waveguide analysis programme, and this exhibited several unsatisfactory features. It was necessary to re-run the waveguide programme for each crystal orientation or plateback considered, which involved a significant amount of computation, and, due to the large number of adaptive procedures in the programme, failures could occur in certain situations. In addition there were formal difficulties concerning the basis for constructing approximate single mode equations and their possible generalisations.

Considerable effort has therefore been devoted to resolving these problems, and a normal mode expansion has been devised which will allow approximate equations to be constructed in a systematic way and also provide analytic expressions for the essential parameters. The equations for a given resonator geometry are then solved by the variational method developed previously [2], but with a number of detailed improvements; calculation of Q and motional inductance is now included.

CONSTRUCTION OF APPROXIMATE EQUATIONS

The analysis of a trapped energy resonator starts by considering the behaviour of acoustic modes in an infinite plate of uniform thickness and of the same material, orientation and plateback as the resonator itself. The resonator mode is very similar to that of a particular plate mode at cutoff, but small transverse dependences do exist and are vital to the problem; it is therefore necessary to determine not only the cutoff frequency, but the dispersion relations as well. In a single mode analysis it is assumed that the functional dependence on thickness (z) of the acoustic displacements is constant, but that their amplitude varies slowly in the transverse directions (x and y); the scalar amplitude function is then described by a differential equation which must be solved for the specified geometry. This technique describes the main mode and its anharmonic overtones very well, but does not account for phenomena such as 'bandbreaks' which involve coupling to other modes with quite different thickness dependences. An attempt has been made, by means of a normal mode expansion, to place the construction the approximate equations on a quite general footing, and also to allow clear generalisations to multi-mode analysis, should these be necessary.

The constitutive equations of linear piezoelectric theory are

$$T_{ij} = c_{ijkl} S_{kl} - e_{kij} E_k$$

$$D_i = e_{ijk} S_{jk} + \varepsilon_{ij} E_j \qquad (1)$$

T_{ij} = stress

S_{ij} = strain

E_i = electric field

c_{ijkl} = stiffness constants

D_i = electric displacement

e_{ijk} = piezoelectric constants

ε_{ij} = dielectric permittivities

CH2186-0/85/0000-0392$1.00©1985IEEE

It is assumed that the equations have already been rotated to the cartesian axes of the plate. Using the quasi-static approximation $E_i = -\phi,_i$, the equations of motion become

$$\begin{aligned} \rho\ddot{u}_i &= c_{ijkl}\, u_{k,lj} + e_{kij}\, \phi,_{kj} \\ 0 &= e_{jik}\, u_{i,kj} - \varepsilon_{ij}\, \phi,_{ij} \end{aligned} \qquad (2)$$

There are four independent variables u_1, u_2, u_3 and ϕ, but it can be seen from Equation (2) that ϕ enters on a different footing from u_i, in that the left hand side of the last equation is zero. In the development of the normal mode expansion it is helpful to introduce a more compact notation. The components of a 4-D displacement vector $\tilde{u}_i$ will be defined by

$$\begin{aligned} \tilde{u}_i &= u_i \quad i = 1,2,3 \\ \tilde{u}_4 &= \phi \end{aligned} \qquad (3)$$

a generalised stiffness matrix $\tilde{c}_{ijkl}$ with dimensions 4x3x4x3 may be defined by

$$\begin{aligned} \tilde{c}_{ijkl} &= c_{ijkl} \quad i,j,k,l = 1,2,3 \\ \tilde{c}_{4jkl} &= e_{jkl} \quad j,k,l = 1,2,3 \\ \tilde{c}_{ij4l} &= e_{lij} \quad l,i,j = 1,2,3 \\ \tilde{c}_{4i4j} &= -\varepsilon_{ij} \quad i,j = 1,2,3 \end{aligned} \qquad (4)$$

A generalised stress $\tilde{T}_{ij}$ may be defined by

$$\begin{aligned} \tilde{T}_{ij} &= T_{ij} \quad i,j = 1,2,3 \\ \tilde{T}_{4j} &= D_j \quad j = 1,2,3 \end{aligned} \qquad (5)$$

and a mass tensor ρ_{ij} by

$$\rho_{ij} = \rho[\delta_{ij} - \delta_{i4}\delta_{j4}] \quad i,j = 1,2,3,4 \qquad (6)$$

with this notation the constitutive equations and the equations of motion become

$$\tilde{T}_{ij} = \tilde{c}_{ijkl}\ \tilde{u}_{k,l} \qquad (7)$$

$$\text{and } \rho_{ij} \frac{\partial^2 \tilde{u}_j}{\partial t^2} = \tilde{c}_{ijkl}\, \tilde{u}_{k,lj} \qquad (8)$$

The objective is to find an expansion of the form

$$\tilde{u}_i(x,y,z,t) = \sum_m A_m(x,y,t)\, \tilde{u}_i^m(z) \qquad (9)$$

for a general displacement field; $\tilde{u}_i^m(z)$ may be regarded as the m'th cutoff mode. With such a formula the equations of motion may be expressed in terms of the amplitudes $A_m(x,y,t)$ alone.

If we look for a mode at cutoff with an $e^{j\omega t}$ time dependence, the eigenvalue equation becomes

$$-\omega^2\, \rho_{ij}\, \tilde{u}_j = \tilde{c}_{i3k3}\, \tilde{u}_{k,33} \qquad (10)$$

The thickness direction, is taken to be z (3), and the plate boundaries are at z=0 and l. A variety of boundary conditions may occur in practice, but the most important are those describing massive conducting electrodes

$$\begin{aligned} T_{i3} &= \sigma \frac{\partial^2 u_i}{\partial t^2} \quad \text{at } z=0 \\ T_{i3} &= -\sigma \frac{\partial^2 u_i}{\partial t^2} \quad \text{at } z=l \\ \phi &= 0 \text{ at } z=0 \text{ and } l \\ \sigma &= \text{mass loading/unit area/side.} \end{aligned} \qquad (11)$$

It is well known that equation (10) provide three families of eigenmodes, one corresponding principally to overtones of the quasi-longitudinal wave, and the other two to overtones of the quasi-shear waves. The solution of (10) is quite straightforward, and simple transcendental equations for the mode frequencies are known [3]. If one of the eigensolutions is denoted by $\tilde{u}_i^m(z)$ and has a frequency ω_m, then it is possible to establish certain orthogonality relationships. As

$$-\omega_m^2\, \rho_{ij}\, \tilde{u}_j^m = \tilde{c}_{i3k3}\, \tilde{u}_{k,33}^m \qquad (12)$$

$$-\omega_m^2 \int_0^l \rho_{ij}\, \tilde{u}_i^n\, \tilde{u}_j^m\, dz = \int_0^l \tilde{c}_{i3k3}\, \tilde{u}_{k,33}^m\, \tilde{u}_j^n\, dz \qquad (13)$$

Integrating by parts and using the boundary conditions (11), which apply to both m and n modes

$$-\omega_m^2 \int_0^l \rho\, u_i^n\, u_i^m\, dz = \left[\tilde{c}_{i3k3}\, \tilde{u}_{k,3}^m\, \tilde{u}_i^n \right]_0^l - \int_0^l \tilde{c}_{i3k3}\, \tilde{u}_{k,3}^m\, \tilde{u}_{i,3}^n\, dz \qquad (14)$$

$$\omega_m^2 \left[\sigma[u_i^m(l)\, u_i^n(l) + u_i^m(0)\, u_i^n(0)] + \int_0^l \rho\, u_i^m\, u_i^n\, dz \right] = \int_0^l \tilde{c}_{i3k3}\, \tilde{u}_{k,3}^m\, \tilde{u}_{i,3}^n\, dz \qquad (15)$$

[The implied summation for $\tilde{u}_i$ runs over i = 1,2,3,4; for u_i it runs over i = 1,2,3 only].

If the indices n and m in equation (15) are exchanged, then, comparing the new equation with (15), it is clear that if $\omega_m \neq \omega_n$ then both sides of equation (15) are identically zero. Normalisation equations may therefore be introduced

$$\sigma(u_i^n(l)\, u_i^m(l) + u_i^n(0)\, u_i^m(0)) + \int_0^l \rho\, u_i^m\, u_i^n\, dz = \rho_n \delta_{nm} \qquad (16)$$

$$\int_0^l \tilde{c}_{i3j3}\, \tilde{u}_{i,3}^m\, \tilde{u}_{j,3}^n\, dz = \nu_n\, \delta_{nm} \quad (\nu_n = \rho_n\, \omega_n^2) \qquad (17)$$

ρ_n is defined to be the total area mass density $(\rho l + 2\sigma)$ and $\nu_n = \rho_n\, \omega_n^2$. Having found orthogonality relationships for the cutoff modes, this sort of treatment usually proceeds by claiming that the eigensolutions $\tilde{u}_i^m(z)$ form a complete set of functions

for expanding any set of displacements in z; unfortunately, in this case, such a claim is quite unfounded. The reason for this can be seen by considering a non-piezoelectric plate where the mechanical and electrical variables are quite separate. In this case the above theory would be essentially unaltered, in particular, the same three families of eigenmodes would be identified, but obviously these modes would be useless for describing an arbitrary electrostatic potential in the plate. When the piezoelectricity is 'turned on' the basic situation is unchanged; it cannot be assumed that the normal cutoff modes will be complete as far as the quasi-static potential is concerned. To overcome this it is necessary to introduce an additional family of 'auxiliary electrostatic modes', which are, for the short circuit boundary conditions,

$$u_i = 0$$

$$\phi = \sqrt{2}\sin\left(\frac{r\pi z}{\ell}\right) \qquad r = 1,2,3,.... \tag{18}$$

obviously these modes satisfy

$$\int_0^\ell \tilde{u}_i^n \, \tilde{u}_i^m \, dz = \ell\delta_{nm} \tag{19}$$

In addition it is easily shown that equation (17) is universally valid for any two modes whatsoever if we define

$$\nu_n = -\tilde{\epsilon}_{33}\left(\frac{r\pi}{\ell}\right)^2 \tag{20}$$

for an auxiliary electrostatic mode n described by the index r. Furthermore equation (16) is universally valid if $\rho_n=0$ for an electrostatic mode. Having now identified a complete set of functions with general orthogonality and normalisation relations, it is possible to deduce the equations of motion for A_n.

Returning to the equations of motion (8),

multiplying by $\tilde{u}_i^n$ and integrating over z

$$\int_0^\ell \rho_{ij}\frac{\partial^2\tilde{u}_j}{\partial t^2}\,\tilde{u}_i^n\,dz = \int_0^\ell \tilde{c}_{ijk\ell}\,\tilde{u}_{k,\ell j}\,\tilde{u}_i^n\,dz \tag{21}$$

The obvious procedure would now be to substitute the expansion of equation (9) and use the orthogonality relations; however this would require double derivatives of (9) with respect to z, and due to problems with the uniformity of convergence of the expansion a naive term by term differentation would produce erroneous results. This delicate problem can be sidestepped by integrating (21) by parts <u>before</u> substituting the expansion (9). Separating the derivatives, and using Greek letters α and β to indicate implied summations over 1 and 2 alone,

$$\int_0^\ell \rho_{ij}\frac{\partial^2\tilde{u}_j}{\partial t^2}\,\tilde{u}_i^n\,dz = \int_0^\ell \tilde{c}_{i3j3}\,\tilde{u}_{j,33}\,\tilde{u}_i^n \; +$$

$$\tilde{c}_{i\alpha j3}\,\tilde{u}_{j,\alpha 3}\,\tilde{u}_i^n + \tilde{c}_{i3j\alpha}\,\tilde{u}_{j,\alpha 3}\,\tilde{u}_i^n + \tag{22}$$

$$\tilde{c}_{i\alpha j\beta}\,\tilde{u}_{j,\alpha\beta}\,\tilde{u}_i^n\,dz$$

$$= \int_0^\ell \frac{d}{dz}\left[(\tilde{c}_{i3j3}\,\tilde{u}_{j,3} + \tilde{c}_{i3j\alpha}\,\tilde{u}_{j,\alpha})\,\tilde{u}_i^n\right]$$

$$- \tilde{c}_{i3j3}\,\tilde{u}_{j,3}\,\tilde{u}_{i,3}^n + \tilde{c}_{i\alpha j3}\,\tilde{u}_{j,\alpha 3}\,\tilde{u}_i^n \tag{23}$$

$$- \tilde{c}_{i\alpha j3}\,\tilde{u}_{j,\alpha}\,\tilde{u}_{i,3}^n + \tilde{c}_{i\alpha j\beta}\,\tilde{u}_{j,\alpha\beta}\,\tilde{u}_i^n\,dz$$

$$= \left[\tilde{T}_{i3}\,\tilde{u}_i^n\right]_0^\ell + \int_0^\ell -\tilde{c}_{i3j3}\,\tilde{u}_{j,3}\,\tilde{u}_{i,3}^n$$

$$+ \tilde{c}_{i\alpha j3}\,\tilde{u}_{j,\alpha 3}\,\tilde{u}_i^n - \tilde{c}_{i3j\alpha}\,\tilde{u}_{j,\alpha}\,\tilde{u}_{i,3}^n \tag{24}$$

$$+ \tilde{c}_{i\alpha j\beta}\,\tilde{u}_{j,\alpha\beta}\,\tilde{u}_i^n\,dz$$

Using the boundary conditions (11) to remove $\tilde{T}_{i3}$, the expansion (9) may be substituted without more ado. Using equations (16) and (17)

$$\rho_n\frac{\partial^2 A_n}{\partial t^2} + \nu_n A_n + \sum_m (a_{nm}^\alpha - a_{mn}^\alpha)\frac{\partial A_m}{\partial x_\alpha}$$

$$- \sum_m b_{nm}^{\alpha\beta}\frac{\partial^2 A_m}{\partial x_\alpha \partial x_\beta} = 0 \tag{25}$$

where

$$a_{nm}^\alpha = \int_0^\ell \tilde{c}_{i3j\alpha}\,\tilde{u}_j^m\,\tilde{u}_{i,3}^n\,dz \tag{26}$$

$$b_{nm}^{\alpha\beta} = \int_0^\ell \tilde{c}_{i\alpha j\beta}\,\tilde{u}_j^m\,\tilde{u}_i^n\,dz \tag{27}$$

Despite the rather complex derivation, the equations of motion are fairly straightforward coupled second order differential equations with constant coefficients; these coefficients being expressed in terms of the cutoff modes alone.

Having deduced the coupled equations for the amplitudes A_m, we are now in a position to develop approximate single mode equations. At cutoff, where all x and y dependences vanish, the solutions to equation (25) are trivial

$A_n = a_n e^{j\omega_n t}$ for an ordinary mode

$A_n = 0$ for an auxiliary electrostatic mode

We are concerned with a particular mode n, and in the solution of (25) for which A_n is the dominant amplitude close to cutoff. If x and y dependences are 'slow', then modes m, m≠n, will be mixed in only to first order in $\partial/\partial x_\alpha$

$$A_m \propto \frac{\partial A_n}{\partial x_\alpha}$$

Under these conditions it is easily shown that the approximate equation for A_n is

$$\rho_n\frac{\partial^2 A_n}{\partial t^2} + \nu_n A_n - s_{nn}^{\alpha\beta}\frac{\partial^2 A_n}{\partial x_\alpha \partial x_\beta} = 0 \tag{28}$$

where

$$s_{nn}^{\alpha\beta} = b_{nn}^{\alpha\beta} + \Sigma' \frac{(a_{nm}^{\alpha} - a_{mn}^{\alpha})(a_{mn}^{\beta} - a_{nm}^{\beta})}{\nu_m - \rho_m\nu_n/\rho_n} \qquad (29)$$

(The ' on the summation indicates that the term m=n is omitted)

and

$$A_m = \frac{(a_{nm}^{\alpha} - a_{mn}^{\alpha})}{\nu_m - \rho_m\nu_n/\rho_n} \cdot \frac{\partial A_n}{\partial x_\alpha} \qquad (30)$$

Equation (25) provides an effective uncoupled differential equation for the amplitude A_n, and gives a precise measure of how the dispersion is affected by mode mixing; clearly large values of

$$a_{mn}^{\alpha} ,$$

or a mode with a close cutoff frequency ($\nu_m \approx \nu_n$), could produce large coupling effects.

It might be supposed that although this normal mode expansion provides a neat formal approach to the problem it has little to offer from the point of view of practical calculation. However this is not true; the summation in the formula for $s_{nn}^{\alpha\beta}$ can be evaluated analytically in a number of cases, including that of the short circuited plate with massive electrodes. The resulting closed form expression is highly complex and involves quantities such as the mode frequency at cutoff, and the polarisations and velocities of waves propagating along the z axis. Despite this the formula is very easily evaluated by a computer, and is very much more satisfactory than the general waveguide analysis programme originally used.

SOLUTION OF APPROXIMATE EQUATIONS FOR A SPECIFIC RESONATOR GEOMETRY

In the previous section we deduced approximate equations for an infinite uniform plate which describe the behaviour of a particular mode close to cutoff; these will now be applied to the resonator geometry shown in Figure 1. There are now two distinct regions, one with metal electrodes, one without, each described by a slightly different equation; in addition there is the complication that plate thickness varies with position. The local thickness h´is related to the maximum thickness by

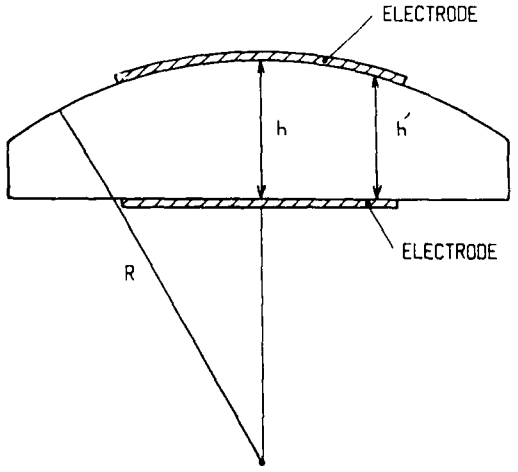


Figure 1 : Geometry of a contoured resonator.

$$h' = h - \tfrac{1}{2}\frac{(x^2 + y^2)}{R} \qquad (31)$$

If ω_F is the frequency of a free plate of thickness h then a 'local frequency', ω_L, may be defined by

$$\omega_L = \omega_F \frac{h}{h'} = \omega_F\left(1 + \frac{(x^2 + y^2)}{2Rh}\right) \qquad (32)$$

similarly for the metallised region

$$\omega_L = \omega_M\left(1 + \frac{(x^2 + y^2)}{2Rh}\right) \qquad (33)$$

where ω_M is the frequency for a plate of thickness h. The equations of motion become, in the metallised region

$$-d^{\alpha\beta}\frac{\partial^2 A}{\partial x_\alpha \partial x_\beta} + \omega_M^2\left(1 + \frac{(x^2 + y^2)}{Rh}\right)A = \omega^2 A \qquad (34)$$

and in the free region

$$-d^{\alpha\beta}\frac{\partial^2 A}{\partial x_\alpha \partial x_\beta} + \omega_F^2\left(1 + \frac{(x^2 + y^2)}{Rh}\right)A = \omega^2 A \qquad (35)$$

The coefficients $d^{\alpha\beta} = s_{nn}^{\alpha\beta}/\rho_n$ are assumed to be the same in the two regions; this is not strictly correct, but it is a good approximation and avoids considerable additional complications; as only one mode is involved the index n on A has been dropped. The modification for a contoured plate makes an assumption that the thickness variation is sufficiently slow that it is only necessary to allow the coefficients in the equation to vary with thickness as they would for an infinite plate ($d^{\alpha\beta}$ is independent of thickness); this sort of approximation is vary hard to justify formally, but it seems to work well enough in practice [4].

To complete the formulation of the problem it is necessary to choose boundary matching conditions for the electrode edges. An exact match would require all the mode amplitudes, and this constitutes a mechanism for coupling to other modes. If however only one mode amplitude is used, the most reasonable choice is that A and its normal derivative should be continuous across the boundary. It is assumed that trapping is sufficiently good that the crystal boundary does not have to be considered.

As a preliminary the term in $\partial^2/\partial x \partial y$ in eqations (34) and (35) will be removed by a rotation to principal axis x´and y´. The equations now become, in the metalised region,

$$-\nu_1\frac{\partial^2 A}{\partial x'^2} - \nu_2\frac{\partial^2 A}{\partial y'^2} + \nu_3 A + \nu_4(x'^2 + y'^2)A = \omega^2 A \qquad (36)$$

ν_1 and ν_2 are the eigenvalues of $d^{\alpha\beta}$

$$\nu_3 = \omega_M^2 \quad , \quad \nu_4 = \omega_M^2/Rh \qquad (37)$$

and in the free region

$$-\nu_1\frac{\partial^2 A}{\partial x'^2} - \nu_2\frac{\partial^2 A}{\partial y'^2} + \nu_5 A + \nu_6(x'^2 + y'^2)A = \omega^2 A \qquad (38)$$

$$\nu_5 = \omega_F^2 \quad , \quad \nu_6 = \omega_F^2/Rh \qquad (39)$$

From now on the primes on x and y will be omitted and it will be assumed that they refer to the principal axes. If a separation of the variables is performed in (36) or (38) the solutions may be expressed in terms of parabolic cylinder functions, but matching these across the boundaries analytically is impracticable. For this reason the following approach has been adopted. First a countably infinite set of orthogonal functions which is complete with respect to arbitrary functions of x and y is chosen. An expansion of A is then formed in terms of these functions and, using equations (36) and (38), the eigenvalue problem in terms of differential equations is replaced by an equivalent problem in terms of an infinite matrix. This in itself is not particularly helpful, but if the solution can be approximated by a truncated expansion in terms of the basis functions, then one is left with a finite matrix eigenvalue problem which is very easily solved by a computer. A finite number of basis functions can indeed provide an excellent approximation, and using this method the problem can be solved for any reasonable electrode geometry.

The set of basic functions which will be chosen will be of the form

$$\psi_n(\alpha x).\psi_m(\beta y)$$

where

$$\psi_n(x) = 2^{-n/2}\,\pi^{-1/4}\,(n!)^{-1/2}\,e^{-x^2/2}\,H_n(x) \qquad (40)$$

These are the familiar 'harmonic oscillator' functions, and $H_n(x)$ is a Hermite polynomial. The normalisation is chosen so that

$$\int_{-\infty}^{\infty} \psi_m(x)\psi_n(x)dx = \delta_{nm} \qquad (41)$$

For simplicity we will use the Dirac notation

$$|p\rangle \equiv \alpha^{1/2}\,\psi_n(\alpha x).\beta^{1/2}\,\psi_m(\beta y) \qquad (42)$$

where the index p corresponds to a particular ordered pair of indices n and m. In this notation

$$\langle p'|p\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi_{n'}(\alpha x)\psi_{m'}(\beta y)\psi_n(\alpha x)\psi_m(\beta y)\alpha\beta dxdy$$

$$= \delta_{nn'}.\delta_{mm'} = \delta_{pp'} \qquad (43)$$

strictly speaking the 'function' $\langle p'|$ is the complex conjugate of $|p'\rangle$, but as the basis functions here are real, this distinction is not important. if we introduce the operators

$$H \equiv -\nu_1\frac{\partial^2}{\partial x^2} - \nu_2\frac{\partial^2}{\partial y^2} + \nu_5 + \nu_6(x^2+y^2) \qquad (44)$$

$$V \equiv (\nu_3 - \nu_5) + (\nu_4 - \nu_6)(x^2+y^2) \text{ in the electrode region}$$

$$\equiv 0 \text{ elsewhere} \qquad (45)$$

Then the eigenvalue problem may be stated

$$(H+V)|A\rangle = \omega^2|A\rangle \text{ where } |A\rangle \equiv A(x,y) \qquad (46)$$

The procedure is to form an expansion

$$|A\rangle = \sum_p a_p|p\rangle \qquad (47)$$

which must exist as the basis functions are complete, and to substitute into equation (46)

$$\sum_p a_p(H+V)|p\rangle = \omega^2\sum_p a_p|p\rangle \qquad (48)$$

using equation (39)

$$\sum_p \langle p'|H+V|p\rangle\, a_p = \omega^2\, a_{p'} \qquad (49)$$

which is a straightforward matrix eigenvalue problem, provided the matrix elements $\langle p'|H+V|p\rangle$ can be calculated. This matrix element is, of course, understood to mean the operator H+V acting on the basis function $|p\rangle$, multiplied by the basis function $\langle p'|$, and integrated over the whole xy plane. Clearly if the eigenvalue equation (49) can be solved, then the resonant frequencies ω are determined, and the mode is defined in terms of its 'Fourier' coefficients a_p. The boundary matching conditions do not appear to have been used in this derivation; in fact they are included implicitly in equation (49), and any solution is automatically continuous across all boundaries.

In the actual calculation a truncated set of functions is employed, p=1,...N, and the solution will be estimated from the equation

$$\sum_{p=1}^{N} \langle p'|H+V|p\rangle a_p = \omega^2 a_{p'} \qquad (50)$$

As the fundamental mode is the lowest eigenstate of H+V it is easily seen that this solution would minimise the expression,

$$\frac{\langle A|H+V|A\rangle}{\langle A|A\rangle} \qquad (51)$$

it is also easily seen that the lowest eigenvalue of (50) will give the minimum value of (51) obtainable with a superposition of the specified set of basis functions p=1,...N; so in this sense the method is a variational one. It is possible to make an even stronger statement; if the first N exact eigenvalues are ω_i, i=1,...N, arranged in ascending order of magnitude, and the N frequencies from (46) are ω_i^A also in ascending order, then,

$$\omega_i^A > \omega_i \qquad i = 1,...N \qquad (52)$$

In other words the effect of (50) is to generate approximations to the first N modes and the true eigenfrequency of each mode is a lower bound on the appropriate approximation. Increasing N improves the estimates for all modes, and a fairly modest number of basis functions can provide excellent estimates for the lower modes. Up till now we have not specified the scale factors α and β. Obviously the method will work for any non-zero values, but the convergence of (47) could be seriously affected by an improper choice. The fundamental mode is the most interesting, and ideally it should be fairly well approximated by the first of the basis functions, p=1 (equivalent to m=n=o), with only relatively small contributions from the higher order functions. This implies that a good procedure for the choice of α and β would be to form the variational expression

$$\langle 1|H+V|1\rangle \qquad (53)$$

and minimise it with respect to α and β; thus choosing α and β so that the first basis function forms as good an approximation to the fundamental mode as possible.

It is now necessary to consider the evaluation of the matrix elements

$$\langle p'|H+V|p\rangle \tag{54}$$

$$\langle p'|H+V|p\rangle = \langle p'|H|p\rangle + \langle p'|V|p\rangle \tag{55}$$

The first term in H involves only integrals over the whole xy plane, and is relatively easy, the second term in V involves finite integrals over the electrode area and is rather more difficult. In the present implementation of the program consideration is restricted to two electrode types

$$\begin{aligned} &\text{Rectangular } \{(x,y): -a<x<a \text{ and } -b<y<b\} \\ &\text{Elliptical } \{(x,y): \frac{x^2}{a^2} + \frac{y^2}{b^2} < 1\} \end{aligned} \tag{56}$$

In the rectangular case the dimensions are 2ax2b, and in the elliptical case the major and minor axes are of length 2a and 2b respectively. For both these geometries it is possible to separate the modes into four classes.

(1) Even in x, even in y
(2) Odd in x, even in y
(3) Even in x, odd in y
(4) Odd in x, odd in y

The basis functions may be grouped into the same four classes, and if, for instance, we wish to approximate a mode in class 1 then only basis functions from class 1 are required; this considerably improves the efficiency of the procedure as it is possible to exclude three quarters of the basis functions in the calculation of any particular mode. In practical terms the modes in class 1 are far and away the most important, but the program can compute all the four types.

The integrals $\langle p'|H|p\rangle$ can be computed analytically using the standard properties of the Hermite polynomials [5]; the integrals $\langle p'|V|p\rangle$ can all be reduced to expressions of the form

$$\int\int_{\text{electrode area}} \Psi_n'(\alpha x)\Psi_m'(\beta y)\Psi_n(\alpha x)\Psi_m(\beta y)\ .\ \alpha\beta dxdy \tag{57}$$

For the case of the rectangular electrode this becomes

$$\int_{-\alpha a}^{\alpha a} \Psi_n'(x)\Psi_n(x)dx\ .\ \int_{-\beta b}^{\beta b} \Psi_m'(y)\Psi_m(y)dy \tag{58}$$

and these finite integrals may also be evaluated analytically. For the elliptical case (57) becomes

$$\int_{-\alpha a}^{+\alpha a} \Psi_n'(x)\Psi_n(x) \int_{-\beta d}^{\beta d} \Psi_m'(y)\Psi_m(y)dy.dx \tag{59}$$

where $d = b(1-x^2/\alpha^2a^2)^{1/2}$

The integral over y may be performed analytically as before, but the subsequent integral over x is not amenable to such a procedure, and a numerical quadrature must be employed. It is easy to show that the x integral is of the form

$$\int_{-\alpha a}^{+\alpha a} \left(1 - \frac{x^2}{\alpha^2a^2}\right)^{1/2} f(x)dx \tag{60}$$

where f(x) is a smooth function of x. This type of integral can best be evaluated by a Gauss-Jacobi whole interval quadrature [6], which is only slightly less efficient than the analytic formulae used in the rectangular case.

Having calculated the matrix elements the eigenvalue problem is very easily solved using a standard numerical routine.

The normal mode theory furnishes expressions for surface charge density (Q) and the energy density (E)

$$Q = q_0 A$$

$$E = e_0 |A|^2$$

where q_0 and e_0 are constants; in fact Q and E depend on derivatives of A as well, but these corrections are too small to justify the additional complications which they produce. Having found expressions for A in terms of the basis functions, it is relatively easy to integrate the expressions for Q and E to compute the total electrode charge and the total energy of the mode. This of course gives a value for the motional inductance, and it is also possible to estimate the effect on the Q factor of imperfect energy trapping. The theory takes no account of the boundaries of the crystal plate, and a small amount of the mode energy, $\delta\varepsilon$, would theoretically be outside the physical boundary; ideally $\delta\varepsilon$ should be as small as possible to avoid damping by the crystal mountings. As a simple model, it may be assumed that this effect may be described by considering the infinite plate approximation, but regarding the region outside the actual crystal boundary as extremely viscous. If the total mode energy is ε, and the amount outside the physical boundary is $\delta\varepsilon$, then, as the mode energy converts from kinetic to potential energy twice per cycle, the rate of energy dissipation of $2f\delta\varepsilon$. The Q would therefore be $\omega\varepsilon/2f\delta\varepsilon = \pi\varepsilon/\delta\varepsilon$. Clearly this argument is vague in the extreme, but it does describe the essential physical problem, and this contribution to the loss may be included with others by adding the *inverse Q's of each mechanism.*

DESIGN EXAMPLES

The present computer programmes have been applied very successfully to a number of resonator designs. A typical problem might be to determine suitable values of blank contour, electrode diameter, and plating thickness for a resonator of prescribed crystal orientation and blank diameter. An interesting example is that of the classic 5 MHz 50T AT-cut quartz resonator used in many frequency standards. Figure 2 shows the theoretical Q factor against curvature (curvature = 1.0/radius of curvature in metres) for the fundamental mode and the first symmetric unwanted mode,

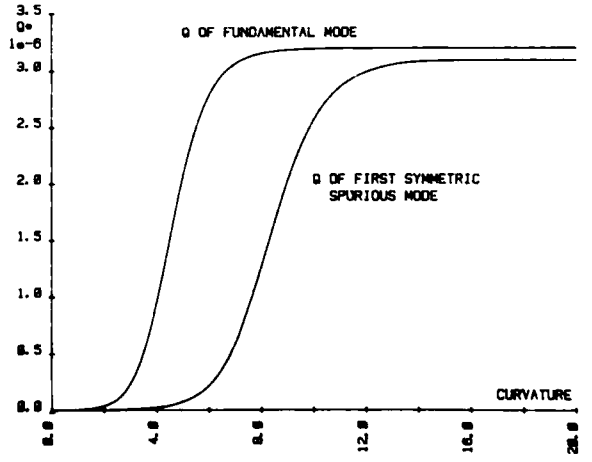


Figure 2 : Theoretical Q vs curvature for 5 MHz 50T AT-quartz resonator.

including only the effects of energy trapping and intrinsic Q. For this 5 MHz design the specified blank diameter was 13.7 mm, and it can be seen that in this case there is no advantage in using a curvature of more than 10.0 (radius of 100 mm). The inductance of the main mode is shown in Figure 3 as a function of electrode diameter for two values of the curvature, 10.0 and 20.0. It can be seen that in both cases the inductance curve flattens at diameters of 8-9 mm, and no significant reduction in inductance can be obtained by the use of larger electrodes; in all the calculations a nominal plating of 0.1 μ of gold on each side was assumed.

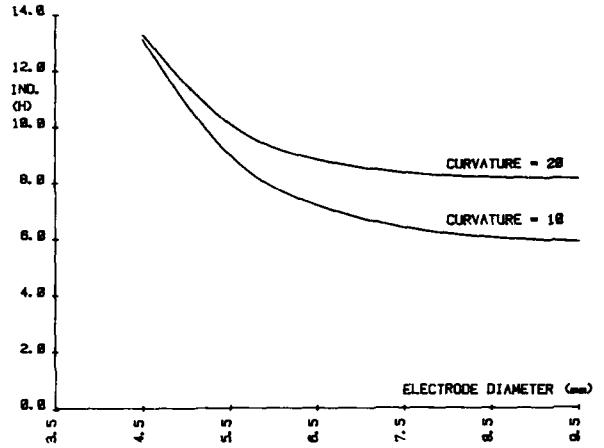


Figure 3 : Inductance vs electrode diameter for 5 MHz 50T AT-quartz resonator.

A recent problem was the design of a 10 MHz fundamental mode SC-cut quartz resonator. As in the case of the AT design the contour was chosen to give full energy trapping, and the electrode diameter was then selected to be on the flat portion of the inductance/diameter curve. This design was then produced experimentally, as closely as was possible with existing bevelling cups and plating masks. The performance of those resonators was very good, and, even though they had a lapped surface, Q's of greater than 800,000 were easily obtained. Table 1 shows a

Parameter	Mode	Measured Result	Theoretical Result
Frequency	(0,0,1)	10.003847 MHz	10.003847 MHz
Inductance	"	113.8 mH	114.5 mH
Q	"	823 x 10^3	1581 x 10^3
Frequency	(0,1,1)	10.145780 MHz	10.157649 MHz
Inductance	"	533 H	∞
Q	"	701 x 10^3	1512 x 10^3
Frequency	(1,0,1)	10.201188 MHz	10.205012 MHz
Inductance	"	59 H	∞
Q	"	551 x 10^3	1369 x 10^3
Frequency	(0,2,1)	10.287002 MHz	10.308330 MHz
Inductance	"	376.6 mH	276.1 mH
Q	"	576 x 10^3	1203 x 10^3
Frequency	(2,0,1)	10.400037 MHz	10.405353 MHz
Inductance	"	497.0 mH	331.6 mH
Q	"	203 x 10^3	407 x 10^3

Table 1 : Comparison of results for SC resonator.

comparison of measured and predicted parameters for this design including both the main mode and four unwanted modes. The modes are identified by the number of nodes in the x,y and z directions; for example the fundamental mode is (0,0,1). The theoretical value of the blank thickness was adjusted to give exact agreement on the fundamental mode frequency, and for this mode the predicted motional inductance is extremely close to that found experimentally. For the unwanted modes the agreement on frequency is reasonable, but the motional inductances are underestimated; the reasons for this have not yet been established. There is considerable disagreement on Q factor caused by neglecting the effect of the lapped surfaces, but the general trend is reasonably well predicted.

The computer programmes have the capability of determining the frequency-temperature characteristics of resonators, and this has been tested using the experimental results of Vig et al [7]; these include values of the frequency temperature slope at the inflexion point determined as a function of curvature for several resonators. The theoretical F-T curves for the case of the 5 MHz fundamental mode SC resonator considered by Vig are shown in Figure 4, using the same values of blank contour. Vig claimed that the relationship between F-T slope and curvature was approximately linear with a coefficient of

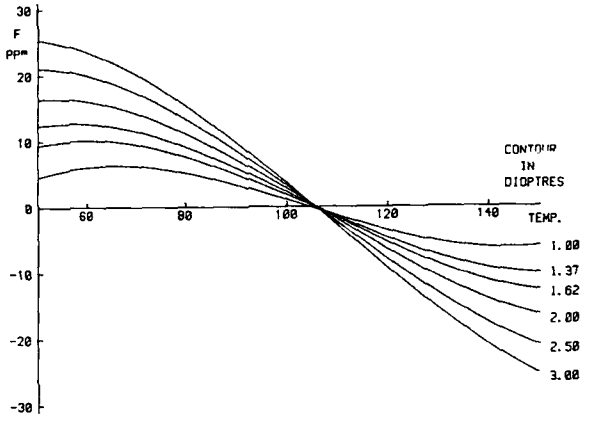


Figure 4 : SC-quartz (21°56', 33°54') F-T curves for various curvatures.

$$K = -21.6 \times 10^{-8} (\pm 1.1 \times 10^{-8})/°C/\text{dioptre}$$

A least squares fit to the theoretical points gives a slope of

$$K = -20.3 \times 10^{-8}/°C/\text{dioptre}$$

which is in excellent agreement with the experimental value. The curvature in dioptres is 0.53/(radius of curvature in metres). It can be seen from Figure 5 that

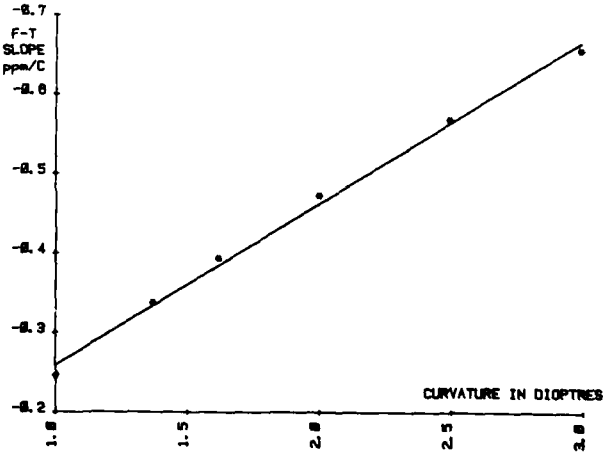


Figure 5 : Least squares fit to theoretical F-T slope against curvature data.

the linear relationship is reasonably good, but a quadratic law is considerably better; in this particular case a parabolic fit gives a relationship between slope (S parts/°C) and curvature (D dioptres)

$$S = S_0 - (29.14 \times 10^{-8})\, D + (2.194 \times 10^{-8})\, D^2$$

where S_0 is the value of the slope extrapolated to zero curvature.

All of the calculations in this section were performed using the elastic constant data of Bechmann et al [8], and although all the examples are for contoured crystals, the system will of course work equally well for plane parallel designs.

CONCLUSIONS

A new normal mode expansion for piezoelectric plates has been developed which provides a systematic procedure for constructing approximate equations of motion. A variational method of solution is used which can solve such equations for any of the commonly used resonator and electrode geometries. The system has been found to work very satisfactorily in practice and new resonator designs can normally be developed with few, if any, experimental iterations. The computer programmes can also predict frequency-temperature characteristics, and these predictions are also in good accord with experiment.

ACKNOWLEDGEMENTS

The author would like to acknowledge SEI Ltd for financial support.

REFERENCES

1 J.F. Werner and A.J. Dyer, 'The relationship between plateback, mass loading and electrode dimensions for AT-cut quartz crystals', Proc. 30th AFCS, pp 40-53, 1976

2 R.C. Peach, 'The design of partially contoured quartz crystal resonators', Proc. 36th AFCS, pp 22-28, 1976

3 T. Yamada and N. Niizeki, 'Admittance of piezoelectric plates vibrating under the perpendicular field exitation', Proc. IEEE, pp 941-942, 1970

4 H.F. Tiersten and R.C. Symthe, 'An analysis of contoured crystal resonators operating in overtones of coupled thickness shear and thickness twist', J. Acoust. Soc. Am., 65, 6, pp 1455-1460, 1979

5 E.D. Rainville, 'Special functions', Macmillan, 1960

6 A. Ralston and P. Rabinowitz, 'A first course in numerical analysis', 2nd Ed., McGraw-Hill, 1978

7 J.R. Vig, W. Washington and R.L. Filler, 'Adjusting the frequency vs temperature characteristics of SC resonators by contouring', Proc. 35th AFCS, pp 104-109, 1981

8 R. Bechmann, A.D. Ballato and T.J. Lukazek, 'Higher-order temperature coefficients of the elastic stiffnesses and compliances of alpha-quartz', Proc. IRE, 50, pp 1812-1822, 1962

FURTHER RESULTS ON THE CONTOUR DEPENDENCE OF THE FREQUENCY-TEMPERATURE CHARACTERISTIC OF SC-CUT RESONATORS

John A. Kosinski
US Army Electronics Technology and Devices Laboratory (ERADCOM)
Fort Monmouth, New Jersey 07703-5302

ABSTRACT

Vig, Filler, and Washington have reported the change in slope of the frequency-temperature characteristic of plano-convex SC-cut resonators, measured at the inflection temperature, as a function of blank contour. This change in slope can be thought of in terms of an "apparent angle shift." The change, however, is only apparent as the uncontoured side of the blank remains unchanged and retains its original orientation relative to the crystal axes. Stevens and Tiersten, on the other hand, have calculated the real changes in the angles of cut necessary to produce zero-temperature coefficient resonators at 25°C for designs of different radii of curvature. In this paper, the best available approximation for the angle dependence of the SC-cut frequency-temperature characteristic is applied to compare Vig's empirical results and Stevens and Tiersten's theoretical results. In the process, the empirical studies which were begun by Vig et al. were continued and expanded.

Key words: Quartz crystal, quartz resonator, SC-cut, frequency-temperature characteristic, blank contour, radius of curvature, frequency-temperature coefficients, angle gradients.

INTRODUCTION

Vig et al.[1] have reported the change in the slope of the frequency-temperature characteristic of plano-convex SC-cut resonators, measured at the inflection temperature, as a function of blank contour. This change in slope can be thought of in terms of an "apparent angle shift." This change, however, is only apparent as the uncontoured side of the blank remains unchanged, and retains it original orientation relative to the crystal axes. Stevens and Tiersten,[2,3] on the other hand, have calculated the real changes in the angles of cut necessary to produce zero-temperature coefficient resonators at 25°C for designs of different radii of curvature. In this work, the correlation between these two sets of data will be established.

In the course of establishing this correlation, some related topics of interest will also be discussed. First, the need to reference the recontouring coefficients to some chosen reference temperature will be established. Second, the changes in the inflection temperature during recontouring will be presented. Third, the recontouring coefficients will be shown to be a function of blank contour. Finally, the change in the slope of the SC-cut frequency-temperature characteristic as a result of simple thickness changes will be examined.

METHOD OF COMPARISON

Denoting the slope of the frequency temperature characteristic as S, we see that Vig has reported $(dS/dC)_{Ti}$, where C is the contour in diopters. Stevens and Tiersten presented graphs of vs. R and O vs. R, the angles of cut necessary to produce zero-temperature coefficient resonators at 25°C when different radii of curvature are used. To analyze the two sets of data we use[4]

$$\left(\frac{dS}{dR}\right)_{Ti} = \frac{-0.530}{R} \cdot \left(\frac{dS}{dC}\right)_{Ti} \qquad (1)$$

and

$$\left(\frac{dS}{dR}\right)_{25} = \left(\frac{\partial S}{\partial \varphi}\right)_{25} \cdot \left(\frac{d\varphi}{dR}\right)_{25} + \left(\frac{\partial S}{\partial \Theta}\right)_{25} \cdot \left(\frac{d\Theta}{dR}\right)_{25}, \qquad (2)$$

where $(d\varphi/dR)_{25}$ and $(d\Theta/dR)_{25}$ are the slopes of the curves published by Stevens and Tiersten,[2,3] and $(\partial S/\partial \varphi)_{25}$ and $(\partial S/\partial \Theta)_{25}$ are the sensitivities of the slope of the frequency-temperature characteristic to changes in the angles of cut. The slopes of Stevens and Tiersten's curves may be approximated by a second order polynomial obtained through least-squares analysis over the range of interest. The angle sensitivities of the frequency-temperature characteristic must be obtained from the approximation for the frequency-temperature characteristic[5] and the angle gradients of the coefficients.

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

Using Stevens and Tiersten's data, we obtain dS/dR for resonators at 25°C, while using Vig's data, we obtain dS/dR for resonators at the inflection temperature, approximately 95°C. Since the angle gradients of the second and third order temperature coefficients are given as zero,[6] these values should be directly comparable.

INITIAL RESULTS

The values of $(dS/dR)_{T_i}$ obtained from Vig and $(dS/dR)_{25}$ obtained from Stevens and Tiersten are listed in Figure 1. We see that for both designs investigated, the theoretical and empirical values thus obtained differ by approximately a factor of two. Interestingly, the ratio of the 5 MHz to 10 MHz value is the same for both sets of data.

To account for this factor of two, we first find that the data, both of Vig and of Stevens and Tiersten, has a limited range of radii of curvature over which they are valid. For Vig's data, the value given for 10 MHz 3rd overtone resonators is valid over the range of R = 0.27 to 0.08 meters (2.0 to 7.0 diopters), while the value for 5 MHz 3rd overtone resonators is valid over the range of R = 0.13 to 0.08 meters (4.0 to 7.0 diopters). In all cases, Stevens and Tiersten's data are most accurate at small radii of curvature and are less accurate for large radii of curvature.

Second, the values of the angle gradients of the frequency-temperature coefficients which are used here are given as approximate values for the case of the infinite flat plate. The resonators used here are neither infinite nor flat, and the angle gradients of the second and third order frequency-temperature coefficients could be nonzero, as we find to be the case.

Based on these facts, an expanded set of data was obtained for both types of resonators. Five 10 MHz resonators were manufactured with initial contours of 1.0 diopter, plano-convex and their frequency-temperature behavior was tested. They were then reprocessed with 3.0 diopter and their frequency-temperature behavior retested. This procedure was repeated for contours of 5.0, 7.0, and 9.0 diopters. Similarly, five 5 MHz resonators were processed and tested with contours of 5.0, 7.0, 9.0, 11.0, and 13.0 diopters. All of the blanks were 14 mm in diameter and, in all cases, the uncontoured side remained unchanged throughout the experiment.

RESULTS OF EXPANDED DATA SET

Typical data from a 10 MHz 3rd overtone unit, which was part of the expanded data set, are shown in Figure 2. Here we see the inflection temperature and the slope of the frequency-temperature characteristic, at both the inflection temperature and at 25°C, listed along with each corresponding blank contour.

Looking first at the slope of the frequency-temperature characteristic, we find that the slope has changed by 1.64×10^{-7} at the inflection temperature but only 1.0×10^{-7} at 25°C. Figure 3 illustrates this a bit more clearly, where the different slope changes at each step in the recontouring process are compared. In general, the slope of the frequency-temperature characteristic at the inflection temperature is changing approximately twice as fast as it is at 25°C.

Using this information, the recontouring coefficients at 25°C were calculated, along with revised empirical values for $(dS/dR)_{25}$. Figure 4 lists these revised values. The values for 10 MHz 3rd overtone units now agree within 26% and the values for the 5 MHz 3rd overtone units agree to better than 7%.

CHANGES IN INFLECTION TEMPERATURE

Looking back at the inflection temperatures listed in Figure 2, we find an almost linear increase with increasing contour. Figure 5 illustrates this fact, showing the inflection temperatures of all the units tested as a function of blank contour. Since $T_i = -b/3c$ and T_i is changing, the contour gradients of the second and third order temperature coefficients are seen to be nonzero. As a result, any determination of recontouring coefficients is valid only at the reference temperature chosen for the determination.

CONTOUR DEPENDENCE OF RECONTOURING COEFFICIENTS

We have thus far assumed that the slope of the frequency-temperature characteristic changes linearly with the blank contour. Figure 6, based on curves published by Stevens and Tiersten,[2,3] shows that the angles of cut of the resonator are a nonlinear function of the blank contour. Since the slope of the frequency-temperature characteristic changes almost linearly with the angle,[5,6] these curves imply that the slope of the frequency-temperature characteristic is also a nonlinear function of the blank contour.

Figures 7, 8, and 9 use the extended data set to show this phenomenon. It should be noted how much more linear the data are for 5 MHz 3rd overtone units than for 10 MHz 3rd overtone units. This would appear to explain why the correlation between "theoretical" and "empirical" values of $(dS/dR)_{25}$ is much closer for the 5 MHz units. Figures 10, 11, and 12 show the values of the recontouring coefficients as a function of blank contour for the designs used here. The expressions for the values of the coefficients listed with each figure are based on a least-squares analysis of the expanded data set.

the free surface charge density on the central electrode associated with the pure thickness-shear vibration at ω_o.

In the above model, the tab can be interpreted as a perturbation of acoustical parameters in the surrounding unelectroded region for a trapped energy

Fig.2. Measured and calculated results of (a)frequencies and (b)equivalent inductances as a function of central electrode width-to-plate thickness w/h for the first group, where $4wl=1.13mm^2$, $2w_t=0.8mm$, R=0.01 and the equivalent inductance of pure thickness-shear vibration $L_o=20mH$.

THICKNESS EFFECTS

The effect of simple thickness changes on the frequency-temperature characteristic of 3rd overtone SC-cut resonators has been investigated. Ten units were fabricated with 5 MHz nominal frequency and their temperature behavior was recorded. These units were then processed to 5.5 MHz, using the same blank contour and maintaining the plano side unchanged. The slopes of the frequency-temperature curves, measured at the inflection temperature, were found to change by $+ 3.0 \times 10^{-8}/^{o}C\text{-MHz}$.

CONCLUSIONS

Close correlation has been established between theoretical and empirical studies of the effects of recontouring on SC-cut resonators. The nonlinear recontouring predicted by the theoretical model has been verified. In the course of establishing this correlation, the need to choose a reference temperature for recontouring measurements has been demonstrated.

The use of recontouring as a method of adjusting the frequency-temperature characteristics of SC-cut resonators is seen to be somewhat more complicated than initially reported. The required change in contour for a particular blank is a function not only of the required slope change, but also of the absolute contour of the blank. When used over a wide range of contours, the recontouring process will be successful only if these effects are taken into account.

ACKNOWLEDGMENTS

The author thanks Mr. Edward Simon and Mr. William Washington for their considerable assistance in fabricating and testing the resonators used in this experiment. The author also thanks Dr. Arthur Ballato of ERADCOM and Professor Harry Tiersten of Rensselaer Polytechnic Institute for their guidance in preparing this report.

REFERENCES

1. J. Vig, W. Washington, and R. Filler, "Adjusting the Frequency vs. Temperature Characteristics of SC-cut Resonators by Contouring," Proc. 35th ASFC* (1981), pp. 104-105.

2. D. S. Stevens and H. F. Tiersten, "Temperature Induced Frequency Changes in Electroded Contoured SC-cut Quartz Crystal Resonators," Proc. 36th ASFC* (1982), pp. 46-54.

3. D. S. Stevens and H. F. Tiersten, "On the Change in Orientation of the Zero-Temperature, Contoured SC-cut Quartz Resonator with the Radius of the Contour," Proc. 38th ASFC* (1984), pp. 176-183.

4. Quartz Resonator Handbook - Manufacturing Guide for AT-Type Units, edited by R. E. Bennett, prepared for the Department of the Army under Contract No. DA36-039-SC-71061 by Union Thermoelectric Div., 1960, p. 62, AD 251289.

5. R. Bechmann, "Frequency-Temperature & Angle Characteristics of AT- and BT-type Quartz Oscillators in an Extended Temperature Range," Proc. IRE, Vol. 48, August 1960, p. 1494.

6. A. Ballato, "Doubly Rotated Thickness Mode Plate Vibrators," in Physical Acoustics, Volume 13 (W. P. Mason & R. N. Thurston, eds.), Academic Press, New York, 1977, pp. 115-181.

*Annual Symposium on Frequency Control

Frequency	$(dS/dR)_{Ti}$	$(dS/dR)_{25}$	$\frac{(dS/dR)_{Ti}}{(dS/dR)_{25}}$
5 MHz	$1.7\times10^{-8}R^{-2}$	$9.4\times10^{-9}R^{-2}$	1.8
10 MHz	$1.4\times10^{-8}R^{-2}$	$7.3\times10^{-9}R^{-2}$	1.9
$\frac{5\text{ MHz}}{10\text{ MHz}}$	1.21	1.29	RATIO

Figure 1. Initial Results

Contour	Ti	S_{Ti}	S_{25}
1.0	93.19	2.92×10^{-7}	1.10×10^{-6}
3.0	92.52	2.37×10^{-7}	1.05×10^{-6}
5.0	94.78	1.91×10^{-7}	1.03×10^{-6}
7.0	95.84	1.53×10^{-7}	1.01×10^{-6}
9.0	96.37	1.28×10^{-7}	1.00×10^{-6}

Figure 2. Typical data for 10 MHz 3rd O.T. resonator during recontouring.

Range	ΔS_{Ti}	ΔS_{25}	$\frac{\Delta S_{Ti}}{\Delta S_{25}}$
1 - 3	-5.6×10^{-8}	-5.2×10^{-8}	1.1
3 - 5	-4.6×10^{-8}	-1.6×10^{-8}	2.7
5 - 7	-3.8×10^{-8}	-2.0×10^{-8}	1.9
7 - 9	-2.4×10^{-8}	-1.4×10^{-8}	1.8

Figure 3. Comparison of change in slope at inflection and at 25°C.

Frequency	Theoretical	Empirical
5 MHz	$9.4\times10^{-9}R^{-2}$	$10.1\times10^{-9}R^{-2}$
10 MHz	$7.3\times10^{-9}R^{-2}$	$5.8\times10^{-9}R^{-2}$

Figure 4. Revised values of $(dS/dR)_{25}$.

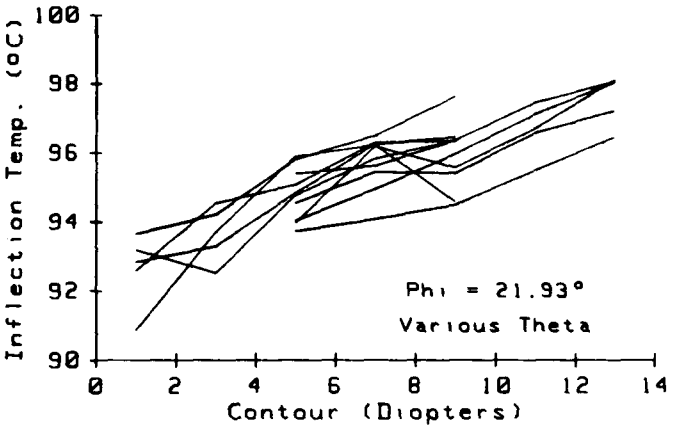


Figure 5. Change in inflection temperature during recontouring for 5 & 10 MHz 3rd O.T. SC-cut resonators.

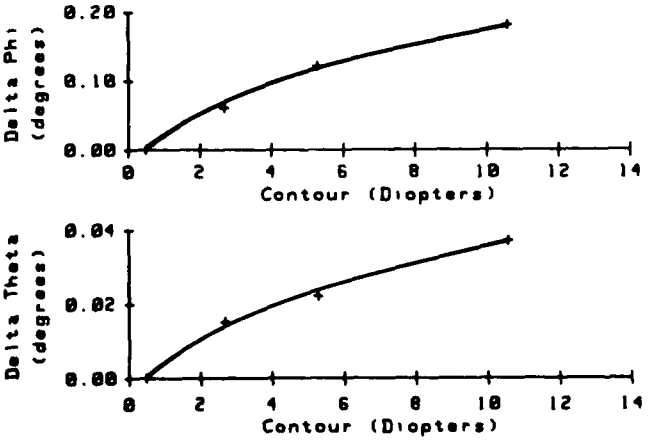


Figure 6. Based on Stevens and Tiersten. Relative changes in angles necessary to produce zero-temperature coefficient resonators at 25°C as a function of blank contour.

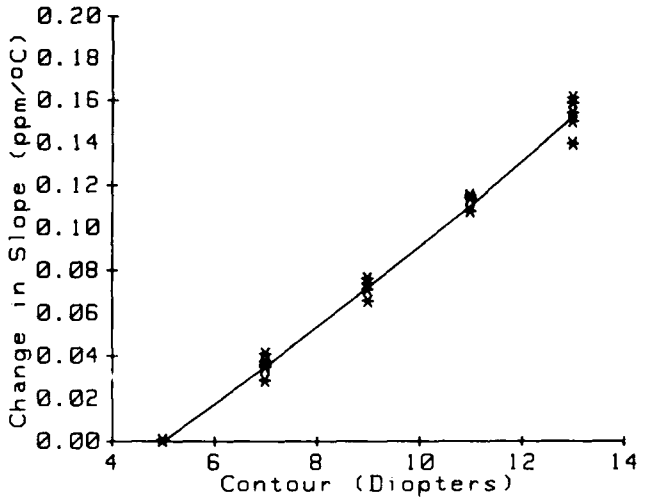


Figure 7. Change in slope at inflection vs. contour - 5 MHz 3rd O.T.

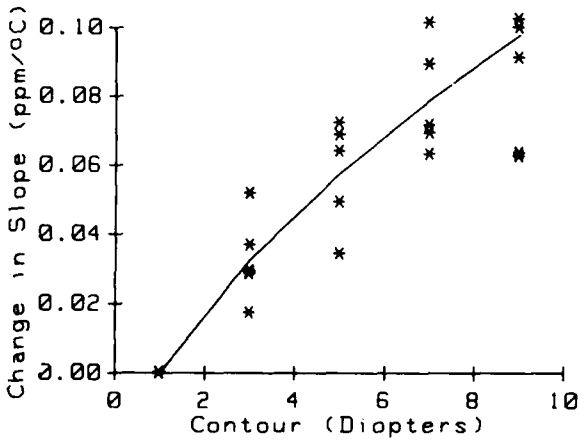


Figure 8. Change in slope at inflection vs. contour - 10 MHz 3rd O.T.

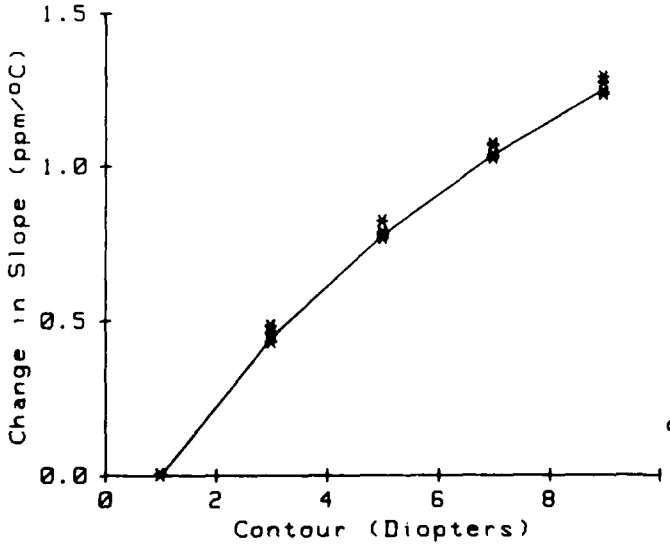


Figure 9. Change in slope at inflection vs. contour - 5 MHz fundamental.

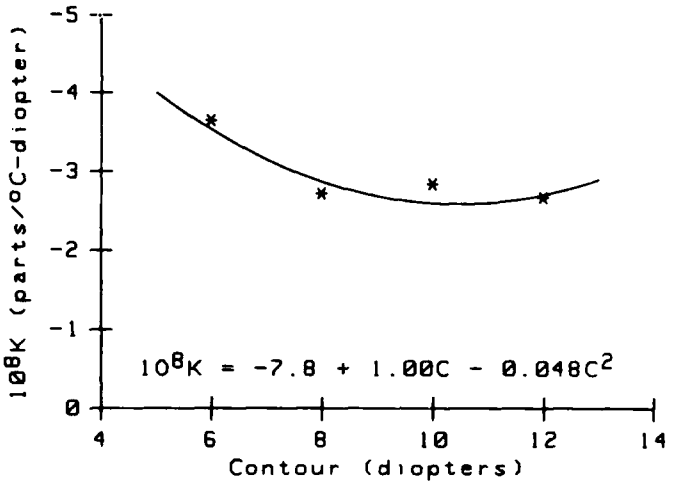


Figure 10. Recontouring coefficient vs. contour - 5 MHz 3rd O.T.

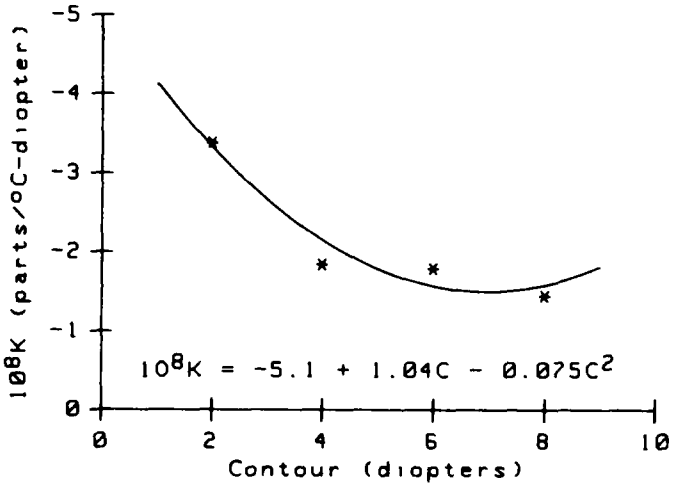


Figure 11. Recontouring coefficient vs. contour - 10 MHz 3rd O.T.

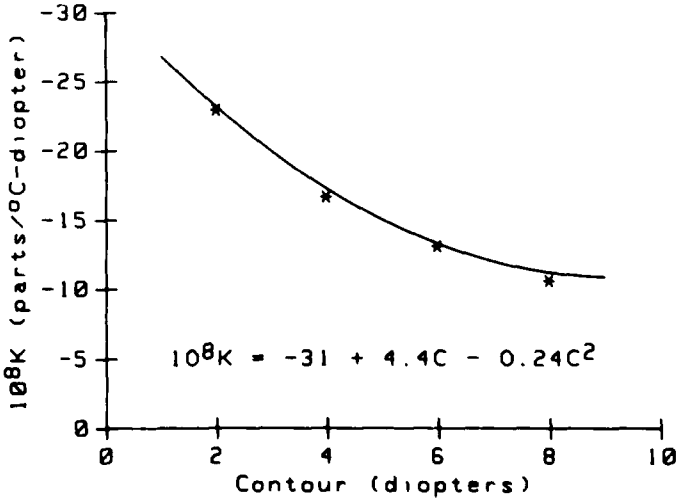


Figure 12. Recontouring coefficient vs. contour - 5 MHz fundamental.

MECHANICAL COUPLINGS INVOLVING DISCONTINUITIES OF THE FREQUENCY-TEMPERATURE CURVES OF CONTOURED QUARTZ RESONATORS

R. Bourquin, R. Dulmet, G. Genestier

Laboratoire de Chronométrie, Electronique et Piézoélectricité

E.N.S.M.M., Route de Gray, La Bouloie, 25030 BESANCON Cédex - FRANCE.

Abstract

In this paper, frequency jumps occuring in the frequency temperature curves of contoured resonators are studied. Besides vibration trapping effect, contouring the plate involves some mechanical couplings which may be strongly dependent on temperature. This will occur whenever two or several coupled modes are very close in frequency, while exhibiting different thermal sensitivities. Because contouring generates anharmonic modes in large frequency range, such an event frequently occurs. We have investigated some examples of such couplings in an AT-cut resonator vibrating in slow shear C-mode (seventh overtone) coupled with fast shear B-mode sixth overtone). Each coupling is very strong in a range of a few degrees. Our theoretical analysis is built on Green's formalism : the mode pattern of the actual vibration is assumed to be a mixing of simple mode patterns. Jumps of the frequency temperature curves are induced by the thermal dependency of this mixing. Use of separated X-rays topographs for the components of the mechanical displacement enables us to verify the predicted mode patterns at various temperatures.

1. Introduction

Two key-concepts are very useful for a comprehensive analysis of the vibration modes in contoured resonators : the quasi-thickness approximation [1], widely introduced by H.F. Tiersten and his coworkers, and the slowly varying thickness approximation, first proposed by R.D. Mindlin [2] for the treatment of shear modes in an AT-cut contoured resonator. Mindlin's model correctly shows the creation of trapped confined solutions in the vicinity of the plate center. In spite of later generalization to doubly rotated cuts [3], this model can be now considered as insufficient because many elastic coefficients are neglected. As shown by H.F. Tiersten [4], the variations of the vibration amplitude in the plane of the plate involve couplings, analysis of which leads to solve the dispersion equation of the quasi-thickness modes. But such an analysis runs into special difficulties whenever two or more coupled modes are very close in frequency, even if the coupling elastic coefficient is small. An effective means to improve the analysis is provided by Green's formalism [5,6,7]. Some terms are taken into the right hand members of the partial derivatives equations of dynamical elasticity. Ignoring these terms, we define a "starting model", solutions of which are obtained following Tiersten's procedure or similarly. These solutions constitute a quasi rigorus mathematical basis. In order to take into account the previously ignored terms, it is then allowed to perform Green's decomposition over this basis. After that, by use of orthogonality properties, we obtain an eigenvalues problem instead of the initial partial derivatives coupled equations. The actual coupled modes must be considered as mixings of the starting analytical solutions. Because of their mechanical character, the studied couplings may affect some modes being usually piezoelectrically driven or not. Moreover, if the couplings deal with modes very close in frequency, but having different thermal sensitivities, the coefficients of the mixing strongly depend on temperature. We have isolated an example of this phenomenon in an AT-cut contoured resonator [8]. The studied modes belong to slow shear C mode, seventh overtone (piezoelectrically driven) coupled with fast shear B mode, sixth overtone (not piezoelectrically driven). In a few degrees range, both approximate starting solutions for B and C modes equally contribute to the mixing wich defines the actual mode of the resonator. The corresponding resistance of each actual mode depends on the coefficients of the mixing and thus on temperature. The temperature exactly corresponding to equal contributions of B and C components in each of both coupled modes defines the "strongest coupling point". At this point, the ratios of the piezoelectric component to the piezoelectric component are inverted by a [illegible] temperature. As a consequence, a jump of the oscillator frequency can be observed at this point, from one actual mode to another.

Use of separated X-ray patterns for two orthogonal components of the mechanical displacement (the B and C components of the actual mode pattern) enables to verify the theoretically predicted mode patterns at various temperatures, near the strongest coupling point or far from it. A powerful tool for studying this kind of problems related to "activity dips" field has been thus obtained by combining theoretical and experimental analysis.

2. Basic equations

The contoured resonators of our interest are usually plano-convex. Their convex surface is actually spherical at room temperature, but do not remain spherical when temperature is varying. This effect results from anisotropy of thermal expansion. To avoid special resulting difficulties, we use a Lagrangian formulation which enables to introduce temperature effect while considering boundary surface shape and mass density at constant reference temperature T_0.

Following a well known procedure [11], we distinguish three states of the body, corresponding to three states of material coordinates, which are the three different fields of variables we can choose :

* First the natural state is defined as

CH2186-0/85/0000-0405$1.00©1985IEEE

the static equilibrium at T_o. The reference temperature is cho sen at T_o = 25°C.

* The initial state is then obtained from the natural one by a temperature variation ($T_o \rightarrow T$), very slow and reversible.

* The actual state later results from the superposition of an adiabatic vibration onto the natural state.

The material coordinates a_i which characterize the natural state constitute the field of variables in terms of which we express the equations of dynamic elasticity. Influence of piezoelectricity is neglected. Then the fundamental equations are :

$$\frac{\partial P_{ij}}{\partial a_i} = \rho_o \ddot{u}_j \qquad \text{in } V_o \tag{1}$$

where V_o denotes the volume of the body, and ρ_o its mass density, in the natural state, u is the dynamical displacement, while P is Piola's stress tensor. Balance equations (1) must be completed by associated boundary conditions :

$$P_{ij} N_i \, dS_o = T_j \, dS \tag{2}$$

where dS_o is a surface element of the external surface in the natural state. This element becomes dS in the actual state. T_j denotes a component of the external traction on dS, while N_i indicates a component of the unit normal related to dS_o. In fact Piola's stress tensor can be divided into two parts : the static one, only depending on temperature, and denoted by the superscript -, and the dynamic one, denoted by the superscript ~ and defined as the increase in P_{ij} due to the vibration :

$$\tilde{P}_{ij} = P_{ij} - \bar{P}_{ij} \tag{3}$$

By eliminating the static part in the equations of dynamical balance, the following equations are obtained :

$$\frac{\partial \tilde{P}_{ij}}{\partial a_i} = \rho_o \ddot{u}_j \quad \text{in } V_o \tag{4}$$

$$\tilde{P}_{ij} N_i = 0 \quad \text{on } S_o \tag{5}$$

Boundary conditions (5) hold for traction free boundaries only. We are interested in studying small dynamic fields superposed on the stress free bias which results from a very slow temperature variation. Thus we are allowed to expand $\tilde{P}_{ij}$ in terms of strain gradients, only retaining the first term :

$$\tilde{P}_{ij} = A_{ijk\ell} \frac{\partial u_\ell}{\partial a_k} \tag{6}$$

where the A tensor is related to the usual effective B tensor as follows :

$$A_{ijk\ell} = \frac{\rho_o}{\rho_1} \frac{\partial a_j}{\partial x_m} \frac{\partial a_k}{\partial x_n} B_{mjn\ell} \tag{7}$$

x_i and ρ_1 respectively indicate material coordinate and mass density of the initial state (static, but at $T \neq T_o$). Numerical values of the $B_{\lambda\mu}$ coefficients ($mj \rightarrow \lambda$, $n\ell \rightarrow \mu$) and their temperature derivatives were early presented in the well known BBL's paper [12].

Because the static strain gradients $\partial x_i / \partial a_k$ and the ρ_1 / ρ_o ratio are only depending on thermal expansion coefficients and on temperature, the temperature derivatives of $A_{\lambda\mu}$ coefficients ($T^n A_{\lambda\mu} = \frac{1}{n!} \frac{d^n A_{\lambda\mu}}{dT^n}$) can be deduced from $T^n B_{\lambda\mu}$ values. Because of asymmetry of relation (7), the $A_{\lambda\mu}$ matrix must be tabulated in a 9 x 9 matrix instead of a 6 x 6 matrix for $B_{\lambda\mu}$. For more details, see on refs [9,10]. When the reference frame differs from crystallographic, the rotated $\hat{A}_{\lambda\mu}$ matrix can be obtained by means of classical tensor rotations. In this way, we obtain a set of effective elastic coefficients at any temperature in the suitable frame associated to any doubly rotated contoured resonator. Next we decompose the mechanical displacement on the triad of pure thickness solutions which would be occuring in an infinite thin plate with the same orientation. This procedure early presented by Tiersten and Stevens [13] is very convenient for studying quasi thickness vibration.

Considering a plane wave travelling along the normal to the plate, its velocity is obtained as one of the three solutions of Christoffel's eigenvalue problem, expressed in the rotated axes:

$$\left[\hat{A}_{2i2\ell} - C^\alpha \delta_{i\ell}\right]\left[V_\ell^\alpha\right] = \left[0\right] \qquad \alpha = 1,2,3 \tag{8}$$

where V_ℓ^α is a component of the normalized polarization vector, while C^α indicates one of the three effective stiffnesses for pure thickness modes.

α = 1 for a slow shear wave, α = 2 for an extensional one,

α = 3 for a fast shear wave, Let a_i and $\hat{u}_i$ be the variables (coordinates) and unknowns (displacement) fields expressed in the rotated axes (see on Fig. 1). In the case of the singly rotated AT-cut of our interest, the decomposition on the pure thickness components consists of replacing the $\hat{u}_i$ unknown field by a new u_i field, which must obey the following equations :

$$\frac{\partial \tilde{T}_{tp}}{\partial a_t} = \rho_o \ddot{u}_p \quad \text{in } V_o \tag{9}$$

$$\tilde{T}_{tp} = \tilde{A}_{tprs} \frac{\partial u_r}{\partial a_s} \tag{10}$$

$$\tilde{A}_{tprs} = Q_{pj} Q_{r\ell} \hat{A}_{tjs\ell} \tag{11}$$

where :

$$Q_{ij} = \delta_i^\alpha V_j^\alpha \,, \quad u_i = Q_{ir} \hat{u}_r \tag{12}$$

The partial derivatives equations (9) are logically equivalent to the usual form, despite from different meaning. For homogeneous boundary conditions, linearity and orthogonality of the transformation from $\tilde{P}_{ij}$ into $\tilde{T}_{tp}$ allow the use of transformed boundary conditions :

$$N_t \tilde{T}_{tp} = 0 \quad \text{on } S_o \tag{13}$$

The symmetry of the transformed coefficients is weak. We can only observe

$$\tilde{A}_{tprs} = \tilde{A}_{srpt} \tag{14}$$

Obviously $\tilde{A}_{2pr2} = \delta_{pr} C^{(r)}$. For convenience we shall omit the superscript $\sim$ in all the following. The subscripts (tprs) may be compressed into ($\lambda\mu$) in the following manner :

11 → 1		22 → 2		33 → 3		
12 → 6		23 → 4		31 → 5		(15)
21 → 9		32 → 7		13 → 8		

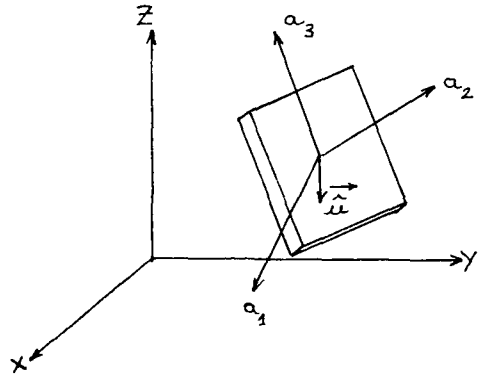


FIG. 1

3. Analytical model and Green's formalism

It may be observed that 7-th overtones of C mode and 6-th overtones of the not piezoelectrically driven B mode present very near frequencies in AT-cut. Remembering the vanishing stiffnesses in such a cut, we obtain the following set of transformed equations :

$$C^{(1)}u_{1,22}+A_{11}u_{1,11}+A_{58}u_{1,33}+2A_{56}u_{1,23}+(A_{12}+A_{66})u_{2,12}+(A_{14}+A_{56})u_{2,13}+\rho\omega^2u_1 = -(A_{13}+A_{55})u_{3,13}-(A_{17}+A_{86})u_{3,12} \quad (16)$$

$$C^{(2)}u_{2,22}+A_{69}u_{2,11}+A_{74}u_{2,33}+2A_{72}u_{2,23}+(A_{12}+A_{66})u_{1,12}+(A_{14}+A_{56})u_{1,13}+(A_{23}+A_{44})u_{3,23}+A_{65}u_{3,11}+A_{73}u_{3,33}+\rho_o\omega^2u_2 = 0 \quad (17)$$

$$C^{(3)}u_{3,22}+A_{85}u_{3,11}+2A_{37}u_{3,23}+(A_{23}+A_{44})u_{2,23}+A_{33}u_{3,33}+A_{65}u_{2,11}+A_{34}u_{2,33}+\rho_o\omega^2u_3 = -(A_{13}+A_{55})u_{1,13}-(A_{17}+A_{86})u_{1,12} \quad (18)$$

In these equations we have taken the liberty of using symmetry relation (14) in order to present each $A_{\lambda\mu}$ coefficient with the lowest possible value of λ. The relevant coupling terms between the C-modes (u_1) and the B-modes (u_3) are restricted to $(A_{17}+A_{86})u_{3,12}+(A_{13}+A_{55})u_{3,13}$ in (16), and the same with u_1 instead of u_3 in (18). Because of small magnitude of A_{17} and A_{86}, respectively close to usual C_{14} and C_{56}, these two elastic moduli can be ignored when elaborating a simple analytical model for the AT-cut. For instance, the X-Ray pattern of the small u_3 component of C-mode, fifth overtone has been found to be essentially governed by $A_{13}+A_{55}$ [14]. But, in spite of their small values, A_{17} and A_{86} involve a more important influence on the particular seventh overtone of C-mode, because of proximity of sixth overtones of B mode. This assumption will be successfully verified in what follows. It leads to ignore A_{13} and A_{55} in Eqs (16 - 18).

Our first approach consists in approximately solving these equations, while omitting the coupling terms located in the right hand members. In the case of thin plates, the quasi-thickness approximation (only $u_{i,2k}$ are retained in j-th equations, when $i \neq j$) enables to obtain very suitable equations of dispersion. In this view, the solutions are first considered in the following form :

$$u_i = (A_i \sin\eta a_2 + B_i \cos\eta a_2)\exp(-\xi a_1 - \nu a_3 + j\omega t) \quad (19)$$

Following Tiersten's and Steven's procedure [13], this form leads to simple equations of dispersion, after considering the boundary conditions on the major surfaces of the plate :

$$\frac{C^{(1)}n^2\pi^2}{4h^2} - \rho_o\omega^2 + M_n\xi^2 + P_n\nu^2 = 0 \text{ for C modes (major } u_1) \quad (20)$$

$$\frac{C^{(3)}n^2\pi^2}{4h^2} - \rho_o\omega^2 + R_n\xi^2 + T_n\nu^2 = 0 \text{ for B modes (major } u_3) \quad (21)$$

According to previous works of Tiersten and coworkers, it can be obtained :

$$M_n = C_{11} + r(C_{12}+C_{66}) + \frac{4}{n\pi\sqrt{C^{(1)}C^{(2)}}}\frac{(rC^{(2)}+C_{12})(rC^{(1)}-C_{66})}{\mathrm{tg}(n\pi\sqrt{C^{(1)}/C^{(2)}}/2)} \quad (22)$$

$$P_n = C_{58} - \frac{C_{56}^2}{C^{(1)}} \quad (23)$$

$$R_n = C_{85} \quad (24)$$

$$T_n = C_{33} - \frac{C_{23}^2}{C^{(3)}} + S(C_{44}+C_{23}) + S^2C_{22} + \frac{(rC^{(2)}+C_{23})(rC^{(3)}-C_{44}S^2)}{n\pi\sqrt{C^{(2)}C^{(3)}}\,\mathrm{tg}(n\pi\sqrt{C^{(3)}/C^{(2)}}/2)C_2} \quad (25)$$

$$r = (C_{12}+C_{66})/(C^{(1)}-C^{(2)}) \quad (26)$$

$$S = (C_{77}+C_{23})/(C^{(3)}-C^{(2)}) \quad (27)$$

After expanding $1/h^2$ $(1/h_o^2)(1 + (a_1^2 + a_3^2)/2Rh_o)$ in terms of radius of contour R and the distance to the plate center, and assimilating ξ^2 to $-\partial^2/a_j^2$ and ν^2 to $-\partial^2/a_3^2$ the equations (20,21) are transformed into separable partial derivatives equations where the thickness variable a_2 has been eliminated from. They lead to Hermite functions repartitions along a_1 and a_3 axes. If ω_i^n denotes the angular frequency of n-th overtone of the i-th pure thickness family, the ω_i^{nmp} eigenfrequencies corresponding to trapped u_i^{nmp} solutions are given by:

$$\frac{(\omega_i^{nmp})^2-(\omega_i^n)^2}{(\omega_i^n)^2} = \frac{1}{n\pi}\sqrt{\frac{2h_o}{R}}\left[(2m+1)\sqrt{\frac{K^{(i)}}{C^{(i)}}} + (2p+1)\sqrt{\frac{L^{(i)}}{C^{(i)}}}\right] \quad (28a)$$

$$u_i^{nmp} = N_i^{nmp} H_m(\sqrt{\alpha_n^{(i)}}\, a_1)\, H_p(\sqrt{\beta_n^{(i)}}\, a_3)\, \exp\left(\frac{-\alpha_n^{(i)} a_1^2 - \beta_n^{(i)} a_3^2}{2}\right) \quad (28b)$$

with

$$K^{(1)} = M_n,\ K^{(3)} = R_n,\ L^{(1)} = P_n,\ L^{(3)} = T_n \quad (28c)$$

$$\alpha_n^{(i)} = n\pi\sqrt{\frac{C^{(i)}}{8h_o^3\, R\, K^{(i)}}} \quad ,\quad \beta_n^{(i)} = n\pi\sqrt{\frac{C^{(i)}}{8h_o^3\, R\, L^{(i)}}} \quad (28d)$$

In addition, $2h_o$ denotes the maximum thickness (at the plate center), while H_m and H_p are Hermite polynomials, degrees m,p of which respectively define "anharmonicity numbers" along a_1 and a_3 axes. N_i^{nmp} is norm coefficient further precised. Thanks to use of Lagrangian formulation, $2h_o$ and R remain considered at T_o, when temperature varies. Parities of the u_i^{nmp} solutions are provided by n,m,p integers. The appoximate solutions (28b) only give us the major component of the u_i^{nmp} modes (u_1 for C, u_3 for B family). Assuming A_{14} and A_{86} govern most important coupling effects between both thickness modes families we are interested in, we have to consider respectively B and C modes with different parities along a_1 and a_2 axes, as a consequence of the concerned partial derivatives in Eqs (16 18). Thus we have selected a set of theoretical frequencies for an AT-cut resonator (θ = 35.4167°, $2h_o$ = 1.675 mm, R = 150 mm) at room temperature (results of Table I).

Mode	Freq (kHz)
C 7,0,0	7010.3
C 7,0,2	7111.6
C 7,2,0	7123.4
B 6,3,0	7027.9
B 6,1,2	7045.3
B 6,5,0	7127.3
B 6,3,2	7144.0
B 6,1,4	7160.8

TABLE I

When temperature increases, some B modes, which have got a negative thermal sensitivity about -250 Hz/K may cross the C modes, thermal slope of which is very small. In such cases, the coupling terms involved by A_{17} + A_{86} in equations (16)-(18) cannot be longer ignored. In a short form, those equations can be written :

$$A^o_{tprs}\, u_{r,st} + \rho_o\omega^2\, u_p = -A^+_{tprs}\, u_{r,st} = f_p \quad (29)$$

where the superscripts + and o respectively denote the coupling terms of our interest and the others. The displacement field at any material point P of the body is assumed to result from the action of elementary sources repartitions $f_j(P)$ whose "effectiveness" is determined by Green's tensor $G^i_j(P,Q)$[5]. When the boundary conditions do not play a particular part, the basic form is :

$$u_i(P) = \int_{V_o} G^i_j(P,Q)\, f_j\, dV_o(Q);\ \forall\ P,Q \text{ in } V_o \quad (30)$$

Replacing (30) into (29) and using the principle of reciprocity[5], we find the fundamental equation earlier presented by Tiersten[6] :

$$A^o_{tprs}\, G^k_{r,st} - \rho_o\omega^2\, G^k_p = \delta^k_p\, \delta(P-Q) \quad (31)$$

This equation remains available when the boundary conditions cannot be ignored. Following the clearly explained processes in ref[6], and taking note of the essential part of the symmetry relations (14) and of the fact that both starting and complete solutions verify homogeneous boundary conditions, we obtain :

$$u_i(P) = \int_{V_o} A^+_{tprs}\, u_{r,s}\, G^i_{p,t}\, dV_o \quad (32)$$

This form is more complete than (30) because it includes boundary conditions effect. Then we have to look for a precise expression of G^i_p. In this wiew we expand G^i_p on the solutions of the starting problem (without A^+_{tprs}). These solutions can be symbolically identified by Greek superscript like α or β. Since the starting problem obeys homogeneous boundary solutions, its solutions u_k have got orthogonality properties[5], use of which leads to :

$$G^k_p(P,Q) = \sum_\alpha K^k_\alpha(P)\, u^\alpha_p(Q) \quad (33)$$

where

$$K^k_\alpha = \frac{u_k}{\omega^2 - \omega_\alpha^2} \quad (34)$$

Equations (32, 33, 34) are rigorous if the u_k solutions are *exactly the solutions of (29) when* A^+_{tprs} *are* ignored. As a matter of fact, those exact solutions may be satisfactorily approximated by u_k^{nmp} modes of equation (28b).

Because the minor components are ignored in (n,m,p) modes, G^k_p may be neglected whenever k differs from p. Thus we obtain the following integral equations, for the treatment of our special problem :

$$u_1 = \sum_\gamma \frac{u_1^\gamma}{\omega^2 - \omega_\gamma^2}\int_{V_o}(A_{86}\, u^\gamma_{1,2}\, u_{3,1} + A_{17}\, u^\gamma_{1,1}\, u_{3,2})dV_o \quad (35\text{-a})$$

$$u_3 = \sum_\beta \frac{u_3^\beta}{\omega^2 - \omega_\beta^2}\int_{V_o}(A_{17}\, u^\beta_{3,2}\, u_{1,1} + A_{86}\, u^\beta_{3,1}\, u_{1,2})dV_o \quad (35\text{-b})$$

The γ or β indexes symbolyse any (n,m,p) set with suitable parities appearing on Table I. In order to solve (35), we have choosen to decompose the u_1 and u_3 component of the unknown solution on the basis of the u_1 and u_3 solutions of the simplified starting model :

$$u_k = C_\lambda\, u^\lambda_k \quad (36)$$

where C_λ are unknown coefficients.

In this manner the general following equations is obtained :

$$C_\lambda\, u^\lambda_k = \frac{u^\alpha_k}{\omega^2 - \omega_\alpha^2}\, C_\mu\, Q_{\alpha\mu} \quad (37)$$

where all repeated indexes must be summed. $Q_{\alpha\mu}$ terms are defined by :

$$Q_{\alpha\mu} = Q_{\mu\alpha} = \int_{V_o} (A_{86}\, u^{\alpha}_{1,2}\, u^{\mu}_{3,1} + A_{17}\, u^{\alpha}_{1,1}\, u^{\mu}_{3,2})\, dV_o \qquad (38)$$

We have already explained that solutions of the starting model are orthogonal. Provided that a suitable norm N_i^{nmp} has been chosen for trapped modes given by relation (28b), we can recall here the well known orthogonality relation :

$$\int_{V_o} \rho_o\, u^{\alpha}_k\, u^{\mu}_k\, dV_o = \delta_{\alpha\mu} \qquad (39)$$

where δ denotes Kronecker's symbol. Multiplying (37) by u_k, integrating on V_o and with help of (39), it can easily be obtained :

$$C_\gamma = \frac{C_\mu Q_{\gamma\mu}}{\omega^2 - \omega_\gamma^2} \qquad (40)$$

In this way, we have converted the integral equations (35) into a homogeneous linear system :

$$[S_{\gamma\mu}][C_\mu] = [0] \qquad (41)$$

in which :

$$S_{\gamma\mu} = \frac{Q_{\gamma\mu}}{\omega^2 - \omega_\gamma^2} - \delta_{\gamma\mu} \qquad (42)$$

All $Q_{\gamma\gamma}$ elements vanish[7]. Since we are expecting non vanishing frequency shifts, each γ-th row of the S-matrix may be multiplied by $(\omega^2 - \omega_\gamma^2)$. So we are obtaining a classical eigenvalues/eigenvectors problem :

$$[E_{\gamma\mu}][C_\mu] = \omega^2\, [C_\mu] \qquad (43)$$

with

$$E_{\gamma\mu} = Q_{\gamma\mu} + \omega_\gamma^2\, \delta_{\gamma\mu} \qquad (44)$$

Normalized eigenvectors C_μ will give us the contributions of each u_1^{nmp} and u_3^{qrs} solution in the mixing which constitutes each actual mode.

4. Results and Experiment

The parameters of the resonator we have studied are those corresponding to Table I. We have solved the system (43) including the eight modes appearing on Table I. Remember that the indexes (n,m,p) respectively indicate the number of nodal lines along (a_2, a_1, a_3) axes.

Since the starting model and the T^n_{CIJ} measured values are not perfectly accurate, we had to look for the strong couplings in a relatively large temperature range. But for convenience in mounting a heating system on X-Ray topographs apparatus, we had better investigate some couplings terms at reasonably high temperature (from 20°C to 120°C). Among the above mentionned modes, the two following couplings were expected to occur in such a range : The C (7,2,0) and B (6,5,0) modes were first predicted to interact about 41°C. A C (7,0,0) - B (6,3,0) coupling was then expected about 99°C.
The corresponding experimental temperatures respectively amount to 60°C and 78°C.

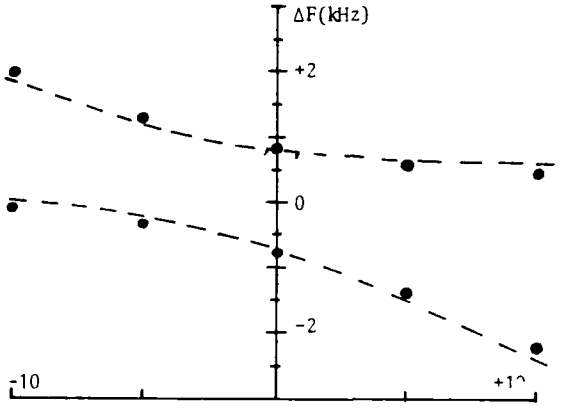


FIG. 2

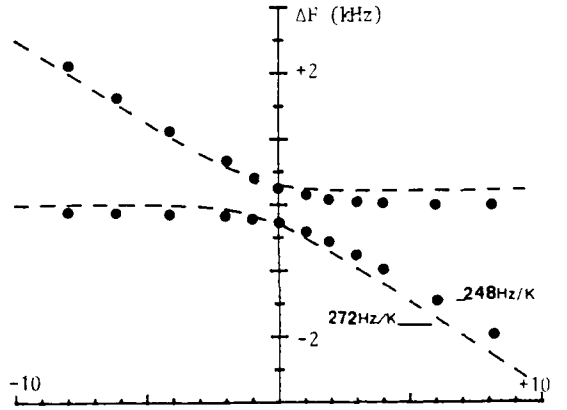


FIG. 3

We are presenting on Fig 2 and Fig 3 theoretical (dashed lines) and measured (dots) frequency temperature curves of each coupling. In order to make the comparison easier, we have chosen to place origins of both theoretical and observed results at the temperatures and frequencies of the "crossing point", which would be observed without coupling. In the case of Figure 2, the measured minimum gap between the two modes amounts to 1520 Hz in agreement with the predicted value, amounting to 1523 Hz. In the case of Figure 3, owing to smaller value of $Q_{\alpha\beta}$ integral, the coupling is weaker : the frequencies gaps amount then to 557 Hz (measured) and 562 Hz (predicted). In addition we give on Figs 2 and 3 the theoretical normalized contribution of the relevant C mode to each of both actual modes. As complement we give on Tables II and III theoretical values of both main components at various temperatures.

		C_α coefficient	
T°C	Freq(Hz)	λ = (C,7,2,0)	μ = (B,6,5,0)
30	7126 265	- 0.261	0.965
	7123 321	0.962	0.262
41.5	7124 291	0.706	- 0.706
	7122 768	0.704	0.709
50	7123 813	0.941	- 0.331
	7121 331	0.330	0.944

TABLE II

T°C	Freq(Hz)	λ = (C,7,0,0,)	μ = (B,6,3,0)
90	7012 756	- 0.117	0.990
	7010 375	0.993	0.116
98.8	7010 748	0.712	- 0.701
	7010 186	0.702	0.710
110	7010 613	0.995	- 0.092
	7007 486	0.092	0.993

TABLE III

In each case, since $(C_\lambda)^2 + (C_\mu)^2 \simeq 1$, the effect of the other modes appearing in Table I remains negligible in the studied temperature range. Until now, influence of piezoelectricity has been neglected in the statement. Usually the C modes seventh overtone number can be piezoelectrically driven but not the B modes (with any overtone number). This fact remains available far from the strongest coupling point. At the inverse, in the region of strong coupling, because actual modes must be considered as mixings of B and C simple modes, an actual mode having a major B component can be piezoelectrically driven as a consequence of its minor C component. The larger will be the C component, the lower will be the resistance of the actual mode. At the strongest coupling point, it becomes quite impossible to distinguish the actual modes : both coupled actual modes are then identical, with equal contributions of B and C components. When the resonator is inserted in an oscillator, a frequency jump from one mode to the other can be observed, in relation with the temperature dependence of the motional resistance.

In order to verify the occuring modal shapes, we performed some X-Ray patterns, by use of reflections on $(2\bar{1},0)$ and $(0\bar{1},2)$ planes of the crystal lattice. The former kind of reflection is only affected by the u_1 mechanical displacement and give us the modal pattern of the "C" component of the mode. The latter kind is essentially affected by the mechanical displacement along Z' axis of the Y + θ rotated plate, and also by the displacement along the normal to the plate, but in lower proportion. Therefore it is absolutely insensitive to u_1 component of the vibration. This reflection is thus quite suitable for investigating "B" components, close to Z' axis (3.45°) in AT-cut.

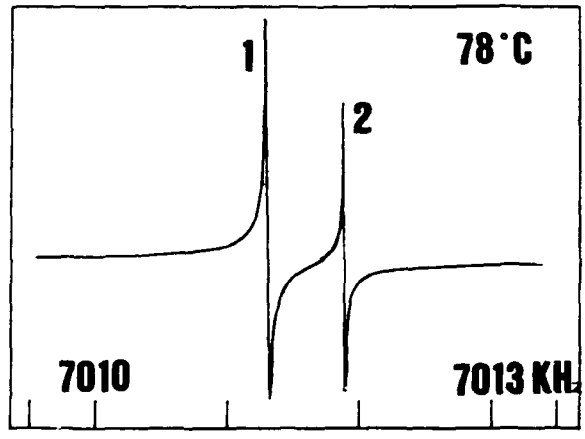


FIG. 4

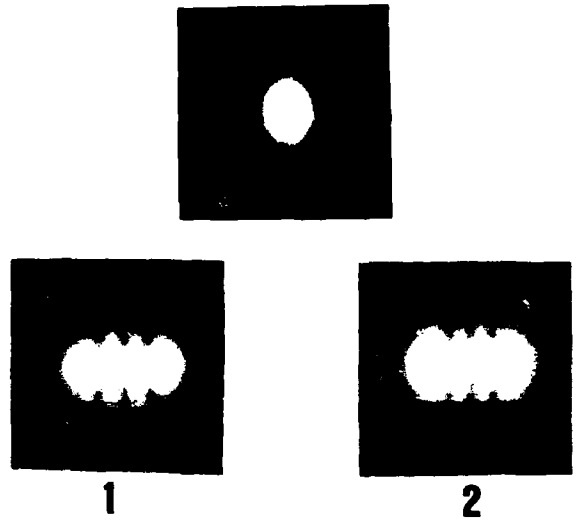


FIG.5

On Fig 4 we show the frequency response for the C(7,0,0) - B(6,3,0) coupling near the strongest coupling point. Corresponding topographs are put underneath on Fig 5. The crystallographic X (or a_1) axis is kept horizontal on the presented topographs, while a_3 axis is vertical. The upper topograph shows the u_1 pattern (absolutely identical for both resonances) while the lower ones show the u_3 patterns, in agreement with m = 3, p = 0 in the u_{3mnp} identification. In addition, reviewing Fig 3 it can be observed that the thermal slope of the frequency, far from the strongest coupling point, denotes a good agreement between theoretical and experimental values. This fact confirms the modal identification.

At last, figures 6a, b, c are showing frequency responses for C(7,2,0) - B(6,5,0). Coupling at various temperatures. X-Rays patterns corresponding to the strongest coupling point (case of Fig. 6b) are given on Fig 7. (See also on Ref[8]). As a complement, we have performed (Fig 8) some section topographs giving the displacement repartition along the plate thickness. On Fig. 8 we

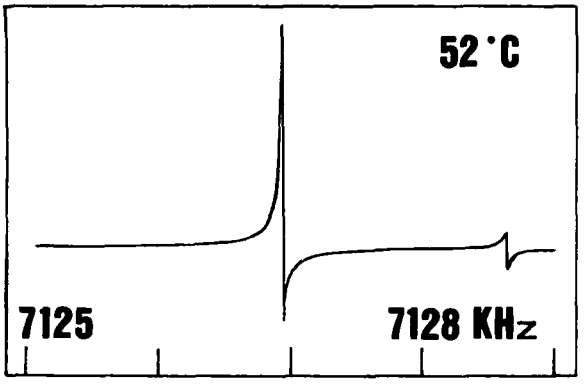


FIG. 6a

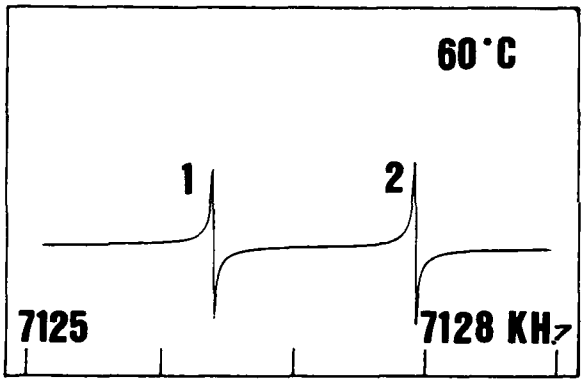


FIG. 6b

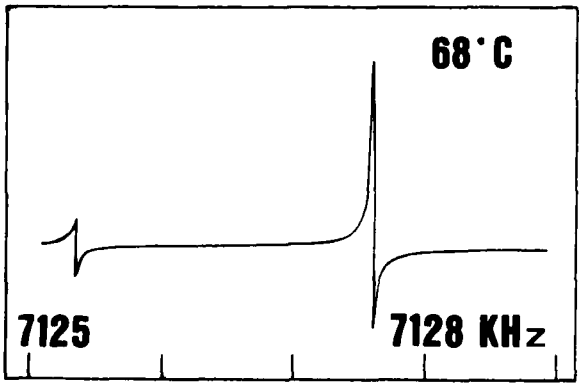


FIG.6c

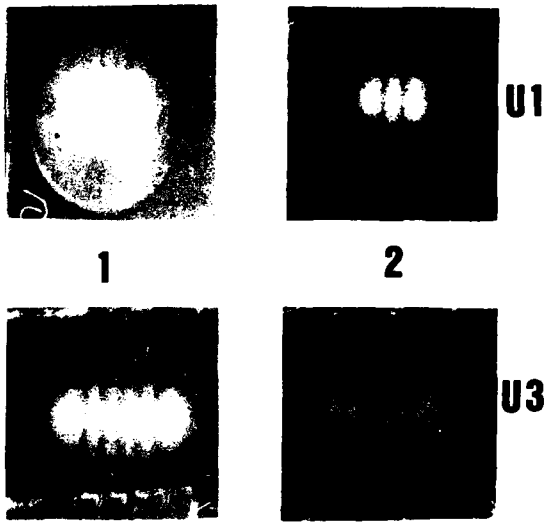


FIG. 7

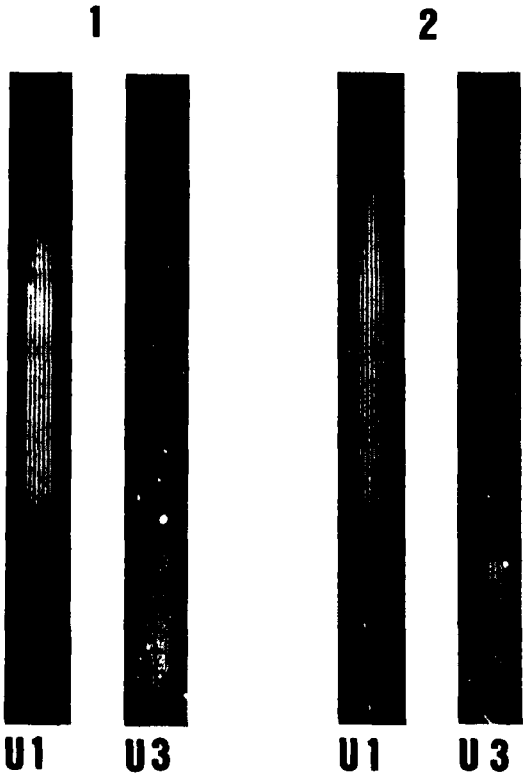


FIG. 8

keep a_3-axis vertical for u_1 component but a_1-axis vertical for u_3 component. Those section topographs confirm the seventh overtone structure for u_1 component and the sixth overtone structure for u_3 component. Because of enlargement we must restrict those here presented section-topographs to a section close to the center of the plate.

Far from the special couplings of our interest, the u_3 patterns of C modes can look different from the here presented ones. As control samples, we are showing on Fig 9 u_1 (upper) and u_3 (lower) topographs for the metrological C(5,0,0) mode and C(5,2,0) anharmonic mode, in the same resonator. For these modes no frequency splitting can be observed when temperature is varying. That is the most frequently occuring case. The observed u_3 patterns, governed by $(A_{13} + A_{55})$ elastic coefficient, are in good agreement[14] with theoretical analysis of Ref[13]. Including both kinds of coupling terms involved by $(A_{13} + A_{55})$ and $(A_{17} + A_{86})$, it seems possible to extend the system (43)[55] in order to generalize the method to all possible cases.

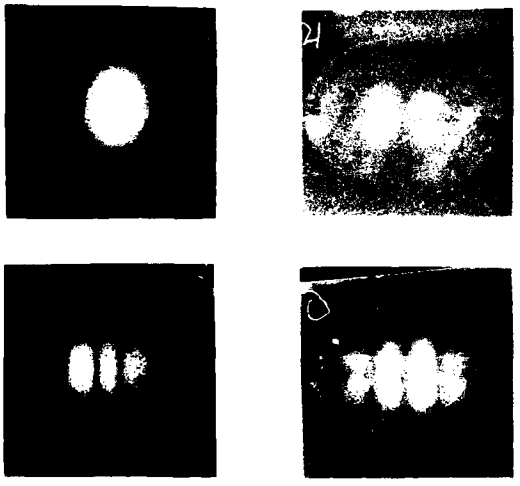


FIG. 9

5. Motional resistance and frequency jumps

According to our special definition of thermal derivatives of elastic stiffnesses, we neglected influence of piezoelectricity in the preceding development. As a matter of fact, piezoelectricity slightly modifies the resonant frequencies but has no influence on integral terms $Q_{\gamma\mu}$. Because the C_λ solutions of the system (43) are only depending on $Q_{\gamma\mu}$ values and on the relative values of the resonant frequencies obtained in the starting model, we may guess that C_λ coefficients are correctly given by the preceding procedure. Then piezoelectric effect only involves small shifts on the location of the "strongest coupling point". This fact provides no difficulty if we choose this point as the origin of frequencies and temperatures.

In the strong coupling region, each actual mode is given by :

$$u_1^\gamma = \lambda M_\gamma \, u_1^{nmp}(a_1,a_2,a_3) \exp(j\omega t) \qquad (45\text{-a})$$

$$u_2^\gamma \text{ neglected} \qquad (45\text{-b})$$

$$u_3^\gamma = \mu M_\gamma \, u_3^{qrs}(a_1,a_2,a_3) \exp(j\omega t) \qquad (45\text{-c})$$

where λ and μ are the relevant values of normalized C_γ coefficients for the actual mode of our interest. M_γ is a parameter, role of which will be explained further. In fact, temperature dependence of C_γ values is considerably higher than temperature dependence of all other occuring physical quantities (especially piezoelectric and elastic coefficients, and the dimensions and mass density of the studied resonator). As a consequence we will get an approximate but reliable model when substituting eqs (45) into classical equations of piezoelectricity in which all temperature dependences are ignored but for λ and μ. Obviously this approach holds only when investigating a selected actual mode of interest.

In the framework of quasi-thickness approximation[1] all electrical variables vary only along the thickness of the plate. If D_i and φ denote the electric displacement and the electrostatic potential, respectively, the following approximate constitutive equation is obtained, for an AT-cut :

$$D_2 = -\varepsilon_{22}\varphi_{,2} + e_{26}\hat{u}_{1,2} \qquad (46)$$

with corresponding equation of electrical charge conservation :

$$e_{26}\hat{u}_{1,22} - \varepsilon_{22}\varphi_{,22} = 0 \qquad (47)$$

Since the decomposition of the mechanical displacement along the triad of pure thickness solutions does not change the C mode of an AT cut, equations (46-47) remain identical when the transformation (12) is performed. Then the pertinent boundary conditions on the major surfaces of the plate are :

$$c^{(1)}u_{1,2} + e_{26}\varphi_{,2} = 0, \quad c^{(3)}u_{3,2} = 0, \varphi = \frac{1}{2}Ve^{j\omega t}, \qquad (48)$$

at $x_2 = \pm h_o$

where V is the driving voltage.

These boundary conditions may be changed into a homogeneous form, by use of a well known transformation :

$$u_1 = \bar{u}_1 + Kx_2, \; u_3 = \bar{u}_3, \; \varphi = \bar{\varphi} + Vx_2/2h_o, \; K = -e_{26}V/2h_oc^{(1)} \qquad (49)$$

When expressed in terms of $\bar{u}$ and $\bar{\varphi}$, boundary conditions become homogeneous, but equations of elasticity become unhomogeneous :

$$\Lambda_{ij}\bar{u}_j - \rho_o\omega^2\bar{u}_i = -\rho_o\omega^2 Kx_2\delta_{i1} = F_i \qquad (50)$$

where Λ is elasticity operator in a matrix form ($c^{(1)}$ has been replaced in it by the stiffened elastic constant $c^{(1)} = c^{(1)} + e_{26}^2/\varepsilon_{22}$).

When following Green's formalism for the problem we are now interested in, it should be observed that,

because F_j does not depend on mechanical displacement, we may use the simple general form (30) corresponding to the case where the sources repartition has no influence on the boundary conditions. Thus, provided that solutions of (50) without the right hand member are orthogonal and normalized, we may use the general decomposition (33, 34) for Green's tensor. In this way, the integral equation associated to the driving of actual vibration is obtained as follows (the resonator s assumed to be fully electroded) :

$$\bar{u}_i = \sum_{\gamma} \frac{\bar{u}_i^{\gamma}}{(\omega^2 - \omega_{\gamma}^2)} \int_{V_o} \bar{u}_j . F_j \, dV_o \qquad (51)$$

where $\bar{u}_i^{\gamma}$ is obtained from eqs (45) and (49) and denotes a free actual mode obtained when the right hand member in (50) remains neglected. Equation (51) holds only if orthogonality relation (39) remains available for here studied $\bar{u}^{\gamma}$ modes. In this view we have to renormalize $\bar{u}^{\gamma}$ solutions. If u_i^{rst} theirselves obey orthogonality relation (39), we can easily obtain :

$$M_{\gamma} = \frac{1}{\sqrt{\lambda^2 + \mu^2}} \qquad (52)$$

To compute (51), we may notice that in the vicinity of a free actual mode of eigenfrequency , the α-th term of the sum dominates in (51). Then the mechanical displacement of a driven solution is given by :

$$\bar{u}_i \simeq \frac{\bar{u}_i^{\alpha}}{\omega^2 - \omega_{\alpha}^2} \int_{V_o} - \rho_o \omega^2 \, Kx_2 \, \bar{u}_1^{\alpha}(Q) \, dV(Q) \; ; \; i = 1,3 \qquad (53)$$

As a consequence, the driven actual solutions are presenting the same shape as the free actual solutions in (45), but with a different amplitude factor A_{α} instead of the normalization factor M_{α}. We have found :

$$\bar{u}_1 \simeq \lambda \, A_{\alpha} \, (u_1^{nmp} - Kx_2) \qquad (54\text{-a})$$

$$\bar{u}_3 \simeq \mu \, A_{\alpha} \, u_3^{qrs} \qquad (54\text{-b})$$

with

$$A_{\alpha} = \frac{\lambda \, e_{26} \, V \, \rho_0 \int_{V_o} K x_2 \, u_1^{nmp} \, dV_o}{(\lambda^2 + \mu^2) \, 2h_o C^{(1)} \left(1 - \frac{\omega_{\alpha}^2}{\omega^2}\right)} \qquad (55)$$

At this point, the displacement amplitude remains infinite when $\omega = \omega_{\alpha}$. We may introduce damping effect due to internal losses in the resonator by use of complex notation for the angular frequency :

$$\omega^* = \omega(1 + j/2Q) \qquad (56)$$

We may calculate the mechanical Q factor from the ratio of stored energy to lost energy during a period. Since the actual mode is a mixing of u_1^{nmp} and u_3^{qrs} components, mainly depending on a_2, it can be written :

$$Q = \frac{\lambda^2 C^{(1)} \int_{V_o} (\bar{u}_{1,2})^2 dV_o + \mu^2 C^{(3)} \int_{V_o} (\bar{u}_{3,2})^2 dV_o}{\omega \left[\lambda^2 \eta^{(1)} \int_{V_o} (\bar{u}_{1,2})^2 \, dV_o + \mu^2 \eta^{(3)} \int_{V_o} (\bar{u}_{3,2})^2 dV_o \right]} \qquad (57)$$

This form leads to :

$$Q = \frac{1}{\omega} \frac{\lambda^2 n^2 C^{(1)} + \mu^2 q^2 C^{(3)}}{\lambda^2 n^2 \eta^{(1)} + \mu^2 q^2 \eta^{(3)}} \quad (\text{here } n = 7 \; , \; q = 6) \qquad (58)$$

In this equation, $\eta^{(1)}$ and $\eta^{(3)}$ are the damping coefficients associated with slow shear and fast shear waves. They are obtained from a tensorial rotation performed on values of Ref[15] and assuming that the very weak losses do not modify the polarization vector of plane waves solutions of (18). $\eta^{(3)}$ is then obtained from the rotated η_{ij} coefficients, by use of a transformation similar to (11) :

$$\eta^{(3)} = Q_{3j} \, Q_{3l} \eta_{2j2l} \qquad (59)$$

Finally, the motional resistance is obtained from the calculated value of dissipated power :

$$R = \frac{V_o^2}{2 P_{lost}} \qquad (60)$$

Remembering that $\lambda^2 + \mu^2 \simeq 1$, we can obtain the variation of resistance due to coupling effect, in a very simple form :

$$K = \frac{R}{R_o} = \frac{n^2(\lambda^2 n^2 + \mu^2 q^2)}{\lambda^2(\lambda^2 n^2 + \mu^2 q^2 C)^2} \qquad (61)$$

with :

$$\eta = \eta^{(3)}/\eta^{(1)} \quad \text{and} \quad C = C^{(3)}/C^{(1)}$$

R_o denotes the resistance of the corresponding C_o mode when its coupling with B modes is ignored. We can deduce from (61) the ratio S of the resistance of one coupled mode (upper frequency for instance) to the other (lower frequency). We show on Table IV some theoretical and experimental values for C(7,0,0) - B(6,3,0) coupling.

T°C	S_{th}	T_{exp}	S_{exp}
96	5,85	75,5	6,48
98	1,72	77,5	1,82
99	0,852	78,5	1,08
100	0,426	79,5	0,612
102	0,136	81,5	0,202

TABLE IV

When inserted in an oscillator, the resonator is first running on the frequency of C mode at room temperature. When temperature increases, this mode becomes coupled with the relevant B mode, then the oscillator go on running on the same frequency temperature curve, but as the B mode contribution becomes very strong, the resistance of the actual mode quickly increases and the oscillator may then jump to the other mode of lower resistance. For instance, such discontinous frequency-temperature curve is shown on fig. 10.

6. Conclusion

The here studied effect does not be considered as a rare occurence. First, contouring generates anharmonic modes in large frequency ranges. In addition, some $Q_{\gamma\mu}$ integrals like (38) but eliminated by symmetry considerations may involve some spurious couplings in case of defects in the resonator geometry (radius of contour, location of electrodes...). As a general result, this study clearly implies that many defects of the frequency temperature curves are related to vibration modes and spectra, even if the material is considered to be perfect. The method we explained here will permit further progress in the study of some doubly rotated cuts in which "activity dip" phenomenon is stronger than in here presented AT cut.

Acknowledgements

This work was supported by the "Direction des Recherches, Etudes et Techniques" office (groupe 1, Contrat N° 84-012).

References

(1) 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).

(2) C.J. Wilson "Vibration modes of AT cut convex quartz resonators" J. Phys. D : Appl. Phys., Vol. 7, p 2449, (1974).

(3) R. Bourquin, D. Nassour, D. Hauden, "Amplitude frequency effect of SC-cut quartz trapped energy resonators" 36th Annual Freq. Cont. Symp. pp 200-207 (1982).

(4) H.F. Tiersten and R.C. Smythe, "An Analysis of overtone modes in contoured crystal resonator" Proc. 31th A.F.C.S., pp 44-47, (1977).

(5) P.M. Morse, H. Feshbach, "Methods of theoretical Physics" Mc Graw Hill, New-York, (1953).

(6) H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a bias" J. Ac. Soc. Am 64 (3), pp 832-837 (1978).

(7) B. Dulmet, "Application d'une méthode de perturbation à l'étude de résonateurs à quartz présentant des modes d'épaisseur à énergie piégée" Revue Phys. Appl. 19, pp 839-849 (1984).

(8) B. Bourquin, B. Dulmet, G. Genestier, "Jumps in frequency temperature response of contoured resonators : an analysis performed with a perturbation model and X-Ray patterns" Proc. 1984 Ultrasonics Symposium IEEE, pp 394-399 (1984).

(9) B. Dulmet, R. Bourquin, "Méthodes de détermination des courbes fréquence-température d'un résonateur à quartz vibrant en mode d'épaisseur" Revue Phys. Appl. 18, pp 619-624 (1983).

(10) R. Bourquin, B. Dulmet, "Méthode rapide de calcul des courbes fréquence-température des résonateurs à quartz : Application au barreau vibrant en flexion" XIème Congrès International de Chronométrie, Besançon (1984).

(11) R.N. Thurston, "Waves in Solids" Handbuch der Physik VI a/4 Springer Verlag (Berlin) (1974).

(12) R. Bechmann, A. Ballato, T. Lukaszed, "High Order Temperature Coefficients of the Elastic Stiffnesses and Compliances of quartz" Proc. I.R.E. p. 1812 (1962).

(13) H.F. Tiersten, D.S. Stevens, "An Analysis of contoured SC cut Quartz Crystal Resonators" Proc. 36th A.F.C.S. (1982).

(14) R. Bourquin, D. Nassour, "Etude tridimensionnelle d'un résonateur à quartz de coupe AT, de forme plan-convexe" C.R. Acad. Sci. Paris, t. 298, Série II n°12, p.517 (1984).

(15) J. Lamb, J. Richter, "Anisotropic Acoustic Attenuation with New Measurement for quartz at room temperature", Proc. Roy. Soc. London 293 A, pp 479-492 (1966).

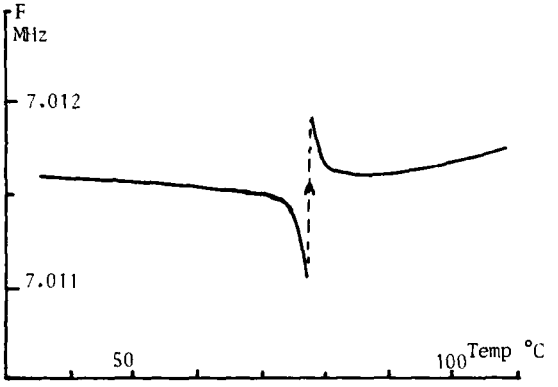


FIG. 10

FREQUENCY-TEMPERATURE BEHAVIOR OF FLEXURAL AND THICKNESS-SHEAR VIBRATIONS OF RECTANGULAR ROTATED Y-CUT QUARTZ PLATES

Y.K. Yong
Department of Civil and Environmental Engineering
Rutgers University
P.O. Box 909
Piscataway, New Jersey 08854
(201) 932-3219

P.C.Y. Lee
Department of Civil Engineering
Princeton University
Princeton, New Jersey 08544
(609) 452-4610

Abstract

While the classical frequency equation for flexural vibrations of beams or plates gives good predictions of resonances at low frequencies, it has been found that far less accurate results are obtained when the same frequency equation is used with temperature coefficients for material properties to predict the frequency-temperature behavior. One of the causes of this discrepancy is usually attributed to the lack of accuracy in the values of temperature coefficients and hence various modifications of temperature coefficients of elastic stiffness has been introduced.

The classical frequency equation ignores the lateral dimension or boundary conditions on the lateral faces. Since the quartz plate is subjected to a thermally biased deformation, the lateral dimension or lateral boundary conditions should have significant influence on the frequency-temperature behavior. In the present paper, rectangular, rotated Y-cut plates are considered. In order to take into account the lateral boundary conditions and allow simple closed-form solutions, a set of coupled one-dimensional equations is derived from the two-dimensional plate equations for incremental vibrations in a thermally biased plate by expanding the displacements in a power series of the width coordinate and integrating through the width of the plate.

A set of four coupled one-dimensional equations are extracted from the infinite set to study the frequency-temperature behavior of flexural and thickness-shear vibrations of rectangular, rotated Y-cut plates. The AT cut plate is found relatively insensitive to changes in the width to thickness ratio. For low-frequency flexural vibrations, the frequency-temperature behavior is sensitive to changes in the width to thickness ratio. For a Y-cut plate with length to thickness ratio equal to 20, the turn-over temperature increases by 40°C when the width to thickness ratio is reduced from 20 to 2.5. Good agreement between the theoretical results and experimental data by Paros and Singer is obtained.

I. Introduction

Although the classical frequency equation for flexural (or extensional) vibrations of beams or plates gives good predictions of resonances at low frequencies, it is often found that far less accurate results are obtained when the same frequency equations is used with temperature coefficients for material properties to predict the frequency-temperature behavior. One of the causes of this discrepancy is usually attributed to the lack of accuracy in the values of temperature coefficients and hence various modifications of temperature coefficients of elastic stiffness have been introduced [1-6].

We note that the frequency equations employed for predicting resonance frequencies and frequency-temperature behavior are usually derived from the one-dimensional governing equation or straight-crested wave solution of the two-dimensional plate equations, and, in either case, the lateral dimension or boundary conditions on the lateral faces are ignored. Mindlin and Gazis [7] demonstrated that strong resonances can be accurately predicted by straight-crested wave solutions ignoring the boundary conditions on the lateral faces. For the frequency-temperature behavior of a quartz plate, the plate is subjected to a steady and uniform temperature increase and due to the thermally induced deformations, the lateral dimension or lateral boundary conditions should have significant influence on frequency changes. Hence, the lateral dimension or lateral boundary conditions should not be ignored.

In the present paper, rectangular, rotated Y-cut quartz crystal plates vibrating predominantly in the length direction are considered. In order to take into account the lateral boundary conditions and allow simple closed-form solutions, the two-dimensional plate equations for incremental vibrations in a thermally biased plate [8] with traction-free face conditions are reduced to an infinite set of one-dimensional equations by expanding the displacements in a power series of the width coordinate and integrating through the width of the plate. The zeroth and first order two-dimensional plate equations are used.

To study the frequency-temperature behavior of flexural and thickness-shear vibrations of rectangular, rotated Y-cut quartz plates, four coupled one-dimensional equations are extracted from the infinite set. These equations accommodate the thickness-length flexure (LF), thickness-width flexure (WF) and thickness shear (TS) modes of vibrations. Rotated Y-cut quartz plates have monoclinic symmetry and hence is uncoupled from the extensional or torsional vibration. Closed-form solutions are obtained which satisfy traction-free conditions on all four edges of the plate, or one pair of free edges with a pair of fixed edges. Dispersion relations and frequency spectra are calculated. The changes of frequencies of flexural as well as thickness-shear modes as functions of length to thickness ratio, width to thickness ratio and angle of plate orientation are computed.

For low-frequency flexural vibrations, it is found that the frequency-temperature of the plate is, indeed, sensitive to the changes of the width to thickness ratios. For example, for a Y-cut quartz plate with length to thickness ratio equal to 20, the turn-over temperature can increase about 40°C when the width to

CH2186-0/85/0000-0415$1.00©1985IEEE

thickness ratio is changed from 20 to 2.5. Results of the computed frequency changes as a function of temperature increment are compared to the experimental data by Paros and Singer and good agreement is obtained.

II. One-Dimensional Equations of Motion for Small Amplitude Waves Superposed on Homogeneous Thermal Strain

Consider a rectangular plate with the dimensions and coordinate axes shown in Fig. 1. The major surface of the plate and the lateral or width edges at $x_3 = \pm c$ are traction free. The length edges at $x_1 = \pm a$ can be traction free, fixed or simply supported, depending on any particular problem of interest. Since the one-dimensional equations are derived from the two-dimensional equations, we first review from reference [8] and [9] the two-dimensional plate equations for incremental vibrations in a thermally biased plate. The zeroth and first order variational stress equations of motion for the plate subjected to a thermally biased homogeneous strain [8] are:

$$\int_{-a}^{a}\int_{-c}^{c}[\beta_{ki}t^{(0)}_{kj,j}-2b\rho\ddot{u}^{(0)}_i]\delta u^{(0)}_i dx_3 dx_1 = 0 \tag{1a}$$

and

$$\int_{-a}^{a}\int_{-c}^{c}[\beta_{ki}t^{(1)}_{kj,j}-\beta_{ki}t^{(0)}_{k2}-\frac{2b^3}{3}\rho\ddot{u}^{(1)}_1]\delta u^{(1)}_i dx_3 dx_1 = 0 \tag{1b}$$

where the indices i,j or k = 1,2,3; dummy indices signify summation; and a comma between the indices indicates partial differentiation with respect to the spatial variables, x_i. The terms $t^{(n)}_{kj}$ and $u^{(n)}_i$ are respectively the two-dimensional, n-th order incremental stress and displacement. The term β_{ki} is given by the relation [8,9]:

$$\beta_{ki} = \delta_{ki} + \alpha^{\theta}_{ki} \tag{2}$$

where δ_{ki} is a Kronecker delta and

$$\alpha^{\theta}_{ki} = \alpha^{(1)}_{ki}\cdot\theta + \alpha^{(2)}_{ki}\cdot\theta^2 + \alpha^{(3)}_{ki}\cdot\theta^3 \tag{3}$$

The n-th order linear thermal expansion coefficients, $\alpha^{(n)}_{ki}$, are measured by Bechmann, Ballato and Lukaszek [1]. The temperature change from the reference temperature, T_0, is $\theta = (T-T_0)°C$. Associated with the variational stress equations of motion are the variational equations which yields the necessary and sufficient boundary conditions for the plate:

$$\int_C[p^{(0)}_i - n_j\beta_{ik}t^{(0)}_{kj}]\delta u^{(0)}_i dC = 0 \tag{4a}$$

$$\int_C[p^{(1)}_i - n_j\beta_{ik}t^{(1)}_{kj}]\delta u^{(1)}_i dC = 0 \tag{4b}$$

where C defines the edges of the plate and $p^{(n)}_i$ is the two-dimensional, n-th order incremental surface traction.

The two-dimensional zeroth and first order incremental strain-displacement relations are respectively:

$$e^{(0)}_{ij} = \frac{1}{2}[\beta_{kj}u^{(0)}_{k,i}+\beta_{ki}u^{(0)}_{k,j}+\delta_{2i}\beta_{kj}u^{(1)}_k+\delta_{2j}\beta_{ki}u^{(1)}_k] \tag{5a}$$

and

$$e^{(1)}_{ij} = \frac{1}{2}[\beta_{kj}u^{(1)}_{k,i}+\beta_{ki}u^{(1)}_{k,j}+2(\delta_{2i}\beta_{kj}u^{(2)}_k+\delta_{2j}\beta_{ki}u^{(2)}_k)] \tag{5b}$$

The relevant constitutive relations, or the two-dimensional stress-strain-temperature relations are

$$t^{(0)}_{ij} = 2bD_{ijk\ell}e^{(0)}_{k\ell} \tag{6a}$$

$$t^{(1)}_{ij} = \frac{2b^3}{3}D_{ijk\ell}e^{(1)}_{k\ell} \tag{6b}$$

where [9]

$$D_{ijk\ell} = C_{ijk\ell} + D^{(1)}_{ijk\ell}\cdot\theta + D^{(2)}_{ijk\ell}\cdot\theta^2 + D^{(3)}_{ijk\ell}\cdot\theta^3 \tag{7a}$$

and

$$D^{(1)}_{ijk\ell} = C^{(1)}_{ijk\ell} + C_{ijk\ell mn}\alpha^{(1)}_{mn} \tag{7b}$$

$$D^{(2)}_{ijk\ell} = \frac{1}{2}\tilde{C}^{(2)}_{ijk\ell} + C_{ijk\ell mn}\alpha^{(2)}_{mn} \tag{7c}$$

$$D^{(3)}_{ijk\ell} = \frac{1}{6}\tilde{C}^{(3)}_{ijk\ell} + C_{ijk\ell mn}\alpha^{(3)}_{mn} \tag{7d}$$

$C_{ijk\ell}$ and $C_{ijk\ell mn}$ are the second and third order elastic stiffness of quartz; while $C^{(1)}_{ijk\ell}$, $\tilde{C}^{(2)}_{ijk\ell}$ and $\tilde{C}^{(3)}_{ijk\ell}$ are respectively the first temperature derivatives, second effective temperature derivatives and third effective temperature derivatives. Values of the second and third order elastic stiffness of quartz were measured respectively by Bechmann, Ballato and Lukaszek [1] and Thurston, McSkimin and Andreatch [10]. Values of the temperature derivatives were calculated and reported in references [8] and [9]. The magnitudes of $C^{(1)}_{pq}$, $\tilde{C}^{(2)}_{pq}$, and $\tilde{C}^{(3)}_{pq}$ are given in Table 1 below:

Table 1 Temperature Derivatives of Elastic Stiffness for Alpha Quartz at 25°C

pq	First Temperature Derivatives $C^{(1)}_{pq}$, $10^6 N/M^2/°C$	Effective Second Temperature Derivatives $\tilde{C}^{(2)}_{pq}$, $10^3 N/m^2/°C$	Effective Third Temperature Derivatives $\tilde{C}^{(3)}_{pq}$, $N/m^2/°C$
66	5.0747	11.066	-55.824
22	1.5976	-12.989	-38.145
44	-5.3780	-26.505	-20.468
14	0.91675	4.2849	85.773
13	-2.1983	-18.806	-8.9302
33	-6.5255	-20.835	46.255

In frequency-temperature problems, the tensors $D^{(n)}_{pq}$, which are the temperature coefficients of elastic stiffness, are very useful and can be calculated using relations (7b) to (7d). Their values are given in Table 2 below:

Table 2 $D^{(n)}_{pq}$, Temperature Coefficients of Elastic Stiffness for Alpha Quartz at 25°C

pq	First Temperature Coefficient $D^{(1)}_{pq}$, $10^6/m^2/°C$	Second Temperature Coefficient $D^{(2)}_{pq}$, $10^3/m^2/°C$	Third Temperature Coefficient $D^{(3)}_{pq}$, $N/m^2/°C$
66	6.3083	6.0238	-9.5420
22	-5.9217	-10.119	-5.3702
44	-10.780	-15.777	-2.6439
14	-1.5087	0.97458	14.615
13	-8.3983	-12.176	-0.51647
33	-21.177	-16.918	10.044

Consider the vibration of a plate with strong resonance of waves travelling predominantly in the length, or x_1, direction. Equations (1) to (7) are the requisite relations for developing a system of one-dimensional equations of motion for small amplitude waves superposed on homogeneous thermal strain. The two-dimensional displacements are expanded in a power series of the width coordinate of the plate:

$$u_i^{(0)}(x_1,x_3,t) = \sum_{n=0}^{\infty} x_3^n v_i^{(n)}(x_1,t) \tag{8a}$$

$$u_i^{(1)}(x_1,x_3,t) = \sum_{n=0}^{\infty} x_3^n \psi_i^{(n)}(x_1,t) \tag{8b}$$

and

$$u_i^{(n)} = 0 \quad \text{for} \quad n > 1$$

$v_i^{(n)}$ and $\psi_i^{(n)}$ are the one-dimensional displacement components of, respectively, the two-dimensional, zeroth order and first order displacement. The geometric meaning of the zeroth, first and second order one-dimensional displacement components are given in Fig. 2. Note that if the width edges at $x_3 = \pm c$ were fixed, a different series expansion of one-dimensional displacement components would be needed. When equations (8) are substituted into equations (1) and the resulting variational equations are integrated through the width coordinate, we obtain:

$$\int_{-a}^{a} \sum_{n=0}^{\infty} [\beta_{ki}\sigma^{(n)}_{kj,j} - n\beta_{ki}\sigma^{(n-1)}_{k3} + \beta_{ki}H^{(n)}_k - 2b\rho \sum_{n=0}^{\infty} B_{mn}\ddot{v}_i^{(m)}]$$

$$\delta v_i^{(n)} dx_1 = 0 \tag{9a}$$

where

$$\sigma^{(n)}_{kj} = \int_{-c}^{c} t^{(0)}_{kj} x_3^n dx_3$$

$$H^{(n)}_k = [t_{k3}x_3^n]^c_{-c}$$

and

$$B_{mn} = \frac{2c^{m+n+1}}{m+n+1} \quad \text{when } m+n \text{ is even}$$

$$= 0 \quad \text{when } m+n \text{ is odd.}$$

$$\int_{-a}^{a} \sum_{n=0}^{\infty} [\beta_{ki}\bar{\sigma}^{(n)}_{kj,j} - n\beta_{ki}\bar{\sigma}^{(n-1)}_{k3} +$$

$$\beta_{ki}\bar{H}^{(n)}_k - \beta_{ki}\sigma^{(n)}_{k2} - \frac{2b^3}{3}\rho \sum_{n=0}^{\infty} B_{mn}\ddot{\psi}_i^{(m)}]\delta\psi_i^{(n)} dx_1 = 0 \tag{9b}$$

where

$$\bar{\sigma}^{(n)}_{kj} = \int_{-c}^{c} t^{(1)}_{kj} x_3^n dx_3$$

$$\bar{H}^{(n)}_k = [t^{(1)}_{k3} x_3^n]^c_{-c}$$

The terms $\sigma^{(n)}_{k1}$ and $\bar{\sigma}^{(n)}_{k1}$ are, respectively the one-dimensional stress components of the zeroth and first order two-dimensional stress. The n-th order face tractions, $H^{(n)}_k$ and $\bar{H}^{(n)}_k$, on the width edges are equal to zero for traction free width edges. Since the coefficients of $\delta v_i^{(n)}$ and $\delta\psi_i^{(n)}$ in the integrand of equations (9a) and (9b) must vanish independently, they yield the n-th order, one-dimensional, incremental stress-temperature equations of motion:

$$\beta_{ki}\sigma^{(n)}_{k1,1} - n\beta_{ki}\sigma^{(n-1)}_{k3} = 2b\rho \sum_{m=0}^{\infty} B_{mn}\ddot{v}_i^{(m)} \tag{10a}$$

$$\beta_{ki}\bar{\sigma}^{(n)}_{k1,1} - n\beta_{ki}\bar{\sigma}^{(n-1)}_{k3} - \beta_{ki}\sigma^{(n)}_{k2} = \frac{2b^3}{3}\rho \sum_{m=0}^{\infty} B_{mn}\ddot{\psi}_i^{(m)}$$

$$i,k = 1,2,3 \tag{10b}$$

where we have set $H_k^{(n)}$ and $\bar{H}_k^{(n)}$ to zero for traction-free width edges at $x_3 = \pm c$. The term $\sigma_{kj,j}^{(n)}$ becomes $\sigma_{k1,1}^{(n)}$ because $\sigma_{kj}^{(n)}$ is a function x_1 and t only.

The variational equations (4a-b) for boundary conditions of a plate have taken into account that the major surface at $x_2 = \pm b$ are traction-free. For a rectangular plate, the equations can be written as:

$$\int_{-a}^{a}[(p_i^{(0)}-\beta_{ik}t_{k3}^{(0)})\delta u_i^{(0)}]_{-c}^{c}dx_1 + \int_{-c}^{c}[(p_i^{(0)}-\beta_{ik}t_{k1}^{(0)})\delta u_i^{(0)}]_{-a}^{a}dx_3 = 0 \tag{11a}$$

$$\int_{-a}^{a}[(p_i^{(1)}-\beta_{ik}t_{k3}^{(1)}]_{-c}^{c}dx_1 + \int_{-c}^{c}[(p_i^{(1)}-\beta_{ik}t_{k1}^{(1)})\delta u_i^{(1)}]_{-a}^{a}dx_3 = 0 \tag{11b}$$

Substitution of the series expansions ((8a) and (8b)) into the integrals will yield the following:

$$\int_{-a}^{a}[\sum_{n=0}^{\infty}(p_i^{(0)}x_3^n-\beta_{ik}t_{k3}^{(0)}x_3^n)\delta v_i^{(n)}]_{-c}^{c}dx_1 + \int_{-c}^{c}[\sum_{n=0}^{\infty}(p_i^{(0)}x_3^n-\beta_{ik}t_{k1}^{(0)}x_3^n)\delta v_i^{(n)}]_{-a}^{a}dx_3 = 0 \tag{12a}$$

$$\int_{-a}^{a}[\sum_{n=0}^{\infty}(p_i^{(1)}x_3^n-\beta_{ik}t_{k3}^{(1)}x_3^n)\delta\psi_i^{(n)}]_{-c}^{c}dx_1 + \int_{-c}^{c}[\sum_{n=0}^{\infty}(p_i^{(1)}x_3^n-\beta_{ik}t_{k1}^{(1)}x_3^n)\delta\psi_i^{(n)}]_{-a}^{a}dx_3 = 0 \tag{12b}$$

Using the terms defined in equations (9), the abo.e equations can be rewritten as:

$$\int_{-a}^{a}[\sum_{n=0}^{\infty}(p_i^{(0)}x_3^n-\beta_{ik}t_{k3}^{(0)}x_3^n)\delta v_i^{(n)}]_{-c}^{c}dx_1 + [\sum_{n=0}^{\infty}(q_i^{(n)}-\beta_{ik}\sigma_{k1}^{(n)})\delta v_i^{(n)}]_{-a}^{a} = 0 \tag{13a}$$

$$\int_{-a}^{a}[\sum_{n=0}^{\infty}(p_i^{(1)}x_3^n-\beta_{ik}t_{k3}^{(1)}x_3^n)\delta\psi_i^{(n)}]_{-c}^{c}dx_1 + [\sum_{n=0}^{\infty}(\bar{q}_i^{(n)}-\beta_{ik}\bar{\sigma}_{k1}^{(n)})\delta\psi_i^{(n)}]_{-a}^{a} = 0 \tag{13b}$$

where

$$q_i^{(n)}(x_1,t) = \int_{-c}^{c} p_i^{(0)}(x_1,x_3,t)x_3^n dx_3$$

$$\bar{q}_i^{(n)}(x_1,t) = \int_{-c}^{c} p_i^{(1)}(x_1,x_3,t)x_3^n dx_3$$

which are the one-dimensional, n-th order incremental surface traction. Equations (13) yield the following conditions:

1) Along edges at $x_3 = \pm c$, specify $\beta_{ik}t_{k3}^{(0)}x_3^n$ or $v_i^{(n)}$ (14a)

and specify $\beta_{ik}t_{k3}^{(1)}x_3^n$ or $\psi_i^{(n)}$ (14b)

2) Along edges at $x_1 = \pm a$, specify $q_i^{(n)}=\beta_{ik}\sigma_{k1}^{(n)}$ or $v_i^{(n)}$ (15a)

and specify $\bar{q}_i^{(n)}=\beta_{ik}\bar{\sigma}_{k1}^{(n)}$ or $\psi_i^{(n)}$ (15b)

Since the terms $H_k^{(n)}$ and $\bar{H}_k^{(n)}$ have been set to zero in the stress equations of motion (10), conditions (14) are satisfied identically. For a well posed eigenvalue problem, the stress equations of motion (10) are used with the boundary conditions given in equations (15).

The one-dimensional geometric relations are formed by putting the power series expansion (8) into equations (5) to yield:

$$e_{ij}^{(0)} = \sum_{m=0}^{\infty} x_3^m \varepsilon_{ij}^{(m)} \tag{16a}$$

$$e_{ij}^{(1)} = \sum_{m=0}^{\infty} x_3^m \bar{\varepsilon}_{ij}^{(m)} \tag{16b}$$

where

$$\varepsilon_{ij}^{(m)} = \frac{1}{2}[\beta_{ki}v_{k,j}^{(m)}+\beta_{kj}v_{k,i}^{(m)}+(m+1)(\delta_{3j}\beta_{ki}v_k^{(m+1)}+\delta_{3i}\beta_{kj}v_k^{(m+1)}) + \delta_{2j}\beta_{ki}\psi_k^{(m)} + \delta_{2i}\beta_{kj}\psi_k^{(m)}] \tag{17a}$$

and

$$\bar{\varepsilon}_{ij}^{(m)} = \frac{1}{2}[\beta_{ki}\psi_{k,j}^{(m)}+\beta_{kj}\psi_{k,i}^{(m)}+(m+1)(\delta_{3j}\beta_{ki}\psi_k^{(m+1)}+\delta_{3i}\beta_{kj}\psi_k^{(m+1)})] \tag{17b}$$

The one-dimensional constitutive relations are obtained by substituting equations (17) into equations (6) and integrating through the width coordinate:

$$\sigma_{ij}^{(n)} = 2bD_{ijk\ell}\sum_{m=0}^{\infty} B_{mn}\varepsilon_{k\ell}^{(m)} \tag{18a}$$

$$\bar{\sigma}_{ij}^{(n)} = \frac{2b^3}{3} D_{ijk\ell} \sum_{m=0}^{\infty} B_{mn} \bar{\varepsilon}_{k\ell}^{(m)} \qquad (18b)$$

The tensor β_{ij} is predominantly a diagonal tensor [8,9] with the off-diagonal terms smaller by a magnitude of 10^{-6}. Therefore, the off-diagonal terms are neglected:

$$\beta_i = \beta_{ii} = 1 + \alpha_{ii}^{\theta} \qquad (19)$$

$$\beta_{ij} = 0 \quad \text{for } i \neq j$$

This brings about considerable simplifications in the system of one-dimensional equations. In summary, we have the following one-dimensional model for small amplitude waves superposed on a thermally biased plate:

1) Stress equations of motion

$$\sigma_{i1,1}^{(n)} - n\beta_i \sigma_{i3}^{(n-1)} = 2b\rho \sum_{m=0}^{\infty} B_{mn} \ddot{v}_i^{(m)} \qquad (20a)$$

$$\bar{\sigma}_{i1,1}^{(n)} - n\beta_i \bar{\sigma}_{i3}^{(n-1)} - \beta_i \sigma_{i2}^{(n)} = \frac{2b^3}{3} \rho \sum_{m=0}^{\infty} B_{mn} \ddot{\psi}_i^{(m)}$$

$$(i = 1,2,3; \text{ no sum on } i) \qquad (20b)$$

2) Boundary conditions

$$\text{specify } q_i^{(n)} = \sigma_{i1}^{(n)} \text{ or } v_i^{(n)} \qquad (21a)$$

$$\text{and specify } \bar{q}_i^{(n)} = \bar{\sigma}_{i1}^{(n)} \text{ or } \psi_i^{(n)}$$

$$(i = 1,2,3; \text{ no sum on } i) \qquad (21b)$$

3) Geometric Relations

$$\varepsilon_{ij}^{(m)} = \frac{1}{2}[\beta_i v_{i,j}^{(m)} + \beta_j v_{j,i}^{(m)} + (m+1)(\delta_{3j}\beta_i v_i^{(m+1)} + \delta_{3i}\beta_j v_j^{(m+1)})$$

$$+ \delta_{2j}\beta_i \psi_i^{(m)} + \delta_{2i}\beta_j \psi_j^{(m)}] \qquad (22a)$$

$$\bar{\varepsilon}_{ij}^{(m)} = \frac{1}{2}[\beta_i \psi_{i,j}^{(m)} + \beta_j \psi_{j,i}^{(m)} + (m+1)(\delta_{3j}\beta_i \psi_i^{(m+1)} + \delta_{3i}\beta_j \psi_j^{(m+1)})]$$

$$(i,j = 1,2,3; \text{ no sum on } i \text{ and } j) \qquad (22b)$$

4) Constitutive Relations

$$\sigma_{ij}^{(n)} = 2bD_{ijk\ell} \sum_{m=0}^{\infty} B_{mn} \varepsilon_{k\ell}^{(m)} \qquad (23a)$$

$$\bar{\sigma}_{ij}^{(n)} = \frac{2b^3}{3} D_{ijk\ell} \sum_{m=0}^{\infty} B_{mn} \bar{\varepsilon}_{k\ell}^{(m)} \qquad (23b)$$

III. Flexural and Thickness-Shear Vibrations of Rectangular Rotated Y-Cut Quartz Plates

Rotated Y-cut quartz plates have monoclinic symmetry, and the flexural and thickness-shear vibrations for waves traveling predominantly in the length direction can be accommodated using the three displacements: $u_2^{(0)}$, $u_1^{(1)}$ and $u_3^{(1)}$. These displacements allow for length flexural (LF), width flexural (WF) and thickness-shear (TS) vibrations. To allow for free development of the strains $e_{22}^{(0)}$, $e_{22}^{(1)}$, $e_{23}^{(1)}$ and $e_{12}^{(1)}$, we set [11]

$$t_{22}^{(0)} = 0 \qquad (24a)$$

and

$$t_{22}^{(1)} = t_{23}^{(1)} = t_{12}^{(1)} = 0 \qquad (24b)$$

Equation (6a) can be written as

$$t_p^{(0)} = 2bD_{pq}e_q^{(0)} \qquad (p,q = 1,2,3,\ldots,6)$$

with the following indicial convention:

$$ij \rightarrow p$$
$$11 \rightarrow 1$$
$$22 \rightarrow 2$$
$$33 \rightarrow 3$$
$$23 \rightarrow 4$$
$$13 \rightarrow 5$$
$$12 \rightarrow 6$$

It is understood that the strain $e_q^{(0)} = e_{ij}^{(0)}$ for $i = j$ and $e_q^{(0)} = 2e_{ij}^{(0)}$ for $i \neq j$. From condition (24a) and equation (6a) we form the relation:

$$\frac{D_{2p}t_2^{(0)}}{D_{22}} = \frac{2bD_{2p}D_{2q}e_q^{(0)}}{D_{22}} = 0 \qquad (pq = 1,2,\ldots,6)$$

The difference between equation (6a) and the above equation will yield a relation which satisfies condition (24a):

$$t_p^{(0)} = 2b\bar{D}_{pq}e_q^{(0)} \qquad (25a)$$

where

$$\bar{D}_{pq} = D_{pq} - \frac{D_{2p}D_{2q}}{D_{22}}$$

By a similar process, we obtain from equation (6b) an equation which satisfies conditions (24b):

$$t_p^{(1)} = \frac{2b^3}{3} \bar{D}_{pq}e_q^{(1)} \qquad (25b)$$

where

$$\bar{D}_{pq} = D_{pq}^4 - \frac{D_6^4 D_{6q}^4}{D_{66}^4}$$

$$D_{pq}^4 = D_{pq}^2 - \frac{D_4^2 D_{4q}^2}{D_{44}^2}$$

and

$$D_{pq}^2 = D_{pq} - \frac{D_{2p}D_{2q}}{D_{22}}$$

Therefore equations (23) will in turn become:

$$\sigma_p^{(n)} = 2b\bar{D}_{pq} \sum_{m=0}^{\infty} B_{mn}\varepsilon_q^{(m)} \tag{26a}$$

$$\bar{\sigma}_p^{(n)} = \frac{2b^3}{3} D_{pq} \sum_{m=0}^{\infty} B_{mn}\bar{\varepsilon}_q^{(m)} \tag{26b}$$

The three displacements $u_2^{(0)}$, $u_1^{(1)}$ and $u_3^{(1)}$ will by the series expansion (8) yield only $v_2^{(n)}$, $\psi_1^{(n)}$ and $\psi_3^{(n)}$ (n = 0,1,2,...,) which consequently simplify the zeroth and first order one-dimensional strains (22)

$$\begin{aligned}
\varepsilon_{12}^{(0)} &= \frac{1}{2}(\beta_2 v_{2,1}^{(0)} + \beta_1\psi_1^{(0)}) \\
\varepsilon_{23}^{(0)} &= \frac{1}{2}(\beta_2 v_2^{(1)} + \beta_3\psi_3^{(0)}) \\
\varepsilon_{12}^{(1)} &= \frac{1}{2}(\beta_2 v_{2,1}^{(1)} + \beta_1\psi_1^{(1)}) \\
\varepsilon_{23}^{(1)} &= \frac{1}{2}(\beta_2 v_2^{(2)} + \beta_3\psi_3^{(1)})
\end{aligned} \tag{27a}$$

$$\begin{aligned}
\bar{\varepsilon}_{11}^{(0)} &= \beta_1\psi_{1,1}^{(0)} \\
\bar{\varepsilon}_{13}^{(0)} &= \frac{1}{2}(\beta_3\psi_{3,1}^{(0)} + \beta_1\psi_1^{(1)}) \\
\bar{\varepsilon}_{33}^{(0)} &= \beta_3\psi_3^{(1)} \\
\bar{\varepsilon}_{11}^{(1)} &= \beta_1\psi_{1,1}^{(1)} \\
\bar{\varepsilon}_{13}^{(1)} &= \frac{1}{2}(\beta_3\psi_{3,1}^{(1)} + 2\beta_1\psi_1^{(2)}) \\
\varepsilon_{33}^{(1)} &= 2\beta_3\psi_3^{(2)}
\end{aligned} \tag{27b}$$

For LF-WF-TS vibrations, we need the strains $\varepsilon_{12}^{(0)}$, $\varepsilon_{23}^{(1)}$ and $\bar{\varepsilon}_{11}^{(0)}$ which involve the displacement components $v_2^{(0)}$, $v_2^{(2)}$, $\psi_1^{(0)}$ and $\psi_3^{(1)}$. The relevant stress equations of motion (20) for these four displacements are:

$$\begin{aligned}
\beta_2\sigma_{12,1}^{(0)} &= 4bc\rho\ddot{v}_2^{(0)} + \frac{4bc^3}{3}\rho\ddot{v}_2^{(2)} \\
\beta_2\sigma_{12,1}^{(2)} - 2\beta_2\sigma_{23}^{(1)} &= \frac{4bc^3}{3}\rho\ddot{v}_2^{(0)} + \frac{4bc^5}{5}\rho\ddot{v}_2^{(2)} \\
\beta_1\bar{\sigma}_{11,1}^{(0)} - \beta_1\sigma_{12}^{(0)} &= \frac{4b^3c}{3}\rho\ddot{\psi}_1^{(0)} \\
\beta_3\bar{\sigma}_{13,1}^{(1)} - \beta_3\bar{\sigma}_{33}^{(0)} - \beta_3\sigma_{23}^{(1)} &= \frac{4b^3c^3}{9}\rho\ddot{\psi}_3^{(1)}
\end{aligned} \tag{28}$$

Stress-displacement relations are formed by substituting the geometric relations (27) into equations (26), taking into account that we only consider the four components $v_2^{(0)}$, $v_2^{(2)}$, $\psi_1^{(0)}$ and $\psi_3^{(1)}$:

$$\begin{aligned}
\sigma_{12}^{(0)} &= 4bck^2\bar{D}_{66}(\beta_2 v_{2,1}^{(0)} + \beta_1\psi_1^{(0)}) + \frac{4bc^3}{3}k^2\bar{D}_{66}\beta_2 v_{2,1}^{(2)} \\
\sigma_{12}^{(2)} &= \frac{4bc^3}{3}k^2\bar{D}_{66}(\beta_2 v_{2,1}^{(0)} + \beta_1\psi_1^{(0)}) + \frac{4bc^5}{5}k^2\bar{D}_{66}\beta_2 v_{2,1}^{(2)} \\
\sigma_{23}^{(1)} &= \frac{4bc^3}{3}k^2\bar{D}_{44}(2\beta_2 v_2^{(2)} + \beta_3\psi_3^{(1)}) \\
\bar{\sigma}_{11}^{(0)} &= \frac{4bc^3}{3}(\beta_1\tilde{D}_{11}\psi_{1,1}^{(0)} + \beta_2\tilde{D}_{13}\psi_3^{(1)}) \\
\bar{\sigma}_{33}^{(0)} &= \frac{4b^3c}{3}(\beta_1\tilde{D}_{11}\psi_{1,1}^{(0)} + \beta_3\tilde{D}_{33}\psi_3^{(1)}) \\
\bar{\sigma}_{13}^{(1)} &= \frac{4b^3c^3}{9}\beta_3\tilde{D}_{55}\psi_{3,1}^{(1)}
\end{aligned} \tag{29}$$

where $k^2 = \pi^2/12$ is the shear correction factor [11] for the fundamental thickness shear mode. In the first two equations of (29), the strain $\varepsilon_{12}^{(2)} = 1/2\,\beta_2 v_{2,1}^{(2)}$ was used in addition to the strains in equations (27).

We set $\varepsilon_{23}^{(1)}$ to zero to obtain a relation between $\psi_3^{(1)}$ and $v_2^{(2)}$:

$$\psi_3^{(1)} = \frac{-2\beta_2 v_2^{(2)}}{\beta_3} \tag{30}$$

By neglecting the inertia term in the fourth equation of (28), we may use the resulting relation to determine $\sigma_{23}^{(1)}$:

$$\sigma_{23}^{(11)} = \bar{\sigma}_{13,1}^{(1)} - \bar{\sigma}_{33}^{(0)}$$

$$\sigma_{23}^{(1)} = \frac{4b^3c^3}{9}\beta_3\tilde{D}_{55}\psi_{3,11}^{(1)} - \frac{4b^3c}{3}(\beta_1\tilde{D}_{11}\psi_{1,1}^{(0)} + \beta_3\tilde{D}_{33}\psi_3^{(1)})$$

When (30) is substituted, the resulting stress-displacement for $\sigma_{23}^{(1)}$ becomes:

$$\sigma_{23}^{(1)} = \frac{-8b^3c^3}{9}\ \beta_2\tilde{D}_{55}v_{2,11}^{(2)} - \frac{4b^3c}{3}\left(\beta_1\tilde{D}_{11}\psi_{1,1}^{(0)} - 2\beta_2\tilde{D}_{33}v_2^{(2)}\right) \tag{31}$$

Substitution of (31) and the relevant equations of (29) into the first three equations of (28) and using (30) yields the displacement equations of motion of LF-WF-TS vibrations of a quartz plate.

$$\beta_2k^2\bar{D}_{66}v_{2,11}^{(0)} + \beta_1\beta_2k^2\bar{D}_{66}\psi_{1,1}^{(0)} + \frac{c^2}{3}\beta_2^2k^2\bar{D}_{66}v_{2,11}^{(2)} = \rho\ddot{v}_2^{(0)} + \frac{c^2}{3}\rho\ v_2^{(2)}$$

$$\frac{5}{2}c^2\left(\beta_2^2k^2\bar{D}_{66}v_{2,11}^{(0)} + \beta_1\beta_2k^2\bar{D}_{66}\psi_{1,1}^{(0)}\right) + \beta_2^2k^2\bar{D}_{66}v_{2,11}^{(2)} + \frac{20}{9}\frac{b^2}{c^2}\beta_2^2\tilde{D}_{55}v_{2,11}^{(2)} + \frac{10}{3}\frac{b^2}{c^4}\left(\beta_1\beta_2\tilde{D}_{13}\psi_{1,1}^{(0)} - 2\beta_2\beta_3\tilde{D}_{33}v_2^{(2)}\right) = \frac{5}{3c^2}\rho\ddot{v}_2^{(0)} + \rho\ddot{v}_2^{(2)}$$

$$\beta_1^2\tilde{D}_{11}\psi_{1,11}^{(0)} - 2\beta_1\beta_2\tilde{D}_{13}v_{2,1}^{(2)} - \frac{3}{b^2}\left(\beta_1\beta_2k^2\bar{D}_{66}v_{2,1}^{(0)} + \beta_1^2k^2\bar{D}_{66}\psi_1^{(0)}\right) - \frac{c^2}{b^2}\beta_1\beta_2k^2\bar{D}_{66}v_{2,1}^{(2)} = \rho\ddot{\psi}_1^{(0)} \tag{32}$$

The boundary conditions are, from (21):

$$\text{specify}\quad q_2^{(0)} = \beta_2\sigma_{12}^{(0)}\ \text{ or }\ v_2^{(0)}$$

$$q_2^{(2)} = \beta_2\sigma_{12}^{(2)}\ \text{ or }\ v_2^{(2)}$$

$$\text{and}\quad \bar{q}_1^{(0)} = \beta_1\bar{\sigma}_{11}^{(0)}\ \text{ or }\ \psi_1^{(0)}$$

For low frequency vibrations, we set [14] $t_{11}^{(0)}$, $t_{23}^{(0)}$ and $t_{13}^{(0)}$ to zero to allow for free development in the strains $e_{11}^{(0)}$, $e_{33}^{(0)}$ and $e_{13}^{(0)}$; so that by a similar process used in deriving the expression (25a) the material tensor D_{pq} is now written as:

$$\bar{D}_{pq} = D_{pq}^3 - \frac{D_{5p}^3D_{5q}^3}{D_{55}^3}$$

where

$$D_{pq}^3 = D_{pq}^1 - \frac{D_{3p}^1D_{3q}^1}{D_{33}^1}$$

and

$$D_{pq}^1 = D_{pq}^3 - \frac{D_{1p}^2D_{1q}^2}{D_{11}^2}$$

and

$$D_{pq}^2 = D_{pq} - \frac{D_{2p}D_{2q}}{D_{22}} \tag{33}$$

IV. Length-Flexural and Thickness Shear Vibrations of Rotated Y-Cut Beams and Strips

If we consider only the strains $\varepsilon_{12}^{(0)}$ and $\varepsilon_{11}^{(0)}$, neglecting $\varepsilon_{23}^{(1)}$, we would obtain a model which is similar to the classical flexural strip with rotary inertia. This is a reduction from the model in the previous section. Since we are only considering the displacement components $v_2^{(0)}$ and $\psi_1^{(0)}$, the first and third equations of (28) are reduced to:

$$\beta_2\sigma_{12,1}^{(0)} = 4bc\rho\ddot{v}_2^{(0)}$$

$$\beta_1\bar{\sigma}_{11,1}^{(0)} - \beta_1\sigma_{12}^{(0)} = \frac{4b^3c}{3}\ \rho\ddot{\psi}_1^{(0)} \tag{34}$$

The required stress-displacement relations from (29) are

$$\sigma_{12}^{(0)} = 4bck^2\bar{D}_{66}\left(\beta_2v_{2,1}^{(0)} + \beta_1\psi_1^{(0)}\right)$$

$$\bar{\sigma}_{11}^{(0)} = \frac{4b^3c}{3}\ \beta_1\tilde{D}_{11}\psi_{1,1}^{(0)} \tag{35}$$

From (34) and (35), the displacement equations of motion are:

$$\beta_2^2k^2\bar{D}_{66}v_{2,11}^{(0)} + \beta_1\beta_2k^2\bar{D}_{66}\psi_{1,1}^{(0)} = \rho\ddot{v}_2^{(0)}$$

$$\beta_1^2\tilde{D}_{11}\psi_{1,11}^{(0)} - \frac{3}{b^2}\left(\beta_1\beta_2k^2\bar{D}_{66}v_{2,1}^{(0)} + \beta_1^2k^2\bar{D}_{66}\psi_1^{(0)}\right) = \rho\ddot{\psi}_1^{(0)} \tag{36}$$

The boundary conditions are:

$$\text{specify}\quad q_2^{(0)} = \beta_2\sigma_{12}^{(0)}\ \text{ or }\ v_2^{(0)}$$

$$\text{and}\quad \bar{q}_1^{(0)} = \beta_1\bar{\sigma}_{11}^{(0)}\ \text{ or }\ \psi_1^{(0)} \tag{37}$$

Note that the displacement equations of motion (36) are now independent of the width dimension, c. The model is good only for straight-crested waves traveling in the length, or x_1, direction.

The equations for a flexural beam has the same mathematical form as a flexural strip. When we set the moment $t_{33}^{(1)}$ to zero, the material tensor D_{pq} is changed to

$$\bar{D}_{pq} = D_{pq}^6 - \frac{D_{3p}^6 D_{3q}^6}{D_{33}^6} \tag{38}$$

where

$$D_{pq}^6 = D_{pq}^4 - \frac{D_{6p}^4 D_{6q}^4}{D_{66}^4}$$

The use of (33) and (38) with equations (35-37) would adequately model a low frequency flexural beam with no width effects.

V. Dispersion Relations and Frequency Spectrum for the Length Flexural (LF), Width Flexural (WF) and Thickness Shear (TS) Vibrations of a Rotated Y-Cut Quartz Plate

The dispersion relations are obtained by substituting the following assumed displacement functions into the equations (32):

$$v_2^{(0)} = A_1 b \cos \xi x_1 e^{i\omega t}$$

$$v_2^{(2)} = A_2 \frac{b}{c^2} \cos \xi x_1 e^{i\omega t} \tag{39}$$

and $$\varphi_1^{(0)} = A_3 \sin \xi x_1 e^{i\omega t}$$

With some algebraic manipulations, the following dispersions are obtained:

$$\begin{vmatrix} B_1 & B_2 & B_3 \\ B_2 & B_4 & B_5 \\ B_3 & B_5 & B_6 \end{vmatrix} \begin{vmatrix} A_1 \\ A_2 \\ A_3 \end{vmatrix} = 0$$

where

$$B_1 = 3(\beta_2^2 D_{66}^a \bar{\xi}^2 - 3\Omega^2)$$

$$B_2 = (\beta_2^2 D_{66}^a \bar{\xi}^2 - 3\Omega^2)$$

$$B_3 = -3\beta_1\beta_2 D_{66}^a \bar{\xi}$$

$$B_4 = 3/5(\beta_2^2 D_{66}^a \bar{\xi}^2 + \frac{20}{9S^2}\beta_2^2 D_{55}\bar{\xi}^2 + \frac{20}{3S^4}\beta_2^2 D_{33}^b - 3\Omega^2)$$

$$B_5 = -\beta_1\beta_2 D_{66}^a \bar{\xi} + \frac{2}{S^2}\beta_1\beta_2 D_{13}^b \bar{\xi})$$

$$B_6 = (\beta_1^2 D_{11}^b \bar{\xi}^2 + 3\beta_1^2 D_{66}^a - 3\Omega^2)$$

and

$$\bar{\xi} = \xi b, \quad S = c/b, \quad \Omega^2 = \omega^2/\omega_1^2, \quad \omega_1^2 = \frac{3k^2 C_{66}}{\rho b^2} \tag{40}$$

$$D_{66}^a = \bar{D}_{66}/C_{66} \quad \text{and} \quad D_{pq}^b = \frac{\bar{D}_{pq}}{k^2 C_{66}}$$

Equations (40) would be of the same form if we had assumed symmetric shear $\varphi_1^{(0)} = A_3 \cos \xi x_1 e^{iwt}$ and antisymmetric $v_2^{(0)}$ and $v_2^{(2)}$. The determinant of (40) must be equal to zero to satisfy the dispersion relations. Three wave numbers are obtained by solving the polynomial associated with the determinant of (40). Figure 3 shows the dispersion relations for a Y-cut plate with c/b = 7.69. Since the rank of (40) is two, we obtained two independent amplitude ratios for each wave number:

$$\gamma_{1q} = A_{2q}/A_{1q} \quad \text{and} \quad \gamma_{2q} = A_{3q}/A_{1q} \qquad q = 1,2,3 \tag{41}$$

a) Frequency Spectrum for Predominantly Thickness Shear Vibrations

The AT-cut ($\theta = 35.25°$) plate with four free edges is considered. Symmetric shear across the plate is assumed:

$$v_2^{(0)} = A_1 b \sin \xi x_1 e^{i\omega t}$$

$$v_2^{(2)} = A_2 \frac{b}{c^2} \sin \xi x_1 e^{i\omega t} \tag{42}$$

$$\varphi_1^{(0)} = A_3 \cos \xi x_1 e^{i\omega t}$$

For free edges at $x_1 = \pm a$, we specify

$$\beta_2 \bar{\sigma}_{12}^{(0)} = \beta_2 \sigma_{12}^{(2)} = \beta_1 \bar{\sigma}_{11}^{(0)} = 0 \tag{43}$$

The frequency matrix is obtained by using (42) in the stress-displacement relations (29) for $\bar{\sigma}_{12}^{(0)}$, $\sigma_{12}^{(2)}$ and $\bar{\sigma}_{11}^{(0)}$ and setting the resulting equations to zero as in conditions (43):

$$\sum_{q=1}^{3} (3\beta_2^2 \bar{D}_{66}\bar{\xi}_q + \gamma_{1q}\beta_2^2 \bar{D}_{66}\bar{\xi}_q + 3\gamma_{2q}\beta_1\beta_2 \bar{D}_{66}) A_{1q} \cos \bar{\xi}_q r = 0$$

$$\sum_{q=1}^{3} (\beta_2^2 \bar{D}_{66}\bar{\xi}_q + \gamma_{1q}\beta_2^2 \bar{D}_{66}\bar{\xi}_q + \gamma_{2q}\beta_1\beta_2 \bar{D}_{66}) A_{1q} \cos \bar{\xi}_q r = 0$$

$$\sum_{q-1}^{3} (\frac{2}{S^2}\gamma_{1q}\beta_1\beta_2 \bar{D}_{13} + \gamma_{2q}\beta_1^2 \bar{D}_{11}\bar{\xi}_q) A_{1q} \sin \bar{\xi}_q r = 0 \tag{44}$$

where relation (30) was used in the first term of the third equation, and r = a/b.

The resonant frequency is the frequency which satisfies simultaneously the frequency matrix and the dispersion relations. Figure 4 shows the frequency spectrum of an AT-cut plate with c/b = 20 and a/b ranges from 17 to 23. The symbols TS-n, LF-n and WF-n denote, respectively, the thickness-shear (TS), flexure in the x_1-x_2 plane (LF) and flexure in the x_3-x_2 plane (WF) with n the number of phase reversals in the x_1 direction.

b) Frequency Spectrum for Predominantly Flexural Vibrations

The Y-cut ($\theta = 0°$) plate with free edges at $x_3 = \pm c$ and fixed edges at $x_1 = \pm a$ is considered. For symmetric flexural vibrations, the displacement functions (39) are used to obtain the dispersion relations (40). Since the edges are fixed at $x_2 = a$, we specify the boundary conditions:

$$v_2^{(0)} = v_2^{(2)} = \phi_1^{(0)} = 0 \tag{45}$$

The frequency matrix is obtained from (39) and (41) satisfying conditions (45):

$$\sum_{q=1}^{3} A_{1q}\cos\bar{\xi}_q r = 0$$

$$\sum_{q=1}^{3} A_{1q}\,{}_{1q}\cos\bar{\xi}_q r = 0 \tag{46}$$

$$\sum_{q=1}^{3} A_{1q}\,{}_{2q}\sin\bar{\xi}_q r = 0$$

The resonant frequency is a function of two parameters: the c/b ratio in the dispersion relations (40) and the a/b ratio in the frequency matrix (46). At low frequencies, the material constants D_{pq} of (33) are used. Figure 5 shows the frequency spectrum of a Y-cut plate with free width edges and fixed edges at $x_1 = \pm a$. The c/b ratio is fixed at 10. Included in the figure are the mode shapes of the respective branches of the frequency spectrum. The symbols LF-n and WF-n are, respectively, flexure in the X_1-X_2 plane (LF) and flexure in the X_3-X_2 plane (WF) with n the number of nodal lines. Figure 6 exhibits the frequency spectrum for the same plate, but this time, the ratio a/b is fixed at 20 and the c/b ratio is allowed to vary. The anticlastic branch WF-2 remain relatively horizontal for fixed c/b, but the frequency decreases rapidly with increasing c/b values, approaching asymptotically a value of $\bar{\omega}$ = 0.042 at c/b = ∞. At ratios of c/b smaller than one, when a/b = 20, there is little antielastic action in the first two symmetric flexural modes of vibration.

VI. Frequency-Temperature Behavior of Flexural and Thickness-Shear Vibrations of Rectangular Rotated Y-Cut Quartz Plates

The usefulness of the equations developed thus far lies in the fact that the material constants D_{pq} and e_{ij} are functions of the temperature change, $\theta = T-25°C$, where T is the ambient temperature of the plate. We concentrate our study on the frequency temperature behavior of the thickness-shear vibrations of an AT-cut plate and the flexural vibrations of a Y-cut plate.

a) Frequency-Temperature Behavior of the Thickness-Shear Vibrations of an AT-Cut plate

Figure 7 shows the frequency-temperature curves of the fundamental thickness-shear vibrations of the AT-cut with increasing a/b ratio. The curves do not quite approach the infinitely extended plate curve using the theory derived by Koga [12]. Since most AT-cut resonators have the a/b ratio less than 50, a finite plate theory would provide better predictions of the frequency-temperature behavior.

Figure 8 gives the results of coupling of the predominant thickness shear vibrations with flexural vibrations. The c/b ratio is fixed at 20, while the a/b ratio is given different values which yield various degrees of interaction of the TS-1 branch with the flexural branches of Fig. 4. Strong interactions with the flexural vibrations affect adversely the frequency-temperature characteristics. Consequently, it is important to isolate the thickness shear mode.

Experimental observations indicated that changes in width dimension have small effects on the frequency temperature behavior of AT-cut plates. Figure 9 exhibits the effects of changing the c/b ratio from 2.5 to 20 at a fixed a/b ratio of 20. The effect of changing width dimension is negligible when c/b is greater than five. Pronounced effects are predicted only when c/b is less than 3.5. Therefore in general, the change in a/b ratio has significant effects on the frequency-temperature curves, while the effects due changing c/b ratio is negligible.

b) Frequency-Temperature Behavior of Length-Flexural Vibrations (LF) of Y-Cut Plates

The low frequency length-flexural vibration of Y-cut plates is found to be much more sensitive than the thickness-shear vibration of AT-cut plates to changes in the c/b ratio. Figure 10 shows the changes in frequency-temperature behavior for the fundamental, symmetric, length-flexural vibrations of a Y-cut plate with an a/b ratio equal to 20 and various c/b ratios ranging from 2.5 to 20. Also shown are two curves of the 2-branch strip and 2-branch beam solution. The 2-branch strip curve is calculated using the displacement equations of motion of (35). It represents a plate infinitely extended in the width direction. The 2-branch beam curve neglects the width dimension and the moment $t_{33}^{(1)}$ is set to zero. The frequency temperature curves progress from right to left as the c/b ratio increases from 2.5 to 20.

Figure 11 gives the effect decreasing a/b ratios from 23 to 10 while c/b is fixed at 10. The turnover temperature, T_{ZTC}, of the curves increases rapidly as the a/b ratio becomes less than 14. We observe that the frequency-temperature curves of length flexural vibrations of Y-cut plates are sensitive to changes in both the a/b and c/b ratios. The T_{ZTC} increases when either the a/b or c/b ratio decreases. The present theoretical results are checked with available experimental data by Paros and Singer [13] for a Y-cut plate with a/b = 33.45 and c/b = 7.69. The theoretical curve agrees very well with the experimental curve by Paros as seen in Fig. 12. The classical 2-branch beam solution predicts a T_{ZTC} which is higher than the experimental data.

It is interesting to compare the present theory based on a lagrangian formulation with Bechmann's theory using the classical theory of linear elasticity and letting all material properties, such as density, elastic stiffness and physical dimensions, to be functions of temperature change. Using Bechmann's

theory, we can formulate a similar one-dimensional model, which takes into account the boundary conditions at the width edges, for the length flexural vibrations of a Y-cut plate. Figure 13 shows the comparison between the present theoretical curve and the curve using Bechmann's formulation. The Bechmann curve agrees well with the present curve and the experimental curve by Paros. Hence, we make a conclusion that the commonly used approach to predicting the frequency-temperature behavior of length-flexural vibrations of Y-cut plates employing the classical beam equation yielded poor results due not to the inaccuracies of the temperature coefficients but to neglecting the width effects. Other causes of the discrepancy may be attributed to electrode plating effects.

References

[1] Bechmann, R.. Ballato, A.D. and Lukaszek, T.J., "Higher-Order Temperature Coefficients of the Elastic Stiffnesses and Compliances of the Alpha-Quartz," Proc. IRE 50, 1962, pp. 1812.

[2] Adams, C.A., Enslow, G.M., Kusters, J.A. and Ward, R.W., "Selected Topics in Quartz Crystal Research," Proc. 24th Ann. Freq. Control Symp., 1970, pp. 55-63.

[3] Kahan, A., "Elastic Constants of Quartz and Their Temperature Coefficients," Proc. 36th Ann. Freq. Control Symp., 1982, pp. 159-167.

[4] Nakazawa, M., et al., "Frequency-Temperature Characteristics of Quartz Crystal Flexure Bars and Quartz Crystal Tuning Forks," IEEE Trans. Sonics and Ultrasonics, Vol. SU-26, No. 5, Sept. 1979, pp. 369-376.

[5] Mason, W.P. and Sykes, R.A., "Low-Frequency Quartz Crystal Cuts Having Low Temperature Coefficients," Proc. IRE, Vol. 32, 1944, pp. 208-215.

[6] Zelenka, J. and Lee, P.C.Y., "On the Temperature Coefficients of the Elastic Stiffness and Compliances of Alpha Quartz," IEEE Trans. on Sonics and Ultrasonics, Vol. SU-18, No. 2, April 1971, pp. 79-80.

[7] Mindlin, R.D. and Gazis, D.C., "Strong Resonances of Rectangular AT-Cut Quartz Plates," Proc. 4th U.S. Nat. Congress Applied Mech., Vol. 1, 1962, pp. 305-310.

[8] Lee, P.C.Y. and Yong, Y.K., "Frequency-Temperature Behavior of Thickness Vibrations of Doubly-Rotated Quartz Plates Affected by Plate Dimensions and Orientations," Proc. 38th Ann. Freq. Control Symp., 1984, pp. 164-175.

[9] Lee, P.C.Y. and Yong, Y.K., "Temperature Derivatives of Elastic Stiffness Derived from the Frequency-Temperature Behavior of Quartz Plates," J. Appl. Phys. 56(5), Sept. 1984, pp. 1514-1521.

[10] Thurston, R.N., McSkimin, H.J. and Andreatch, P., Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 1966, pp. 267.

[11] Mindlin, R.D., "High-Frequency Vibrations of Crystal Plates," Quarterly Appl. Math., Vol. 19, No. 1, 1961, pp. 51-61.

[12] Koga, I., "Thickness Vibrations of Piezoelectric Oscillating Crystals," Physics, Vol. 3, August 1932, pp. 70-80.

[13] Adams, J.W., "Temperature Coefficient for a Vibrating Beam with Trapezoidal Cross Section," M.S. Thesis, Dept. of Mechanical Engineering, M.I.T., June 1978.

[14] Mindlin, R.D., "An Introduction to the Mathematical Theory of Vibrations of Elastic Plates, U.S. Army Signal Corps Engineering Lab, Fort Monmouth, NJ, 1955.

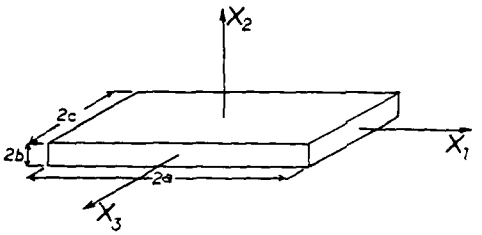


Figure 1 Crystal Plate Coordinates and Dimensions

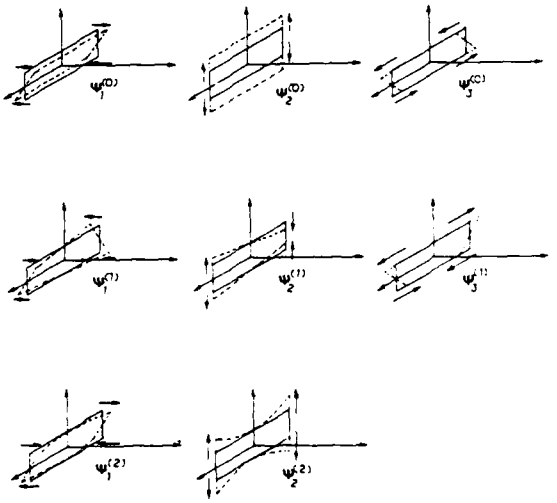
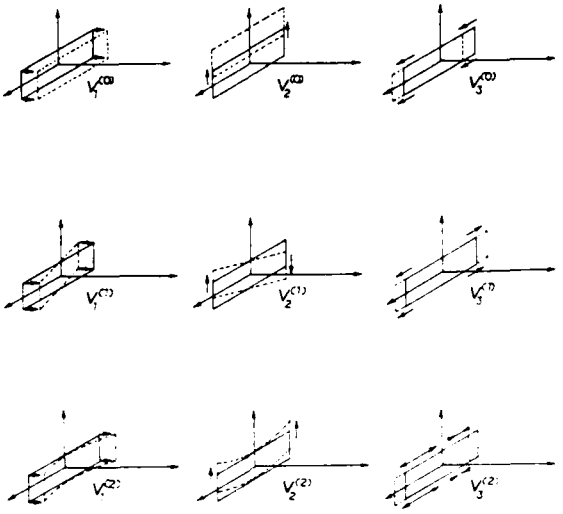


Figure 2 One-Dimensional Displacement Components

continued ...



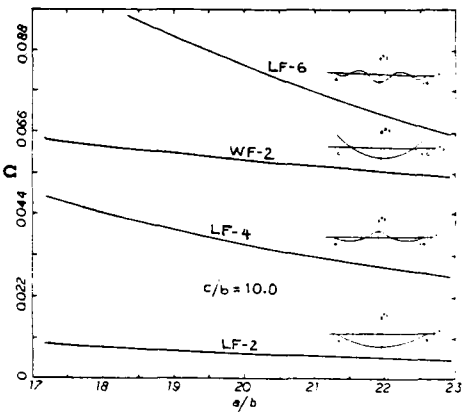
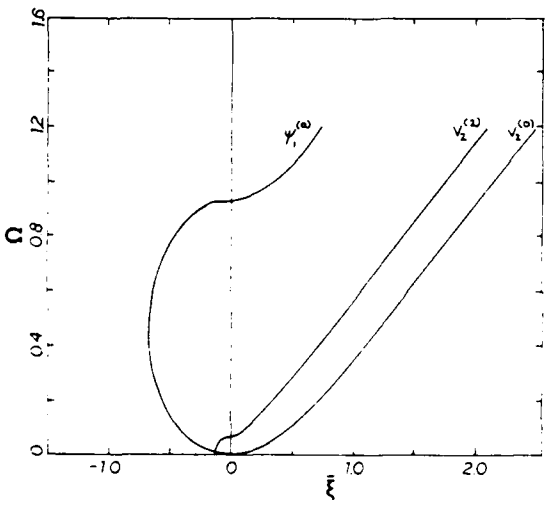


Figure 5 Frequency Spectrum of a Y-cut Plate Ratio c/b fixed.



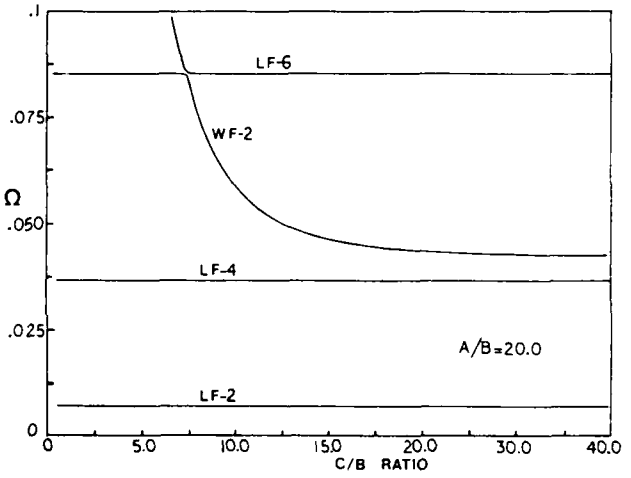


Figure 6: Frequency Spectrum of a Y-cut Plate, Ratio A/B fixed.

Figure 3 Dispersion Relation of a Y-cut Plate c/b = 7.63 a/b = 33.45

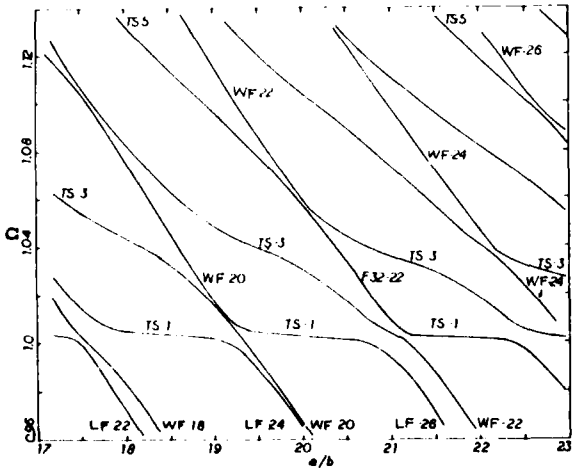


Figure 4 Frequency Spectrum, AT-cut, c/b = 20.0

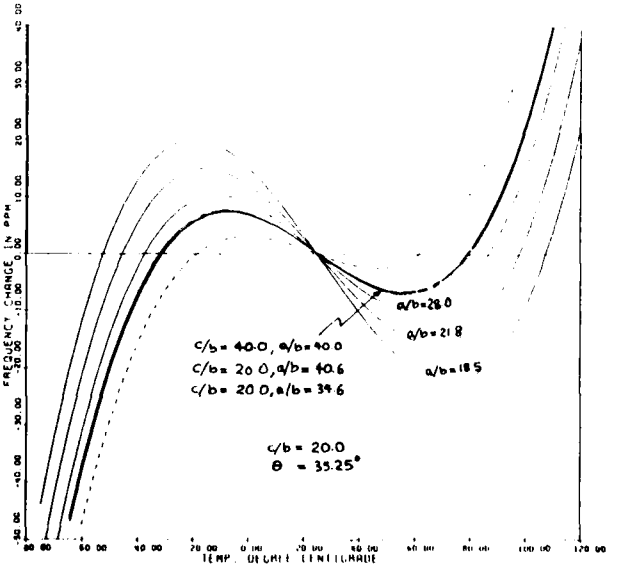


Figure 7: Thickness-Shear Frequency-Temperature Curves for an AT-Cut Plate.

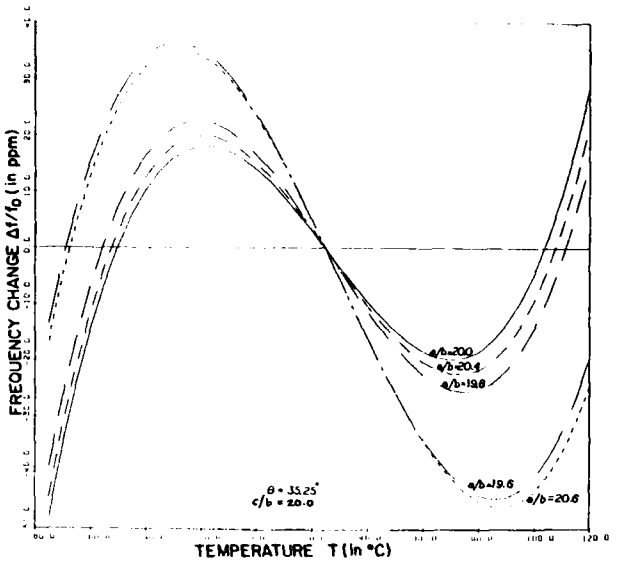


Figure 8 Thickness Shear Frequency-Temperature Curves for an AT-cut Plate with a/b = 19.6, 19.8, 20.0, 20.4 and 20.6.

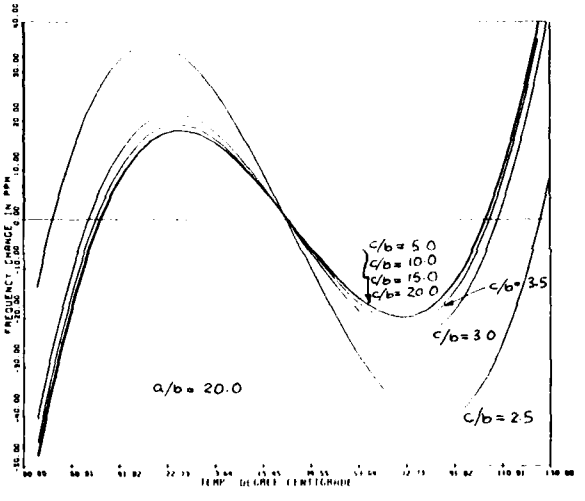


Figure 9: Thickness-Shear Frequency-Temperature Curves for an AT-Cut Plate. Ratio a/b fixed. Ratio c/b variable.

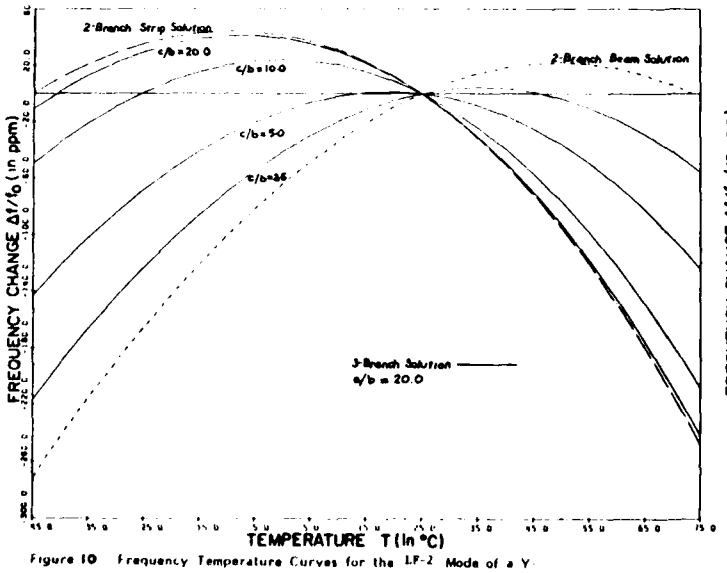


Figure 10 Frequency Temperature Curves for the LF-2 Mode of a Y-cut Plate for Different c/b Ratios

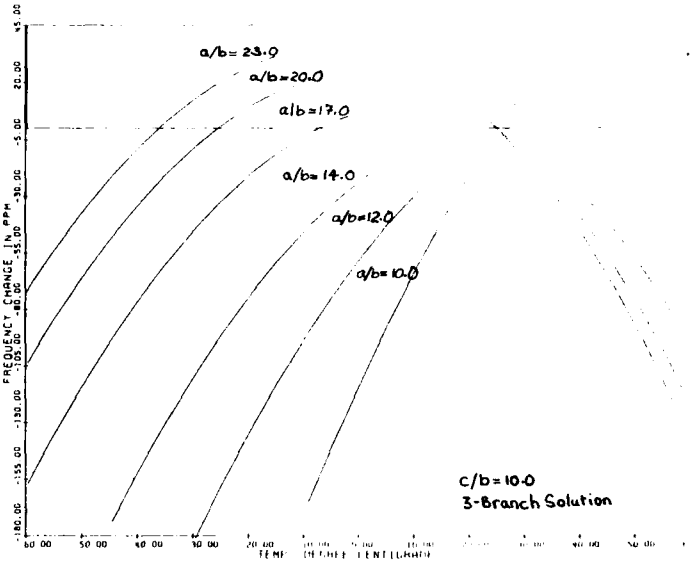


Figure 11: Frequency-Temperature Curves for the LF-2 Mode of a Y-Cut Plate at Different a/b

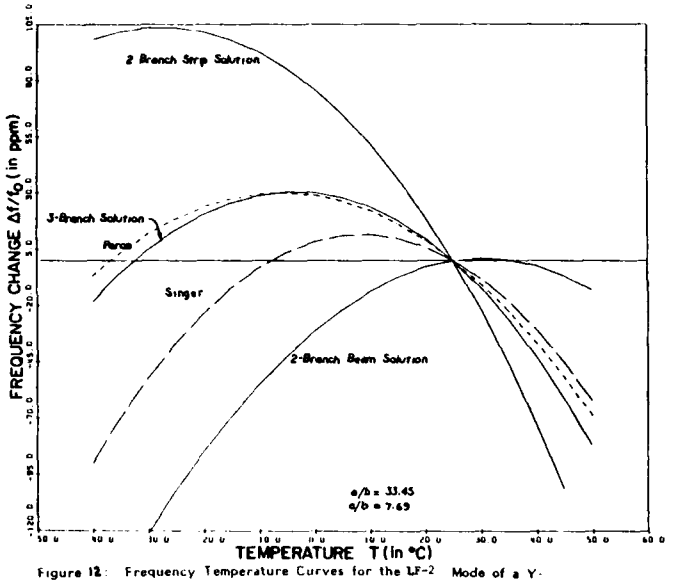


Figure 12: Frequency Temperature Curves for the LF-2 Mode of a Y-cut Plate

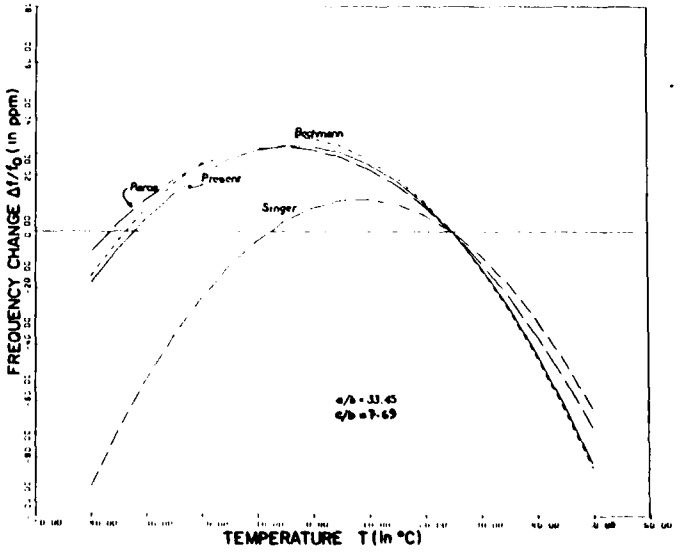


Figure 13: Frequency-Temperature Curves for the Fundamental Symmetric Flexure of a Y-cut Bar

A NOTE ON "BALLATO'S ANGLE INCREMENT"

John H. Sherman, Jr.

2022 Woodcrest Drive
Lynchburg, VA 24503

Abstract

Paper reviews the theory of the design angle increment known loosely as "Ballato's increment", which is the range of orientations in which one finds the designs to achieve a given temperature-frequency characteristic for all capacitive circuit loadings. It uses published information on the AT-cut to explain why the increment occurs, its magnitude, and how it affects resonator design. It gives values of temperature coefficient of the coefficient of electromechanical coupling for the SC-cut and the corresponding "increment".

Introduction

At the 38th Annual Frequency Control Symposium, in the oral presentation of the paper by Stevens and Tiersten [1], mention was made of an arbitrary seeming change in the orientation angle of a practical resonator from that computed by the method of the paper. This seemed of the nature of an interpolation in the presentation. The printed version of the paper credits a 6 minute angle correction to a communication from Lukaszek calling attention to the work of Bechmann referenced as [4] below. Attempting to understand what was going on, I have identified this correction with what Dr. John Vig of the Fort Monmouth Laboratories has called "Art Ballato's Correction Factor". This increment of angle was mentioned casually, but not in sufficient detail that most hearers know what is meant. One hesitates to interrupt an otherwise important discussion to create what is surely only a digression. The purpose of this paper is to explain "Ballato's Angle Increment" once and for all so it will never more have to be a mystery.

Discussion

Even at this late date most of us in crystal work maintain a familiarity with the resonator theory developed by McSkimin and Sykes, originally published in BSTJ, which makes up Chapters VI and VII of the compilation by Heising [2]. The theory developed in these two monographs is quite simplified, but is still a good introduction to the behavior of isolated resonators. By isolated resonators is meant just that, resonators which are not affected in their resonance by any effect of the environment. This description is of a highly idealized entity; a real resonator is in intimate conjunction with an electrical environment in any conceivable use. This is a very useful view of the resonator, nevertheless, as it defines a readily visualized and significant limiting case for the application of resonators. At this limit certain naively expected regularities are observed, and under no other circumstances.

In particular, resonance effects here become consequences of geometry and material properties only. The overtone frequencies of a purely thickness mode resonator, such as an AT-cut resonator of large width-thickness ratio, are rationally related. The frequencies of the modes are exact integral multiples of a lowest, or fundamental, frequency. They become simply a property of the geometry. Similarly the temperature-frequency behavior of all modes is the same. It is a *material property only*. Again, these statements apply only to absolutely isolated resonators, a limiting case.

This very simplified analysis was almost immediately too incomplete for adequate understanding of phenomena or for resonator design. The theory was concurrently bringing the environment into the discussion, even at the cost of greatly increased complexity of the treatment. Resonator theory had to bring the electric field into the picture. Doing this introduced consideration of the reasons why real overtone frequencies are observed not to be rationally related and why temperature-frequency behavior is observed to be strongly dependent upon overtone order.

By 1940 Bechmann [3] had published an expression for the depression of the resonant frequency by the field. His expression, to which he assigned the symbol gamma, correctly stated the variation of the effect inversely with the square of the overtone order. Bechmann [4] also investigated the variation of temperature-frequency behavior with overtone and in 1955 published an expression for that. In the simplest terms the difference in frequency is a function of the electromechanical coupling of the quartz to the electric field, and the difference in the temperature-frequency characteristic of the temperature coefficient of the coupling.

The piezoelectric resonator is modeled by a number of equivalent circuits. The most accurate

CH2186-0/85/0000-0427$1.00©1985IEEE

of the assemblage, in which the body of the quartz is represented by a transmission line analogue, was elaborated by Ballato [5] to illustrate a large number of significant relationships in resonators. Development of the more conventionally used equivalent circuits follows from this by the expansion of the transmission line impedance using Foster's reactance theorem into an infinite parallel assemblage of LC arms. A well-known equivalent, originally due to Mason, is shown in Figure 1. This contains an element, a negative capacitance, which is not physically realizable, but this equivalent has certain advantages here over the more familiar representation due to Butterworth and Van Dyke shown in Figure 2. For the purposes of this paper the Mason equivalent is useful, as it models the antiresonant frequency at 1/2 /LC and the series resonant frequency as the result of a frequency lowering from this. This generates an expression for the frequencies of resonance in a form similar to the expressions derived by Bechmann from the study of the physics of quartz.

In the equivalents it is necessary to recognize that the two capacitances are not arbitrary. The capacitance C_0 is a very real capacitor, the capacitor formed by the electrodes separated by the thickness of the quartz. The other capacitance, C_1 is a fictitious capacitor which does not exist anywhere as a structure, but which is, rather, a model of an energy storage mechanism. When the real capacitor, C_0, is charged, there is energy stored in the electric field and also energy stored in the strain field of the piezoelectrically distorted quartz. This strain energy is modeled as the energy stored in the "motional capacitance", C_1, charged to the same potential. The ratio of the two capacitors is given by the familiar expression:

$$C_0/C_1 = \pi^2/8k^2 \qquad (1)$$

where k is the electromechanical coupling coefficient. A coefficient can be considered appropriate for each overtone. If k_1 is presumed to be the appropriate coefficient for the fundamental, then $k_n = k_1/n$. The coupling coefficient will hereafter be noted simply as k and the ratio of the capacitances, C_0/C_1, as ρ. The ratio on the overtone, ρ_n, is then $n^2\rho$. When this is done the antiresonant frequency of the equivalent models the resonant frequency of the isolated resonator and the series resonant frequency of the equivalent models the frequency of the quartz in the electric field as lowered by "Bechmann's gamma".

For the equivalent of Figure 1,

$$(f_{sn}/f_{an})^2 = (1-1/n^2\rho) \qquad (2)$$

Taking logarithmic derivatives of both sides with respect to temperature in order to obtain temperature coefficients,

$$2[T'(f_{sn})-T'(f_{an})] = T'(\rho)/(n^2\rho-1) \qquad (3)$$

The antiresonant frequency models the frequency of the isolated resonator for which the temperature-frequency behavior is independent of overtone order. We can therefore write $T'(f_a)$ for $T'(f_{an})$ and then write an expression for the difference in the series resonant temperature-frequency behavior of a resonator on two different overtones, as the antiresonant behavior subtracts out exactly on every pair of modes. In particular, between the fundamental and third overtone we get:

$$T'(f_{s3})-T'(f_{s1}) = \frac{T'(\rho)}{2}\left[\frac{1}{9\rho-1} - \frac{1}{\rho-1}\right] \qquad (4)$$

Since for all cuts of quartz ρ is in excess of 100, these expressions can be simplified safely by ignoring the subtracted unity in each denominator. Doing this and rearranging terms yields:

$$T'(\rho) = \frac{9\rho}{4}\left[T'(f_{s1}) - T'(f_{s3})\right] \qquad (5)$$

To obtain the difference between the fundamental temperature-frequency characteristics at series and anti- resonance one writes (from (2)):

$$T'(f_{s1}) - T'(f_a) = T'(\rho)/2(\rho-1) \qquad (6)$$

Bechmann's result

In Bechmann [4] a value of -0.66×10^{-6}/K is given for the difference of slope of the temperature-frequency characteristic of an AT-cut plate between the fundamental and a harmonic of infinite order. Reference to (2) shows that the difference between the characteristics vanishes as the order increases without limit. Since the antiresonant characteristic is invariant with overtone and the infinite order series resonant characteristic merges with the antiresonant characteristic, on the fundamental the difference between the series resonant characteristic and the antiresonant characteristic is equal to the difference between the fundamental and the infinite order at series resonance. When this is combined in (6) with the plane-wave value of ρ, 159.2, a value of -210×10^{-6} is found for $T'(\rho)$ for the AT-cut.

Ballato and Iafrate [6] contains a plot, Figure 32, of the partial derivative of $T(f_a)$ with respect to θ, the (second) angle of rotation, around the x'-axis, vs the first angle, ϕ, around the z-axis, for the three thickness modes of motion. For the entire series of zero-coefficient thickness mode cuts on the principal locus, the mode is c. For the AT-cut, ϕ=0, the value given is -5.2×10^{-6}/K/degree θ. From these two values one finds that the difference between the slopes at the inflection points of the characterisic curves is equivalent to 0.1269 degrees or 7.6 minutes of rotation.

The fractional frequency difference is approximated by $1/2n^2\rho$. The difference of the slopes and the difference of the frequencies track together as the crystal is operated through a range of load capacitance [7]. If the angle for a desired slope of characteristic is known for series resonance, then the same slope can be

attained at a new higher angle when operated with a load capacitance:

$$\Delta f/f = [1/(2n^2\rho)][C_0/(C_0+C_1)] \quad (7a)$$

$$\Delta\theta = [7.6/(n^2)][C_0/(C_0+C_1)] \text{ min.} \quad (7b)$$

The SC-Cut

Information was not available for the SC-cut to determine the equivalent angle difference or the temperature coefficient of the electromechanical coupling coefficient. For this reason the Piezo Crystal Company fabricated a test crystal especially for the necessary measurements to be made. The resonator is a plano-plano unit at 10 Mhz fundamental frequency. It runs well at the fundamental and third mode, but the fifth mode could not be tracked with the available test equipment. The temperature-frequency characteristic was measured on the fundamental every five degrees from -50 to +150 degrees C and on the third mode every two degrees from +50 to +130 degrees. Both curves were of excellent symmetry and fit to a cubic equation. The slopes at the respective inflection points were taken as the values of the temperature coefficients of frequency to be compared.

The values taken from the least squares fitted curves are:

$$T'(f_{s1}) = -0.673408476\times10^{-6}/K$$

and

$$T'(f_{s3}) = -0.002383876\times10^{-6}/K \quad (8)$$

The difference is $-0.6710246\times10^{-6}/K$. Ballato's best value for the electromechanical coupling coefficient for SC-cut quartz is 0.0499, which corresponds to a capacitance ratio of 495.5. Using (5) one computes:

$$T'(\rho) = -748.005\times10^{-6}/K \quad (9a)$$

whence

$$T'(k) = 374.02\times10^{-6}/K \quad (9b)$$

From Figure 32 of [6] we read $T'(f_a) \approx 0$ to have an SC-cut value of $-3.8\times10^{-6}/K/\text{degree}^a 0$. Applying (6) to the foregoing difference yields a value of $-0.75637\times10^{-6}/K$ for the difference of slope between series and anti- resonance on the fundamental. When divided by $-3.8\times10^{-6}/K$ this yields:

$$\Delta\theta = 0.19905 \text{ degrees}$$

or:

$$\Delta\theta = 11.94 \text{ minutes} \quad (10)$$

This, again, gives the full range of the angle increment which must be added to the series resonant orientation in order to obtain the same slope as the series resonant curve if the resonator is to be used with a load capacitance. The formulas used in (7) applied with changed and 0 give the division of the frequency change and angle increment with load capacitance for the SC-cut.

Further Discussion

Reference to [6] shows a complex of curves between Figures 32 and 36 in which these ideas are almost implicit. This investigation of the AT- and SC- cuts would seem to be driving toward two points on curve c of Figure 36. This figure must be judged as unreliable, however, for it does not contain Bechmann's value of $-0.66\times10^{-6}/K$ at zero or a value which even approximates the value of $-0.756\times10^{-6}/K$ found in this investigation for the SC-cut. On the other hand, Figure 32 is correct at zero, and is probably reliable for the SC-cut also. The value of $374\times10^{-6}/K$ for T'(k) of the SC-cut is not a bad confirmation of the value 230×10^{-6}~/K in Figure 33, agreeing to better than a factor of two.

Recapitulation of Results

Values of certain numbers of interest have been developed for the SC-cut. They are tabulated here along with other standard values for both the AT- and SC- cuts:

Quantity	AT-cut value	SC-cut value
Coupling Coefficient, k	0.087958	0.0499
Temperature Coefficient T'(k)	105x10~-6~/K	375x10~-6~/K
Capacitance ratio, ρ	159.2	495.5
Temperature Coefficient T'(ρ)	-210x10~-6~/K	-750x10~-6~/K
$T'(f_{s1})-T'(f_a)$	-0.66x10~-6~/K	-0.755x10~-6~/K
Equiv. angle difference (min)	7.6	12

Acknowledgements

To Piezo Crystal Co. of Carlisle, PA, for the excellent test resonator from which the determination of SC-cut properties was made, and to Dr. Arthur Ballato from whom the significance of the absolute properties of the antiresonant resonator was gradually learned through many conversations.

References

[1] D. S. Stevens and H. F. Tiersten, "On the change in orientation of the zero-temperature contoured SC-cut quartz resonator with the radius of the contour", Proc 38th AFCS, 1984, pp 176-183

[2] R. A. Heising editor, "Quartz Crystals for Electrical Circuits", Van Nostrand, 1946, Electrical Industries Assn., 1978

[3] W. G. Cady, "Piezoelectricity", McGraw-Hill, 1946, Dover, 1964, section 250

[4] R. Bechmann, "Influence of the order of overtone on the temperature coefficient of frequency of AT-type quartz resonators", Proc IRE, Vol 43, no 11, 1955, pp 1667-68

[5] A. D. Ballato, "Transmission line analogues for stacked piezoelectric crystal devices", Proc 26th AFCS, 1972, pp 86-91

A. Ballato and T. Lukaszek, "Mass effects on crystal resonators with arbitrary piezo-coupling", Proc 27th AFCS, 1973, pp 20-29

[6] A. Ballato and G. J. Iafrate, "The angular dependence of piezoelectric plate frequencies and their temperature coefficients", Proc 30th AFCS, 1976, pp 141-156

[7] J. H. Sherman, Jr., "Orientation of an AT-cut blank from specifications of temperature-frequency characteristic", Proc 4th Quartz Crystal Conference, 1982, pp 288-299

Notes: AFCS = Annual Frequency Control Symposium

Proceedings of Quartz Crystal Conferences available from Electronic Industries Association, 2001 Eye St., NW, Washington, DC 20006

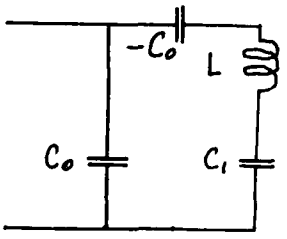


Figure 1. Equivalent network due to Mason

$$(f_s/f_a)^2 = 1 - C_1/C_0$$

$$C_0/C_1 = f_a^2/(f_a^2-f_s^2)$$

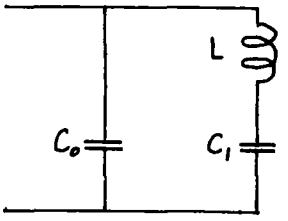


Figure 2. Equivalent network due to Butterworth and Van Dyke

$$(f_a/f_s)^2 = 1 + C_1/C_0$$

$$C_0/C_1 = f_s^2/(f_a^2-f_s^2)$$

VIBRATIONS OF PIEZOELECTRIC DISCS UNDER INITIAL STRESSES

M. Cengiz Dökmeci

Faculty of Aeronautics and Astronautics
Istanbul Technical University, P.K.9, Istanbul

ABSTRACT

A system of two-dimensional equations is derived to govern high frequency motions of piezoelectric discs (plates) under initial stresses. The approximate governing equations are deduced from a three-dimensional quasi-variational principle of piezoelectricity by expanding the electric potential and the incremental components of mechanical displacement in a series of Jacobi's polynomials. These equations in an invariant form are applicable to all the types of extensional, flexural and torsional motions of piezoelectric strained discs. Besides, they incorporate as many higher order effects as deemed necessary in any case of interest by a proper truncation of the series. Further, some special cases, and in particular, the case of piezoelectric unstrained discs and the uniqueness for its solutions are indicated.

Key Words: piezoelectricity, quasi-variational principles, plate vibrations, initial stresses, discs.

INTRODUCTION

The mathematical modelling of the dynamic response of piezoelectric plates was extensively studied in the literature, and it was reviewed by several authors (e.g.,/1-3/). Most recently, Gerber and Ballato /4/ provided almost a complete list of pertinent publications dealing with dynamic problems of the piezoelectric elements. In view of these, it appears that the application of initial stresses or strains may be utilized to control the performance of certain piezoelectric devices. However, the effect of initial stresses in piezoelectric elements was treated only in a few particular cases. Especially, the propagation of surface acoustic waves was investigated both analytically and experimentally in a piezoelectric continuum with initial stresses /5,6/. In addition, a quasi-variational principle was recently derived to govern the motions of piezoelectric strained continua /7/. Now, an attempt is made to develop consistently the two-dimensional equations in an invariant form, of successively higher orders of approximation for piezoelectric strained discs of any geometrical shape.

The presence of initial stresses or strains may significantly change the static and dynamic behavior of structures (e.g./8,9). Revealing this fact, Thurston /10/ studied the wave propagation in stressed crystals under hydrostatic pressure. Herrmann and Armenakas /11/ investigated the vibrations and stability of elastic plates under initial stresses. Further, Lee and his colleagues (e.g.,/12,13/) treated the high-frequency vibrations of crystal plates so as to predict changes in the resonant frequencies due to initial stresses. Additional references dealing with the effect of initial stresses in plates were compiled by the author /14/. Moreover, in the absence of initial stresses, one should mention the recent works of Bogy and his students /15,16/, Karlash /17/, Zaretskii-Feoktistov /18/, Baboux and his colleagues /19/ and Pan-fu /20/ for various problems of piezoelectric discs. As for piezoelectric plates or discs with initial stresses, this is precisely the topic of this paper.

In this paper, the method of reduction due to Mindlin /21/ is applied to derive a system of two-dimensional governing equations of piezoelectric plates (discs) under initial stresses. In the first stage, the three-dimensional differential equations of piezoelectric strained continua are expressed by means of a quasi-variational principle /7/. Then, the geometry of a piezoelectric disc is described, certain regularity assumptions are introduced, and the electric potential and the incremental components of disc are expanded in a series of Jacobi's polynomials. Also, the higher orders components of stress, electric displacements and surface loads are defined in consistent with the series expansions. In the next stage, the governing equations of piezoelectric strained discs are consistently and systematically formulated by using the quasi-variational principle together with the series expansions of field quantities. The governing equations incorporate as many higher order effects as deemed desirable, and they take into account for the coupling between extensional, flexural and torsional modes. Lastly, special cases and in particular, the case of piezoelectric unstrained discs are pointed out, and the results are briefly discussed.

NOTATION

In the paper, standard tensor notation is used in a Euclidean 3-space. Accordingly, Einstein's summation convention is implied for all repeated Latin indices (1,2,3) and Greek indices (1,2). Superposed dots are assigned for time differentiations, primes for partial differentiations with respect to the thickness coordinate x^3, and commas and semicolons for partial and covariant differentiations with respect to space coordinates, respectively. Further, a piezoelectric region B with its boundary surface $\partial B(=S_t \cup S_u = S_d \cup S_p)$ is referred to by a fixed, right-handed system of curvilinear coordinates x^K in the space. The symbol B(t) refers to the region B at time t and n_k to the unit outward vector normal to ∂B. Asterisks are used to indicate prescribed quantities. The time interval is denoted by $T=(t_0,t_1)$ and the thickness interval by $H=(-h,h)$.

THREE-DIMENSIONAL EQUATIONS OF PIEZOELECTRICITY

The three-dimensional fundamental equations to govern the motions of a piezoelectric strained continuum are summarized in differential form as follows /1,9,5,6/.

Divergence Equations:

$$T^{ij}{}_{;i} - \rho b^j = 0 \; ; \; T^{ij} = \sigma_o^{ik} u^j{}_{;k} + \sigma^{ij}$$

$$D^i{}_{;i} = 0 \qquad (1)\text{-}(2)$$

Gradient Equations

$$S_{ij} = \frac{1}{2}(u_{i;j} + u_{j;i}) \qquad (3)$$

$$E_i = -\phi_{,i} \qquad (4)$$

Constitutive Equations

$$\sigma^{ij} = C^{ijkl} S_{kl} - C^{kij} E_k \qquad (5)$$

$$D^i = C^{ijk} S_{jk} + C^{ij} E_j \qquad (6)$$

CH2186-0/85/0000-0431$1.00©1985IEEE

Boundary Conditions

$$T_*^j - n_i T^{ij} = 0 \quad \text{on } S_t \tag{7}$$

$$u_i^* - u_i = 0 \quad \text{on } S_u \tag{8}$$

$$\sigma^* - n_i D^i = 0 \quad \text{on } S_d \tag{9}$$

$$\phi^* - \phi = 0 \quad \text{on } S_p \tag{10}$$

Initial Conditions

$$u_i(x^j,t_o) - v_i^*(x^j) = 0 \ ,$$

$$\dot{u}_i(x^j,t_o) - w_i^*(x^j) = 0 \ ,$$

$$\phi(x^j,t_o) - \Psi^*(x^j) = 0 \quad \text{in } B(t_o) \tag{11}$$

In the above equations, T^{ij} is the stress tensor, u_i the incremental displacement vector, ρ the mass density, $b_j(=\ddot{u}_j)$ the acceleration vector, $\sigma_o{}^{ij}$ and σ^{ij} the initial and incremental stress tensors, D^i the electric displacement vector, E_i the quasi-static electric field vector, ϕ the electric potential, S_{ij} the incremental stain tensor, $Tj(=n_i T^{ij})$ the stress vector and $\sigma(=n_i D^i)$ the surface charge. Also, C^{ijkl}, C^{ijk} and C^{ij} denote the elastic, piezoelectric and dielectric material constants with their usual symmetry properties in the form

$$C^{ijkl} = C^{jikl} = C^{klij}, \ C^{ijk} = C^{ikj}, \ C^{ij} = C^{ji} \tag{12}$$

A QUASI-VARIATIONAL PRINCIPLE

The fundamental differential equations (1)-(11) can be alternatively expressed by means of a quasi-variational principle in the form.

$$\delta I = \delta I^{\alpha}_{\alpha} + \delta J^{\alpha}_{\alpha} + \delta L^{\alpha}_{\alpha} + \delta K^{\alpha}_{\alpha} + \delta N^{\alpha}_{\alpha} = 0 \tag{13}$$

with

$$\delta I_1^1 = \int_T dt\{\int_B [(\sigma_o{}^{ik}u^j{}_{;k} + \sigma^{ij})_{;i} - \rho b^j]\delta u_j dV + \int_{S_t} T_*^j - n_i(\sigma_o{}^{ik}u^j{}_{;k} + \sigma^{ij})\ \delta u_j dS\} \tag{14}$$

$$\delta I_2^2 = \int_T dt\{\int_B -D^i{}_{;i}\delta\phi dv + \int_{S_d} (n_i D^i - \sigma^*)\delta\phi dS\} \tag{15}$$

$$\delta J_1^1 = \int_T dt\int_B [S_{ij} - \tfrac{1}{2}(u_{i;j} + u_{j;i})]\delta T^{ij} dV \tag{16}$$

$$\delta J_2^2 = \int_T dt\int_B -(E_i + \phi_{,i})\delta D^i dV \tag{17}$$

$$\delta L_1^1 = \int_T dt\int_B [\sigma^{ij} - (C^{ijkl}S_{kl} - C^{kij}E_k)]\delta S_{ij} dV \tag{18}$$

$$\delta L_2^2 = \int_T dt\int_B [D^i - (C^{ijk}S_{jk} + C^{ij}E_j)]\delta E_i dV \tag{19}$$

$$\delta K_1^1 = \int_T dt\int_{S_u} (u_i^* - u_i)\delta T^i ds \tag{20}$$

$$\delta K_2^2 = \int_T dt\int_{S_p} (\phi^* - \phi)\delta\sigma ds \tag{21}$$

and

$$\delta N_1^1 = \int_B \rho\{[\dot{u}_i(x^j,t_o) - w_i^*(x^j)]\delta u_i(x^j,t_o) + [u_i(x^j,t_o) - v_i^*(x^j)]\delta\dot{u}^i(x^j,t_o)\}dV \tag{22}$$

$$\delta N_2^2 = \int_B \rho[\phi(x^j,t_o) - \Psi^*(x^j)]\delta\phi(x^j,t_o)dV \tag{23}$$

The quasi-variational principle (12) is fully unconstrained, and it evidently leads to all the fundamental differential equations of piezoelectricity (1)-(12) as the appropriate Euler-Lagrange equations; and conversely, if the fundamental differential equations are met, the quasi-variational principle is clearly satisfied. This principle is recently deduced from Hamilton principle by the author /7/, and it can be similarly obtained from the principle of virtual work as will be reported in a forthcoming communication.

GEOMETRY OF A PIEZOELECTRIC DISC

Consider a piezoelectric disc of any geometrical shape, embedded in the Euclidean 3-space. The piezoelectric disc of thickness 2h is referred to the system of curvilinear coordinates x^k, with the faces, of area A, at $x^3=\pm h$ and with x^α the coordinates on the midplane which intersects the right cylindrical boundary of the disc in a Jordan curve C. The disc is coated with perfectly conducting electrodes on both its faces. Further, one should recall the fundamental assumption of the form

$$(2h/d)<<1 \tag{24}$$

where d is a characteristic length of disc. This allows one to treat the disc (plate) as a two-dimensional mathematical model of a three-dimensional body.

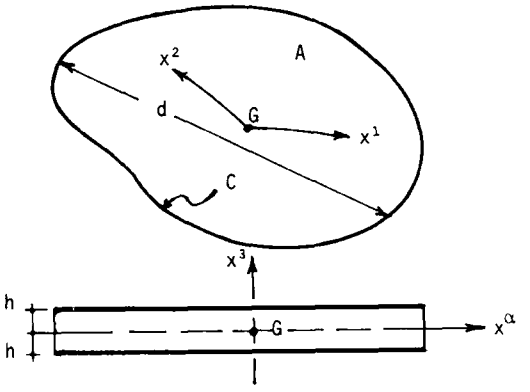


Figure 1. Disc geometry

SERIES OF ELECTRIC POTENTIAL AND INCREMENTAL DISPLACEMENTS

The fundamental assumption (24) and the absence of any kind of singularities as well as the suitable regularity and smoothness assumptions are considered for the piezoelectric disc region B+∂B. In addition, all the field functions together with their derivatives are assumed to exist and to be continuous in the closure of disc $\bar{B}(=B\cup\partial B)$, and not to vary widely across the disc thickness. Thus, the electric potential and the incremental components of displacements of disc are approximated by a series in the thickness coordinate as

$$u_i(x^j,t) = \sum_{n=0}^{N=\infty} Q_n^{(i)}(x^3)\, u_i^{(n)}(x^\alpha,t), \tag{25}$$

$$\phi(x^j,t) = \sum_{n=0}^{N=\infty} P_n(x^3)\,\phi_n(x^\alpha,t) \tag{26}$$

Here, the functions Q_n^i and P_n are consistently chosen as Jacobi's polynomials /22,23/ in the form

$$Q_n^i = P_n = J_n(x^3) \tag{27}$$

with

$$J_n(z) = 1, z, 1-2z^2, z-\tfrac{3}{2}z^3, \ldots \tag{28}$$

The choice of Eqs.(25)-(27) as a starting point leads to the governing equations of piezoelectric disc in a consistent and tractable manner; this will be shown below. Moreover, in lieu of Jacobi's polynomials, Legendre's polynomials, power series and trigonometric series can be similarly chosen /1-3/. The present

choice, however, is more fruitful in the case of circular and elliptical discs.

HIGHER ORDER COMPONENTS OF STRESS AND ELECTRIC DISPLACEMENTS AND SURFACE LOADS

In accordance with the foregoing assumptions and the series expansions (25)-(27), the two-dimensional stress and electric displacement components and surface loads of order n in the form

$$[T^{\alpha j}_{(n)}, D^{\alpha}_{(n)}, N^{j}_{(n)}, D^{3}_{(n)}] = \int_H [(\sigma^{\alpha j}_o, D^{\alpha}) J_n ; (\sigma^{3j}_o, D^3) J_n'] dx^3$$

$$[T^{i\alpha}_{o(m+n)}, T^{\alpha 3}_{o(m+n)}] = \int_H [(\sigma^{j\alpha}_o J_m, \sigma^{\alpha 3}_o J_m') J_n] dx^3$$

$$[N^{\alpha}_{o(m+n)}, N^{3}_{o(m+n)}] = \int_H [(\sigma^{3\alpha}_o J_m, \sigma^{33}_o J_m') J_n'] dx^3$$

$$[F^{\alpha}_{o(m+n)}, F^{3}_{o(m+n)}] = [(\sigma^{3\alpha}_o J_m, \sigma^{33}_o J_m') J_n]_H$$

$$[F^{j}_{(n)}, G_{(n)}] = [(\sigma^{3j}_o, D^3) J_n]_H$$

$$[T^{j}_{\star(n)}, D_{\star(n)}] = \int_H [(T^{j}_{\star}, \sigma^{\star}) J_n dx^3] \quad (29)\text{-}(32)$$

and

$$(I_m, I_{mn}) = \int_H (J_m, J_m J_n) dx^3 \quad (33)$$

are defined.

HIGHER ORDER, EQUATIONS OF MOTION AND ASSOCIATED NATURAL TRACTION BOUNDARY CONDITIONS

To derive the higher order equations of motion of piezoelectric strained disc. the variational integral (14) is splitted into the area integral over the midplane A and the integral across the thickness, namely,

$$\delta I^1_1 = \int_T dt \{ \int_A dA \int_H [(\sigma^{ik}_o u^j_{;k} + \sigma^{ij})_{;i} - \rho b^j] \delta u_j dx^3 + \oint_C ds \int_H T^{j}_{\star} - n_i(\sigma^{ij} + \sigma^{ik}_o u^j_{;k}) \delta u_j \, dx^3 \} \quad (34)$$

Here, the tractions are taken to be prescribed on the edge boundary surface of disc, and accordingly, the displacements are specified on the faces. Following the method of reduction as in /21/, the series expansions of incremental displacement components (25)-(27) are inserted into Eq.(34), and it is integrated with respect to the thickness coordinate. Then, using the two-dimensional divergence theorem and regrouping the higher order components of stress and surface loads, one obtains

$$\delta I^1_1 = \int_T dt \sum_{n=0}^{N} \{ \int_A \chi^{j}_n \delta u^{(n)}_j dA + \oint_C \chi^{\star j}_n \delta u^{(n)}_j dx^3 \} \quad (35)$$

This equation leads to the equations of motion and the natural boundary conditions, of order n for arbitrary and independent variations $\delta u_j^{(n)}$ in the quasi-variational equation (13). Thus, the equations of motion of order n are expressed by

$$\chi^{j}_n = T^{\alpha j}_{n ;\alpha} - N^{j}_n + F^{j}_n + \sum_{m=0}^{N} (T^{\beta\alpha}_{o(m+n)} u^{j}_{m ;\alpha})_{;\beta} + T^{\beta_3}_{o(m+n)} u^{j}_{m ;\beta} - \sum_{m=0}^{N} (N^{\alpha}_{o(m+n)} u^{j}_{m ;\alpha} + N^{3}_{o(m+n)} u^{j}_{m}) + \sum_{m=0}^{N} (F^{\alpha}_{o(m+n)} u^{j}_{m ;\alpha} + F^{3}_{o(m+n)} u^{j}_{m}) - \rho \sum_{m=0}^{N} I_{mn} \ddot{u}^{j}_{m} = 0 \; ; \; n=1,2,\ldots,N \text{ on } A \quad (36)$$

Besides, the natural boundary conditions of order n are written in the form

$$\chi^{\star j}_n = T^{\star j}_n - n_\alpha [T^{\alpha j}_n + \sum_{m=0}^{N} (T^{\alpha\beta}_{o(m+n)} u^{j}_{(m) ;\beta} + T^{\alpha 3}_{o(m+n)} u^{j}_{(m)})] = 0, \quad n=1,2,\ldots,N \text{ along } C \quad (37)$$

HIGHER ORDER, EQUATIONS OF ELECTROSTATICS AND ASSOCIATED NATURAL BOUNDARY CONDITIONS OF SURFACE CHARGE

By paralelling to the derivation above, the variational integral (15) is evaluated, the definitions of Eqs. (29)-(32) are used, and then the equations of electrostatics of order n are expressed by

$$\chi_n = D^{\alpha}_{(n) ;\alpha} + G_{(n)} - D^{3}_{(n)} = 0 \; ; \; n=1,2,\ldots,N \text{ on } A \quad (38)$$

and the natural boundary conditions of surface charge of order n by

$$\chi^{\star}_n = D^{\star}_{(n)} - n_\alpha D^{\alpha}_{(n)} = 0 \; ; \; n=1,2,\ldots,N \text{ along } C \quad (39)$$

where the surface charges are taken to be prescribed only on the edge boundary of disc.

HIGHER ORDER, DISPLACEMENT AND ELECTRIC POTENTIAL BOUNDARY CONDITIONS

The electric potential and the mechanical displacements are considered to be given on the faces. Accordingly, the variational integrals (20) and (21) are carried out on A. Then, the n-th order boundary conditions of displacements are obtained as

$$u^{(n)}_i - u^{\star(n)}_i = 0 \; ; \quad n=1,2,\ldots,N \quad \text{on } A \quad (40)$$

and those of electric potential as

$$\phi^{(n)} - \phi^{(n)}_{\star} = 0 \; ; \quad n=1,2,\ldots,N \quad \text{on } A \quad (41)$$

DISTRIBUTIONS OF INCREMENTAL STRAIN AND ELECTRIC FIELD

By inserting the series expansions (25)-(27) into the variational equations (16) and (17), and then integrating with respect to the thickness coordinate and using the higher order components of stress and electric displacements, one readily arrives at the distributions of incremental strain and those of electric field in the form

$$(S_{\alpha\beta}, S_{\alpha 3}; E_\alpha, E_3) = \sum_{n=0}^{N} (S^{(n)}_{\alpha\beta} J_n, S^{(n)}_{\alpha} J_n' + S^{(n)}_{\alpha_1} J_n, S_{33} J_n'; E^{(n)}_{\alpha} J_n, E^{(n)}_{3} J_n') \quad (42)$$

where

$$S^{(n)}_{\alpha\beta} = \frac{1}{2}(u^{(n)}_{\alpha ;\beta} + u^{(n)}_{\beta ;\alpha}) \; ; \; S^{(n)}_{\alpha 3} = S^{(n)}_{\alpha} + S^{(n)}_{\alpha_1}$$

$$S^{(n)}_{\alpha} = \frac{1}{2} u^{(n)}_{\alpha} \; , \; S^{(n)}_{\alpha_1} = \frac{1}{2} u^{(n)}_{3 ;\alpha} \; ; \; S^{(n)}_{33} = u^{(n)}_3 \quad (43)$$

and

$$E^{(n)}_{\alpha} = - \phi^{(n)}_{,\alpha} \; , \; E^{(n)}_{3} = - \phi^{(n)} \quad (44)$$

are introduced.

MACROSCOPIC CONSTITUTIVE EQUATIONS

With the help of the distributions of electric field and incremental strain (42)-(44), the variational equations (18) and (19) are evaluated and hence the macroscopic constitutive equations of order n are obtained as follows.

$$T^{\alpha j}_{(n)} = \sum_{m=0}^{N} [A^{\alpha j\beta\gamma}_{m+n} S^{(m)}_{\beta\gamma} + 2(B^{\alpha j\beta_3}_{m+n} S^{(m)}_{\beta} + A^{\alpha j\beta_3}_{m+n} S^{(m)}_{\beta^1}) + B^{\beta\alpha j_{33}}_{m+n} S^{(m)}_{3} - (A^{\beta\alpha j}_{m+n} E^{(m)}_{\beta} + B^{3\alpha j}_{m+n} E^{(m)}_{3})]$$

$$N^{j}_{(n)} = \sum_{m=0}^{N} [B^{3j\beta\gamma}_{n+m} S^{(m)}_{\beta\gamma} + 2(C^{3j\beta_3}_{m+n} S^{(m)}_{\beta} + B^{3j\beta_3}_{n+m} S^{(m)}_{\beta^1}) + C^{3j_{33}}_{m+n} S^{(m)}_{3} - (B^{\alpha_3 j}_{n+m} E^{(m)}_{\alpha} + C^{33j}_{m+n} E^{(m)}_{3})]$$

$$D^{\alpha}_{(n)} = \sum_{m=0}^{N} [A^{\alpha\beta\gamma}_{m+n} S^{(m)}_{\beta\gamma} + 2(B^{\alpha\beta_3}_{m+n} S^{(m)}_{\beta} + A^{\alpha\beta_3}_{m+n} S^{(m)}_{\beta^1})$$

$$+ B^{\alpha 33}_{m+n}\ S_3^{(m)} + (A^{\alpha\beta}_{m+n}\ E_\beta^{(m)} + B^{\alpha 3}_{m+n}\ E_3^{(m)})]$$

$$D^3_{(n)} = \sum_{m=0}^{N} [B^{3\beta\gamma}_{n+m}\ S^{(m)}_{\beta\gamma} + 2(C^{3\beta 3}_{m+n}\ S^{(m)}_{\beta} + B^{3\beta 3}_{n+m}\ S^{(m)}_{\beta^1})$$
$$+ C^{3\ 33}_{m+n}\ S_3^{(m)} + (B^{3\beta}_{n+m}\ E_\beta^{(m)} + C^{33}_{m+n}\ E_3^{(m)})] \qquad (45)\text{-}(48)$$

where the higher order components of material constants are defined by

$$(A^{i\ldots j}_{m+n},\ B^{i\ldots j}_{m+n}) = \int_H (C^{i\ldots j}{}_{J_m J_n},\ C^{i\ldots j}{}_{J_m' J_n})dx^3$$
$$(C^{i\ldots j}_{m+n}) = \int_H (C^{i\ldots j}{}_{J_m' J_n'})dx^3 \qquad (49)$$

HIGHER ORDER INITIAL CONDITIONS

As before, the variational equation (23) is evaluated, and the initial conditions of displacements and electric potential of order n are expressed by

$$u_i^{(n)}(x^\alpha, t_o) - v_i^{*(n)}(x^\alpha) = 0\ ,$$
$$\dot{u}_i^{(n)}(x^\alpha, t_o) - w_i^{*(n)}(x^\alpha) = 0 \qquad (50)$$
$$\phi^{(n)}(x^\alpha, t_o) - \psi^{*(n)}(x^\alpha) = 0 \qquad (51)$$

where $v_i^{*(n)}$, $w_i^{*(n)}$ and $\psi^{*(n)}$ are the given functions of incremental displacements and electric potentials at time $t=t_o$.

GOVERNING EQUATIONS OF PIEZOELECTRIC STRAINED DISCS

The system of two-dimensional equations of piezoelectric discs (plates) under initial stresses consists of the series expansions of incremental strain and electric potential (25)-(27), the higher order equations of motion and electrostatics (36) and (38), the associated natural boundary conditions (37) and (39)-(41), the distributions of incremental strain and electric field (42)-(44), the macroscopic constitutive equations (45)-(48) and the natural initial conditions (50) and (51).

SPECIAL CASES

The approximate, successively higher orders governing equations of piezoelectric strained discs are formulated in an invariant form, and hence they are readily applicable to an arbitrarily shaped disc using a particular coordinate system most suitable for its geometrical configuration. Among those, the resulting equations for a circular disc can be given by using the system of polar coordinates, that is, $x^1=r$, $x^2=\theta$ and $x^3=z$. Likewise, the system of elliptical coordinates can be selected for an elliptical disc under initial stresses.

In the absence of initial stresses, the two-dimensional equations derived may be reduced to those of piezoelectric unstrained discs. These linear governing equations accommodate high frequency motions of a piezoelectric plate (disc), and have a unique solution in each case of interest. The boundary and initial conditions (37) and (39)-(41) which now exclude the terms involving initial stresses and hence become linear are sufficient for the uniqueness. To prove this, utilizing the technique due to Knops and Payne /24/ and following the author /25/, the existence of two possible solutions is considered and the homogeneous governing equations are formed for the difference of solutions. Then, a logarithmic function is introduced, and it is calculated in terms of the field quantities of disc. By using the convexity of the function and Schwartz's inequality, it is shown that the homogeneous problem may possess only a trivial solution. Accordingly, the linearized version of the initial and boundary conditions (37)-(41) are found to be sufficient for the uniqueness /26/.

CONCLUDING REMARKS

Presented herein is the system of two-dimensional equations of successively higher orders of approximation for all the types of extensional, flexural and torsional motions of piezoelectric discs (plates) under initial stresses. These governing equations are systematically and consistently deduced from the three-dimensional equations of piezoelectricity by means of a quasi-variational principle together with the series expansions of the field quantities. The effects of elastic stiffeness and inertia of electrodes are omitted, but those of shear and normal strains, full anisotropy and heterogeneity are all taken into account. Then, some special cases, and in particular, the case of unstrained piezoelectric disc and the uniqueness for its solution are pointed out. In closing, the case of circular discs with and without initial stresses is reported in the reference /27/, and the detailed analyses of certain motions of strained elliptical discs and the extension of the present results to those of composite discs with initial stresses will be studied in a forthcoming memoir.

ACKNOWLEDGEMENTS

The author would like to express his cordial thanks to Dr. Arthur Ballato and Professor Harry F. Tiersten for stimulating fruitful discussions, and to acknowledge the typing skills of Miss Ayşe Inci and the financial support provided in part by the Army Research Office (U.S.A.R.D.S.G. - U.K., London).

REFERENCES

1. H.F.Tiersten, Linear Piezoelectric Plate Vibrations, Plenum Press, New York, 1969.

2. M.C.Dökmeci, "Recent Advances : Vibrations of Piezoelectric Crystals", Int.J.Engng. Sci., Vol. 18, March 1980, pp. 431-448.

3. M.C.Dökmeci, "Dynamic Applications of Piezoelectric Crystals; Part I: Fundamentals, Part II: Theoretical Studies and Part III: Experimental Studies", The Shock and Vibration Digest, Vol. 15, March-May 1983, pp. 9-20, 15-26 and 11-22.

4. A.Gerber and A.Ballato, Eds., Precision Frequency Control, Vol. 1: Acoustic Resonators and Filters, and Vol. 2: Oscillators and Standards, Academic Press, New York, 1985.

5. A.L.Nalamwar and M.Epstein, "Surface Acoustic Waves in Strained Media", J.Appl. Phys., Vol. 47, January 1976, pp. 43-48.

6. A.L.Nalamwar and M.Epstein, "Effects of Applied Strain in ZnO Thin-Film SAW Devices", IEEE Trans. Sonics Ultrason., Vol. SU-23, May 1976, pp. 144-147.

7. M.C.Dökmeci, "Quasi-variational Principles for Strained Piezoelectric Continua", Proc. Int. Symp. Strength Mater. Struct. Components at Sonic Ultrason. Freq., Acad. Sci. USSR and Ukr. SSR, Kiev, 1984, pp. 1-8.

8. M.A.Biot, Mechanics of Incremental Deformations, J.Wiley, New York, 1965.

9. V.V.Bolotin, Nonconservative Problems of the Theory of Elastic Stability, Pergamon Press,

London, 1963.

10. R.N.Thurston, "Effective Elastic Coefficients for Wave Propagation in Crystals under Stress", J.Acoust. Soc. Am., Vol. 37, February 1965, pp. 348-356.

11. G.Herrmann and A.E.Armenakas, "Vibrations and Stability of Plates Under Initial Stress", ASCE EM, Vol. 86, 1960, pp. 65-94.

12. P.C.Y.Lee, Y.S.Wang and X.Markenscoff, "High-frequency Vibrations of Crystal Plates under Initial Stresses", J.Acoust. Soc. Am., Vol. 57, January 1975, pp. 95-105.

13. P.C.Y.Lee and K-M.Wu, "In-plane Acceleration and Forces on Frequency Changes in Doubly Rotated Quartz Plates", J.Acoust. Soc. Am., Vol. 75, April 1984, pp. 1105-1117.

14. M.C.Dökmeci, Vibration and Stability Analysis of Laminated Composites, AEK Tek: 2940-3-77, November 1979.

15. D.B.Bogy and D.K-K. Miu, "Transient Voltage across Axisymmetrically Loaded Piezoelectric Disks with Electroded Faces", J.Acoust. Soc. Am., Vol. 71, Feb. 1982, pp. 487-497.

16. D.B.Bogy and S.E.Bechtel, "Electromechanical Analysis of Nonaxisymmetrically Loaded Piezoelectric Disks with Electroded Faces", J.Acoust. Soc. Am., Vol. 72, November 1982, pp. 1498-1507.

17. V.L.Karlash, "Energy Dissipation during Vibrations of thin Circular Piezoceramic Plates", Sov. Appl. Mechs., Vol. 20, 1984, pp. 460-464.

18. G.G.Zaretskii-Feoktistov, "Vibrations of thin Piezoceramic Disk", Sov. Appl. Mechs., Vol. 20, 1985, pp. 832-836.

19. J.C.Baboux, F.Lakestani, and M.Perdrix, "Theoretical and Experimental study of the Contribution of Radial Modes to the Pulsed Ultrasonic Field Radiated by a thick Piezoelectric Disk", J.Acoust. Soc. Am., Vol. 75, June 1984, pp. 1722-1731.

20. Q.Pan-fu, "The Frequency Spectrum of Coupled-Vibrations of Finite Circular Piezoelectric Ceramic Disks", Acta Acustica, Vol. 9, 1984, pp. 47-54.

21. R.D.Mindlin, Theory of Beams and Plates, Lecture notes at Columbia University, New York, 1968.

22. W.Magnus and F.Oberhettinger, Special Functions of Mathematical Physics, Chelsea Publishing Company, New York, 1949.

23. R.D.Mindlin and H.D.McNiven, "Axially Symmetric Waves in Elastic Rods", J.Appl. Mechs., Vol. 27, March 1960, pp. 145-151.

24. R.J.Knops and L.E.Payne, Uniqueness Theorems in Linear Elasticity, Springer-Verlag, Berlin, 1972.

25. M.C.Dökmeci, "An Isothermal Theory of Anisotropic Rods", J. Engng. Math., Vol. 9, 1975, pp. 311-322.

26. M.C.Dökmeci, "A Piezoelectric Model for Vibrations of Bones", Proceed. 19th Midwestern Mechanics Conference, September 9-11, 1985.

27. M.C.Dökmeci, "Dynamics of a Piezoelectric Circular Disc under Initial Stresses", to appear.

AN ANALYSIS OF DOUBLY-ROTATED CONTOURED QUARTZ CRYSTAL RESONATORS

D.S. Stevens and H.F. Tiersten
Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, New York 12180-3590

Abstract

The analysis presented is an extension of the recent treatment of the doubly-rotated contoured SC-cut quartz resonator, in which a number of relatively small transformed elastic constants were neglected. In this work no transformed elastic constants are neglected and only the fundamental assumptions of small piezoelectric coupling and small wavenumbers along the plate are made. The resulting equations are applicable to any potentially practical contoured quartz resonator independent of orientation. In particular, calculations performed for the SC-cut reveal that the earlier treatment, which neglected certain relatively small terms, was quite accurate in the prediction of the resonant frequencies. However, since the differential equation for each harmonic family depends on the order of the harmonic and in this more general treatment contains mixed derivatives in the plane of the plate, a different transformation is required for each harmonic family to obtain the coordinate system in which the mixed derivatives do not appear and, hence, the equation is separable. An interesting consequence of this transformation is that since the nodal planes of the anharmonics of each harmonic family of the contoured SC-cut quartz resonator are oriented along the transformed coordinate system for that harmonic family, they are oriented differently for each harmonic family. Calculations performed for the SC-cut and for a number of other doubly-rotated orientations are shown to be in excellent agreement with experiment.

1. Introduction

In a recent analysis of doubly-rotated contoured SC-cut quartz crystal resonators a number of relatively small transformed elastic constants were neglected[1]. In the treatment presented here no transformed elastic constants are neglected and only the fundamental assumptions of small piezoelectric coupling and small wavenumbers along the plate are made. However, these fundamental assumptions significantly reduce the complexity of the equations. The resulting description is applicable to any orientation for which the simple thickness-frequencies are sufficiently well separated.

As in the earlier work[1] the mechanical displacement is decomposed along the eigenvector triad of the pure thickness solution[2] for the orientation under consideration. Since the piezoelectric coupling is small in quartz and we are interested in obtaining the dispersion relations in the vicinity of the thickness-frequencies of interest for small wavenumbers along the plate, only the thickness-dependence of all electrical variables is included in the treatment. The resulting system of transformed differential equations and boundary conditions on the major surfaces of the plate are employed in the determination of the asymptotic dispersion relation in the vicinity of each thickness-frequency of interest. However, the asymptotic analysis cannot proceed from the pure thickness solutions as in certain earlier work[3] because a degeneracy occurs, which is circumvented by proceeding from an appropriate simple solution: containing wavenumbers along the plate. This simple solution is a restricted generalization to three dimensions of the solution for thickness-twist waves in rotated Y-cut quartz plates obtained by Mindlin[4]. The asymptotic dispersion relation obtained contains a term with mixed wavenumbers along the plate. Consequently, the single scalar differential equation describing the mode shape along the plate that we construct from the dispersion equation contains mixed derivatives in the plane of the plate. As in earlier work the differential equation for the flat plate[5] is extended to the contoured plate with slowly varying thickness[6,7]. Since the influence of the contouring on the trapping is much greater than the influence of the electrode[6], the edge of the electrode is ignored in the determination of the eigensolutions. Since the mode is sharply confined to the vicinity of the center of the contoured plate, the edge of the plate is ignored in the analysis.

An inhomogeneous differential equation is obtained for each harmonic thickness mode of interest for the case of a driving voltage applied across surface electrodes. The differential equation obtained depends on the order of the odd harmonic and contains mixed derivatives in the plane of the plate. Consequently, for each odd harmonic a transformation in the plane of the plate is employed to eliminate the mixed derivative and determine the planar Cartesian coordinate system in which the equation is separable. An important consequence of this transformation is that the nodal planes of the anharmonics of each harmonic family of a doubly-rotated contoured resonator are oriented differently in order to correspond to the different planar transformations. The resulting system of equations is applied in the determination of the harmonic and anharmonic resonant frequencies and associated motional capacitances of doubly-rotated plano-convex contoured quartz crystal resonators. Numerical results are presented for a number of different orientations and shown to be in excellent agreement with experiment[8,9]

2. Transformation of Equations

As in the earlier treatment[1] the stress equations of motion and charge equation of electrostatics may be written in the form

$$\hat{T}_{mn,m} = \rho \ddot{\hat{u}}_n, \quad D_{m,m} = 0, \tag{2.1}$$

and the linear piezoelectric constitutive equations are given by

$$\hat{T}_{mn} = \hat{c}_{mnrs}\hat{u}_{r,s} + \hat{e}_{rmn}\varphi_{,r}, \tag{2.2}$$

$$D_m = \hat{e}_{mrs}\hat{u}_{r,s} - \epsilon_{mn}\varphi_{,n}. \tag{2.3}$$

The notation is conventional[10] and carets have been introduced over the mechanical variables and the elastic and piezoelectric constants because we will be decomposing the mechanical displacement vector in a very special Cartesian coordinate system that facilitates the analysis. Moreover, in the original coordinate system we consistently use the tensor indices m, n, r, s. The substitution of (2.2) and (2.3) into (2.1) yields

$$\hat{c}_{mnrs}\hat{u}_{r,sm} + \hat{e}_{rmn}\varphi_{,rm} = \rho\ddot{\hat{u}}_n, \tag{2.4}$$

$$\hat{e}_{mrs}\hat{u}_{r,sm} - \epsilon_{mn}\varphi_{,mn} = 0, \tag{2.5}$$

CH2186-0/85/0000-0436$1.00©1985IEEE

which are the differential equations of linear piezoelectricity.

The coordinate axes are oriented with x_2 normal to the major surfaces of the plate, which are at $x_2 = \pm h$, and x_1 directed along the axis of second rotation of the doubly-rotated quartz plate. Since in the modes of interest the spatial rate of variation of the dependent variables in the plane of the plate is much less than in the thickness direction, and the piezoelectric coupling is small, we can ignore the x_w dependence of φ and D_v and of any variable in the expression for D_2, where we have introduced the convention that the subscripts v,w take the values 1 and 3, but skip 2. Therefore, in place of Eqs.(2.3) and (2.5), respectively, we have

$$D_2 = \hat{e}_{2r2}\hat{u}_{r,2} - \epsilon_{22}\varphi_{,2}, \tag{2.6}$$

$$\hat{e}_{2r2}\hat{u}_{r,22} - \epsilon_{22}\varphi_{,22} = 0. \tag{2.7}$$

Moreover, it is advantageous for us to write (2.2) and (2.4) in the forms

$$\hat{T}_{mn} = \hat{c}_{2mnr2}\hat{u}_{r,2} + \hat{e}_{2mn}\varphi_{,2} + \hat{c}_{2vrnm}\hat{u}_{r,v}, \tag{2.8}$$

$$\hat{c}_{2nr2}\hat{u}_{r,22} + \hat{e}_{22n}\varphi_{,22} + (\hat{c}_{2rnv} + \hat{c}_{2nrv})\hat{u}_{r,2v} + \hat{c}_{2vnrw}\hat{u}_{r,wv} = \rho\ddot{\hat{u}}_n, \tag{2.9}$$

since we are interested in modes in the vicinity of the frequencies of thickness vibration. Substituting from (2.7) into (2.9), we obtain

$$\bar{\hat{c}}_{2nr2}\hat{u}_{r,22} + (\hat{c}_{2rnv} + \hat{c}_{2nrv})u_{r,2v} + \hat{c}_{2vnrw}\hat{u}_{r,wv} = \rho\ddot{\hat{u}}_n, \tag{2.10}$$

where

$$\bar{\hat{c}}_{2nr2} = \hat{c}_{2nr2} + \frac{\hat{e}_{22n}\hat{e}_{22r}}{\epsilon_{22}}, \tag{2.11}$$

are the usual piezoelectrically stiffened elastic constants.

For thickness vibrations only x_2 dependence is considered, and (2.10) takes the form

$$\bar{\hat{c}}_{2nr2}\hat{u}_{r,22} = \rho\ddot{\hat{u}}_n. \tag{2.12}$$

In the case of the unelectroded plate, the antisymmetric thickness vibration solution can be written in the form

$$u_r = A_r \sin \eta x_2 e^{i\omega t}, \tag{2.13}$$

which satisfies (2.12), provided

$$(\bar{\hat{c}}_{2nr2} - \bar{c}\delta_{nr})A_r = 0, \tag{2.14}$$

where

$$\bar{c} = \rho\omega^2/\eta^2. \tag{2.15}$$

For a nontrivial solution, the determinant of the coefficients of A_r in (2.14) must vanish, i.e.,

$$|\bar{\hat{c}}_{2nr2} - \bar{c}\delta_{nr}| = 0, \tag{2.16}$$

which yields a cubic equation in $\bar{c}$. Since the $\bar{\hat{c}}_{2nr2}$ constitute a symmetric, positive-definite matrix[11], Eq.(2.16) yields 3 real, positive roots[12] $\bar{c}^{(i)}$ $(i = 1,2,3)$ which yield mutually orthogonal eigenvectors $A_r^{(i)}$ when substituted into (2.14), and we may write

$$A_r^{(i)}A_r^{(j)} = N_{(i)}^2\delta_{ij}, \tag{2.17}$$

where the $N_{(i)}$ are the normalization factors. If we normalize the $A_r^{(i)}$ thus

$$q_r^{(i)} = A_r^{(i)}/N_{(i)}, \tag{2.18}$$

and write the $q_r^{(i)}$ as a 3×3 orthogonal transformation matrix Q_{ir}, from (2.17) we have the orthogonality relations

$$Q_{ir}Q_{jr} = \delta_{ij}, \tag{2.19}$$

and, of course, the other orthogonality relations

$$Q_{ir}Q_{is} = \delta_{rs}, \tag{2.20}$$

hold.

We may now transform the components of the vector $\hat{u}_r$ in the original coordinate system to the components u_i in the thickness solution eigenvector system and vice-versa, thus

$$u_i = Q_{ir}\hat{u}_r, \quad \hat{u}_r = Q_{ir}u_i. \tag{2.21}$$

A schematic diagram showing the original coordinate system along with the eigenvector coordinate system is shown in Fig.1. For a pure thickness solution for an unelectroded plate, one of the thickness eigendisplacements, say u_1, exists, and the other two vanish identically. Moreover, for a mode of vibration in the vicinity of the u_1 thickness solution, although u_2 and u_3 exist, they are one or more orders of magnitude smaller than u_1. This is the reason there is a significant advantage in decomposing the mechanical displacement along the eigenvector triad of the pure thickness solution when describing this type of mode.

Transforming (2.10) with Q_{jn}, substituting from $(2.21)_2$ for $\hat{u}_r$ and $\hat{u}_n$ and employing (2.14) with (2.18) and (2.19), we obtain

$$\bar{c}^{(j)}u_{j,22} + (c_{aij2} + c_{aji2})u_{i,2a} + c_{ajib}u_{i,ab} = \rho\ddot{u}_j, \tag{2.22}$$

where

$$c_{aij2} = Q_{jn}Q_{ir}\hat{c}_{2arn2},$$

$$c_{ajib} = Q_{jn}Q_{ir}\hat{c}_{2arnb}, \tag{2.23}$$

and we have introduced the convention that the subscripts a,b take the values 1 and 3, but skip 2.

In the case of the thickness vibrations of the electroded plate with shorted electrodes, the three solutions of the differential equations (2.12) are coupled[13-15] in the transcendental frequency equation. However, since the piezoelectric coupling is small in quartz and the three $\bar{c}^{(n)}$ are widely separated, the transcendental frequency equation approximately uncouples, and we effectively have three independent transcendental frequency equations[11], one for each thickness eigendisplacement u_j. Consequently, (2.22) is a useful form for the treatment of modes of vibration in the vicinity of the pure thickness solution u_j for the electroded as well as the unelectroded plate.

For the electroded plate with shorted electrodes the electrical boundary condition is

$$\varphi = 0 \quad \text{at} \quad x_2 = \pm h \,. \tag{2.24}$$

Substituting from $(2.21)_2$ into (2.7), integrating and satisfying (2.24), we obtain

$$\varphi = (e_{22j}/\epsilon_{22})u_j + Cx_2 \,, \tag{2.25}$$

where

$$e_{22j} = Q_{jr}\hat{e}_{22r} \,, \quad C = -(e_{22j}/\epsilon_{22})u_j(h)/h \,. \tag{2.26}$$

Since in the case of the unelectroded plate the source $\hat{e}_{22r}\hat{u}_{r,22}$ in (2.7) is antisymmetric, the electric field vanishes at $|x_2| = \infty$ and, hence, in this one-dimensional case the electric field vanishes everywhere outside the plate and we have the electrical condition

$$D_2 = 0 \quad \text{at} \quad x_2 = \pm h \,. \tag{2.27}$$

Hence, from (2.27) with (2.6) and $(2.26)_1$, we obtain

$$\varphi = (\epsilon_{22j}/\epsilon_{22})u_j \,, \tag{2.28}$$

which is the same as Eq. (2.25), but with $C = 0$.

The traction boundary conditions on the major surfaces of the plate are obtained from the constitutive equations in (2.8) with $m = 2$. Transforming (2.8) for $m = 2$ with Q_{jn}, substituting from Eqs. $(2.21)_2$, (2.25), $(2.26)_1$, and (2.11), and employing Eqs. (2.14) with (2.18), (2.19), and $(2.23)_1$, we obtain

$$T_{2j} = \bar{c}^{(j)}u_{j,2} + e_{22j}C + c_{aij2}u_{i,a} \,, \tag{2.29}$$

which holds for the electroded plate and for the unelectroded plate, provided $C = 0$. As in the earlier work[1], the traction continuity conditions along the surfaces separating the electroded from the unelectroded regions and at the free-edges of the plate are not needed in this work.

3. Scalar Differential Equations for Transversely Varying Thickness Modes

Although in this work we are primarily interested in the dispersion equation for the electroded region of the plate, we obtain the dispersion equation for the unelectroded region because the treatment is somewhat simpler than that for the electroded region and it has been shown[3] that the dispersion equation for the electroded plate may readily be inferred from that for the unelectroded plate provided certain details depending only on the thickness (x_2) coordinate, i.e., the pure thickness solution are incorporated.

The differential equations expressed in terms of the thickness eigendisplacements are given in (2.22). The boundary conditions on the major surfaces of the unelectroded plate are

$$T_{2j} = 0 \,, \quad \text{at } x_2 = \pm h \,, \tag{3.1}$$

where the constitutive equations for T_{2j} expressed in the same way are given in (2.29), with $C = 0$. Since we are interested in the solution for plate waves in the vicinity of the thickness frequencies for the thickness eigendisplacement u_1 with wave, or decay, numbers in the plane of the plate that are much smaller than the thickness wavenumbers, a large number of terms in the differential equations (2.22) and the constitutive equations (2.29) are negligible for any values of the associated transformed elastic constants[11]. On account of this Eqs. (2.22) and (2.29) reduce to the considerably simpler forms[11]

$$\begin{aligned}
&c_{11}u_{1,11} + 2c_{16}u_{1,12} + 2c_{56}u_{1,23} + 2c_{51}u_{1,13} + c_{58}u_{1,33} \\
&\quad + (c_{12}+c_{66})u_{2,12} + (c_{52}+c_{76})u_{2,23} + (c_{17}+c_{86})u_{3,12} \\
&\quad + (c_{36}+c_{57})u_{3,23} + \bar{c}^{(1)}u_{1,22} = \rho\ddot{u}_1 \,, \\
&(c_{12}+c_{66})u_{1,12} + (c_{52}+c_{76})u_{1,23} + 2c_{29}u_{2,12} \\
&\quad + 2c_{24}u_{2,32} + \bar{c}^{(2)}u_{2,22} = \rho\ddot{u}_2 \,, \\
&(c_{17}+c_{86})u_{1,12} + (c_{36}+c_{57})u_{1,23} + 2c_{45}u_{3,12} \\
&\quad + 2c_{43}u_{3,32} + \bar{c}^{(3)}u_{3,22} = \rho\ddot{u}_3 \,,
\end{aligned} \tag{3.2}$$

$$\begin{aligned}
T_{21} &= \bar{c}^{(1)}u_{1,2} + c_{66}u_{2,1} + c_{16}u_{1,1} + c_{56}u_{1,3} \\
&\quad + c_{76}u_{2,3} + c_{86}u_{3,1} + c_{36}u_{3,3} \,, \\
T_{22} &= \bar{c}^{(2)}u_{2,2} + c_{12}u_{1,1} + c_{52}u_{1,3} \,, \\
T_{23} &= \bar{c}^{(3)}u_{3,2} + c_{17}u_{1,1} + c_{57}u_{1,3} \,,
\end{aligned} \tag{3.3}$$

where we have introduced the compressed index notation defined in Table I, and employed the relations

$$c_{ajib} = c_{bija} \,, \tag{3.4}$$

which are a consequence of $(2.23)_2$ and the standard symmetries of the elastic constants $\hat{c}_{2vnrw}$.

We now proceed to obtain the asymptotic solutions to (3.2) and (3.1) with (3.3) in the vicinity of odd u_1-thickness frequencies for small wavenumbers along the plate. Since solutions decaying in both directions along the plate are somewhat simpler to treat than those exhibiting trigonometric behavior in one or both directions, and the dispersion equations for trigonometric behavior may readily be obtained[1,3] from those for decaying behavior, we consider the decaying solutions. Accordingly, as a solution of (3.2) consider

$$\begin{aligned}
u_1 &= (A\sin\eta x_2 + B\cos\eta x_2)e^{-\xi x_1}e^{-\nu x_3}e^{i\omega t} \,, \\
u_2 &= (C\sin\eta x_2 + D\cos\eta x_2)e^{-\xi x_1}e^{-\nu x_3}e^{i\omega t} \,, \\
u_3 &= (E\sin\eta x_2 + F\cos\eta x_2)e^{-\xi x_1}e^{-\nu x_3}e^{i\omega t} \,,
\end{aligned} \tag{3.5}$$

which satisfies (3.2) provided a system of six homogeneous linear algebraic equations in A, B, C, D, E and F are satisfied, which are too lengthy to present here[11].

In earlier work[3,5] on a simpler problem of this nature the asymptotic dispersion relation was obtained starting with the pure thickness solution of interest. However, in the situation under consideration here, the thickness coupling appearing in (3.5) is caused by the terms containing c_{16} and c_{56} in $(3.2)_1$, which vanish in the case of the pure thickness solution. Since under these circumstances ($\xi = \nu = 0$) the thickness terms in each of (3.5) are uncoupled in $(3.2)_1$, the asymptotic dispersion relation cannot be obtained from the pure thickness solution of interest because there is no way of determining the relation between A and B when $\xi = \nu = 0$. However, if only those terms containing c_{16} and c_{56} that cause the aforementioned coupling are

retained in addition to the purely thickness dependent terms, the full solution for u_1 ($u_2 = u_3 = 0$) splits the degeneracy and yields simple analytical expressions and fixes the relation between A and B. Consequently, the solution to this reduced problem[11] then serves as a convenient zeroth iterate for obtaining the asymptotic dispersion relation by means of a straightforward iterative procedure analogous to the one used earlier[3]. As noted in the Introduction the solution to this reduced problem is a restricted generalization to three dimensions of the solution for thickness-twist waves in rotated Y-cut quartz plates obtained by Mindlin[4]. Thus, in consideration of the foregoing and following the procedure used in Refs.3 and 5, to lowest order in ξ and ν for large A and B, we obtain[11]

$$\eta_1^{\pm} = \frac{n\pi}{2h} \pm \frac{(c_{16}\xi + c_{56}\nu)i}{\bar{c}^{(1)}}, \quad B_{\pm}^{(1)} = \pm iA_{\pm}^{(1)}, \tag{3.6}$$

$$C_{\pm}^{(1)} = \pm i\,\frac{r_2\xi + r_4\nu}{\eta_1^{\pm}} A_{\pm}^{(1)}, \quad D_{\pm}^{(1)} = \frac{r_2\xi + r_4\nu}{\eta_1^{\pm}} A_{\pm}^{(1)},$$

$$E_{\pm}^{(1)} = \mp i\,\frac{r_3\nu + r_5\xi}{\eta_1^{\pm}} A_{\pm}^{(1)}, \quad F_{\pm}^{(1)} = \frac{r_3\nu + r_5\xi}{\eta_1^{\pm}} A_{\pm}^{(1)}, \tag{3.7}$$

where

$$r_2 = \frac{c_{12} + c_{66}}{\bar{c}^{(1)} - \bar{c}^{(2)}}, \quad r_3 = \frac{c_{36} + c_{57}}{\bar{c}^{(1)} - \bar{c}^{(3)}},$$

$$r_4 = \frac{c_{52} + c_{76}}{\bar{c}^{(1)} - \bar{c}^{(2)}}, \quad r_5 = \frac{c_{17} + c_{86}}{\bar{c}^{(1)} - \bar{c}^{(3)}}. \tag{3.8}$$

Proceeding in a similar way for large C and D and large E and F, respectively, we obtain

$$\eta_2^{\pm} = \kappa_2\eta_1^{\pm}\left[1 \mp i\,\frac{(c_{29}\xi + c_{24}\nu)}{\bar{c}^{(2)}\kappa_2\eta_1^{\pm}}\right], \quad \kappa_2 = \sqrt{\frac{\bar{c}^{(1)}}{\bar{c}^{(2)}}}, \tag{3.9}$$

$$D_{\pm}^{(2)} = \pm iC_{\pm}^{(2)}, \quad A_{\pm}^{(2)} = \pm i\,\frac{r_2\xi + r_4\nu}{\eta_2} C_{\pm}^{(2)},$$

$$B_{\pm}^{(2)} = -\frac{r_2\xi + r_4\nu}{\eta_2} C_{\pm}^{(2)}, \quad E_{\pm}^{(2)} = F_{\pm}^{(2)} = 0, \tag{3.10}$$

$$\eta_3^{\pm} = \kappa_3\eta_1^{\pm}\left[1 \mp i\,\frac{(c_{45}\xi + c_{43}\nu)}{\bar{c}^{(3)}\kappa_3\eta_1^{\pm}}\right], \quad \kappa_3 = \sqrt{\frac{\bar{c}^{(1)}}{\bar{c}^{(3)}}}, \tag{3.11}$$

$$F_{\pm}^{(3)} = \pm iE_{\pm}^{(3)}, \quad A_{\pm}^{(3)} = \pm i\,\frac{r_3\nu + r_5\xi}{\eta_3} E_{\pm}^{(3)},$$

$$B_{\pm}^{(3)} = -\frac{r_3\nu + r_5\xi}{\eta_3} E_{\pm}^{(3)}, \quad C_{\pm}^{(3)} = D_{\pm}^{(3)} = 0, \tag{3.12}$$

in which the amplitudes that have been shown to vanish are actually of order higher than linear in ξ or ν, and hence will be negligible to order ξ^2 and ν^2.

In order to satisfy the boundary conditions (3.1) with (3.3), we take a sum of the six asymptotic solutions of the differential equations in the form

$$u_1 = \Big[A_+^{(1)}(\sin\eta_1^+X_2 + i\cos\eta_1^+X_2) + A_-^{(1)}(\sin\eta_1^-X_2 - i\cos\eta_1^-X_2) + i\,\frac{r_2\xi + r_4\nu}{\eta_2} C_+^{(2)}(\sin\eta_2X_2 + i\cos\eta_2X_2) - i\,\frac{r_2\xi + r_4\nu}{\eta_2} C_-^{(2)}(\sin\eta_2X_2 - i\cos\eta_2X_2) + i\,\frac{r_3\nu + r_5\xi}{\eta_3} E_+^{(3)}(\sin\eta_3X_2 + i\cos\eta_3X_2) - i\,\frac{r_3\nu + r_5\xi}{\eta_3} E_-^{(3)}(\sin\eta_3X_2 - i\cos\eta_3X_2)\Big] e^{-\xi X_1} e^{-\nu X_3},$$

$$u_2 = \Big[-i\,\frac{r_2\xi + r_4\nu}{\eta_1^+} A_+^{(1)}(\sin\eta_1^+X_2 + i\cos\eta_1^+X_2) + i\,\frac{r_2\xi + r_4\nu}{\eta_1^-} A_-^{(1)}(\sin\eta_1^-X_2 - i\cos\eta_1^-X_2) + C_+^{(2)}(\sin\eta_2X_2 + i\cos\eta_2X_2) + C_-^{(2)}(\sin\eta_2X_2 - i\cos\eta_2X_2)\Big] e^{-\xi X_1} e^{-\nu X_3},$$

$$u_3 = \Big[-i\,\frac{r_3\nu + r_5\xi}{\eta_1^+} A_+^{(1)}(\sin\eta_1^+X_2 + i\cos\eta_1^+X_2) + i\,\frac{r_3\nu + r_5\xi}{\eta_1^-} A_-^{(1)}(\sin\eta_1^-X_2 - i\cos\eta_1^-X_2) + E_+^{(3)}(\sin\eta_3X_2 + i\cos\eta_3X_2) + E_-^{(3)}(\sin\eta_3X_2 - i\cos\eta_3X_2)\Big] e^{-\xi X_1} e^{-\nu X_3}, \tag{3.13}$$

which satisfies Eqs.(3.1) with (3.3) provided a system of six homogeneous linear algebraic equations in $A_+^{(1)}$, $A_-^{(1)}$, $C_+^{(2)}$, $C_-^{(2)}$, $E_+^{(3)}$ and $E_-^{(3)}$ are satisfied and which are too lengthy to present here[11]. Since we are interested in modes in the vicinity of the pure u_1-thickness solution, for which $A_+^{(1)}$ and $A_-^{(1)}$ are large and η_1 is very near the expression given in $(3.6)_1$, we take

$$\eta_1^{\pm}h = \frac{n\pi}{2} \pm i\,\frac{(c_{16}\xi + c_{56}\nu)}{\bar{c}^{(1)}} h + \hat{\alpha}_n, \quad n = 1,3,5,\ldots, \tag{3.14}$$

where $\hat{\alpha}_n$ is small and is the same[11] for both $\eta_1^{\pm}$, and substitute from (3.14) into the vanishing determinant obtained[11] from the aforementioned six equations obtained from the boundary conditions (3.1) with (3.3), and expand the trigonometric functions and the determinant to obtain[11]

$$\hat{\alpha}_n = -\frac{4h^2\cot\kappa_2\frac{n\pi}{2}}{\bar{c}^{(1)}\bar{c}^{(2)}\kappa_2 n^2\pi^2}\,[(r_2\bar{c}^{(1)} - c_{66})\xi + (r_4\bar{c}^{(1)} - c_{76})\nu][(r_2\bar{c}^{(2)} + c_{12})\xi + (r_4\bar{c}^{(2)} + c_{52})\nu] - \frac{4h^2\cot\kappa_3\frac{n\pi}{2}}{\bar{c}^{(1)}\bar{c}^{(3)}\kappa_3 n^2\pi^2}\,[(r_5\bar{c}^{(1)} - c_{86})\xi + (r_3\bar{c}^{(1)} - c_{36})\nu][(r_5\bar{c}^{(3)} + c_{17})\xi + (r_3\bar{c}^{(3)} + c_{57})\nu], \tag{3.15}$$

to second order in ξ and ν. Moreover, from the six homogeneous linear algebraic equations, we obtain[11]

$$A_-^{(1)} = A_+^{(1)} ,$$

$$C_-^{(2)} = -C_+^{(2)} = i(-1)^{\frac{n+1}{2}}\left[\frac{(r_2\bar{c}^{(2)}+c_{12})\xi}{\bar{c}^{(2)}\eta_2 \sin \eta_2 h} + \frac{(r_4\bar{c}^{(2)}+c_{52})\nu}{\bar{c}^{(2)}\eta_2 \sin \eta_2 h}\right] A_+^{(1)}$$

$$E_-^{(3)} = -E_+^{(3)} = i(-1)^{\frac{n+1}{2}}\left[\frac{(r_5\bar{c}^{(3)}+c_{17})\xi}{\bar{c}^{(3)}\eta_3 \sin \eta_3 h} + \frac{(r_3\bar{c}^{(3)}+c_{57})\nu}{\bar{c}^{(3)}\eta_3 \sin \eta_3 h}\right] A_+^{(1)} \qquad (3.16)$$

Now, substituting from $(3.6)_2$, (3.7), (3.9) - (3.12), (3.14) and (3.15) into either of the two algebraic equations obtained from $(3.2)_1$ for either η_1^+ or η_1^-, for nonzero $A_+^{(1)}$ and $A_-^{(1)}$, we obtain[11]

$$M_n\xi^2 + Q_n\xi\nu + P_n\nu^2 - \frac{n^2\pi^2\bar{c}^{(1)}}{4h^2} + \rho\omega^2 = 0 , \qquad (3.17)$$

where

$$M_n = c_{11} - \frac{c_{16}^2}{\bar{c}^{(1)}} + r_2(c_{12}+c_{66}) + r_5(c_{17}+c_{86}) + \frac{4(r_2\bar{c}^{(1)}-c_{66})(r_2\bar{c}^{(2)}+c_{12})}{\bar{c}^{(2)}\kappa_2 n\pi} \cot \kappa_2 \frac{n\pi}{2} + \frac{4(r_5\bar{c}^{(1)}-c_{86})(r_5\bar{c}^{(3)}+c_{17})}{\bar{c}^{(3)}\kappa_3 n\pi} \cot \kappa_3 \frac{n\pi}{2} ,$$

$$Q_n = 2c_{51} - \frac{2c_{16}c_{56}}{\bar{c}^{(1)}} + r_2(c_{52}+c_{76}) + r_4(c_{12}+ \;;) + r_3(c_{17}+c_{86}) + r_5(c_{36}+c_{57}) + 4\left[\frac{(r_2\bar{c}^{(1)}-c_{66})(r_4\bar{c}^{(2)}+c_{52})}{\bar{c}^{(2)}\kappa_2 n\pi} + \frac{(r_2\bar{c}^{(2)}+c_{12})(r_4\bar{c}^{(1)}-c_{76})}{\bar{c}^{(2)}\kappa_2 n\pi}\right] \cot \kappa_2 \frac{n\pi}{2} + 4\left[\frac{(r_5\bar{c}^{(1)}-c_{86})(r_3\bar{c}^{(3)}+c_{57})}{\bar{c}^{(3)}\kappa_3 n\pi} + \frac{(r_5\bar{c}^{(3)}+c_{17})(r_3\bar{c}^{(1)}-c_{36})}{\bar{c}^{(3)}\kappa_3 n\pi}\right] \cot \kappa_3 \frac{n\pi}{2} ,$$

$$P_n = c_{58} - \frac{c_{56}^2}{\bar{c}^{(1)}} + r_4(c_{52}+c_{76}) + r_3(c_{36}+c_{57}) + \frac{4(r_4\bar{c}^{(1)}-c_{76})(r_4\bar{c}^{(2)}+c_{52})}{\bar{c}^{(2)}\kappa_2 n\pi} \cot \kappa_2 \frac{n\pi}{2} + \frac{4(r_3\bar{c}^{(1)}-c_{36})(r_3\bar{c}^{(3)}+c_{57})}{\bar{c}^{(3)}\kappa_3 n\pi} \cot \kappa_3 \frac{n\pi}{2} . \qquad (3.18)$$

Furthermore, it can be shown[11] that the system of Eqs. (3.1) - (3.3) is satisfied to second order, but not third order in ξ and ν. The dominant algebraic equations for the η_2 and η_3 solutions, respectively, turn out[11] to be of third order in ξ and ν and, hence, negligible.

Equation (3.17) is the asymptotic dispersion relation for solutions decaying in x_1 and x_3 for the unelectroded plate. Hence, past experience[3,5] indicate that for solutions trigonometric in x_1 and x_3 the asymptotic dispersion equations for the electroded plate are given by

$$M_n\bar{\xi}^2 + Q_n\bar{\xi}\bar{\nu} + P_n\bar{\nu}^2 + \frac{n^2\pi^2}{4h^2}\hat{c}^{(1)} - \rho\omega^2 = 0 , \qquad (3.19)$$

where

$$\hat{c}^{(1)} = \bar{c}^{(1)}\left(1 - \frac{8k_1^2}{n^2\pi^2} - 2\hat{R}\right) ,$$

$$k_1^2 = \frac{e_{26}^2}{\bar{c}^{(1)}\epsilon_{22}} , \quad \hat{R} = \frac{2\rho' h'}{\rho h} , \qquad (3.20)$$

and ρ' and $2h'$ denote the mass density and thickness of the electrodes, respectively.

The manner of derivation of the dispersion relation (3.17) reveals that for (3.19) for each n we must have the homogeneous equation

$$\left(M_n\bar{\xi}^2 + Q_n\bar{\xi}\bar{\nu} + P_n\bar{\nu}^2 + \frac{n^2\pi^2\hat{c}^{(1)}}{4h^2} - \rho\omega^2\right) u_1^n = 0 , \qquad (3.21)$$

where u_1^n is the asymptotic solution function for the u_1-displacement for the nth odd harmonic, which is very accurately given by

$$u_1^n = 2A_+^{(1)} \sin \frac{n\pi x_2}{2h} e^{-i\bar{\xi}x_1} e^{-i\bar{\nu}x_3} e^{i\omega t} , \qquad (3.22)$$

for small ξ and ν. Consequently, it is clear that the homogeneous differential equation for u_1^n governing the mode shapes along the surface of the electroded plate may be written in the form

$$M_n\frac{\partial^2 u_1^n}{\partial x_1^2} + Q_n\frac{\partial^2 u_1^n}{\partial x_1 \partial x_3} + P_n\frac{\partial^2 u_1^n}{\partial x_3^2} - \frac{n^2\pi^2\hat{c}^{(1)}}{4h^2}u_1^n - \rho\ddot{u}_1^n = 0 , \qquad (3.23)$$

since the substitution of Eq. (3.22) into (3.23) yields (3.21). Furthermore, Eq. (3.23) holds for the unelectroded plate when $\hat{c}^{(1)}$ is replaced by $\bar{c}^{(1)}$.

When a driving voltage is applied across the surface electrodes it can be shown[3,5,11] that a transformation of the inhomogeneous terms from the boundary conditions into the differential equations causes the equation obtained from $(3.2)_1$ to be inhomogeneous. Now, the homogeneous form of the inhomogeneous equation obtained from $(3.2)_1$ is the equation from which the dispersion relation (3.19) is ultimately obtained and all other equations are homogeneous[11] as in the eigensolutions leading to (3.19). Since the solution of the driven problem will be written as an infinite sum of the eigensolutions of the associated homogeneous problem for the electroded plate with shorted

electrodes, the inhomogeneous differential equation can be written in the form[11,16]

$$\sum_{n=1,3,5}^{\infty}\left[M_n\frac{\partial^2 u_1^n}{\partial x_1^2}+Q_n\frac{\partial^2 u_1^n}{\partial x_1\partial x_3}+P_n\frac{\partial^2 u_1^n}{\partial x_3^2}-\frac{n^2\pi^2\hat{c}^{(1)}}{4h^2}u_1^n-\rho\ddot{u}_1^n\right]=\rho\omega^2\frac{e_{26}Vx_2}{c^{(1)}2h}e^{i\omega t},\tag{3.24}$$

where V is the driving voltage and

$$u_1'=\sum_{n=1,3,5}^{\infty}u_1^n,\quad c^{(1)}=\bar{c}^{(1)}(1-k_1^2).\tag{3.25}$$

Utilizing the orthogonality of $\sin n\pi x_2/2h$ in the interval -h to h, from (3.24) we obtain

$$M_n\frac{\partial^2\tilde{u}_1^n}{\partial x_1^2}+Q_n\frac{\partial^2\tilde{u}_1^n}{\partial x_1\partial x_3}+P_n\frac{\partial^2\tilde{u}_1^n}{\partial x_3^2}-\frac{n^2\pi^2\hat{c}^{(1)}}{4h^2}\tilde{u}_1^n-\rho\ddot{\tilde{u}}_1^n=\rho\omega^2(-1)^{\frac{n-1}{2}}\frac{e_{26}}{c^{(1)}}\frac{4Ve^{i\omega t}}{n^2\pi^2},\tag{3.26}$$

where

$$u_1^n=\tilde{u}_1^n(x_1,x_3,t)\sin\frac{n\pi x_2}{2h}.\tag{3.27}$$

Equation (3.26) is the inhomogeneous differential equation in x_1, x_3, and t, for an electroded driving region with driving voltage V for the nth odd harmonic family of modes, i.e., the harmonic and associated anharmonics.

The mixed derivative term in (3.26) may be removed by the usual[17] planar transformation of the independent coordinates x_1 and x_3, i.e.,

$$x_j=R_{ij}x_i',\quad x_i'=R_{ij}x_j,\tag{3.28}$$

where

$$R_{ij}=\begin{bmatrix}\cos\hat{\beta}_n & 0 & \sin\hat{\beta}_n\\ 0 & 1 & 0\\ -\sin\hat{\beta}_n & 0 & \cos\hat{\beta}_n\end{bmatrix},$$

$$\hat{\beta}_n=\frac{1}{2}\tan^{-1}\left(\frac{-Q_n}{M_n-P_n}\right).\tag{3.29}$$

Thus, P_{ij} represents a rotation about the x_2 axis by the angle $\hat{\beta}_n$. Note that $\hat{\beta}_n$ is dependent on the particular harmonic family being considered. The resulting transformed equation (3.26) now takes the very convenient separable form

$$M_n'\frac{\partial^2\tilde{u}_1^n}{\partial x_1'^2}+P_n'\frac{\partial^2\tilde{u}_1^n}{\partial x_3'^2}-\frac{n^2\pi^2\hat{c}^{(1)}}{4h^2}\tilde{u}_1^n-\rho\ddot{\tilde{u}}_1^n=\rho\omega^2(-1)^{\frac{n-1}{2}}\frac{e_{26}4Ve^{i\omega t}}{n^2\pi^2},\tag{3.30}$$

where

$$M_n'=M_n\cos^2\hat{\beta}_n-Q_n\sin\hat{\beta}_n\cos\hat{\beta}_n+P_n\sin^2\hat{\beta}_n,$$

$$P_n'=M_n\sin^2\hat{\beta}_n+Q_n\sin\hat{\beta}_n\cos\hat{\beta}_n+P_n\cos^2\hat{\beta}_n.\tag{3.31}$$

4. Doubly-Rotated Contoured Quartz Resonators

A schematic diagram of a plano-convex resonator is shown in Fig. 2. It has been shown that the slowly varying thickness of a spherically contoured resonator can be represented in the form[18,6]

$$2h=2h_o[1-(x_1^2+x_3^2)/4Rh_o],\tag{4.1}$$

the substitution of which in (3.30) and expansion to first order in $x_1^2+x_3^2$ yields

$$M_n'\frac{\partial^2 u^n}{\partial x_1^2}+P_n'\frac{\partial^2 u^n}{\partial x_3^2}-\frac{n^2\pi^2\hat{c}^{(1)}}{4h_o^2}\left[1+\frac{(x_1^2+x_3^2)}{2Rh_o}\right]u^n+\rho\omega^2u^n=\rho\omega^2(-1)^{\frac{n-1}{2}}\frac{e_{26}}{c^{(1)}}\frac{4V}{n^2\pi^2},\tag{4.2}$$

where

$$\tilde{u}_1^n=u^n(x_1,x_3)e^{i\omega t}.\tag{4.3}$$

Equation (4.2) is the inhomogeneous differential equation for the anharmonic family of the nth odd harmonic of the doubly-rotated quartz plano-convex resonator. It has been shown that the eigensolutions of the associated homogeneous problem, i.e., of (4.2) with V = 0, can be written in the form[18,6]

$$u_{nmp}=e^{-\alpha_n\frac{x_1^2}{2}}H_m(\sqrt{\alpha_n}x_1)e^{-\beta_n\frac{x_3^2}{2}}H_p(\sqrt{\beta_n}x_3),\tag{4.4}$$

where H_p and H_m are Hermite polynomials and

$$\alpha_n^2=\frac{n^2\pi^2\hat{c}^{(1)}}{8Rh_o^3M_n'},\quad\beta_n^2=\frac{n^2\pi^2\hat{c}^{(1)}}{8Rh_o^3P_n'}.\tag{4.5}$$

The eigenfrequencies ω_{nmp} obtained from the solution (4.4) are given by

$$\omega_{nmp}^2=\frac{n^2\pi^2\hat{c}^{(1)}}{4h_o^2\rho}\left[1+\frac{1}{n\pi}\left(\frac{2h_o}{R}\right)^{1/2}\left(\sqrt{\frac{M_n'}{\hat{c}^{(1)}}}(2m+1)+\sqrt{\frac{P_n'}{\hat{c}^{(1)}}}(2p+1)\right)\right],\tag{4.6}$$

where

$$n=1,3,5,\ldots,\ m,p=0,2,4,\ldots..\tag{4.7}$$

As noted in the Introduction, since the mode is sharply confined to the vicinity of the center of the plate, the edge of the electrode and, of course, the edge of the plate are both ignored in the treatment used in obtaining the solution presented here.

We now write the steady-state solution of (4.2) as a sum of eigensolutions (4.4), thus

$$u^n=\sum_m\sum_p H^{nmp}u_{nmp}.\tag{4.8}$$

Substituting from (4.8) into (4.2) and employing the homogeneous form of (4.2) for each eigensolution, we obtain

$$\sum_m \sum_p H^{nmp}(\omega^2 - \omega_{nmp}^2)u_{nmp} = \omega^2(-1)^{\frac{n-1}{2}} \frac{e_{26}4V}{c^{(1)}{}_n{}^2{}_2} , \quad (4.9)$$

from which with the aid of the orthogonality[19] of the u_{nmp}, we obtain

$$H^{nmp} = \frac{(-1)^{\frac{n-1}{2}} e_{26}\mathcal{L}_{nmp}V4}{c^{(1)}{}_n{}^2\pi^2 L_{nmp}[1-(\omega_{nmp}^2/\omega^2)]} , \quad (4.10)$$

where

$$L_{nmp} = \frac{\pi 2^m m!\, 2^p p!}{\sqrt{\alpha_n}\sqrt{\beta_n}} , \quad \mathcal{L}_{nmp} = 4F_{1nm}F_{3np} , \quad (4.11)$$

and

$$F_{1nm} = \int_0^{\ell_1} e^{-\alpha_n \frac{x_1^2}{2}} H_m(\sqrt{\alpha_n}x_1)dx_1 ,$$

$$F_{3np} = \int_0^{\ell_3} e^{-\beta_n \frac{x_3^2}{2}} H_p(\sqrt{\beta_n}x_3)dx_3 . \quad (4.12)$$

Since the mode is sharply confined to the center of the contoured resonator, in obtaining (4.10) from (4.9) we have replaced the circular electrode of radius ℓ by the circumscribed square of lengths $2\ell_1 = 2\ell_3 = 2\ell$ in order to simplify the integrals and obtain (4.12). Thus, we how have the series representation of the steady-state solution for the linear forced vibrations of the contoured resonator. The representation is given by (4.8) with (4.10) - (4.12), (4.4), (4.3), (3.27), (3.25) and (2.25) with $(2.26)_2$, all in the absence of the terms that result[3,5,11] from the transformation of the driving voltage. In the vicinity of a resonance, say the NMPth, one term in the sums in (4.8) and $(3.25)_1$ dominates and the others are negligible. Thus, from (4.8), (4.3), $(3.25)_1$ and (2.25), in the vicinity of said resonance the steady-state solution may be written in the form

$$u_1 = H^{NMP} \sin\frac{N\pi x_2}{2h} u_{NMP}e^{i\omega t} - \frac{e_{26}Vx_2}{c^{(1)}2h} e^{i\omega t} ,$$

$$\varphi = \frac{Vx_2}{2h} e^{i\omega t} + \frac{e_{26}}{\epsilon_{22}} H^{NMP}u_{NMP}\left(\sin\frac{N\pi x_2}{2h} - (-1)^{\frac{N-1}{2}}\frac{x_2}{h}\right)e^{i\omega t} , \quad (4.13)$$

where we have taken the liberty of introducing the terms resulting[3,5,11] from the transformation of the driving voltage, which have heretofore been omitted and, as usual, ω_{NMP} in (4.13) is to be replaced by

$$\hat{\omega}_{NMP} = \omega_{NMP} + i\omega_{NMP}/2Q_{NMP} , \quad (4.14)$$

in which Q_{NMP} is the unloaded quality factor of the doubly-rotated contoured resonator in the NMPth mode. The admittance Y_{NMP} of this contoured resonator operating in the NMPth mode is obtained by first substituting from (4.13) into

$$D_2 = e_{26}u_{1,2} - \epsilon_{22}\varphi_{,2} , \quad (4.15)$$

which is obtained by substituting from $(2.21)_2$ into (2.6), employing $(2.26)_1$ and ignoring u_2 and u_3, which are negligible for modes in the vicinity of the u_1-thickness mode, and then substituting from (4.15) into

$$I = -\int_{A_e} \dot{D}_2\, dx_1dx_3 , \quad (4.16)$$

with the result

$$Y_{NMP} = \frac{I}{V} = \frac{i\omega\epsilon_{22}}{2h_o}(1+\hat{k}_1^2)\hat{A}_e + \frac{i\omega\epsilon_{22}\hat{k}_1^2\mathcal{L}_{NMP}^2}{[(\omega_{NMP}^2/\omega^2) - 1]N^2\pi^2 L_{NMP}h_o} , \quad (4.17)$$

where

$$\hat{A}_e = A_e(1 + \ell^2/8Rh_o) , \quad (4.18)$$

and in obtaining the second term in (4.17) we have again replaced the circular electrode by the circumscribed square to perform the integrations. The quantities C_o and $\hat{C}_{NMP}$ defined by

$$C_o = \frac{\hat{A}_e\epsilon_{22}(1+\hat{k}_1^2)}{2h_o} , \quad \hat{C}_{NMP} = \frac{4\epsilon_{22}\hat{k}_1^2\mathcal{L}_{NMP}^2}{N^2\pi^2 L_{NMP}h_o} , \quad (4.19)$$

are called the static and motional capacitances, respectively.

Equation (4.6) has been employed in the calculation of the resonant frequencies of some doubly-rotated contoured quartz resonators, for which experimental data is available so that the calculations can be compared with measurement. We consider a number of doubly-rotated orientations, beginning with the SC-cut. The results of calculations of the resonant frequencies obtained from Eq.(4.6) for the BVA resonator treated in Table II of Ref.8 are presented in Table II of this work. For this particular resonator, the cut angles are $(\varphi,\theta) = 22°\,20', 34°\,6')$, R = 30 cm, and the stated center thickness is $2h_o = 1.092$ mm. We have adjusted the center thickness in our calculations to $2h_o = 1.091$ mm in order that $f_{300} = 5$ MHz. Since the resonator is of the BVA electrodeless type, the mass loading $\hat{R} = 0$, but the effects of the electrical shorting term are retained.

Before discussing the results in the table, we note that for the SC-cut orientation used in Table II, $M_n' > P_n'$ for the fundamental, and $M_n' < P_n'$ for the third, fifth, and seventh overtones. An examination of Eq.(4.6) shows that for these three overtones, for m + p = r, (r and n both fixed) ω_{nmp} is largest for m = 0, p = r, and decreases to a minimum at m = r, p = 0. Thus, for example, the following ordering should be observed for the SC-cut, $f_{304} > f_{322} > f_{340}$. Since the modal assignments for the anharmonics is usually conjectural, we base the data in Table II according to the above ordering scheme. It should be noted that the terms in Ref.8 which are used in place of our M_n' and P_n' are the constants c_{11} and c_{55}, respectively. For this particular orientation of quartz, $c_{11} > c_{55}$, and hence, for the fundamental and <u>every</u> overtone, the analysis of Ref.8 predicts the same anharmonic ordering scheme, which differs from that predicted here for the overtones using M_n' and P_n'. Now it is well known that some

harmonic families of certain resonators trap energy while others of the same resonator do not trap. This depends on the curvatures of the dispersion curves, i.e., M_n' or P_n', in the vicinity of zero lateral wavenumbers for each value of n. For energy trapping to occur in convex contoured resonators both M_n' and P_n' must be positive[20]. In addition, for good trapping to be present, the cotangent functions in M_n' and P_n' cannot be too near their infinities. The analysis of Ref.8 can predict only the resonant frequency, and implies that trapping is always present in convex resonators if c_{11} and c_{55} are both positive, since c_{11} and c_{55} are constants. This is a direct result of not satisfying the very important boundary conditions on the major surfaces of the plate.

Another valuable piece of information is contained in the anharmonic frequency spacing. The frequency differences predicted by the analysis of this work show that $|f_{302} - f_{320}| = 7$ kHz, which coincides with the measured value. The analysis of Ref.8 indicates the difference in spacing between the frequencies of the same modes to be 14 kHz. Similarly, $|f_{504} - f_{540}| = 9$ kHz for this work, 7 kHz from experiment and 27 kHz from the reported analytical results of Ref.8. After further examination, we conclude that the present analysis is consistently more accurate in predicting the spacing of the anharmonics than is the analysis of Ref.8. This further supports our claim of the proper modal ordering scheme . A comparison of the results obtained from the two analytical procedures shows that the results obtained by means of the present method generally agree more closely with the measured data than do those obtained by the method of Ref.8.

In Table III Eq.$(4.19)_2$ has been employed in calculating some values of motional capacitances for a few harmonic and anharmonic modes of the SC-cut quartz resonator of Table II. For this resonator, the static capacitance is calculated from Eq.$(4.19)_1$ as $C_o = 6.42$ pF.

Table IV provides some very enlightening information demonstrating the value of the present analytical procedure in predicting the presence or absence of trapping. The three sets of orientations are taken from Ref.9, Table 2. For each orientation, the center thickness of the plate was selected such that the calculated resonant frequency of the one harmonic mode indicated by an equals sign in the table matched the experimental value. Good agreement between experiment and theory is observed for all the orientations. The orientation $(\varphi, \theta) = (34^\circ, -22^\circ)$ has positive M_n' and P_n' for both the fundamental and third harmonics. The spacing between anharmonic frequencies is seen to be in quite close agreement with experiment; for example, $f_{302} - f_{320} = 23$ kHz from the present theory, and 20 kHz from experiment; $f_{304} - f_{340} = 45$ kHz calculated, 46 kHz measured.

For the orientation $(\varphi, \theta) = (46.1^\circ, -23.55^\circ)$, the calculations indicate a lack of energy trapping for the fifth harmonic which is a result of M_n' being negative. However, for the fundamental and third overtone, energy trapping is clearly present, as indicated both by calculation and by the experimental mode spectra plots of Ref.9, Figs.6 and 7. The presence of trapping is characterized by the strong resonance lines. Generally good agreement between theory and experiment is indicated for the trapped modes shown in Table IV.

For the orientation $(\varphi, \theta) = (40.9^\circ, -21^\circ)$ the mode spectra displayed in Figs.6, 7 and 8 of Ref.9 clearly indicate that for this particular orientation, the fundamental harmonic and anharmonic modes exhibit energy trapping. Again, trapping is characterized by strong resonance lines. The third overtone mode spectra are characterized by a host of spurious modes, indicating a lack of energy trapping. The calculations show that this orientation causes one of the cotangent functions in M_n' to be near its infinity, which means that the mode shape does not decay sufficiently sharply for trapping to occur.

For the fifth harmonic, the mode spectra of Ref.9, Fig.8, indicate that energy trapping is again present, as do the calculations. The resonant frequency of this fifth harmonic is not reported in Ref.9. Calculations of the motional capacitances for all three resonators of Table IV are also displayed in the same table. For the three reported orientations $(\varphi, \theta) = (34^\circ, -22^\circ)$, $(46.1^\circ, -23.55^\circ)$, and $(40.9^\circ, -21^\circ)$, the static capacitances C_o from Eq.$(4.19)_1$ turn out to be 6.43 pF, 5.67 pF, and 6.05 pF, respectively.

Acknowledgement

This work was supported in part by the Army Research Office under Contract No. DAAG 29-82-K-0130.

References

1. H.F. Tiersten and D.S. Stevens, "An Analysis of Contoured SC-Cut Quartz Crystal Resonators," Proceedings of the 36th Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey, 37 (1982).
2. Normal coordinates have been used in this manner before in the determination of the dispersion relations at the cutoff frequencies for propagation along the digonal axis in a monoclinic crystal plate by R.K. Kaul and R.D. Mindlin, "Vibrations of an Infinite, Monoclinic Crystal Plate at High Frequencies and Long Wavelengths," J. Acoust. Soc. Am., 34, 1895 (1962).
3. H.F. Tiersten, "Analysis of Intermodulation in Thickness-Shear and Trapped Energy Resonators," J. Acoust. Soc. Am., 57, 667 (1975).
4. R.D. Mindlin, "Thickness-Twist Vibrations of an Infinite Monoclinic Crystal Plate," Intern. J. Solids and Structures, 1, 141 (1965).
5. 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).
6. 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).
7. D.S. Stevens, H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Contoured AT-Cut Quartz Crystal Resonators," J. Appl. Phys., 54, 1704 (1983).
8. R. Bourquin, D. Nassour and D. Hauden, "Amplitude Frequency Effect of SC-Cut Quartz Trapped Energy Resonators," Proceedings of the 36th Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey, 200 (1982).
9. F. Euler and A. Kahan, "Harmonic and Anharmonic Modes of AK-Cut Crystal Resonators," Proceedings of the 38th Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey and Institute of Electrical and Electronics Engineers, New York, IEEE Cat. No.84-CH2062-8, 150 (1984).
10. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.7, Secs.1 and 2.

11. For more detail see D.S. Stevens and H.F. Tiersten, "An Analysis of Doubly-Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.
12. W.L. Ferrar, Algebra (Oxford, London, 1941), Chap.XII, Sec.2.2.
13. Ref.10, Chap.9, Sec.2.
14. H.F. Tiersten, "Electromechanical Coupling Factors and Fundamental Material Constants of Thickness Vibrating Piezoelectric Plates," Ultrasonics, 8, 19 (1970).
15. A. Ballato, "Doubly Rotated Thickness Mode Plate Vibrators," in Physical Acoustics, edited by W.P. Mason and R.N. Thurston (Academic, New York, 1977), Vol.XIII, Sec.III.
16. Usually the series expansion is written after each entire eigensolution is obtained. Here, in order to accommodate different in-plane functional behavior due to different lateral conditions, only thickness dependence has been fixed at this point. Equations (2.14) and $(2.19)_2$ of Ref.5 are written incorrectly in that the sum appearing in (3.24) of this work has been omitted. The remainder of Ref.5 is correct since the sum on n properly appears from Eq.(3.11) on in the forced vibration analysis.
17. P.W. Berg and J.L. McGregor, Elementary Partial Differential Equations (Holden-Day, San Francisco, 1966), Chap.2, Sec.3.
18. C.J. Wilson, "Vibration Modes of AT-Cut Convex Quartz Resonators," J. Phys., D7, 2449 (1974).
19. P.M. Morse,and H. Feshbach, Methods of Theoretical Physics (McGraw-Hill, New York, 1953), p.786.
20. If M_n' and P_n' are both negative trapping will occur in concave contoured resonators. If one is positive and the other is negative trapping will occur for an appropriate saddle shape.

TABLE I

COMPRESSED INDEX NOTATION

ij or kℓ	11	22	33	23	31	12	32	13	21
p or q	1	2	3	4	5	6	7	8	9

TABLE II

COMPARISON BETWEEN CALCULATED AND MEASURED RESONANT FREQUENCIES OF SOME HARMONIC AND ANHARMONIC MODES OF AN SC CUT BVA ELECTRODELESS RESONATOR WITH $R = 30$ cm, $\hat{R} = 0$, AND $2h_o \approx 1.092$ mm

Mode N	M	P	Calculated Frequency (kHz) Ref. 8	Calculated Frequency (kHz) Present Method	Measured Frequency (kHz)†
3	0	0	5000	5000	5000
3	0	2	5082	5085	5096
3	2	0	5096	5078	5089
3	0	4	5163	5168	5194
3	2	2	5176	5161	5185
3	4	0	5190	5154	5178
3	0	6	5242	5250	5293
3	2	4	5255	5243	5278
3	4	2	5269	5236	5268
3	6	0	5282	5230	-
5	0	0	8302	8312	8309
5	0	2	8366	8407	8411
5	2	0	8400	8402	8404
5	0	4	8468	8500	8512
5	2	2	8482	8496	8502
5	4	0	8495	8491	8495
5	0	6	8549	8593	8610
5	2	4	8562	8588	8599
5	4	2	8576	8584	8589
5	6	0	8589	8580	8582
7	0	0	11607	11621	11616
7	0	2	11689	11733	11741
7	2	0	11703	11702	11711
7	0	4	11772	11845	11865
7	2	2	11786	11814	11837
7	4	0	11799	11784	11809

† Mode identification is conjectural. See text.

TABLE III

CALCULATED VALUES OF MOTIONAL CAPACITANCES FOR A FEW MODES OF THE SC-CUT QUARTZ RESONATOR OF TABLE II

Mode N M P	Calculated Motional Capacitance C_{NMP} (fF)
1 0 0	4.87
1 0 2	0.157
1 2 0	0.006
3 0 0	0.265
3 0 2	0.109
3 2 0	0.115
5 0 0	0.066
5 0 2	0.032
5 2 0	0.032

TABLE IV

MEASURED AND CALCULATED FREQUENCIES, AND CALCULATED MOTIONAL CAPACITANCE FOR A FEW SELECT DOUBLY ROTATED QUARTZ RESONATORS

Measured Data Taken from Ref. [9]. For all Resonators R = 30.285 cm, $\hat{R}$ = .0029. The Center Thickness is Adjusted so that the Indicated Harmonic Calculated Frequency Matches the Measured Value.

Orientation φ θ	Mode N M P	Calculated Frequency (kHz)	Measured Frequency† (kHz)	Calculated Motional Capacitance C_{NMP} (fF)
34° -22°	1 0 0	3381	3381	1.80
	1 0 2	3465	3470	0.84
	1 0 4	3546	3557	0.39
	3 0 0	10002 *	10002	0.051
	3 0 2	10102	10113	0.026
	3 2 0	10079	10093	0.026
	3 0 4	10201	10232	0.019
	3 2 2	10179	10206	0.013
	3 4 0	10156	10186	0.019
46.1° -23.55°	1 0 0	3361	3365	10.82
	1 0 2	3441	3451	4.89
	3 0 0	10003 *	10003	0.497
	3 0 2	10086	10088	0.249
	3 2 0	10203	10192	0.239
	3 0 4	10167	10175	0.186
	3 0 6	10249	10264	0.000
	5 0 0	*	-	-
40.9° -21°	1 0 0	3359 *	3359	7.06
	1 0 2	3442	3451	3.23
	3 0 0	10322**	-	0.543
	5 0 0	16568	Not Reported	0.052

† Mode identification is conjectural

* $M_n' < 0$ ** $M_n' \to \infty$

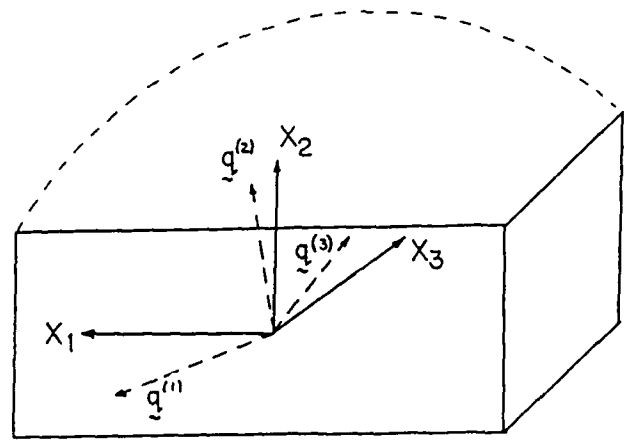


Figure 1 Schematic Diagram Showing the Original Plate Coordinate System and the Thickness Solution Eigendisplacement Coordinate System

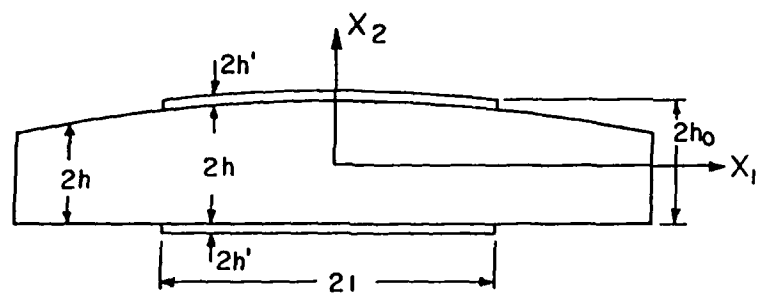


Figure 2 Schematic Diagram Showing a Cross-Section of the Plano-Convex Resonator

GENERALIZED EQUATION FOR THE FORCE-FREQUENCY CHARACTERISTICS OF CIRCULAR QUARTZ PLATES WITH THREE-POINT SUPPORT AND ITS APPLICATION TO SUPPORTING OF AN SC-CUT PLATE

Nobunori OURA, Naimu KURAMOCHI, Yasushi MIYAZAKI,
Masakazu YAMASHINA, and Shigeaki SUZUKI

Tokyo Institute of Technology,
Research Laboratory of Precision Machinery and Electronics,
4259, Nagatsuta, Midori-ku, Yokohama 227, JAPAN

ABSTRACT

A generalized equation which expresses the force-frequency characteristics of a circular quartz plate with unsymmetrical three-point support is derived from the concept of balance of applied forces and the superposition of virtual force pairs. Optimum supporting angles of an SC-cut plate are computed and the validity of the equation is confirmed with experiment. It is proposed that the most probable expansion coefficients of the force-frequency characteristics should be compiled from extensive experimental results.

I. INTRODUCTION

There is frequency change in a quartz plate irradiated with X-rays which is attributable to the changes of elastic constants[1]. This phenomenon has attracted the attention of many researchers and has been examined from various points of view.

One aspect of the research is the utilization of the force-frequency effect to transducers, e.g., a pressure gauge, accelerometer etc.[2,3,4]. Another is the suppression of the force-frequency effect. Researches on the latter aspect led to finding new angles of cut less sensitive to applied forces [5], identified supporting problems of resonator plates [6], frequency aging[7,8,9,10], third order elastic constants[11] etc.

The force-frequency effect was analyzed theoretically and the results compared with the experiment[12]. Ratajski compiled the experimental results of the force-frequency characteristics from many papers and defined the force sensitivity coefficients $K_f(\Psi)$ as a unified expression of the results[13]. Ballato devised a ceramic holder for an AT-cut plate compensating the force-frequency effect arising from a four-point mounting by expressing the edge-force coefficients of an AT-cut plate in a Fourier series [6]. This approach is effective for a practical design of the mounting[14]. The present authors discussed the mounting of circular plano-convex AT- and IT-cut plates with three-point support using a series expansion of the characteristics and the virtual pairs of radial forces[15].

In this paper the authors derive a generalized equation which expresses the force-frequency effect in an unsymmetrical three-point support. The force-frequency effect expressed by this equation is applied to mounting of a circular SC-cut plate supported at three-points. The equation fits experimental results to an extent sufficiently enough for practical use.

II. THEORETICAL

The change in the resonant frequency in a circular plate due to a pair of diametric forces is a function of azimuth angle Ψ between the X-axis and the position of applied force. In this paper, the specimen plate is supported with three points. The azimuth angle Ψ is defined as shown in Fig.1. In an unsymmetrical three-point support, the supporting angles are α and β, and radially applied forces are F, F_1 and F_2. For an SC-cut plate, the angle is measured from X'-axis counterclockwise.

A generalized equation which expresses the force-frequency characteristics in an unsymmetrical three-point support can be derived as follows. The force-frequency characteristics is a function of Ψ with a period of 180°. From the periodicity, or symmetry, Fig.2(a) is valid, and the superpostion [16,12] of the applied forces yields Fig.2(b). Therefore, the three-point support can be reduced to a configuration with three virtual pairs of diametric forces as shown in Fig.2(c). Figure 2 shows a concept to obtain the virtual three pairs of forces by periodicity and superposition. The superposition holds for three-point support as well as two-point support.

In Fig.1, the balance of force along the diameter parallel to the force F is given by

$$F = F_1\cos\alpha + F_2\cos\beta . \qquad (1)$$

On the other hand, the balance of force in the direction normal to that diameter is given by

$$F_1\sin\alpha = F_2\sin\beta . \qquad (2)$$

From (1) and (2), F_1 and F_2 are expressed in terms of F as follows:

$$F_1 = \frac{F}{\cos\alpha + \sin\alpha\cot\beta} , \qquad (3)$$

$$F_2 = \frac{F}{\cos\beta + \sin\beta\cot\alpha} . \qquad (4)$$

The balance of force holds for $0° < \alpha < 90°$ and $0° < \beta < 90°$ or $\alpha = \beta = 0°$. Let the force-frequency characteristic for a circular plate subjected to a single diametric force pair F be

$$\Delta f/f = g(\Psi) , \qquad (5)$$

CH2186-0/85/0000-0448$1.00©1985IEEE

where Δf is the frequency change; f is a resonant frequency perturbed by applied forces. The force-frequency characteristic for the support shown in Fig.1 is expressed by the sum of three pairs of virtual forces in Fig.2(c):

$$\frac{\Delta f}{f} = \frac{1}{2}\left\{g(\Psi) + \frac{g(\Psi-\alpha)}{\cos\alpha+\sin\alpha\cot\beta} + \frac{g(\Psi+\beta)}{\cos\beta+\sin\beta\cot\alpha}\right\}. \quad (6)$$

The factor of 1/2 originates from the virtual forces, whereas the real forces are F, F_1 and F_2. The second and the third term in (6) are frequency changes due to the forces in (3) and (4).

III. EXPERIMENTAL RESULTS FOR AN SC-CUT PLATE

To obtain the force-frequency characteristics for a three-point support, a vertical force is applied at the top edge of the specimen plate which is supported from below at two points as shown in Fig.3

The geometries of the specimens are listed in Table 1. The surfaces of the specimens were polished with a minimum worked layer[9]. Exciting keyhole electrodes, 8 mm in diameter, of Cr (500 A) and Au (700 A) were evaporated onto the central part of the plates, preheated to 200°C in 5×10^{-6} Torr. The resonant frequency changes were measured by the transmission method. The accuracy of angles α, β and Ψ is within $\pm$ 1°.

Figure 4 shows the force-frequency characteristic of a circular plano-convex SC-cut plate supported with a diametric force pair of 50 grams. We express this characteristic by the following polynomial;

$$g(\Psi) = \sum_{n=0}^{\infty}\{A_n\cos(2n\Psi) + B_n\sin(2n\Psi)\}, \quad (7)$$

where A_n and B_n are series expansion coefficients. In this calculation, n = 4 is sufficient. The coefficients are listed in Table 2. The force-frequency characteristics for various supporting angles can be depicted by using (6) and (7).

The frequency changes in an SC-cut plate due to symmetric forces at $\alpha = \beta$ = 30° to 75° are shown in Fig. 5. These characteristics are measured with a weight of 50 grams. The length of error bar corresponds to the standard deviation which was obtained from several measurements. The computed characteristics are the solid lines. For the experiment at supporting angles of 60°, the measurements were made with a weight of 272 grams because the frequency change was too small to measure at 50 grams. The values were reduced to 50 grams for the plot.

The characteristics of the plate supported unsymmetrically are shown in Fig.6. The calculated characteristics are also superposed on the measured characteristics with the solid line.

These results demonstrate that the equation (6) for the force frequency characteristics can be extended to the calculation of the frequency changes of a resonator supported with unsymmetrical multiple points.

IV. DISCUSSION

If the force-frequency characteristics are computed by using the coefficients A_n and B_n of the SC-cut plate obtained from the table in reference[14] for a supporting angle of 60° (symmetrical 120° support), the computed characteristics undulate around the ordinate where $\Delta f/f$ is zero like a sinusoidal wave.

On the contrary, the characteristics computed with the A_n and B_n obtained from the present experiment fit with the measured values and both the characteristics computed and measured do not intersect the $\Delta f/f$ ordinate. This suggests, for the application of equation (6) to the design of supporting the plates, that the coefficients, A_n and B_n, should be compiled from force-frequency characteristics of plates supported with a diametric force pair whose validity has been verified by the symmetrical triplet support of 120° ($\alpha = \beta = 60°$). Otherwise the design may not meet specifications.

A coefficient of sensitivity to planar-stress, $\langle K_f \rangle$ is proposed as follows[14],

$$\langle K_f \rangle = \frac{1}{\pi}\int_0^{\pi} K_f(\Psi)\,d\Psi . \quad (8)$$

This represents the superposition of a continuous distribution of edge stresses. The quantity $\langle K_f \rangle = 0$ means that the normalized frequency change $\Delta f/f$ is zero, whenever hydrostatic pressure is applied to the plate from the periphery, that is, the plate is completely compensated for stress. When $\langle K_f \rangle$ is zero, then the coefficient, A_0, must be zero. However, (8) does not require that any of the coefficients, except A_0, be zero. This is very important from the view point of support. Although the plate is compensated for hydrostatic force, the plate must be supported with a limited numbers of supporting points in reality. In other words, the plate is not compensated completely for radial forces applied from a limited numbers of supporting points, except although each supporting configuration has a few values of Ψ for which the frequency shift is zero.

The more symmetrical the support, the more the stress distribution in the plate resembles the distribution under hydrostatic pressure. In this context, the plate supported with three-points is approximately compensated for stress. Suppose that the force-frequency characteristic is given as a single periodic function of $g(\Psi) = A\cos(2\Psi)$, where A has the same amplitude as that of the real force-frequency characteristic. The quantity $\langle K_f \rangle$ in (8) is zero for this function. If we put this function into (6) with $\alpha = \beta = 60°$, the frequency change is expressed by the same function, $A\cos(2\Psi)$. When the real force-frequency characteristic, which contains higher harmonics, is substituted into (6), the magnitude of the additional frequency change is far smaller than that due to the single periodic function. The distortion of the force-frequency characteristic curve from a sinusoidal wave is instrumental in forming such characteristic as shown in Fig.5.

Therefore, a new criterion for supporting is defined as a root mean square of $\Delta f/f$ as follows:

$$I_f(\Psi) = \sqrt{\frac{1}{\pi}\int_0^{\pi}(\Delta f/f)^2 d\Psi} \quad . \qquad (9)$$

The quantity of $I_f(\Psi)$ was calculated for a symmetrical three-point support for various values of supporting angle, $\alpha = \beta$. $I_f(\Psi)$ reaches the minimum value at an angle of α of about 60° as shown in Fig.7. Incidentally the value calculated with the expansion coefficients A_n and B_n obtained from reference[14] is 0.94×10^{-7} at $\alpha = \beta = 60°$.

The optimum supporting angles, which satisfy the condition,

$$\Delta f(\Psi) = 0 \ , \qquad (10)$$

were computed with respect to the azimuth Ψ. The locus satisfying (10) is shown in Fig.8. This gives the azimuth angles for zero-frequency shift at angles where the criterion is not satisfied. It is helpful for the design of supports. The loci for angles $\alpha >$ 60° are omitted because theory does not match experiment in this region.

V. CONCLUSION

The generalized equation which expresses the force-frequency characteristics of a circular disk plate supported at three unsymmetrical points was derived from the concept of the balance of applied forces with respect to the coordinates whose origin are located on the center of the plate.

The optimum supporting angles, where frequency is relatively immune to radially applied forces for an SC-cut plate, were computed by using the equation.

For application of the equation to the design of supports for circular plates, it is proposed that the most probable expansion coefficients of the force-frequency characteristics, for each angle of cut, be extracted and compiled from extensive experimental results .

ACKNOWLEDGEMENT

The authors wish to extend their gratitude to Professor Emeritus H. Fukuyo for his encouragement and Dr. S. Nonaka and Mr. E. Konishi of NEC Corporation for their offer of specimen plates for the study. They also wish to thank Dr. Joseph Elmer Rhodes for reviewing the manuscript.

REFERENCES

[1] V.E.Bottom,"The effect of X-ray on the elastic constants of quartz," Phys. Rev., Vol.71, p.476, 1947.

[2] S.S.Chuang,"Force sensor using double-ended tuning fork quartz crystals," Proc. 37th AFCS, pp.248-254, 1983.

[3] E.P.EerNisse and J.M.Paros,"Practical consideration for miniature quartz resonator force transducers," Proc. 37th AFCS, pp.255-260, 1983.

[4] M.Onoe et al.,"Quartz crystal accelerometer insensitive to temperature variation," Proc. 31st AFCS, pp.62-70, 1977.

[5] E.P.EerNisse, "Quartz resonator frequency shifts arising from electrode stress," Proc. 29th AFCS, pp.1-4, 1975.

[6] A.Ballato,"Force-frequency compensation applied to four-point mounting of AT-cut resonator," IEEE Trans. Son. Ultrason. vol.SU-25, pp.223-226,1978.

[7] A.W.Warner, D.B.Fraser, and C.D,Stockbridge, "Fundamental studies of aging in quartz resonators," IEEE Trans. Son. Ultrason. vol.SU-12, pp.52-59,1965.

[8] A.E.Wainwright, F.L.Walls, and W.D.McCaa,"Direct measurement of the inherent frequency stability of quartz crystal resonator," Proc. 28th AFCS, pp.177-180, 1974.

[9] N.Oura et al.,"Effect of the worked layer in quartz-crystal plates on their frequency stabilities," IEEE Trans. Instrum. Meas., vol.IM-30, no.2,pp.139-143,1981.

[10]E.Aygun and A.Alalçakir,"An investigation of aging in a quartz crystal by VLF phase comparison method," IEEE Trans. Son. Ultrason. vol.SU-31, no.2, pp.94-100.1984.

[11]R.N.Thurston et al.,"Third-order elastic coefficients of quartz," J. Appl. Phys.,vol.37, no.1, pp.267-275, 1966.

[12]P.C.Y.Lee, Y.S.Wang, and X.Markenschoff, "High-frequency vibrations of crystal plates under initial stresses," J.Acoust. Soc. Amer., vol.57, no.1, pp.95-105,1975.

[13]J.M.Ratajski,"Force-frequency coefficient of singly rotated vibrating quartz crystals," IBM J. Res. Dev.,vol.12,pp.92-99,1968.

[14]A.Ballato and M.Mizan,"Simplified expressions for the stress-frequency coefficients of quartz plates," IEEE Trans. Son. Ultrason. vol.SU-31, no.1,pp.11-18,1984.

[15]N.Oura et al.,"Frequency changes in fifth-overtone 5-MHz circular plano-convex AT- and IT-cut plates due to radially applied three forces," IEEE Trans. Son. Ultrason. vol.SU-30,no.1,pp.111-114, 1983.

[16]C.R.Mingins, L.C.Barcus, and R.W.Perry,"Effect of external forces on the frequency of vibrating crystal plates," Proc. 16th AFCS, pp.46-76,1962.

Table 1. Specification of the specimen crystal plate.

diameter	15 mm
radius of curvature of convex surface	178 mm
maximum thickness	0.674 mm
overtone	3 rd
cut angle	θ = 33°52' ϕ = 22°24' ± 15'
frequency	f = 8.192 MHz

Table 2. Coefficients A_n and B_n.

n	A_n	B_n
0	0.373	0
1	0.42	1.39
2	-5.85E-2	-9.61E-3
3	3.9E-2	-6.59E-2
4	2.09E-2	8.2E-3

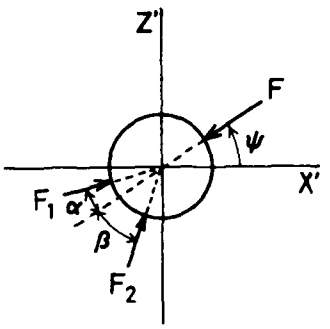


Fig.1. Forces and angles for radially applied three forces.

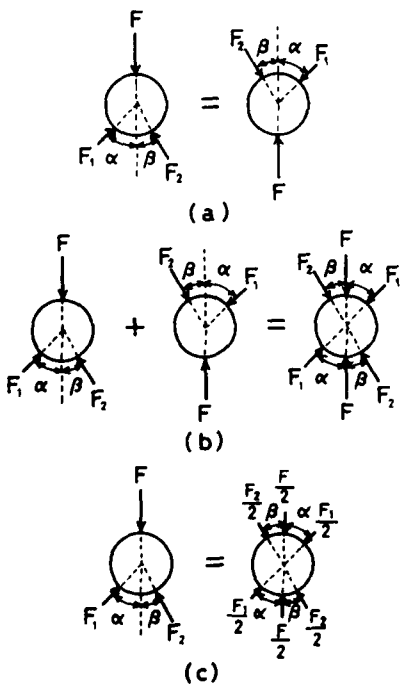


Fig.2. Deduction of the virtual three pairs of forces by periodicity, symmetry, and superposition.

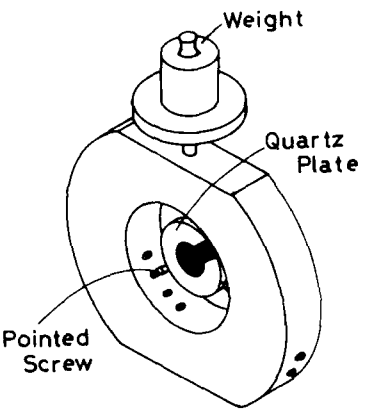


Fig.3. Test equipment setup.

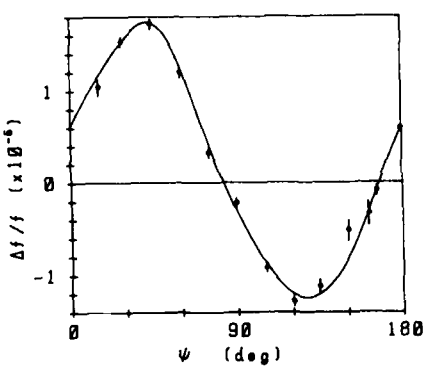


Fig.4. Force-Frequency characteristics for an SC-cut plate. F=50 g. Solid line was obtained from curve fitting by the polynomial expression of eq.(7).

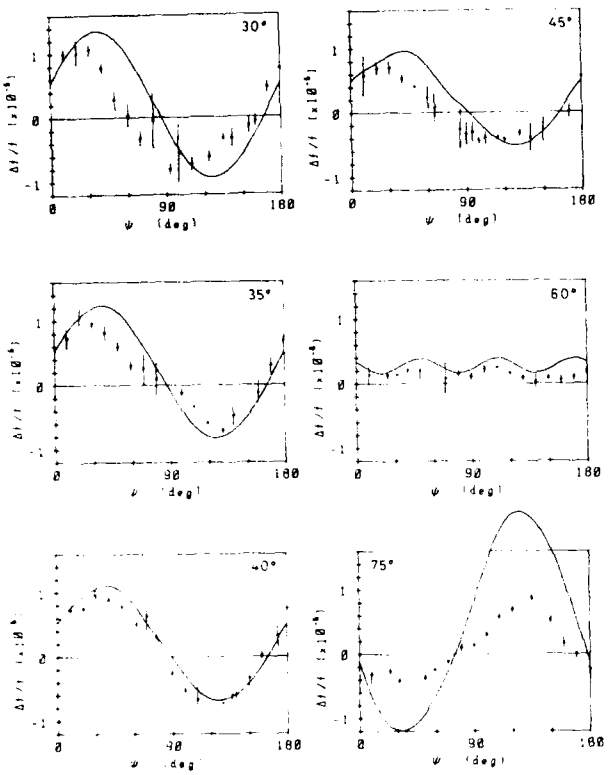


Fig.5. Force-frequency characteristics for an SC-cut plate with three radial forces, two symmetrically applied. F= 50 g.

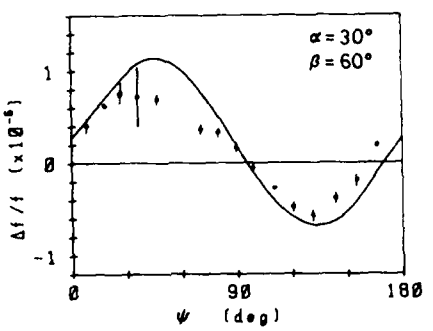


Fig.6 Force-frequency characteristics for an SC-cut plate with three unsymmetrical radial forces.

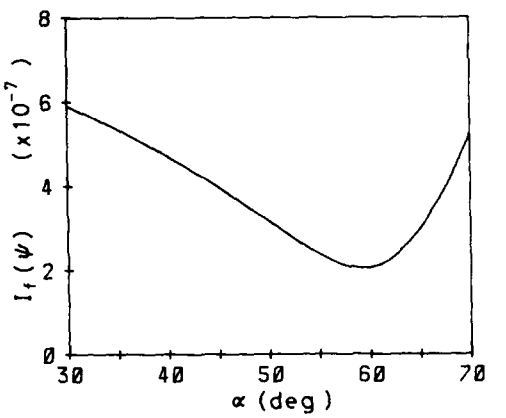


Fig.7. Criterion $I_f(\Psi)$ vs. supporting angle $\alpha = \beta$ in symmetrical three-point support.

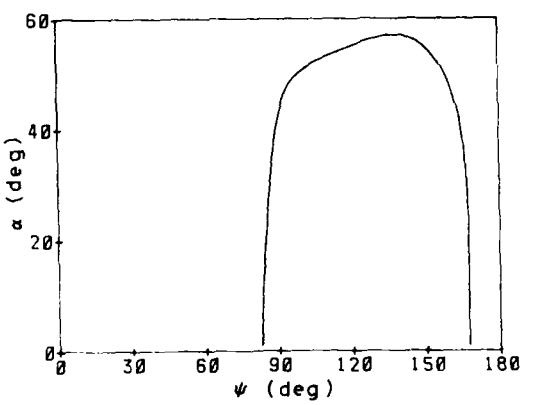


Fig.8. Locus of zero-frequency shift angles satisfying the condition (10) in symmetrical support as a function of Ψ.

Vibrations of Doubly-Rotated Piezoelectric Crystal Strip with a Pair of Electrode-Plated, Traction-Free Edges

P.C.Y. Lee and J.P. Hou

Department of Civil Engineering
PRINCETON UNIVERSITY
Princeton, NJ 08544

Abstract

Two-dimensional equations of motion of piezoelectric crystal plates, derived previously by Syngellakis and Lee, from the three-dimensional equations of the linear piezoelectricity by expansion in trigonometric series of the thickness coordinate of the plate, are employed to study the vibrations driven by the lateral electric field applied to a pair of electrode-plated and traction-free edges of doubly-rotated quartz strip.

Straight-crested vibrational solutions in closed form are obtained for the first eight coupled equations which constitute the two-dimensional, first-order approximate equations of motion of piezoelectric crystal plates.

Dispersion curves, frequency spectrum for electrically shorted and traction-free edge conditions, and mode shapes at different resonances and for various width-to-thickness ratios of the plate are computed and plotted for SC-cut quartz strips.

Introduction

Vibrations of piezoelectric crystal plates in predominantly thickness modes are usually driven by the electric field directed in the thickness direction and hence it is called thickness-field excitation (TE). It is also possible, but less common, for plates to be driven by the electric field applied in the direction parallel to the face of the plate and it is called lateral-field excitation LE).

Lateral excitation of piezoelectric crystal plates in simple thickness modes were studied by Yamada and Niizeki [1] and recently by Hatch and Ballato [2] and Ballato et al [3]. All these studies were based on the solution of the three-dimensional equations of piezoelectricity. Some of the reasons for the recent study and interest in the lateral excitation and a comprehensive list of references on LE are given in [3].

Forced vibrations of piezoelectric crystal plates with finite width (or length) were studied by Tiersten and Mindlin [4], Mindlin [5], [6]. In these studies, two-dimensional plate equations based on power series expansion were employed to investigate thickness-field excitation.

In the present paper, the two-dimensional plate equations based on trigonometric series expansion by Syngellakis and Lee [7] are employed to study the lateral-field excitation of doubly-rotated quartz strip.

General solution in closed form for the eight coupled equations of motion is obtained as the sum of the solutions of homogeneous equations and the solution of the inhomogeneous equations. The electrically forced and traction-free edge conditions are satisfied.

Two-Dimensional Plate Equations

Consider a doubly-rotated crystal plate, referred to a rectangular cartesian coordinate system, with the width denoted by $2a$ in x_1 direction, thickness $2b$ in x_2 direction, and length $2c$ in x_3 direction, as shown in Fig. 1. The plate is called a strip when the length $2c$ is approaching infinite. The faces of the plate are coated with dielectric platings and the edges, at $x_1 = \pm a$, are coated with conducting electrodes.

When the two-dimensional first-order equations for the vibrations of piezoelectric crystal plates (by Syngellakis and Lee are applied to the strips, the mechanical displacements and electric potential are expanded in a two-term trigonometrical series:

$$\begin{aligned} u_1(x_1,x_2,t) &= u_1^{(0)}(x_1,t) + u_1^{(1)}(x_1,t)\sin\left(\frac{\pi x_2}{2b}\right) \\ u_2(x_1,x_2,t) &= u_2^{(0)}(x_1,t) + u_2^{(1)}(x_1,t)\sin\left(\frac{\pi x_2}{2b}\right) \\ u_3(x_1,x_2,t) &= u_3^{(0)}(x_1,t) + u_3^{(1)}(x_1,t)\sin\left(\frac{\pi x_2}{2b}\right) \\ \phi(x_1,x_2,t) &= \phi^{(0)}(x_1,t) + \phi^{(1)}(x_1,t)\sin\left(\frac{\pi x_2}{2b}\right) \end{aligned} \tag{1}$$

Above six displacement components $u_i^{(n)}$ and two potential components $\phi^{(n)}$, $i=1,2,3$ and $n = ,1$, accommodate the vibrations of flexure, extension, face-shear, and fundamental thickness-shear, thickness-twist, and thickness-stretch modes and their anharmonic overtones, and interactions with electric fields $E_1^{(0)} = -\phi^{(0)}_{,1}$ and $E_2^{(0)} = -\phi^{(1)}$.

The stress equations of motion and charge equations which govern the eight components in Equation (1) are:

CH2186-0/85/0000-0453$1.00©1985IEEE

$$
\begin{aligned}
&T_{11,1}^{(1)} - \left(\frac{\pi}{2b}\right)\bar{T}_{21}^{(0)} = \rho(1+2R)u_{1,tt}^{(1)}\\
&T_{21,1}^{(1)} - \left(\frac{\pi}{2b}\right)\bar{T}_{22}^{(0)} = \rho(1+2R)u_{2,tt}^{(1)}\\
&T_{31,1}^{(1)} - \left(\frac{\pi}{2b}\right)\bar{T}_{23}^{(0)} = \rho(1+2R)u_{3,tt}^{(1)}\\
&D_{1,1}^{(1)} - \left(\frac{\pi}{2b}\right)\bar{D}_{2}^{(1)} = 0\\
&T_{11,1}^{(0)} = 2\rho(1+R)u_{1,tt}^{(0)}\\
&T_{21,1}^{(0)} = 2\rho(1+R)u_{2,tt}^{(0)}/\alpha_2\\
&T_{31,1}^{(0)} = 2\rho(1+R)u_{3,tt}^{(0)}\\
&D_{1,1}^{(0)} = 0
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
\bar{D}_2^{(1)} &= \frac{4\alpha_1}{\pi}\left(e_{21}u_{1,1}^{(0)} + e_{25}u_{3,1}^{(0)} + e_{26}u_{2,1}^{(0)}\right)\\
&\quad + \frac{\pi}{2b}\left(e_{22}u_2^{(1)} + e_{24}u_3^{(1)} + e_{26}u_1^{(1)}\right)\\
&\quad - \frac{4\alpha_1}{\pi}\varepsilon_{21}\phi_{,1}^{(0)} - \frac{\pi}{2b}\varepsilon_{22}\phi^{(1)}\\
T_{11}^{(0)} &= 2[C_{11}u_{1,1}^{(0)} + C_{15}u_{3,1}^{(0)} + C_{16}u_{2,1}^{(0)}\\
&\quad + \frac{\alpha_1}{b}\left(C_{12}u_2^{(1)} + C_{14}u_3^{(1)} + C_{16}u_1^{(1)}\right)]\\
&\quad + 2e_{11}\phi_{,1}^{(0)} + \frac{2\alpha_1}{b}e_{21}\phi^{(1)}\\
T_{12}^{(0)} &= 2[C_{61}u_{1,1}^{(0)} + C_{65}u_{2,1}^{(0)} + C_{66}u_{2,1}^{(0)}\\
&\quad + \frac{\alpha_1}{b}\left(C_{62}u_2^{(1)} + C_{64}u_3^{(1)} + C_{66}u_1^{(1)}\right)]\\
&\quad + 2e_{16}\phi_{,1}^{(0)} + \frac{2\alpha_1}{b}e_{26}\phi^{(1)}\\
T_{13}^{(0)} &= 2[C_{51}u_{1,1}^{(0)} + C_{55}u_{3,1}^{(0)} + C_{56}u_{2,1}^{(0)}\\
&\quad + \frac{\alpha_1}{b}\left(C_{52}u_2^{(1)} + C_{54}u_3^{(1)} + C_{56}u_1^{(1)}\right)]\\
&\quad + 2e_{15}\phi_{,1}^{(0)} + \frac{2\alpha_1}{b}e_{25}\phi^{(1)}\\
D_1^{(0)} &= 2[e_{11}u_{1,1}^{(0)} + e_{15}u_{3,1}^{(0)} + e_{16}u_{2,1}^{(0)}\\
&\quad + \frac{\alpha_1}{b}\left(e_{12}u_2^{(1)} + e_{14}u_3^{(1)} + e_{16}u_1^{(1)}\right)\\
&\quad - \varepsilon_{11}\phi_{,1}^{(0)} - \frac{\alpha_1}{b}\varepsilon_{12}\phi^{(1)}]
\end{aligned}
\tag{3}
$$

In Eq. (2), the traction-free and charge-free conditions on faces of the plate, i.e. $T_{2j} = 0$ and $D^{(0)} = D^{(1)} = 0$ at $x_2 = \pm b$, have been taken into account. The presence of the parameter R in Eq. (2) accounts for the mechanical effect of the mass of the dielectric platings. R is defined as the ratio of the mass of dielectric platings per unit area to the mass of the plate per unit area.

The stresses and electric displacements are related to the mechanical displacements and electric potential by the constitutive equations:

$$
\begin{aligned}
T_{11}^{(1)} &= C_{11}u_{1,1}^{(1)} + C_{15}u_{3,1}^{(1)} + C_{16}u_{2,1}^{(1)} + e_{11}\phi_{,1}^{(1)}\\
T_{12}^{(1)} &= \alpha_2(C_{61}u_{1,1}^{(1)} + C_{65}u_{3,1}^{(1)} + \alpha_2 C_{66}u_{2,1}^{(1)} + e_{16}\phi_{,1}^{(1)})\\
T_{13}^{(1)} &= C_{51}u_{1,1}^{(1)} + C_{55}u_{3,1}^{(1)} + C_{56}u_{2,1}^{(1)} + e_{15}\phi_{,1}^{(1)}\\
D^{(1)} &= e_{11}u_{1,1}^{(1)} + e_{15}u_{3,1}^{(1)} + e_{16}u_{2,1}^{(1)} - \varepsilon_{11}\phi_{,1}^{(1)}\\
\bar{T}_{21}^{(1)} &= \frac{4\alpha_1}{\pi}\left(C_{61}u_{1,1}^{(0)} + C_{65}u_{3,1}^{(0)} + C_{66}u_{2,1}^{(0)}\right)\\
&\quad + \frac{\pi}{2b}\left(C_{62}u_2^{(1)} + C_{64}u_3^{(1)} + C_{66}u_1^{(1)}\right)\\
&\quad + \frac{4\alpha_1}{\pi}C_{16}\phi_{,1}^{(0)} + \frac{\pi}{2b}e_{26}\phi^{(1)}\\
\bar{T}_{22}^{(1)} &= \frac{4\alpha_1}{\pi}\left(C_{21}u_{1,1}^{(0)} + C_{25}u_{3,1}^{(0)} + C_{26}u_{2,1}^{(0)}\right)\\
&\quad + \frac{\pi}{2b}\left(C_{22}u_2^{(1)} + C_{24}u_3^{(1)} + C_{26}u_1^{(1)}\right)\\
&\quad + \frac{4\alpha_1}{\pi}e_{12}\phi_{,1}^{(0)} + \frac{\pi}{2b}e_{22}\phi^{(1)}\\
\bar{T}_{23}^{(1)} &= \frac{4\alpha_1}{\pi}\left(C_{41}u_{1,1}^{(0)} + C_{45}u_{3,1}^{(0)} + C_{46}u_{2,1}^{(0)}\right)\\
&\quad + \frac{\pi}{2b}\left(C_{42}u_2^{(1)} + C_{44}u_3^{(1)} + C_{46}u_1^{(1)}\right)\\
&\quad + \frac{4\alpha_1}{\pi}e_{14}\phi_{,1}^{(0)} + \frac{\pi}{2b}e_{24}\phi^{(1)}
\end{aligned}
$$

where c_{pq}, e_{iq} and ε_{ij}, $i,j=1,2,3$, $p,q=1,2,\ldots,6$, are the elastic stiffness coefficients, piezoelectric and dielectric constants of the piezoelectric crystal plate.

Integration of the last equation of Eq. (2) with respect to x_1 leads to the conclusion that $D_1^{(0)}$ must be a function of time only, i.e.

$$
D_1^{(0)} = -\varepsilon_{11}A(t) \tag{4}
$$

By substituting Eq. (4) into the constitutive relation on $D_1^{(0)}$ in Eq. (3), and solving for $\phi_{,1}^{(0)}$, we have

$$
\begin{aligned}
\phi_{,1}^{(0)} &= \frac{1}{\varepsilon_{11}}\left[e_{11}u_{1,1}^{(0)} + e_{16}u_{2,1}^{(0)} + e_{15}u_{3,1}^{(0)}\right.\\
&\quad \left. + \frac{\alpha_1}{b}\left(e_{16}u_1^{(1)} + e_{12}u_2^{(1)} + e_{14}u_3^{(1)} - \varepsilon_{12}\phi^{(1)}\right)\right]\\
&\quad + A(t)
\end{aligned}
\tag{5}
$$

The last equation of Eq. (2) can be eliminated by inserting (5) into (3) and then, in turn, into Eq. (2). The resulting seven equations are the governing equations on the seven functions $u_i^{(0)}$, $u_i^{(1)}$ and $\phi^{(1)}$, $i=1,2,3$, as follows:

$$c_{11}u_{1,11}^{(1)} + c_{15}u_{3,11}^{(1)} + c_{16}u_{2,11}^{(1)} + e_{11}\phi_{,11}^{(1)} - \frac{2\alpha_1}{b}\left(c'_{61}u_{1,1}^{(0)} + c'_{65}u_{3,1}^{(0)} + c'_{66}u_{2,1}^{(0)}\right)$$

$$- \left(\frac{\pi}{2b}\right)^2\left(c'_{62}u_2^{(1)} + c'_{64}u_3^{(1)} + c'_{66}u_1^{(1)} - e'_{26}\phi^{(1)}\right) - \frac{2\alpha_1}{b}e_{16}A(t) = \rho(1 + 2R)u_{1,tt}^{(1)}$$

$$\alpha_2\left(c_{61}u_{1,11}^{(1)} + \alpha_2 c_{66}u_{2,11}^{(1)} + c_{65}u_{3,11}^{(1)} + e_{16}\phi_{,11}^{(1)}\right) - \frac{2\alpha_1}{b}\left(c'_{21}u_{1,1}^{(0)} + c'_{26}u_{2,1}^{(0)} + c'_{25}u_{3,1}^{(0)}\right)$$

$$- \left(\frac{\pi}{2b}\right)^2\left(c'_{22}u_2^{(1)} + c'_{24}u_3^{(1)} + c'_{26}u_1^{(1)} - e'_{22}\phi^{(1)}\right) - \frac{2\alpha_1}{b}e_{12}A(t) = \rho(1 + 2R)u_{2,tt}^{(1)}$$

$$c_{51}u_{1,11}^{(1)} + c_{55}u_{3,11}^{(1)} + c_{56}u_{2,11}^{(1)} + e_{15}\phi_{,11}^{(1)} - \frac{2\alpha_1}{b}\left(c'_{41}u_{1,1}^{(0)} + c'_{45}u_{3,1}^{(0)} + c'_{46}u_{2,1}^{(0)}\right)$$

$$- \left(\frac{\pi}{2b}\right)^2\left(c'_{42}u_2^{(1)} + c'_{44}u_3^{(1)} + c'_{46}u_1^{(1)} - e'_{24}\phi^{(1)}\right) - \frac{2\alpha_1}{b}e_{14}A(t) = \rho(1 + 2R)u_{3,tt}^{(1)}$$

$$e_{11}u_{1,11}^{(1)} + e_{15}u_{3,11}^{(1)} + e_{16}u_{2,11}^{(1)} - \varepsilon_{11}\phi_{,11}^{(1)} - \frac{2\alpha_1}{b}\left(e'_{21}u_{1,1}^{(0)} + e'_{25}u_{3,1}^{(0)} + e'_{26}u_{2,1}^{(0)}\right)$$

$$- \left(\frac{\pi}{2b}\right)^2\left(e'_{22}u_2^{(1)} + e'_{24}u_3^{(1)} + e'_{26}u_1^{(1)} - \varepsilon'_{22}\phi^{(1)}\right) + \frac{2\alpha_1}{b}\varepsilon_{21}A(t) = 0$$

$$c'_{11}u_{1,11}^{(0)} + c'_{15}u_{3,11}^{(0)} + c'_{16}u_{2,11}^{(0)} + \frac{\alpha_1}{b}\left(c'_{12}u_{2,1}^{(1)} + c'_{14}u_{3,1}^{(1)} + c'_{16}u_{1,1}^{(1)} + e'_{21}\phi_{21}^{(1)}\right) = \rho(1 + R)u_{1,tt}^{(0)}$$

$$c'_{61}u_{1,11}^{(0)} + c'_{65}u_{3,11}^{(0)} + c'_{66}u_{2,11}^{(0)} + \frac{\alpha_1}{b}\left(c'_{62}u_{2,1}^{(1)} + c'_{64}u_{3,1}^{(1)} + c'_{66}u_{1,1}^{(1)} + e'_{26}\phi_{,1}^{(1)}\right) = \rho(1 + R)u_{2,tt}^{(0)}/\alpha_2$$

$$c'_{51}u_{1,11}^{(0)} + c'_{55}u_{3,11}^{(0)} + c'_{56}u_{2,11}^{(0)} + \frac{\alpha_1}{b}\left(c'_{52}u_{2,1}^{(1)} + c'_{54}u_{3,1}^{(1)} + c'_{56}u_{1,1}^{(1)} + e'_{25}\phi_{,1}^{(1)}\right) = \rho(1 + R)u_{3,tt}^{(0)} \tag{6}$$

$$c'_{pq} = c_{pq} + \frac{e_{1p}e_{1q}}{\varepsilon_{11}}$$

$$e'_{iq} = e_{ip} - \frac{\varepsilon_{1i}e_{1q}}{\varepsilon_{11}}$$

$$\varepsilon'_{ij} = \varepsilon_{ij} - \frac{\varepsilon_{1j}\varepsilon_{1j}}{\varepsilon_{11}}$$

$$\alpha_1 = \pi/\sqrt{8}$$

$$\alpha_2 = \frac{v_s}{\left[\bar{c}_{66} - \dfrac{\bar{c}_{56}^2\left(\bar{c}_{11} - v_s^2\right) + \bar{c}_{16}^2\left(\bar{c}_{55} - v_s^2\right) - 2\bar{c}_{15}\bar{c}_{16}\bar{c}_{56}}{\left(\bar{c}_{11} - v_s^2\right)\left(\bar{c}_{55} - v_s^2\right) - \bar{c}_{15}^2}\right]^{1/2}}$$

$$\bar{c}_{pq} = c_{pq}/c_{66}\,, \qquad v_s = v/\sqrt{\rho c_{66}} \tag{7}$$

In Eq. (7), c'_{pq}, e'_{iq} and ε'_{ij} are modified material constants, and α_1, α_2 are correction factors introduced in order to match the dispersion relations obtained from the two-dimensional equations more closely to those obtained from the three-dimensional equations of piezoelectricity.

To calculate α_2 by the equation in Eq. (7), one needs to know v, the velocity of surface wave in unit of m/sec, on the face of the plate. It depends on the orientation of the plate as well as the direction of the wave normal of the surface wave.

Let γ denote the angle between the x_3-axis and the wave normal $\underset{\sim}{w}$ of a surface wave as shown in Fig. Fig. 2. For convenience, the surface wave velocities v are computed from the three-dimensional equations of elasticity and plotted in Fig. 2, as a function of γ for various cuts of quartz plate.

For example, a surface wave propagating along the x_1-axis ($\gamma = 90°$) in an SC-cut plate (from Fig. 2) should have a speed of 3250 m/sec. Then by Eq. (7), $\alpha_2 = 0.8714$.

If the second term in the denominator on the right side of the formula for α_2 is neglected, then

$$\alpha_2 = \frac{v}{\sqrt{\rho\, c_{66}}}$$

and for SC-cut and $\epsilon = 90°$, $\alpha_2 = 0.8503$.

Forced Vibrations of Piezoelectric Crystal Strips by Lateral-Field Excitation

We seek the solution of forced vibrations driven by a steady alternating voltage applied to the electrodes coated on the edges of the plate at $x_1 = \pm a$. The edge is also free of traction. Therefore, we require, at $x_1 = \pm a$,

$$T_{11}^{(0)} = T_{12}^{(0)} = T_{13}^{(0)} = T_{11}^{(1)} = T_{12}^{(1)} = T_{13}^{(1)} = 0 \qquad (8)$$

and

$$\phi^{(0)} = \pm \phi_0 e^{i\omega t}, \qquad \phi^{(1)} = 0$$

The general solution of (6) will be obtained as the sum of the solution of homogeneous equations by letting $A(t) = 0$ and the particular solution of the inhomogeneous equations by letting $A(t) = A_0 e^{i\omega t}$. Then, the boundary conditions (8) are imposed on the general solution. From now on, we assume that no dielectric platings on the faces of the plate by setting $R = 0$.

Solutions of Homogeneous Equations and Dispersion Relations

We let $A(t) = 0$ in (6) and assume that steady vibrational form of solution

$$\begin{aligned}
u_1^{(1)} &= bA_1 \cos \xi x_1 e^{i\omega t} \\
u_2^{(1)} &= bA_2 \cos \xi x_1 e^{i\omega t} \\
u_3^{(1)} &= bA_3 \cos \xi x_1 e^{i\omega t} \\
\phi^{(1)} &= (\hat{c}_{66}/\varepsilon_{22})^{1/2} bA_4 \cos \xi x_1 e^{i\omega t} \\
u_1^{(0)} &= bA_5 \sin \xi x_1 e^{i\omega t} \\
u_2^{(0)} &= bA_6 \sin \xi x_1 e^{i\omega t} \\
u_3^{(0)} &= bA_7 \sin \xi x_1 e^{i\omega t}
\end{aligned} \qquad (9)$$

where

$$\hat{c}_{66} = c_{66} + e_{26}^2/\varepsilon_{22} \qquad (10)$$

Substitution of Eq. (9) with $A(t) = 0$ into (6) leads to seven simultaneous and homogeneous equations on seven amplitudes A_n, $n = 1,2,\ldots,7$, as follows

$$\begin{bmatrix}
\tilde{c}_{11}z^2 + c_{66} - \Omega^2 & \tilde{c}_{16}z^2 + \tilde{c}_{62} & \tilde{c}_{15}z^2 + c_{64} & \tilde{e}_{11}z^2 + \tilde{e}_{26} & \frac{4\alpha_1}{\pi}\tilde{c}_{61}z & \frac{4\alpha_1}{\pi}\tilde{c}_{66}z & \frac{4\alpha_1}{\pi}\tilde{c}_{65}z \\
\alpha_2\tilde{c}_{61}z^2 + c_{26} & \alpha_2\tilde{c}_{66}z^2 + c_{22} - \Omega^2 & \alpha_2\tilde{c}_{65}z^2 + \tilde{c}_{24} & \alpha_2\tilde{e}_{16}z^2 + \tilde{e}_{22} & \frac{4\alpha_1}{\pi}\tilde{c}_{21}z & \frac{4\alpha_1}{\pi}\tilde{c}_{26}z & \frac{4\alpha_1}{\pi}\tilde{c}_{25}z \\
\tilde{c}_{61}z^2 + c_{46} & \tilde{c}_{56}z^2 + \tilde{c}_{42} & \tilde{c}_{55}z^2 + \tilde{c}_{44} - \Omega^2 & \tilde{e}_{15}z^2 + \tilde{e}_{24} & \frac{4\alpha_1}{\pi}\tilde{c}_{41}z & \frac{4\alpha_1}{\pi}\tilde{c}_{46}z & \frac{4\alpha_1}{\pi}\tilde{c}_{45}z \\
\tilde{e}_{11}z^2 + c_{26} & \tilde{e}_{16}z^2 + \tilde{e}_{22} & \tilde{e}_{15}z^2 + \tilde{e}_{24} & -\tilde{\varepsilon}_{11}z^2 - \tilde{\varepsilon}_{22} & \frac{4\alpha_1}{\pi}\tilde{e}_{21}z & \frac{4\alpha_1}{\pi}\tilde{e}_{26}z & \frac{4\alpha_1}{\pi}\tilde{e}_{25}z \\
\frac{2\alpha_1}{\pi}c_{16}z & \frac{2\alpha_1}{\pi}\tilde{c}_{12}z & \frac{2\alpha_1}{\pi}\tilde{c}_{14}z & \frac{2\alpha_1}{\pi}\tilde{e}_{21}z & \tilde{c}_{11}z^2 - \Omega^2 & \tilde{c}_{16}z^2 & \tilde{c}_{15}z^2 \\
\frac{2\alpha_1}{\pi}c_{66}z & \frac{2\alpha_1}{\pi}\tilde{c}_{62}z & \frac{2\alpha_1}{\pi}\tilde{c}_{64}z & \frac{2\alpha_1}{\pi}\tilde{e}_{26}z & \tilde{c}_{61}z^2 & \tilde{c}_{66}z^2 - \frac{1}{\alpha_2^2}\Omega^2 & \tilde{c}_{65}z^2 \\
\frac{2\alpha_1}{\pi}\tilde{c}_{56}z & \frac{2\alpha_1}{\pi}\tilde{c}_{52}z & \frac{2\alpha_1}{\pi}\tilde{c}_{54}z & \frac{2\alpha_1}{\pi}\tilde{e}_{25}z & \tilde{c}_{51}z^2 & \tilde{c}_{56}z^2 & \tilde{c}_{55}z^2 - \Omega^2
\end{bmatrix}
\begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad (11)$$

where

$$\tilde{c}_{pq} = c'_{pq}/\hat{c}_{66}\,, \qquad \bar{c}_{pq} = c_{pq}/\hat{c}_{66}\,,$$

$$\tilde{e}_{ip} = e'_{ip}/(\hat{c}_{66}\varepsilon_{22})^{1/2}\,, \qquad \bar{e}_{ip} = e_{ip}/(\hat{c}_{66}\varepsilon_{22})^{1/2}$$

$$\tilde{\varepsilon}_{ij} = \varepsilon'_{ij}/\varepsilon_{22}\,, \qquad \bar{\varepsilon}_{ij} = \varepsilon_{ij}/\varepsilon_{22}\,,$$

$$\Omega = \omega/\left(\frac{\pi}{2b}\sqrt{\frac{\hat{c}_{66}}{\rho}}\right), \quad z = \xi/\left(\frac{\pi}{2b}\right). \tag{12}$$

Vanishing of the determinant of the coefficient matrix in Eq. (11) for nontrivial solutions leads to a seventh-order polynomial in Z^2. For a given value of Ω, it gives seven pairs of roots, $\pm Z_n$, n=1,...,7.

The $\Omega - Z_n$ relation is a seven branch dispersion relation as shown in Fig. 3, for the SC-cut and R = 0.

For each of the seven roots Z_n inserted into (11), there is a set of corresponding amplitudes which are denoted by A_{in}. Then, the ratios among these amplitudes

$$\alpha_{in} \equiv \frac{A_{in}}{A_{1n}}, \qquad i=1,2,\ldots 7 \tag{13}$$

can be solved from (11). We note $\alpha_{1n} = 1$, for n-1,2,...7.

Particular Solution and Cut-Off Frequencies

For the particular solution for the inhomogenous equations, we assume

$$A(t) = \frac{\alpha_1}{b}\left(\frac{\hat{c}_{66}}{\varepsilon_{22}}\right)^{\frac{1}{2}} A_{18}\, e^{i\omega t} \tag{14}$$

for steady vibrations, and let

$$u_1^{(1)} = b\beta_1 A_{18} e^{i\omega t}$$

$$u_2^{(1)} = b\beta_2 A_{18} e^{i\omega t}$$

$$u_3^{(1)} = b\beta_3 A_{18} e^{i\omega t}$$

$$\phi^{(1)} = (\hat{c}_{66}/\varepsilon_{22})^{1/2} b\beta_4 A_{18} e^{i\omega t}$$

$$u_1^{(1)} = u_2^{(0)} = u_3^{(0)} = 0 \tag{15}$$

Substition of Eq. (14) and (15) into (6), yields the following:

$$\begin{bmatrix} \tilde{c}_{66}-\Omega^2 & \tilde{c}_{62} & \tilde{c}_{64} & \tilde{e}_{26} \\ \tilde{c}_{26} & \tilde{c}_{22}-\Omega^2 & \tilde{c}_{24} & \tilde{e}_{22} \\ \tilde{c}_{46} & \tilde{c}_{42} & \tilde{c}_{44}-\Omega^2 & \tilde{e}_{24} \\ \tilde{e}_{26} & \tilde{e}_{22} & \tilde{e}_{24} & -\tilde{\varepsilon}_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} -\bar{e}_{16} \\ -\bar{e}_{12} \\ -\bar{e}_{14} \\ \bar{\varepsilon}_{21} \end{bmatrix} \tag{16}$$

For a given value of Ω, β_m, m=1,2,...4 can be solved from Eq. (16), provided the determinant of the coefficient matrix of β_m not equal to zero. The vanishing of the determinant leads to the values of Ω for the simple-thickness resonance frequencies.

General Solution and Frequency Spectrum

The general solution of Eq. (6) for function A(t) specified in Eq. (14), is the sum of the homogeneous solution (9) and the inhomogeneous solution (15). Therefore, we have

$$u_1^{(1)} = \left[\sum_{n=1}^{7} \alpha_{1n}A_{1n} \cos\left(\frac{\pi}{2}\frac{a}{b}Z_n\right) + \beta_1 A_{18}\right] b\, e^{i\omega t}$$

$$u_2^{(1)} = \left[\sum_{n=1}^{7} \alpha_{2n}A_{1n} \cos\left(\frac{\pi}{2}\frac{a}{b}Z_n\right) + \beta_2 A_{18}\right] b\, e^{i\omega t}$$

$$u_3^{(1)} = \left[\sum_{n=1}^{7} \alpha_{3n}A_{1n} \cos\left(\frac{\pi}{2}\frac{a}{b}Z_n\right) + \beta_3 A_{18}\right] b\, e^{i\omega t}$$

$$\phi^{(1)} = \left[\sum_{n=1}^{7} (\hat{c}_{66}/\varepsilon_{22})^{1/2} \alpha_{4n}A_{1n} \cos\left(\frac{\pi}{2}\frac{a}{b}Z_n\right) + (\hat{c}_{66}/\varepsilon_{22})^{1/2}\beta_4 A_{18}\right] b\, e^{i\omega t}$$

$$u_1^{(0)} = \sum_{n=1}^{7} \alpha_{5n}A_{1n} \sin\left(\frac{\pi}{2}\frac{a}{b}Z_n\right) b\, e^{i\omega t}$$

$$u_2^{(0)} = \sum_{n=1}^{7} \alpha_{6n}A_{1n} \sin\left(\frac{\pi}{2}\frac{a}{b}Z_n\right) b\, e^{i\omega t}$$

$$u_3^{(0)} = \sum_{n=1}^{7} \alpha_{7n}A_{1n} \sin\left(\frac{\pi}{2}\frac{a}{b}Z_n\right) b\, e^{i\omega t} \tag{17}$$

where α_{in}, i=1,2,...7, n-1,2,...7, are defined in (13) and can be calculated from (11), and β_m, m=1,2,3,4, from (16).

By substituting Eq. (17) into (5), and integrating the resulting expression with respect to x_1, we have

$$b^{-1}\phi^{(0)} = \sum_{n=1}^{7} (\hat{c}_{66}/\varepsilon_{22})^{1/2} \alpha_{8n} A_{1n} \sin\xi_n x_1 e^{i\omega t} + (\hat{c}_{66}/\varepsilon_{22})^{1/2} \beta_8 (x_1/b) A_{18} e^{i\omega t} + B_{18} e^{i\omega t} \tag{18}$$

where

$$\alpha_{8n} = \frac{1}{\varepsilon_{11}} \left\{ \frac{2\alpha_1}{\pi} \frac{1}{z_n} [e_{16}\alpha_{1n} + e_{12}\alpha_{2n} + e_{14}\alpha_{3n} - (\hat{c}_{66}/\varepsilon_{22})^{1/2} \varepsilon_{12}\alpha_{4n}] + e_{11}\alpha_{5n} + e_{16}\alpha_{6n} + e_{15}\alpha_{7n} \right\}$$

$$\beta_8 = \frac{\alpha_1}{\varepsilon_{11}} (\hat{c}_{66}/\varepsilon_{22})^{1/2} [e_{16}\beta_1 + e_{12}\beta_2 + e_{14}\beta_3 - (\hat{c}_{66}/\varepsilon_{22})^{1/2} \varepsilon_{12}\beta_4] \tag{19}$$

In order to satisfy the edge conditions, we substitute Eq. (17) and (18) into the constitutive relations (3), and then into (8). The result is a set of eight simultaneous equations on eight amplitudes A_{1n}, n-1,2,...,8, as follows:

$$\begin{bmatrix} & & & & & & & \\ & & & & & & & \\ & & & M_{mn} & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{bmatrix} \begin{bmatrix} A_{11} \\ A_{12} \\ A_{13} \\ A_{14} \\ A_{15} \\ A_{16} \\ A_{17} \\ A_{18} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \phi_0 \end{bmatrix} \tag{20}$$

where, for $n=1,2,...,7$,

$$M_{1n} = \left(\bar{c}_{11}\alpha_{1n} + \bar{c}_{16}\alpha_{2n} + \bar{c}_{15}\alpha_{3n} + \bar{e}_{11}\alpha_{4n}\right) z_n \sin\left(\frac{\pi}{2}\frac{a}{b}z_n\right)$$

$$M_{2n} = \left(\bar{c}_{61}\alpha_{1n} + \alpha_2\bar{c}_{66}\alpha_{2n} + \bar{c}_{65}\alpha_{3n} + \bar{e}_{16}\alpha_{4n}\right) z_n \sin\left(\frac{\pi}{2}\frac{a}{b}z_n\right)$$

$$M_{3n} = \left(\bar{c}_{51}\alpha_{1n} + \bar{c}_{56}\alpha_{2n} + \bar{c}_{55}\alpha_{3n} + \bar{e}_{15}\alpha_{4n}\right) z_n \sin\left(\frac{\pi}{2}\frac{a}{b}z_n\right)$$

$$M_{4n} = \alpha_{4n} \cos\left(\frac{\pi}{2}\frac{a}{b}z_n\right)$$

$$M_{5n} = \left[\frac{2\alpha_1}{\pi}\left(\bar{c}_{16}\alpha_{1n} + \bar{c}_{12}\alpha_{2n} + \bar{c}_{14}\alpha_{3n} + \bar{e}_{21}\alpha_{4n}\right) + \left(\bar{c}_{11}\alpha_{5n} + \bar{c}_{16}\alpha_{6n} + \bar{c}_{15}\alpha_{7n} + \bar{e}_{11}\alpha_{8n}\right) z_n\right] \cos\left(\frac{\pi}{2}\frac{a}{b}z_n\right)$$

$$M_{6n} = \left[\frac{2\alpha_1}{\pi}\left(\bar{c}_{66}\alpha_{1n} + \bar{c}_{62}\alpha_{2n} + \bar{c}_{64}\alpha_{3n} + \bar{e}_{26}\alpha_{4n}\right) + \left(\bar{c}_{61}\alpha_{5n} + \bar{c}_{66}\alpha_{6n} + \bar{c}_{65}\alpha_{7n} + \bar{e}_{16}\alpha_{8n}\right) z_n\right] \cos\left(\frac{\pi}{2}\frac{a}{b}z_n\right)$$

$$M_{7n} = \left[\frac{2\alpha_1}{\pi}\left(\bar{c}_{56}\alpha_{1n} + \bar{c}_{52}\alpha_{2n} + \bar{c}_{54}\alpha_{3n} + \bar{e}_{25}\alpha_{4n}\right) + \left(\bar{c}_{51}\alpha_{5n} + \bar{c}_{56}\alpha_{6n} + \bar{c}_{55}\alpha_{7n} + \bar{e}_{15}\alpha_{8n}\right) z_n\right] \cos\left(\frac{\pi}{2}\frac{a}{b}z_n\right)$$

$$M_{8n} = \alpha_{8n} \sin\left(\frac{\pi}{2}\frac{a}{b}z_n\right)$$

$$M_{18} = 0, \quad M_{28} = 0, \quad M_{38} = 0, \quad M_{48} = \beta_4$$

$$M_{58} = \frac{2\alpha_1}{\pi}\left(\bar{c}_{16}\beta_1 + \bar{c}_{12}\beta_2 + \bar{c}_{14}\beta_3 + \bar{e}_{21}\beta_4\right) + \frac{2}{\pi}\bar{e}_{11}\beta_8$$

$$M_{68} = \frac{2\alpha_1}{\pi}\left(\bar{c}_{66}\beta_{1p} + \bar{c}_{62}\beta_2 + \bar{c}_{64}\beta_3 + \bar{e}_{26}\beta_4\right) + \frac{2}{\pi}\bar{e}_{16}\beta_8$$

$$M_{78} = \frac{2\alpha_1}{\pi}\left(\bar{c}_{56}\beta_1 + \bar{c}_{52}\beta_2 + \bar{c}_{54}\beta_3 + \bar{e}_{25}\beta_4\right) + \frac{2}{\pi}\bar{e}_{15}\beta_8$$

$$M_{88} = \beta_8 \frac{a}{b} \tag{21}$$

For a given value of frequency Ω, the dimensionless wave numbers Z_n and amplitude ratios α_{ij} defined in Eq. (13), can be computed from (11) and β_m from (16). If the width-to-thickness ratio, a/b, is also given, then each member of the coefficient matrix of Eq. (20) can be caluclated, and amplitudes A_{1n}, n=1,2,...,8, can be solved in terms of ϕ_0 for forced vibrations.

If the electrodes are electrically shorted and edges at $X_1 = \pm a$ are traction-free, then $\phi_0 = 0$ on the right hand side of (20). For A_{1n} not all equal zero, the determinant of the coefficient matrix of A_{1n} must be zero. Therefore,

$$\det|M_{mn}| = F(\Omega, \theta, \phi, R, a/b) = 0 \tag{22}$$

The above is the equation for resonance frequencies. The roots Ω of Eq. (22) as a function of a/b are computed for Sc-cut of quartz with R=0. The result is the frequency spectrum, as shown in Fig. 4.

Mode Shapes

Once the resonance frequencies are determined, one may insert a particular value of these resonances back into Eq. (20) and calculate amplitude ratios among A_{1n}. By further substituting these amplitude ratios into (17), mechanical displacements and electric potentials are determined as functions of x_1 and t within a constant of multiplication. These functions are called mode shapes at resonances.

For $\Omega = 0.91$ and $a/b = 17.28$ (corresponding to pt. d_1 in Fig. 4), displacement components are computed from Eq. (17) and plotted as a function of x_1/a from $x_1/a = 0$ to $x_1/a = 1$ in Fig. 5, in which $U10 \equiv u_1^{(0)}$, $U20 \equiv u_2^{(0)}$, ... $U40 \equiv \phi^{(0)}$, and $U41 \equiv \phi^{(1)}$. It can be seen in Fig. 5 that the component U20 is predominant and has 22 anti-nodes across the width of the plate. Therefore, the mode is identified as F22. We note in the same figure, that $U11 (\equiv u_1^{(1)})$ is strongly coupled to U20 and has its maximum value at the edge of the plate.

In the same manner, mode shapes are calculated at the same frequency $\Omega = 0.91$ for $a/b = 18.30$, 19.05, 21.57, and 21.61 (corresponding to pts. d_2, d_4, d_6 and d_7, in Fig. 4, respectively), and they are shown in Figs. 6, 7, 8 and 9, respectively. It can be seen from these figures that, at the same frequency, the predominance of one particular mode over the others can be controlled by the a/b ratio.

Modes at fundamental thickness-shear are given in Fig. 10. It is seen that U11 has only one anti-node across the width of the plate and therefore it is designated by Tsh 1. Fig. 11 shows the modes at the fundamental thickness-twist resonance. Since $U31 (\equiv U_3^{(1)})$ has no phase reversal across the width of the plate, the mode is identified by TT0.

References

[1] T. Yamada and N. Niizeki, "Formulation of admittance for parallel field excitation of piezoelectric plates", J. Appl. Phys., Vol. 41, pp. 3604-3609, 1970.

[2] E. R. Hatch and A. Ballato, "Lateral-field excitation of quartz plates", Proc. IEEE Ultrasonics Symposium, pp. 512-515, 1983.

[3] 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. Control Symp., pp. 206-224, 1984.

[4] H. F. Tiersten and R. D. Mindlin, "Forced Vibrations of piezoelectric crystal plates", Quat. Appl. Math., Vol. 20, pp. 107-119, 1962.

[5] R. D. Mindlin, "High frequency vibrations of piezoelectric crystal plates", Int. J. Solids Structures, Vol. 8, pp. 895-906, 1972.

[6] R. D. Mindlin, "Frequencies of piezoelectrically forced vibrations of electroded, doubly rotated, quartz plates", Int. J. Solids Structures, Vol. 20, pp. 141-157, 1984.

[7] S. Syngellakis, and P. C. Y. Lee, "An approximate theory for the high frequency vibrations of piezoelectric crystal plates", Proc. 30th Ann. Freq. Control Symp., pp 184-190, 1976.

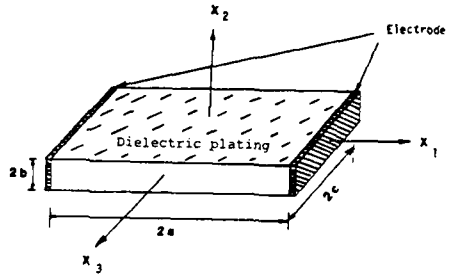


Figure 1. A rectangular piezoelectric crystal plate with dielectric platings on faces and electrodes on edges.

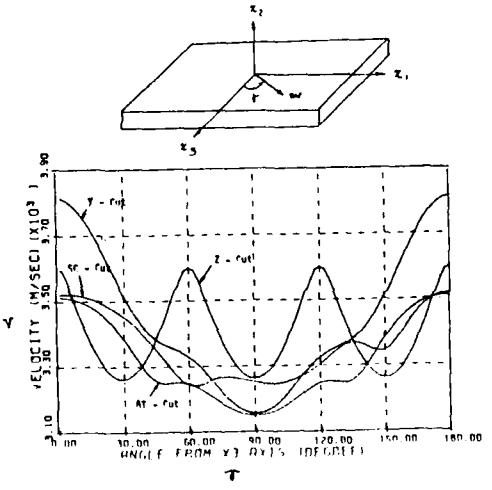


Figure 2. Dependence of surface-wave speed v on the direction of wave propagation for various cuts of quartz plates.

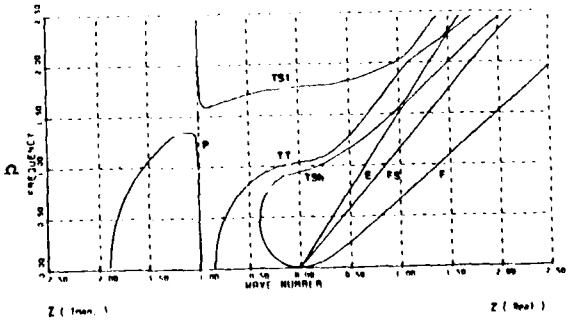


Figure 3. Dispersion curves for Sc-cut of quartz plate, R = 0.

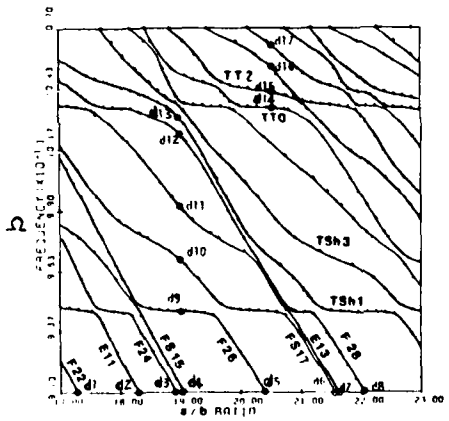


Figure 4

Frequency spectrum for SC-cut of quartz plates, R = 0.

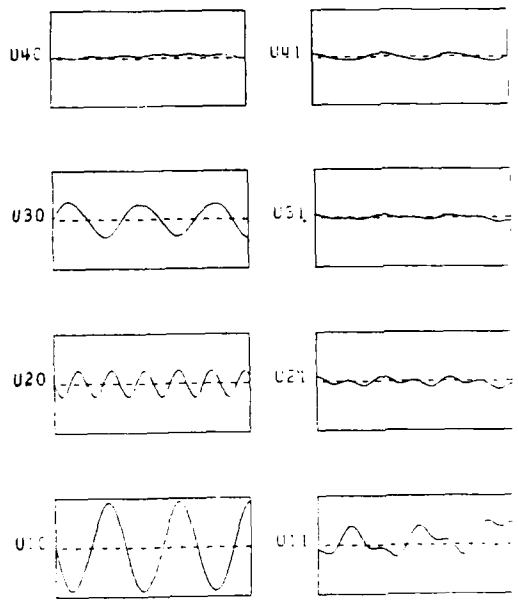


Figure 6. Modes for predominately extensional vibration (E11) at = 0.91, a/b = 18.30.

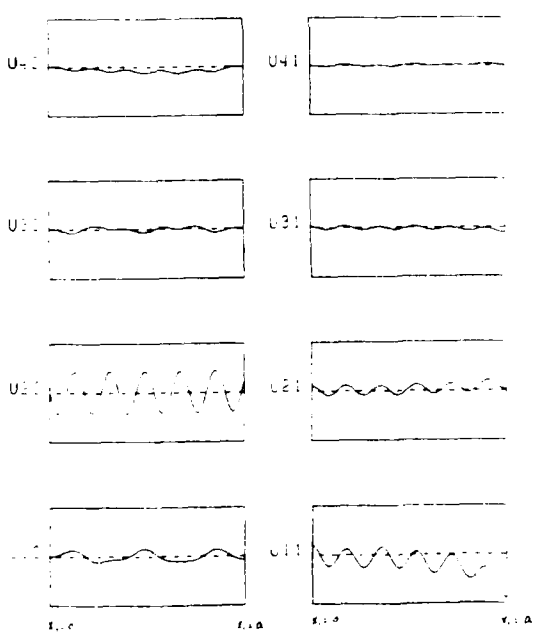


Figure 5. Modes for predominately flexural vibration (F22) at = 0.91, a/b = 17.48.

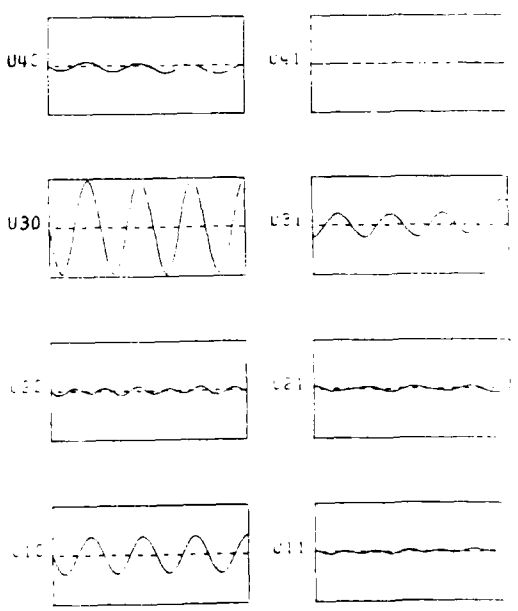


Figure 7. Modes for predominately face-shear vibration (FS15) at = 0.91, a/b = 19.05.

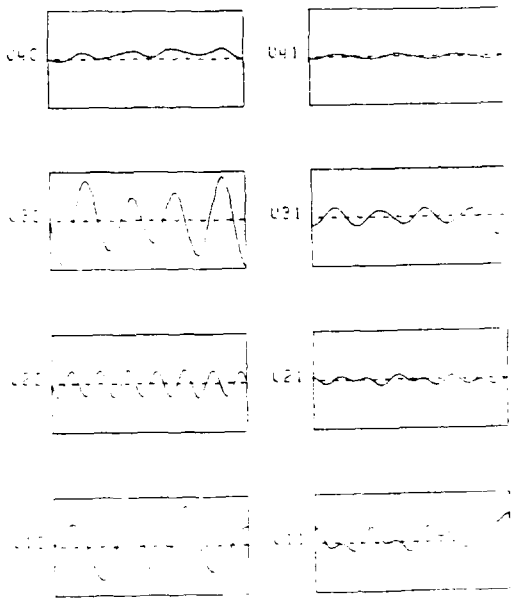


Figure 8. Modes [ ] fundamental [illegible] vibrations [illegible] coupled to [illegible] at [illegible]

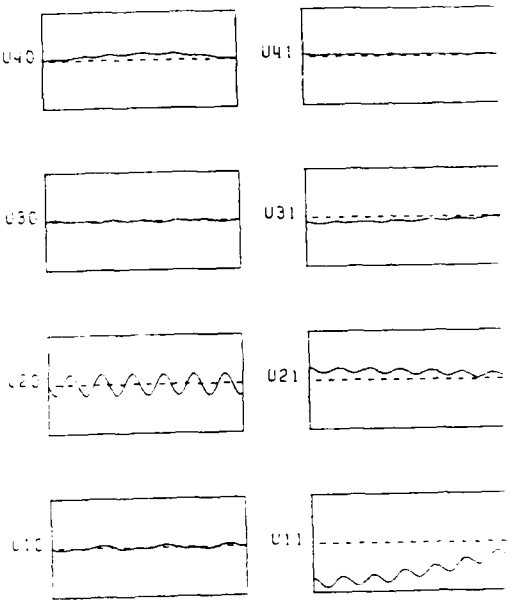


Figure 10. Modes [ ] fundamental thickness-shear vibrations [illegible] at [illegible], a/L = 19.00.

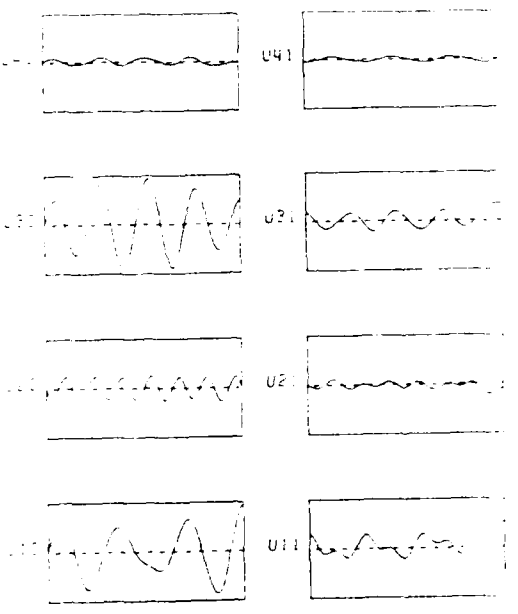


Figure 9. Modes for extensional vibration (E13) strongly coupled to fa e-shear at [illegible] = 0.91, a/t = 21.6.

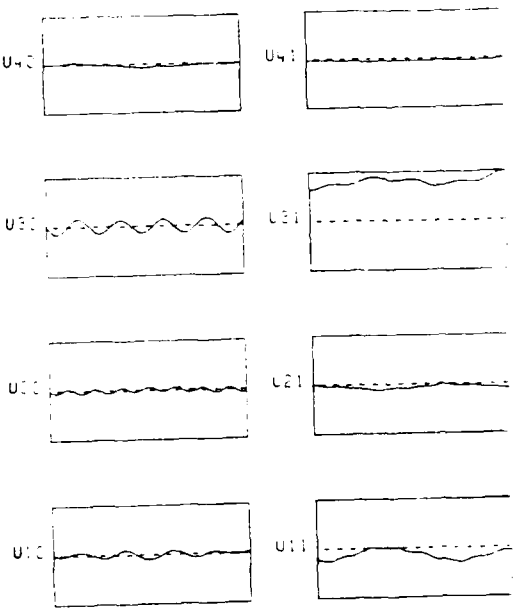


Figure 11. Modes [ ] fundamental thickness-twist vibrations [illegible] at [illegible] = 1.0349, a/b = 20.50.

SIMPLE THICKNESS PLATE MODES DRIVEN BY LATERAL FIELDS

A. Ballato, E.R. Hatch
M. Mizan, T. Lukaszek, & R. Tilton

U.S. Army Electronics Technology & Devices Laboratory
Fort Monmouth, New Jersey 07703

SUMMARY

Lateral field excitation calculations have been carried out for quartz thickness mode plates. The results are presented in terms of piezoelectric coupling coefficients, and interpreted using equivalent networks. Exact circuits using transmission lines are reduced to the more familiar lumped variety valid in the neighborhood of a single resonance. The element values are related to physical parameters of the vibrator. Calculations include coupling coefficients as functions of azimuth, for the three simple thickness modes of quartz plates lying on the zero temperature coefficient of frequency locus from the AT cut to the RT cut.

INTRODUCTION

Crystal plate resonators are usually driven piezoelectrically by electric fields directed along the thickness axis from electrodes placed upon their major surfaces. Lateral-field excitation (LFE) arises from electroding arrangements that produce driving fields parallel to the major surfaces.[1-35] LFE is the subject of current technological interest for a number of reasons:

- Lessened thermal transients.
- Reduced aging, since the electrode is absent from the region of greatest motion.
- Higher Q values, since electrode damping is reduced.
- Ability to eliminate undesired modes, e.g., the b mode in SC-cut quartz.
- Increased stability at a given harmonic, since motional inductances are larger.
- Means of experimentally driving certain plate modes for measurement.

At last year's Frequency Control Symposium, LFE was applied to berlinite.[32] We have now extended the results to include singly and doubly rotated cuts of quartz and have also completed the preliminary investigation reported in Ref. 30.

The distinction between thickness and lateral excitation is illustrated schematically in Fig.1. Also given are the established circuit symbols for TE and the here-proposed symbols for LE. Input immittances for the simple thickness modes of crystal plates of arbitrary anisotropy and piezoelectricity are given in Refs. 30 and 32, and Ref. 32 further gives the exact equivalent network for LE. This network comprises three acoustic transmission lines plus piezotransformers and static capacitance $\underline{C}_o$. Each transmission line accounts for one plate mode. In the following sections we shall give the derivation of the simple Butterworth-Van Dyke (BVD) circuit from the exact form, and present calculations of coupling factor from which the capacitance ratio $\underline{C}_o/\underline{C}_1$, associated with the BVD circuit, may be obtained.

TRANSMISSION LINES

For the case of a single acoustic mode, the expression for LE input admittance is

$$Y_{in} = j\omega \underline{C}_o \left\{ 1 + \underline{k}^2 \tan(\theta/2)/(\theta/2) \right\} . \qquad (1)$$

Underbars are used to indicate quantities associated with LE. See Ref. 32 for further notation and definitions. The resonance and antiresonance frequencies at the fundamental and harmonics may be obtained by the graphical construction of Fig. 2 for both TE and LE. It will be seen that for LE the resonance frequencies are integrally related, while for TE the harmonic antiresonance frequencies are integrally related. The straight lines in the figure have slopes that are inversely proportional to the squares of the piezocoupling factors k and $\underline{k}$. The other curve is a tangent function, its branches determine the harmonic (or overtone) of the resonator. The network realization of the tangent function is the transmission line (TL). The lossless TL is shown in Fig. 3 to be representable in either a tee or pi lumped form.

LUMPED CIRCUITS

Tangent and cotangent functions may be realized in lumped network configurations known as Foster forms.[36] These are shown in Fig. 4 and Fig. 5 along with the element values expressed in terms of the acoustic TL parameters.[37] The partial-fraction expansion about the poles of the tangent function is given at the bottom of Fig. 6; the realizations of Fig. 4 are based on this expansion. Each resonant denominator in Fig. 6 leads to the frequency at which one LC arm in Fig. 4 resonates.

This expansion about the poles may be generalized to take into account loss. The result is the appearance of a tanh function in Eq. (1), and resistances in the tank circuits of Fig. 4. When all three modes are considered, the three tangent functions generally have different periodicities, and the motional circuits representing them appear in parallel,[38] as seen in Fig. 6.

BEHAVIOR AT dc

The complete, lumped, LE network for the three

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

simple thickness modes is that shown at the top of Fig. 6. We relate the piezocoupling factors $\underline{k}_m$ for each mode m to the corresponding capacitance ratios $\underline{C}_o/\underline{C}_1$ for that mode as follows. For simplicity, we take a single mode; the result holds for the three modes individually. We also consider for simplicity the lossless limit. At dc, the input capacitance to the LE crystal resonator is just

$$C_{in} = \underline{C}_o\,(1 + \underline{k}^2). \qquad (2)$$

From the limiting form of the network, shown in Fig. 7, at dc, it is apparent that C_{in} is also equal to $\underline{C}_o$ plus an infinite series of capacitors in parallel, each proportional to $\underline{C}_1$, with a proportionality factor that diminishes with the square of the harmonic order. The infinite series sums to $\pi^2/8$, leading to the relation

$$\underline{C}_o / \underline{C}_1 = \underline{r} = \pi^2 / (8\,\underline{k}^2). \qquad (3)$$

A given electroding arrangement will lead to a fixed value for the lateral field static capacitance $\underline{C}_o$; the LFE motional capacitance $\underline{C}_1$ will then be

$$\underline{C}_1 = 8\,\underline{k}^2\,\underline{C}_o / \pi^2. \qquad (4)$$

The proportionality between $\underline{C}_1$ and $\underline{k}^2$ remains a good approximation for plates when considering an invariant plate and electrode geometry and only changing the azimuth angle psi:

$$\underline{C}_1(\psi) \propto \underline{k}^2(\psi), \qquad (5)$$

where $\underline{k}^2(\psi)$ is calculated from the infinite plate supporting simple thickness modes (examples of which are given for various quartz cuts in a subsequent section), and $\underline{C}_1(\psi)$ is measured on a finite plate with a nonuniform transverse distribution of motion; see the discussion on pp. 168-170 of Ref. 39.

INCLUSION OF LOSS

Inclusion of loss in the form of elastic viscosity leads to a modification of Eq. (1):

$$Y_{in}(s) = \underline{C}_o s \left\{ 1 + \underline{k}^2 \tanh(sh/v)/(sh/v) \right\}. \qquad (6)$$

The quantity $s=\sigma+j$ is complex frequency. With nonzero viscosity, the elastic stiffness, and hence acoustic velocity, wavenumber, and mechanical TL immittance, become complex instead of real.

Viscous loss leads to the introduction of the motional time constant[39] $\underline{\tau}_1$:

$$\underline{\tau}_1 = \eta / \bar{c} = \underline{R}_1 \underline{C}_1 \; ; \qquad (7)$$

η is the modal viscosity and $\bar{c}$ the piezoelectrically stiffened elastic modal stiffness. The time constants for the overtones of simple thickness motion are identical; there is but one time constant. The $\underline{R}_1\,\underline{C}_1$ products are, therefore, all equal, so,

$$\underline{R}_1 = \pi^2 \underline{\tau}_1 / (8\,\underline{C}_o\,\underline{k}^2) \propto \underline{k}^{-2}. \qquad (8)$$

The remarks following Eq. (4) relating to nonuniform motion and dependence on angle Ψ hold here also. The motional distribution is a function of overtone for finite plates and must be accounted for in applications.

Figure 8 gives the complete lumped equivalent circuit for a single simple thickness mode driven by LFE. At the top of the figure is the complex frequency plot of the poles and zeros of both Eq. (1) and Eq. (6). To the right is the graphical construction of Fig. 2 for the lossless case, showing how the pole-zero constellation arises. The lossy case follows from the lossless by a rotation by a fixed angle determined by the quality factor Q_1 of the fundamental; this determines the new zeros of the impedance function $(Y_{in}(s))^{-1}$. The new poles are located at the same imaginary frequency offsets as the zeros, with real frequencies found by the circular arc construction shown.

The motional inductances L are the same for all overtones as for the fundamental, and the motional capacitances diminish with the square of the harmonic number. The motional resistance goes up, accordingly, with the square of the harmonic number. For finite plates with nonuniform distribution of motion, this harmonic dependence will not be followed; bevelling, contouring, and details of electroding must be taken into account.[40]

On the other hand, for a given plate and electroding configuration, the relations for azimuthal dependence, Eqs. (5) and (8), will hold to a degree of approximation that permits practical design conclusions to be drawn.

TIME DOMAIN OPERATION

When the resonator is operated in the immediate vicinity of a single harmonic, the full lumped equivalent network of Fig. 8 is unnecessary. Only the motional arm of the particular resonance (harmonic) of interest need be retained, along with the static capacitance $\underline{C}_o$. We are thus led to the familiar four-element equivalent circuit of Fig. 9: the Butterworth-Van Dyke (BVD) circuit.

In many applications, particularly those using digital circuits, the time domain behavior of the resonator is of importance, irrespective of whether it is driven by LE or TE. In these cases the BVD circuit, which is so convenient for frequency- domain operation, is inadequate. A simple example is shown in Fig. 10 where a step function of voltage is applied to an LE resonator. Assuming that the BVD circuit represents the crystal, yields the time variation of current as

$$i(t) = V_o \underline{C}_o\, \delta(t) + V_o\, \underline{C}_1\, \omega_1 \sin \omega_1 t, \qquad (9)$$

where $\omega_1^2 = (2\pi f_R^{(1)})^2 = (\underline{L}_1 \underline{C}_1)^{-1}$. (10)

The full TL network, represented in lumped form as in Fig. 8, gives instead the correct answer

$$i(t) = V_o C_o\, \delta(t) + V_o C_1 \omega_1 \cdot \frac{\pi}{2} \left\{ \frac{1}{2} U(t) + \sum_{n=1}^{\infty} (-)^n U\left(t - \frac{n2h}{v}\right) \right\}. \qquad (11)$$

The step functions in Eq. (11) arise from the acoustic wave-fronts reflecting from the plate major surfaces.

LATERAL-FIELD COUPLING FACTORS

The procedure for obtaining the lateral-field coupling factors $\underline{k}_m$ for the three simple thickness modes of crystal plates has been described in Refs. 23, 30, and 32. We apply it here to quartz using Bechmann's values[41] for c^E, e, and ϵ^S. The old IEEE angle convention is for the present retained; to convert to the new recommended convention,[28] merely

replace theta by minus theta in the sequel.

Lateral coupling values versus theta for cuts of orientation (YXl) θ for modes m = a, b, and c are shown in Figs. 11 and 12. In Fig. 11, psi equals zero corresponds to the driving electric field in the direction of the digonal (X) axis; in Fig. 12, psi = 90° corresponds to the field in the Z direction. For comparison, the TE coupling is plotted in Fig. 13. Mode a is the quasi-longitudinal (extensional) mode, while modes b and c are the quasi-fast-shear and quasi-slow-shear (transverse) modes, respectively.

In Figs. 14 and 15 are plotted the LE coupling factors versus theta for quartz plates of orientation $(YX\omega\ell)\,\phi/\theta$, with ϕ = 22.4°, the SC cut value. Psi equals zero corresponds to the field along the rotated X axis, and psi = 90° corresponds to the field along the rotated Z axis. Figure 16 is the corresponding TE set of coupling values.

Figures 17 and 18 give the LE coupling factors for rotated-X-cut plates of orientation $(YX\omega\ell)\,\phi = 30°/\theta$, with psi = 0° and 90°. respectively.

The following sequences of figures are for individual quartz cuts. The abscissa is angle psi, the electric field azimuth angle. It, and the cut orientation angles phi and theta, are defined by the IEEE rotational symbol $(YX\omega\ell t)\,\phi/\theta/\psi$. Thus psi is referred to the rotated X axis of the plate.

Figures 19, 20, and 21 give $\underline{k}_m$ versus psi for the X cut, Y cut, and Z cut, respectively. Both shear modes m = b and c are degenerate for the Z cut.

Figures 22 and 23 give $\underline{k}_m$ amd $\underline{r}_m$ for the BT cut versus psi. Capacitance ratios $\underline{r}_m$ are obtained from $\underline{k}_m$ via Eq. (3).

The next sequence is for cuts on the upper zero temperature coefficient of frequency for mode c from the AT cut to the rotated X cut, and then on the lower ZTC c-mode locus to the RT cut. Figures 25 and 25 give $\underline{k}_m$ and $\underline{r}_m$ versus psi for the AT cut. Figures 26, 27, and 28 give $\underline{k}_m$ versus psi for ϕ = 13.9°, 15° (FC cut), and 19.1° (IT cut). Coupling $\underline{k}_m$ and capacitance ratio $\underline{r}_m$ for the SC cut are given in Figs. 29 and 30. respectively, while Fig. 31 shows $\underline{k}_m$ versus psi for the rotated X cut. Because of symmetry, when ϕ = 30°, theta and minus theta are the same.

The lower ZTC locus thus starts with Fig. 31, and continues with Fig. 32 which shows $\underline{k}_m$ for what might be called the "minus SC cut", i.e., the same phi angle as the SC cut, but with the theta angle reversed. The SC cut is on the AT cut side of theta, while the "minus SC cut" is on the BT side. Figure 33 gives $\underline{k}_m$ for the RT cut; this might be called the "minus FC cut", and vice versa.

In Fig. 34 is plotted the square of $\underline{k}_m(\Psi)$, for the SC cut, normalized to the maximum for the c mode. From Eq. (5), this is proportional to $\underline{C}_1(\Psi)$. One sees that the maxima for the a and b modes are only about 34% and 25% that of the desired c mode, and that by adjusting psi one can obtain very good excitation of the c mode while minimizing the excitations of the b and a modes. An angle of psi can always be found such that the excitation of a particular mode for any particular cut will be zero.[7,9,17,30]

Figure 35 shows $\underline{k}_m^2$ versus psi for the thermometric LC cut for LFE, while Fig. 36 shows the corresponding plots for the NLSC cut.[42-44] The NLSC scale is twice that of the LC; the NLSC cut has a stronger LE excitation coefficient, and the other modes are relatively weak, whereas the a mode of the LC is rather strong when driven by LE. The TE and LE excitations of both cuts are further compared in Ref. 44.

Examination of the psi-dependence of $\underline{k}_m$ in Fig. 19 ff discloses a nearly sinusoidal dependence. The curves may therefore be parametized in terms of amplitude and phase. We have fit the $\underline{k}_m(\Psi)$ curves for cuts on the upper ZTC locus of quartz by the relation

$$\underline{k}_m(\Psi) = \underline{k}_{mo} * \sin(\Psi + \Psi_{mo}). \quad (12)$$

Curves of $\underline{k}_{mo}(\phi)$ and $\Psi_{mo}(\phi)$ in the range $0° \leq \phi \leq 30°$ are given for m = a, b, and c in Figs. 37 and 38 respectively. From these curves $\underline{k}_m(\phi,\Psi)$ for cuts on the ZTC locus may be easily obtained.

CONCLUSION

The exact equivalent network for lateral field excitation of simple thickness modes in quartz plates has been given, and reduced to a lumped element circuit. Various simplified versions were obtained, including the ever-popular Butterworth-Van Dyke four element circuit. It was shown that for time-domain operation the exact version must be used, and that the BVD circuit is inadequate.

Values of the lateral field piezoelectric coupling coefficient were obtained for technologically important quartz cuts, and related to the BVD circuit elements. A simple relation was given for the azimuthal dependence of the coupling factors. It was observed that by adjusting the LF azimuth the excitation strengths may be altered for a particular application; in particular, undesired modes may be "tuned out" in a simple fashion, without circuit means.[16,17]

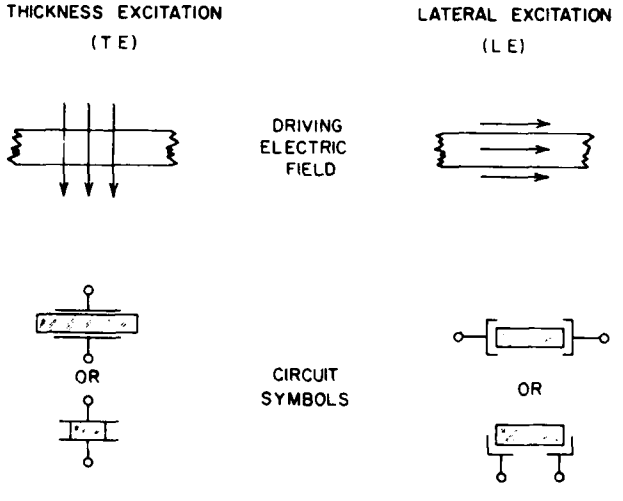


Figure 1. Driving electric field directions and circuit symbols for thickness excitation (TE) and lateral excitation (LE) of thickness modes of crystal plate vibrators.

REFERENCES*

1. J.V. Atanasoff and P.J. Hart, "Dynamical Determination of the Elastic Constants and Their Temperature Coefficients for Quartz," Phys. Rev., Vol. 59, January 1941, pp. 85-96.

2. W.G. Cady, Piezoelectricity, McGraw-Hill, New York, 1946; Dover, New York, 1964.

3. R. Bechmann, Über Dickenschwingungen piezoelektrischer Kristallplatten," Archiv der Elektrischen Übertragung, Vol. 6, September 1952, pp. 361-368.

4. R. Bechmann, Über Dickenschwingungen piezoelektrischer Kristallplatten," Nachtrag, Archiv der Elektrischen Übertragung, Vol. 7, July 1953, pp. 354-356.

5. R. Bechmann, "Filter crystals," Proc. 12th AFCS, May 1958, pp. 437-474.

6. R. Bechmann, "Improved High-Precision Quartz Oscillators Using Parallel Field Excitation," Proc. IRE, Vol. 48, March 1960, pp. 367-368.

7. R. Bechmann, "Parallel Field Excitation of Thickness Modes of Quartz Plates," Proc. 14th AFCS, May-June 1960, pp. 68-88.

8. V. Ianouchevsky, "Parallel Field Excitation," Proc. IRE, Vol. 48, June 1960, p. 1165.

9. R. Bechmann, "Excitation of Piezoelectric Plates by Use of a Parallel Field with Particular Reference to Thickness Modes of Quartz," Proc. IRE, Vol. 48, July 1960, pp. 1278-1280.

10. R. Bechmann, "Piezoelektrische Erregung von dickenschwingenden Quarzoszillatoren mittels eines Parallelfeldes," Archiv der Elektrischen Übertragung, Vol. 14, August 1960, pp. 361-365.

11. W. Ianouchewsky, "Quartz Étalons Lenticulaires Excités par le Champ Parallèle," Annales Françaises de Chronométrie, 31 ème année, 2 ème série, tome xv, 1961 pp. 41-51.

12. W. Ianouchevsky, "The Stability and Q-Factor of Quartz-Crystals Excited by the Parallel Field Technique," (Observatoire de Paris, Paris April 1963), Final Report to U.S. Dept. of Army, Contract No. DA-91-591-EUC-1752, May 1961 to April 1963. Available from NTIS as document No. AD 410296.

13. W. Ianouchevsky, "High Q Crystal Units," Proc. 17th AFCS, May 1963, pp. 233-247.

14. R.W. Warner, "Use of Parallel-Field Excitation in the Design of Quartz Crystal Units," Proc. 17th AFCS, May 1963, pp. 248-266.

15. R.H. Bechmann, "Piezoelectric crystal apparatus," U.S. Patent 3,165,651, issued 12 January 1965.

16. A.D. Ballato and R. Bechmann, "Piezoelectric Crystal Element," U.S. Patent No. 3,202,846, issued 24 August 1965; U.K. Patent No. 1,028,102, issued 4 May 1966.

17. H. Schweppe, "Excitation of Two Adjacent Resonances with a Chosen Frequency Separation in a Ceramic Piezoelectric Resonator," IEEE Trans. Sonics Ultrason., Vol. SU-17, January 1970, pp. 12-17.

18. T. Yamada and N. Niizeki, "Formulation of admittance for parallel field excitation of piezoelectric plates," J. Appl. Phys., Vol. 41, August 1970, pp. 3604-3609.

19. T. Yamada and N. Niizeki, "A new formulation of piezoelectric plate thickness vibration," Rev. Electrical Communication Labs., NTT, Vol. 19, May-June 1971, pp. 705-713.

20. A. Ballato, H.L. Bertoni, and T. Tamir, "Transmission-Line Analogs for Stacked Crystals with Piezoelectric Excitation," Presented at 83rd meeting of the Acoustical Society of America, Buffalo, New York, 18-21 April 1972, paper NN3. Abstract: J. Acoust, Soc. Am., Vol.52, No. 1 (Part 1), July 1972, p. 178, paper NN3.

21. M. Onoe, "General equivalent circuit of a piezoelectric transducer vibrating in thickness modes," Trans. Inst. Electronics and Comm. Engrs. Japan, Vol. 55A, May 1972, pp. 239-244. In Japanese.

22. A. Ballato, "Transmission-Line Analogs for Stacked Piezoelectric Crystal Devices," Proc. 26th AFCS, June 1972, pp. 86-91.

23. A. Ballato, "Transmission-Line Analogs for Piezoelectric Layered Structures." Ph.D. Dissertation, Polytechnic Institute of Brooklyn, June 1972. University Microfilms, 313 N. First Street, Ann Arbor, MI 48106, Order No. 72-28,207; 264 pp.

24. A. Ballato, H.L. Bertoni, and T. Tamir, "Systematic Network Approach for Piezoelectrically Driven Crystal Plates and Stacks," Presented at 1972 IEEE Ultrasonics Symposium, Boston, Massachusetts, 4-7 October 1972, paper B-2. Abstract: IEEE Transactions Sonics Ultrason., Vol. SU-20, No. 1, January 1973, p. 43, paper B-2.

25. A. Ballato, "Transmission-Line Analogs for Stacked Piezoelectric Crystal Devices," Technical Report ECOM-4041, U.S. Army Electronics Command, Fort Monmouth, NJ, October 1972, 21 pp., AD 754 545.

26. A. Ballato, "Networks for crossed-field and inline excitation of bulk and surface acoustic waves," Proc. Symp. Optical and Acoustical Micro-Electronics, Polytechnic Institute of New York, April 1974, pp. 599-615.

27. A. Ballato, "Bulk and Surface Acoustic Wave Excitation and Network Representation," Proc. 28th AFCS, May 1974, pp. 270-279.

28. IEEE Standard on Piezoelectricity. Standard 176-1978, The Institute of Electrical and Electronics Engineers, Inc., 345 East 47th Street, New York, NY 10017.

* AFCS: Annual Frequency Control Symposium, U.S. Army Electronics R&D Command, Fort Monmouth, NJ 07703.

29. T. Uno, "A $LiTaO_3$ Monolithic Crystal Filter by Parallel Field Excitation," Trans. Inst. Electronics and Comm. Engrs. Japan, Vol. E61, November 1978, pp. 915-916.

30. E.R. Hatch and A. Ballato, "Lateral-Field Excitation of Quartz Plates," Proc. IEEE Ultrasonics Symposium, October-November 1983, pp. 512-515.

31. Frequency Electronics, Incorporated, "Lateral Field, SC-Cut, Resonator Development," Contract DAAK20-83-C-0418 with ET&D Laboratory, ERADCOM, Ft. Monmouth, NJ 07703. November 1983 ff. FEI, Mitchel Field, NY 11553.

32. 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 AFCS, May-June 1984, pp. 206-224.

33. A. Ballato, "Microwave Resonator Using Lateral Excitation," Invention Disclosure, US Patent and Trademark Office S/N 664, 267, October 1984.

34. W. Soluch, R. Księżopolski, W. Piekarczyк, M. Berkowski, M.A. Goodberlet, and J.F. Vetelino, "Preliminary Results of Measurement of the Ba La Ga_3 O_7 Piezoelectric Crystal," Proc. IEEE Ultrasonics Symposium, November 1984, pp. 517-522.

35. T.R. Meeker, "Theory and Properties of Piezoelectric Resonators and Waves," in Precision Frequency Control, Vol. 1 (E.A. Gerber and A. Ballato, eds.), Academic Press, New York, 1985, Chapter 2, pp. 47-145.

36. R.M. Foster, "A Reactance Theorem," Bell Syst. Tech. J., Vol. 3, April 1924, pp. 259-267.

37. A. Ballato, "Transduction of Acoustic Surface Waves by Interdigital Arrays, and Equivalent Circuit Representations," Technical Report ECOM - 4359, U.S. Army Electronics Command, Fort Monmouth, NJ, October 1975, 50 pp.

38. B. Parzen, Design of Crystal and Other Harmonic Oscillators, Wiley, New York, 1983. Chapter 3, Figure 3.16.

39. A. Ballato, "Doubly Rotated Thickness Mode Plate Vibrators." In Physical Acoustics, vol. XIII, (W.P. Mason and R.N. Thurston, eds.), pp. 115-181. Academic Press, New York, 1977.

40. H.F. Tiersten. Many papers appearing in Proc. AFCS, J. Acoust. Soc. Amer., J. Appl. Phys., etc.

41. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., Vol. 110, June 1958, pp. 1060-1061.

42. M. Nakazawa, H. Ito, A. Usui, A. Ballato, and T. Lukaszek, "New Quartz Resonators with Precision Frequency Linearity over a Wide Temperature Range," Proc. 36th AFCS, June 1982, pp. 290-296.

43. M. Nakazawa, N. Izuka, A. Ballato, and T. Lukaszek, "A Study of Quartz Crystal Temperature Sensors," 12th Special Committee on the Application Technology Investigation for the Precision Frequency, IEEEJ, No. 12-3, 1983, pp. 1-6.

44. M. Nakazawa, H. Yamaguchi, A. Ballato, and T. Lukaszek, "Stress-Compensated Quartz Resonators Having Ultra-Linear Frequency - Temperature Responses," Proc. 38th AFCS, May-June 1984, pp. 240-244.

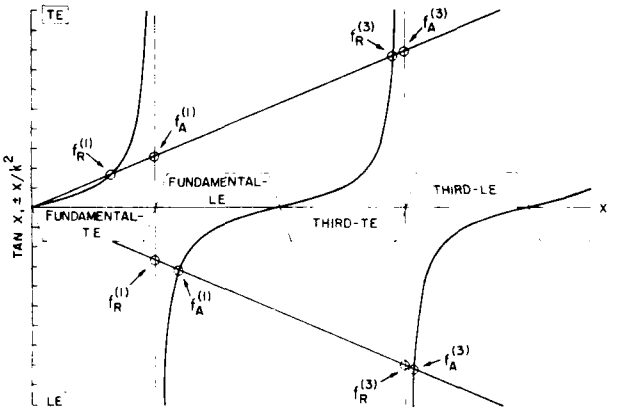
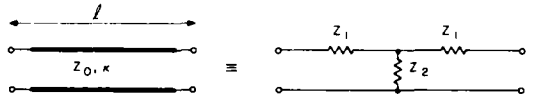


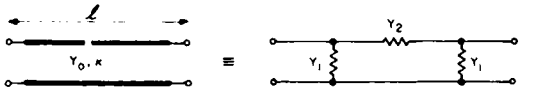
Figure 2. Graphical constructions for determining the resonance (f_R) and antiresonance (f_A) frequencies of TE and LE vibrators.



$$Z_1 = \frac{Z_0}{j \sin\theta}(\cos\theta - 1) = j Z_0 \tan(\theta/2)$$

$$Z_2 = \frac{Z_0}{j \sin\theta}, \quad \theta = \kappa \ell$$

LUMPED, TEE, FORM OF A TRANSMISSION-LINE SECTION.

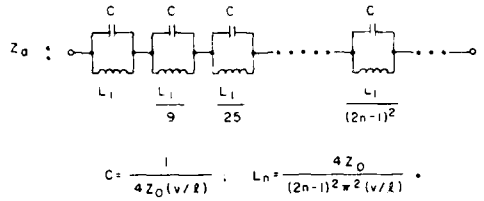


$$Y_1 = \frac{Y_0}{j \sin\theta}(\cos\theta - 1) = j Y_0 \tan(\theta/2)$$

$$Y_2 = \frac{Y_0}{j \sin\theta}, \quad \theta = \kappa \ell$$

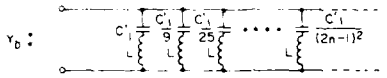
LUMPED, PI, FORM OF A TRANSMISSION-LINE SECTION

Figure 3. Lumped tee and pi forms of a transmission-line section.



$$C = \frac{1}{4Z_0(v/\ell)} \quad , \quad L_n = \frac{4Z_0}{(2n-1)^2\pi^2(v/\ell)} \quad .$$

$$Z_a/Z_o = Y_b/Y_o = +j\tan(\theta/2) \quad ; \quad \theta = \omega\ell/v$$

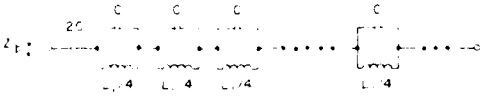


$$L = \frac{1}{4Y_0(v/\ell)} \quad , \quad C_n = \frac{4Y_0}{(2n-1)^2\pi^2(v/\ell)}$$

Figure 4. Foster-form realizations of a tangent function.

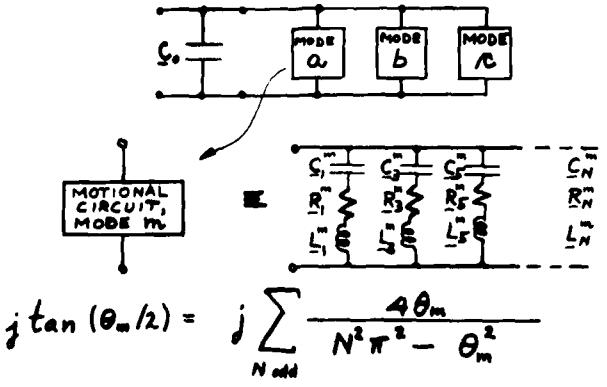
$$L = \frac{1}{4Y_0(v/\ell)} \quad , \quad C_n = \frac{Y_0}{n^2\pi^2(v/\ell)} \quad .$$

$$Y_a/Y_o = Z_b/Z_o = -j\cot(\theta/2) \quad ; \quad \theta = \omega\ell/v$$



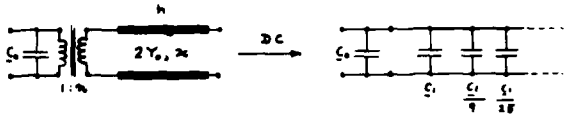
$$C = \frac{1}{4Z_0(v/\ell)} \quad , \quad L_n = \frac{Z_0}{n^2\pi^2(v/\ell)} \quad .$$

Figure 5. Foster-form realizations of a cotangent function.



$$j\tan(\theta_m/2) = j\sum_{N\ \text{odd}} \frac{4\theta_m}{N^2\pi^2 - \theta_m^2}$$

Figure 6. Network for an LE vibrator, and expansion of a tangent function about its poles.



$$Y(s) = C_0 s\left[k^2 \frac{\tanh(sh/v)}{(sh/v)} + 1\right] \quad ; \quad \sum_{n\ \text{odd}} \frac{1}{N^2} = \frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \cdots = \frac{\pi^2}{8}$$

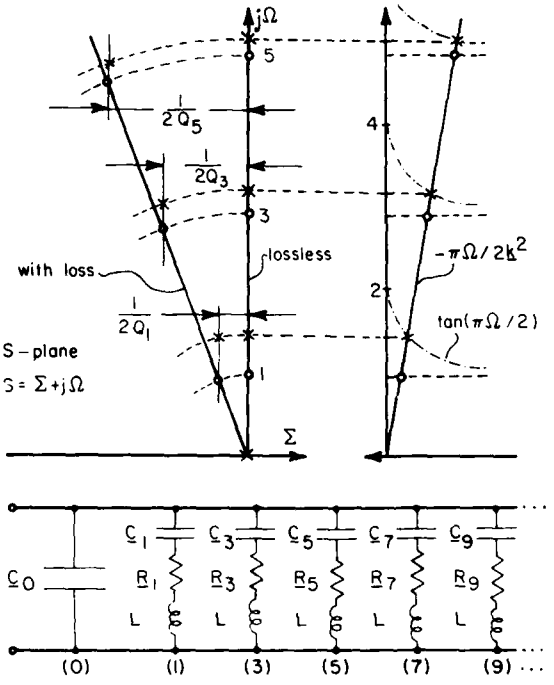
$$s = \sigma + j\omega$$

$$C_{in} = C_0\left[1 + k^2\tan(\theta/k)/(\theta/k)\right] \xrightarrow{DC} C_0(1+k^2) = C_0 + \frac{\pi^2}{8}C_1$$

$$\boxed{C_0/C_1 = \Gamma = \frac{\pi^2}{8}\cdot\frac{1}{k^2} = \frac{C_0}{C_1} \cdot R_1}$$

Figure 7. Behavior of LE network at dc; relation between capacitance ratio and piezocoupling factor.

LATERAL EXCITATION OF A SINGLE THICKNESS MODE



$$C_N = C_1/N^2; \ \tau_1 = R_N C_N = \text{constant};$$

$$R_N^2 Q_N^2 C_N/L = 1; \ (2\pi f_R^{(N)})^2 L C_N = 1; \ \pi^2 C_1 = 8k^2 C_0$$

Figure 8. Complex frequency representation of LE vibrator impedance for lossless and lossy cases; complete lumped equivalent network for LFE.

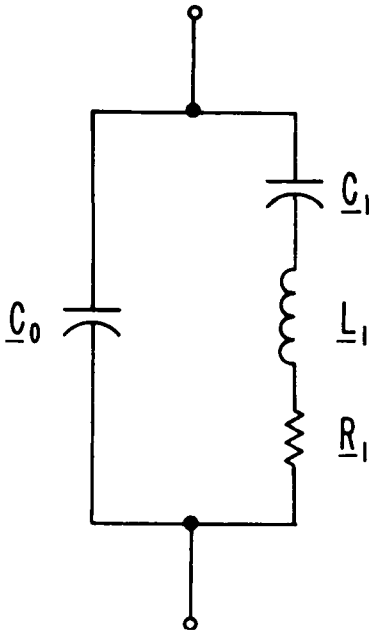


Figure 9. Butterworth-Van Dyke equivalent circuit valid in the vicinity of a single resonance.

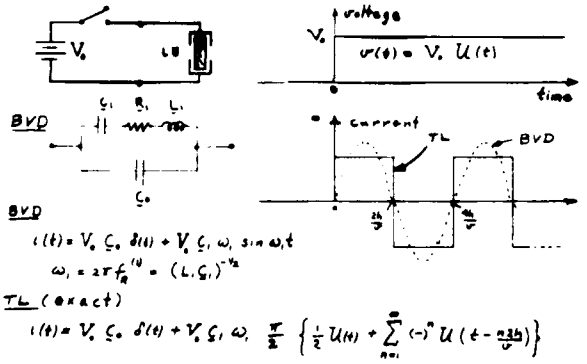


Figure 10. Time-domain behavior of the approximate BVD circuit, and of the exact TL network.

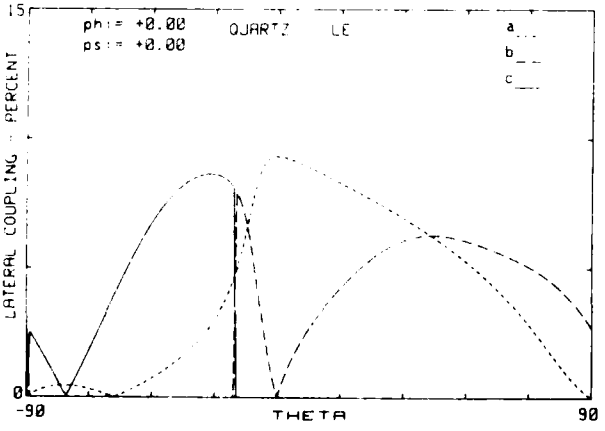


Figure 11. Piezocoupling for rotated Y cuts with LE field along the X axis.

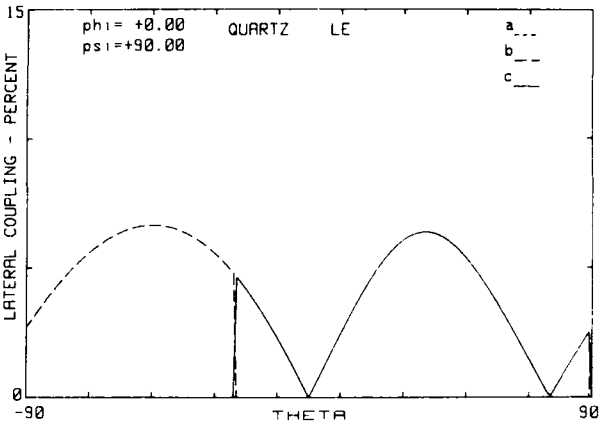


Figure 12. Piezocoupling for rotated Y cuts with LE field along the Z-prime axis.

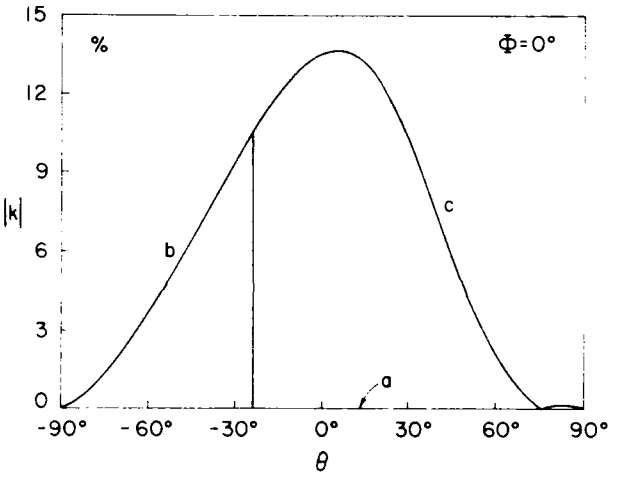


Figure 13. Piezocoupling for rotated Y cuts with TE field.

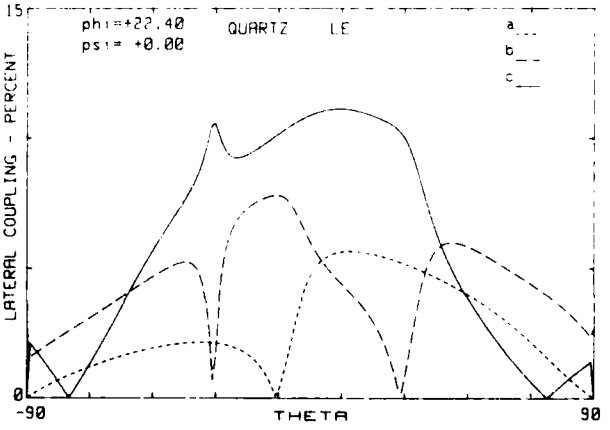


Figure 14. Piezocoupling for doubly rotated cuts having phi = 22.4° and LE field along X.

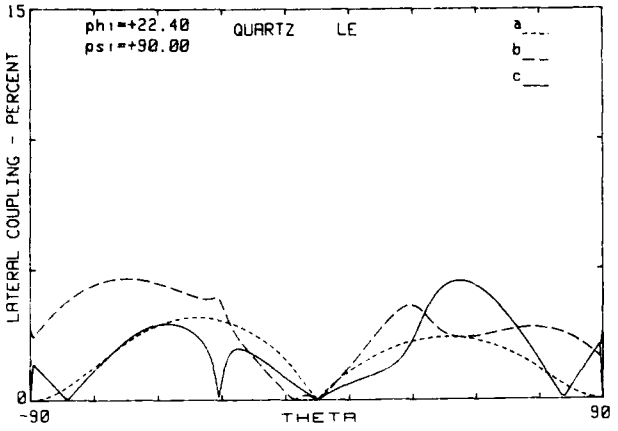


Figure 15. Piezocoupling for doubly rotated cuts having phi = 22.4° and LE field along the rotated Z axis.

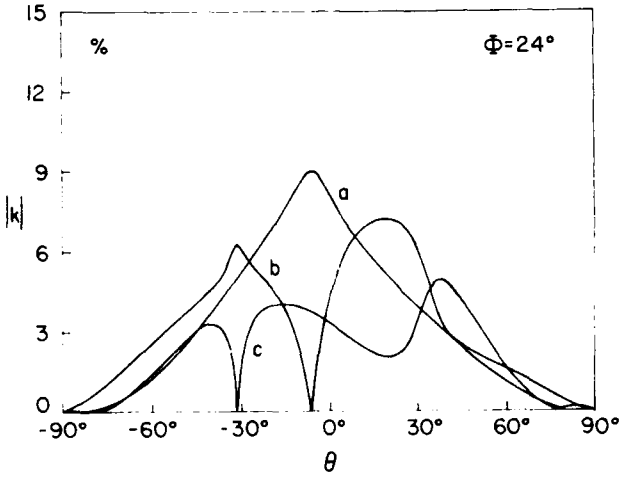


Figure 16. Piezocoupling for doubly rotated cuts having phi = 24° and TE field.

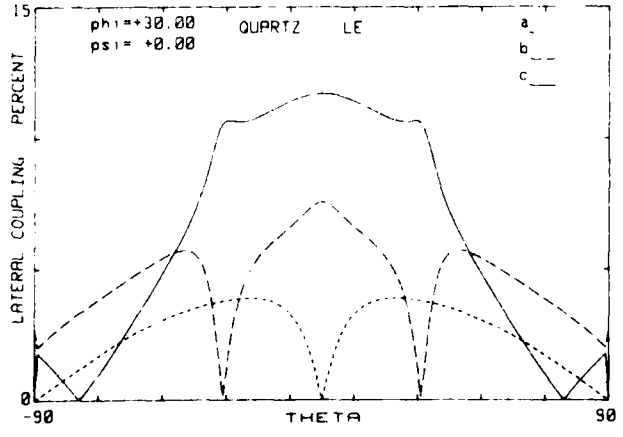


Figure 17. Piezocoupling for rotated X cuts with LE field along the Y axis.

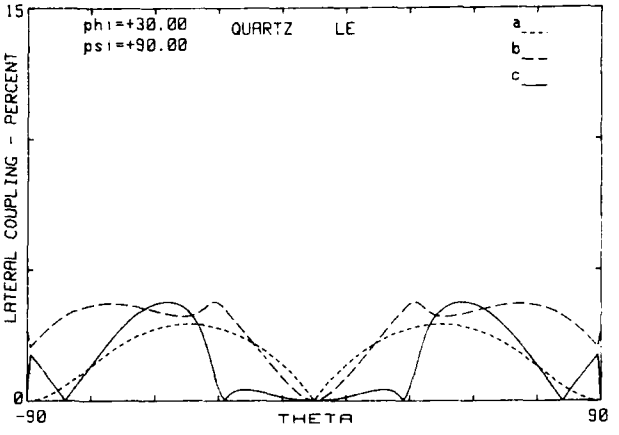


Figure 18. Piezocoupling for rotated X cuts with LE field along the Z-prime axis.

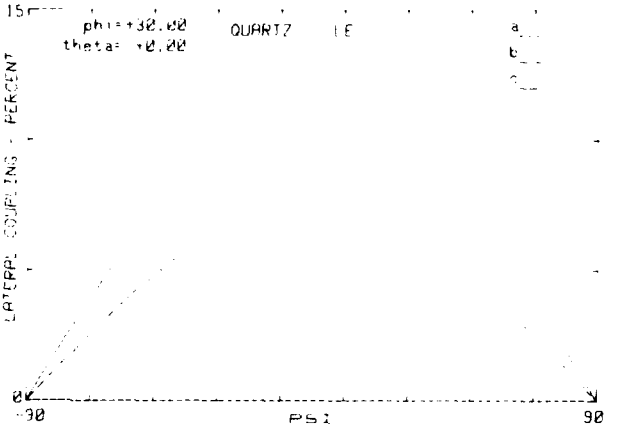


Figure 19. Piezocoupling versus psi for the X cut.

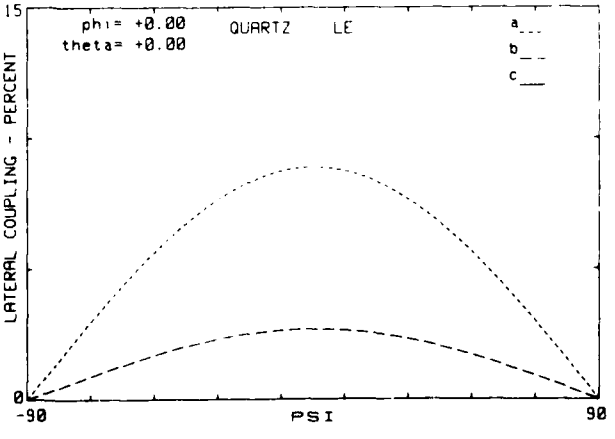


Figure 20. Piezocoupling versus psi for the Y cut.

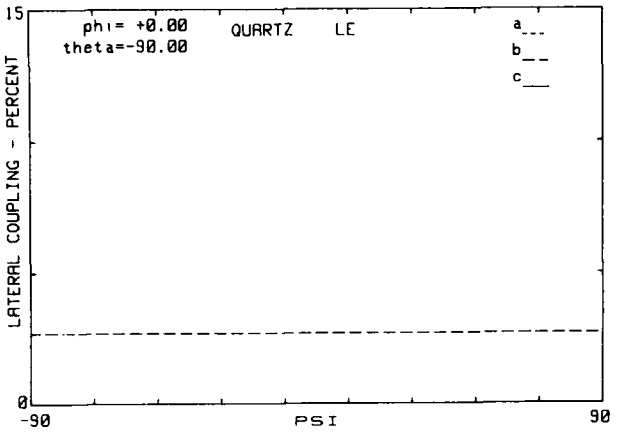


Figure 21. Piezocoupling versus psi for the Z cut.

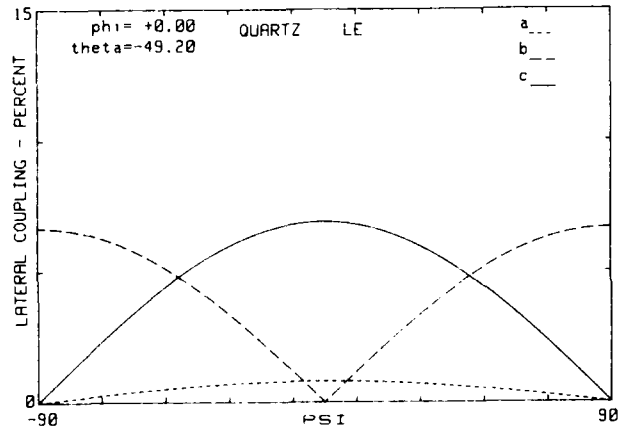


Figure 22. Piezocoupling versus psi for the BT cut.

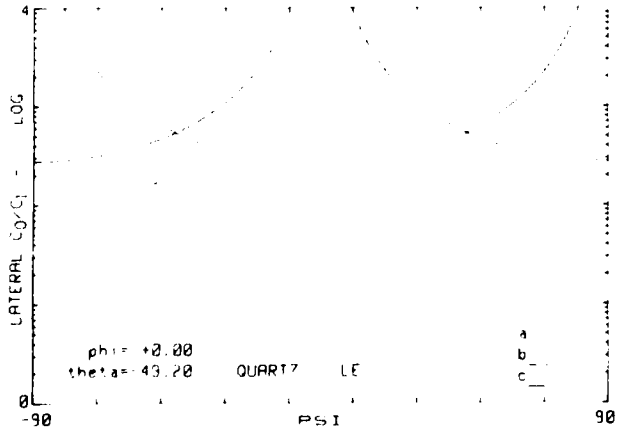


Figure 23. Capacitance ratio versus psi for the BT cut.

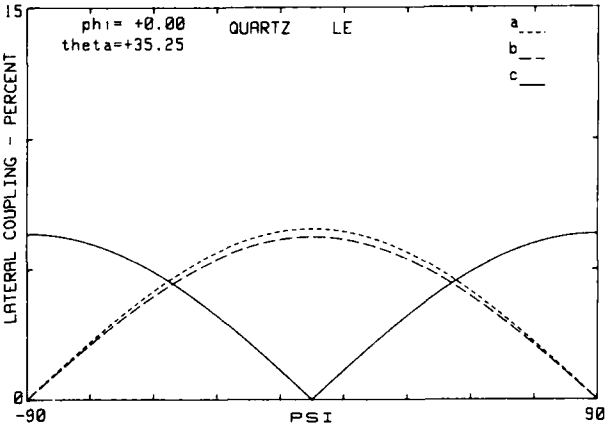


Figure 24. Piezocoupling versus psi for the AT cut.

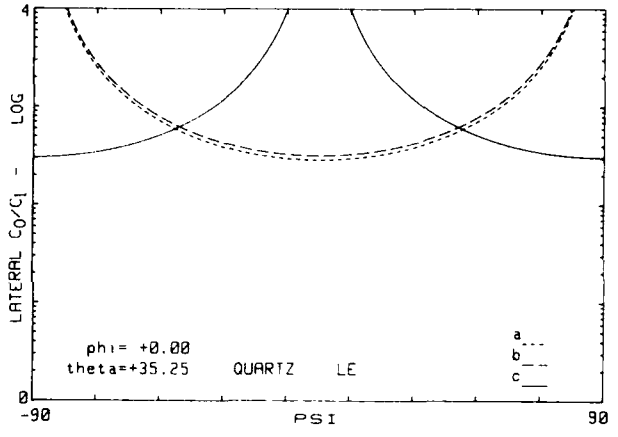


Figure 25. Capacitance ratio versus psi for the AT cut.

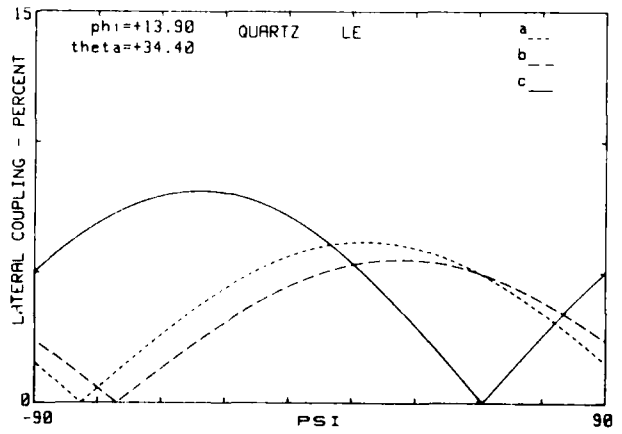


Figure 26. Piezocoupling versus psi for ϕ = 13.9°, θ = +34.4° cut.

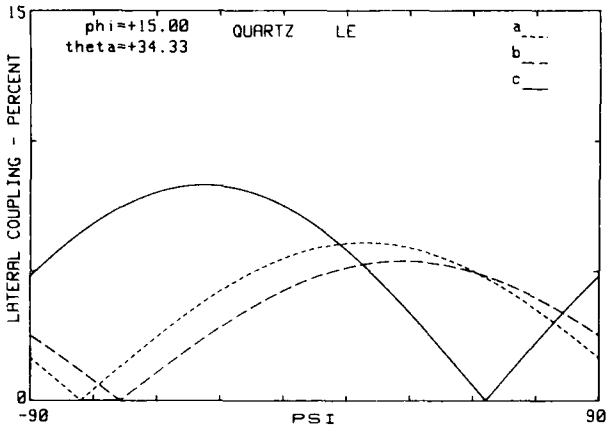


Figure 27. Piezocoupling versus psi for the FC cut.

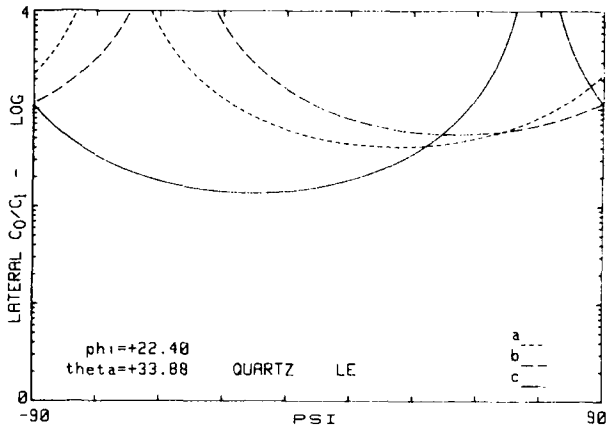


Figure 30. Capacitance ratio versus psi for the SC cut.

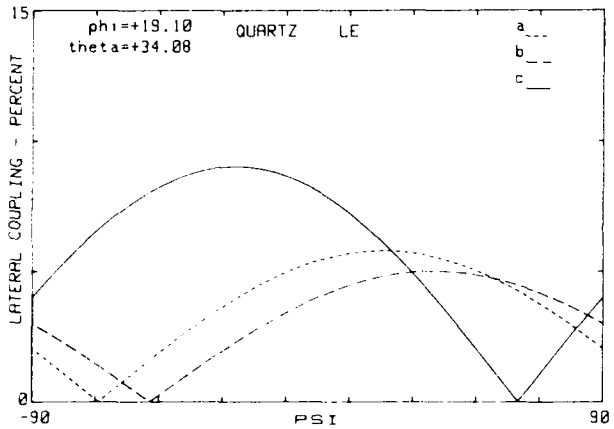


Figure 28. Piezocoupling versus psi for the IT cut.

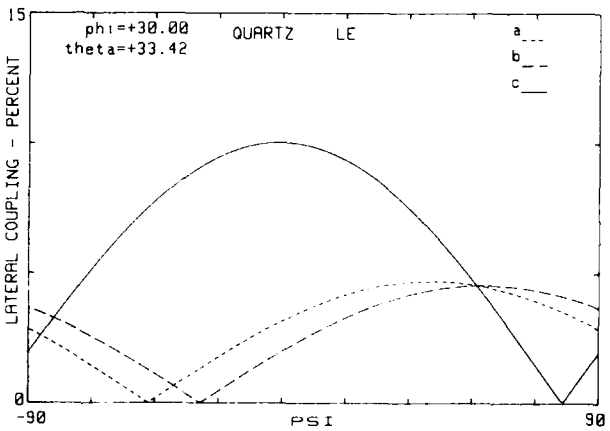


Figure 31. Piezocoupling versus psi for the ZTC rotated X cut.

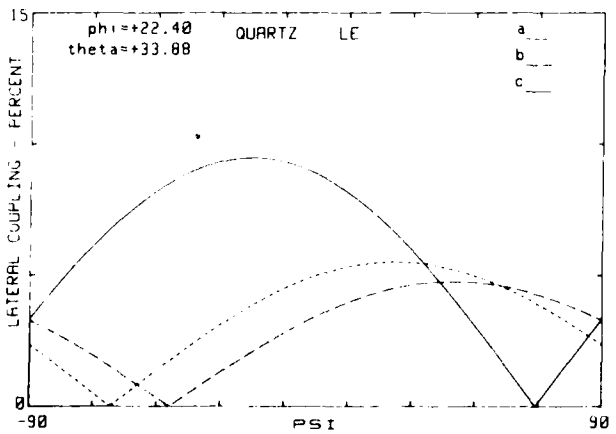


Figure 29. Piezocoupling versus psi for the SC cut.

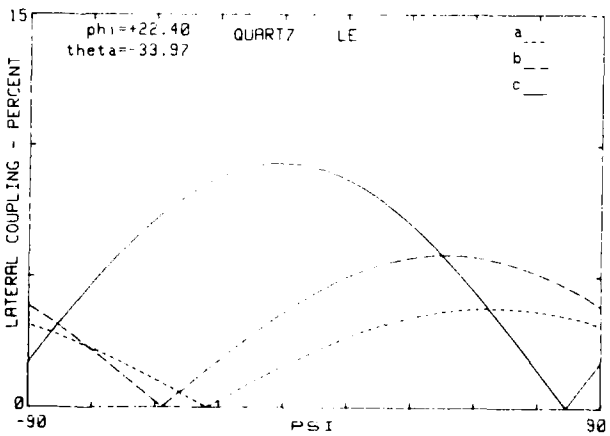


Figure 32. Piezocoupling versus psi for the "minus SC" cut.

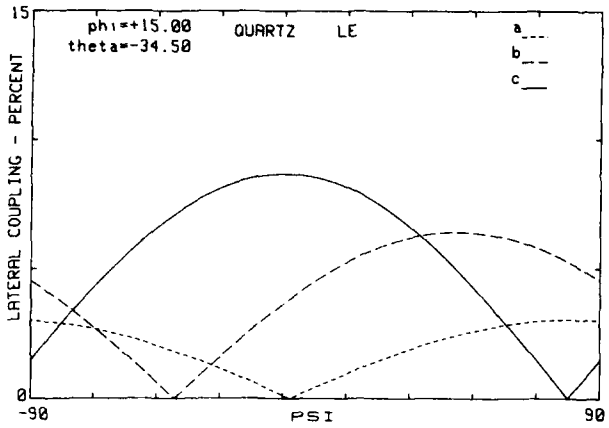


Figure 33. Piezocoupling versus psi for the RT cut.

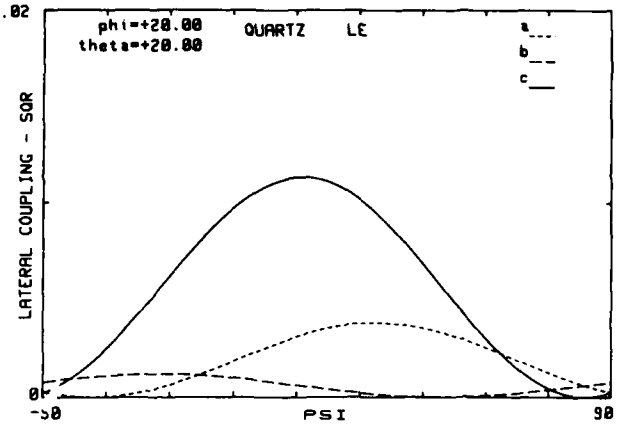


Figure 36. Square of the LE piezocoupling for the NLSC cut. Scale is twice that of Fig. 35.

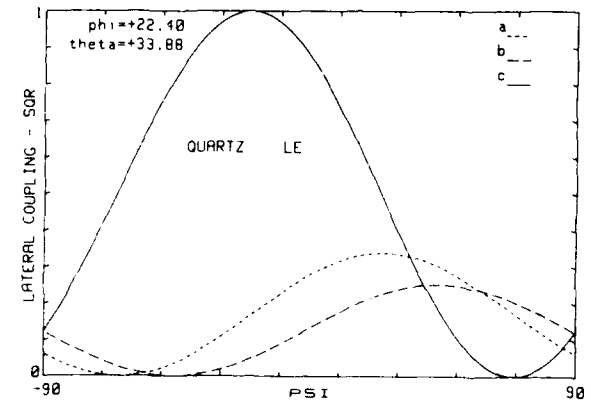


Figure 34. Normalized square of the LE piezocoupling for the SC cut.

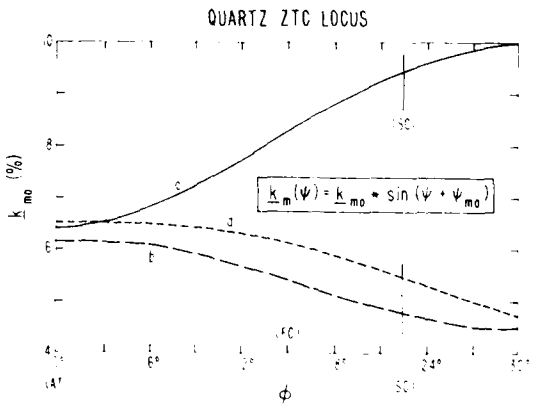


Figure 37. Modulus of piezocoupling versus phi on the upper ZTC locus.

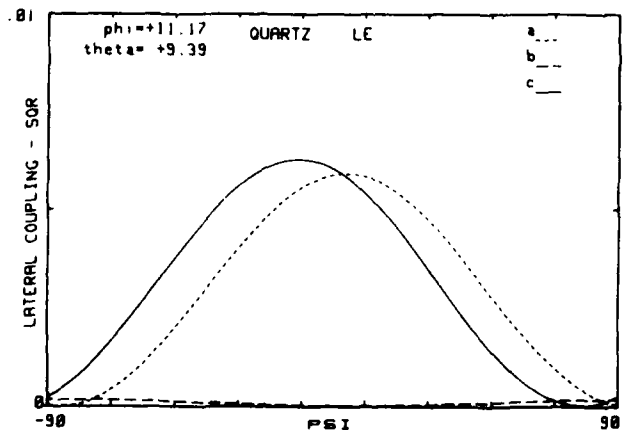


Figure 35. Square of the LE piezocoupling for the LC cut.

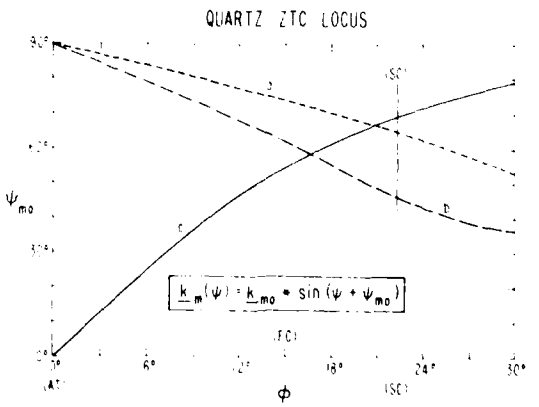


Figure 38. Angle of piezocoupling versus phi on the upper ZTC locus.

LATERAL FIELD RESONATORS

Dr. A. W. Warner, Jr. and B. Goldfrank

Frequency Electronics, Inc.
Mitchel Field, New York 11553

The SC-cut quartz crystal resonator is a doubly rotated Y-cut which has the very valuable characteristic that its frequency is stable while under stress, either from temperature gradients or from shock and vibration. Since it is doubly rotated (i.e., it is first rotated about the Z-axis and then tilted about the X'-axis) the crystal plate has a component of its electrical field in the X- as well as the Y-direction, if it is excited by normal electrodes on the top and bottom major surfaces.

The result is that an unwanted B mode (fast shear mode) is strongly coupled, which in frequency is only 9% above the wanted C mode and usually slightly lower in resistance. It has been necessary to provide a tuned trap in the oscillator circuit to prevent interference from the B mode, with a consequent effect on circuit stability.

Because of the threefold symmetry of quartz, and since the SC-cut is rotated 22 degrees about the Z-axis, there is a Y-axis only 8 degrees from the surface of an SC-cut plate. It has been reasoned that an electrical field along or near this axis would strongly excite the C mode and very weakly excite the B mode, if at all. Such a field can be provided by a special arrangement of electrodes. The crystal units are now referred to as lateral field resonators (LFR).

This need to deal with the B mode would be reason enough to again try the use of a lateral field. The presence of this Y-axis and the relatively high impedance of a normal SC-cut indicates that it may now be advantageous to use overtone mode LFR's. Add to this the infinite variety of field directions and shapes, which was not possible before, to tailor crystal parameters to specific applications. Furthermore, the old reason for the parallel field of 20 years ago,[1] removal of metal from the active area of the quartz plate, is still valid, and actually reduces the amount of frequency adjustment necessary. All of the reasons above make it imperative that we employ some form of lateral field for SC-cut crystal units. The lateral field is produced by putting two electrodes, instead of one, on the plate surface and the electrical field is produced by the gap between the electrodes.

When FEI started on this lateral field project in March of 1982, the effects of contour, overtone, and frequency on the crystal parameters using various electrical field arrangements were virtually unknown. For example, no one could say how inductance varied with overtone. We had to establish the parameters for LFR's by experiment.

Over 100 lateral field crystals have been constructed and measured. The first experiments quickly established that the B mode could indeed be suppressed to any desired degree, even to zero coupling. Many experiments later, the mystery of the crystal parameters began to dissipate. It appears that, to a first degree, inductance L does not change with overtone and furthermore that it is a linear function of blank thickness. Compare this with the normal field case where the ratio of capacitances C_o/C_1 varies with the square of the overtone N, and is independent of frequency. The comparison is shown in Figure 1. Constant thickness is the case where a blank is operated on different overtones. For constant frequency, different thickness blanks are necessary for each overtone.

By making a number of assumptions, such as that the Q is at its maximum and therefore varies with the inverse of frequency, that electrode size is constant, and that inductance equals 17 times the blank thickness in millimeters, a chart, Figure 2, was constructed to visualize the character of the two methods of excitation. One can see now why large, low frequency crystal units below 1 megahertz worked so well in the old parallel field AT units of 20 years ago.[2] Also, it can be shown that the useful frequency range for LFR's is probably below 10 MHz. At frequencies below 5 MHz there is a crossover point where lateral field units are lower in resistance than normal field units.

One design, 5 MHz, 3rd overtone has received most of the attention up to this time. Figure 3 shows a bar graph of responses for this crystal using a simple straight gap in the electrodes. It can be seen that, while the B mode is weak, the satellite modes of the wanted C mode are strongly coupled and in some instances the C_{311} or C_{313} mode was actually lower in resistance than the main C_{311} mode.

By shaping the field the results of Figure 4 were produced. Now the C_{311} mode is clearly dominant. By further modification of the field even better units have been produced. Figure 5 indicates no unwanted responses less than 10,000 ohms. Such a clean spectrum should, of course, be useful in filter designs and, perhaps, will even make possible oscillators with a cleaner output.

Figure 6 shows the characteristics of a group of crystal units, to show the fairly uniform high Q that is possible. Except for crystal #6, Q ranges from 2.6 to 2.9 million and inductance from 27 to 28 henries. The resistance is about 300 ohms, similar to a 5th overtone normal field resonator. Figure 7 shows frequency and resistance vs. temperature for one of the better designs. At 75°C the unit has the flat curve typical of the SC cut.

With the almost infinite variety of designs possible using lateral fields, it would seem that the future is still bright for crystal unit designers.

This work was supported by the U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey, under Contract No. DAAK20-83-C-0418.

References

1. Warner, A. W., "Use of Parallel Field Excitation in the Design of Quartz Crystal Units," Proc. 17th Ann. Symp. Freq. Control, 1963, pp. 248-266.

2. Ianouchevsky, W., "High Q Crystal Units," Proc. 17th AFCS, 1963, pp. 233-247.

CH2186-0/85/0000-0473$1.00©1985IEEE

CRYSTAL TYPE	CONSTANT THICKNESS	CONSTANT FREQUENCY
NORMAL FIELD	$C_0/C_1 \sim N^2$	$L_1 \sim N^3$
LATERAL FIELD	$L_1 \sim N^0$	$L_1 \sim N$

Figure 1. Inductance (L_1) as a Function of Overtone (N)

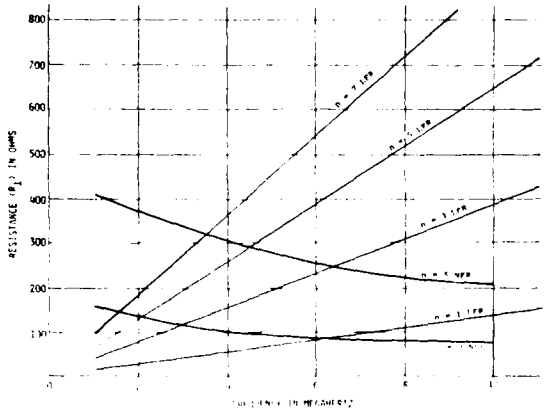


Figure 2. Resistance at Maximum Q for Various Overtones of Lateral Field Resonators and Normal Field Resonators

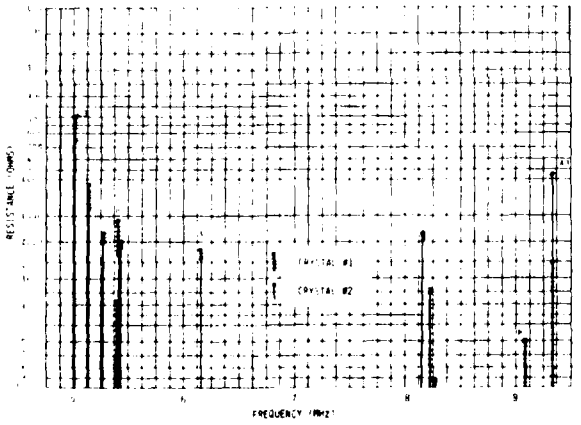


Figure 3. 5 MHz, Third Overtone Original Design

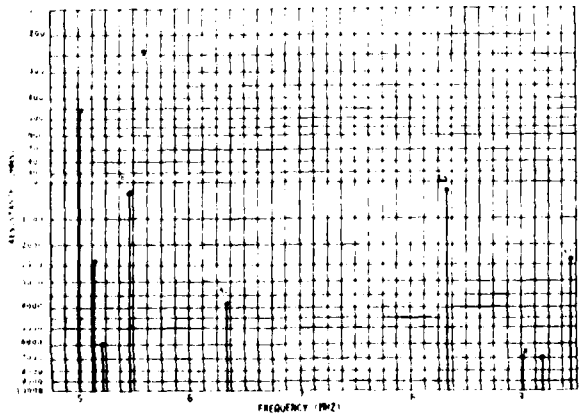


Figure 3. 5 MHz, Third Overtone Second Design

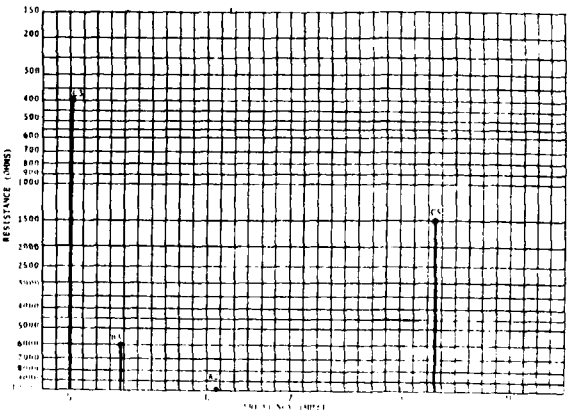


Figure 5. 5 MHz, Third Overtone Third Design

LFR DATA SHEET

SER NO.	Δf (10pf)	C_1 x10^{17}	L_1	Q x10^{-6}	R_1				NOTES
1	8.4	3.7	27.4	2.9	300				
2	8.5	3.7	27.0	2.9	290				
3	8.5	3.7	27.0	2.7	310				
4	8.2	3.6	28.0	2.7	330				
5	8.5	3.7	27.0	2.8	300				
6	8.5	3.7	27.0	0.7	1200				
7	8.3	3.6	27.7	2.6	330				

FREQUENCY	O.T.	θ ANGLE	PLATING THICKNESS	FIELD ORIENTATION	CONTOUR	GAP WIDTH	GAP SHAPE
5.007 MHz	3	34°	2500	110	4-1/8d	75	3

Figure 6. Typical Data from One Lot of Crystals

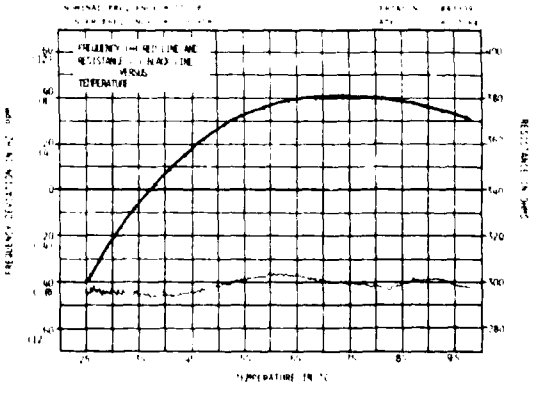


Figure 7. Crystal Slew

APPLICATIONS OF TOTAL PROCESS CONTROL TECHNIQUES IN THE PRODUCTION OF HIGH PRECISION QUARTZ RESONATORS

John A. Kusters and Charles A. Adams

Hewlett Packard Company
Santa Clara, CA 95051

Summary

Statistical quality control is a powerful tool popularized by its success in Japan over the last two decades. However, when combined with further analytic and decision-making techniques, it forms the basis for Total Process Control (TPC). The additional elements include understanding customer requirements, total workforce participation, and heavy use of control charting. The combination form the basis for a formalized program in TPC developed and used at Hewlett Packard for the past two years.

Introduction

Over the last forty years, Japanese manufacturing has risen from virtually total destruction to become a major force in the international market. Originally, products manufactured in Japan meant inferior materials and poor workmanship. Today, "Made in Japan" brings to mind quality, productivity, and good value.

Much has been written about the growth of manufacturing in Japan, the progress shown, and the implementation of modern manufacturing techniques [1-3]. Most of the successes are due to theory originally developed, but ignored, in the United States.

Although heavy use of automation has benefited some Japanese manufacturers, this has occurred only in the largest companies, and even there, only in selected products. The majority of manufacturing plants in Japan can be considered crude by American standards, with old, surplus, or locally developed equipment, and no automation or computer control. Even in these companies, the drive is for outstanding quality and productivity improvement.

Many of the names commonly associated with Japanese quality are American, such as Deming [4] and Juran [5]. The question becomes, if the Japanese can do it with American know-how, why can't we?

American companies have experimented with different components of quality and productivity programs used in Japan, but with mixed or limited success. In some companies and industries, even the mention of "Quality Circles" leads to disparaging remarks.

In our experience, all of these limited experiments were probably doomed from the beginning because they did not attack the whole problem of quality and productivity, but only concentrated on whatever was of interest at that time.

Designing of a total process improvement program which uses the best of theories and practices from a variety of sources, then adapting these to an American work force has proven especially beneficial in our crystal and oscillator production areas. In particular, ideas originally presented in Deming [4], Juran [5], Kepner-Tregoe [6], Ishikawa [7], the work done at Western Electric, now AT&T Technology [8], and others, can be combined into a single cohesive program dedicated to quality at all levels, continuous process improvement, and continued productivity gains. When properly implemented, these can have a tremendous leveraging effect on cost and profit.

The Concept

The basic philosophy of TPC can be stated in four short sentences. These are:

1. Focus on customer needs and expectations.
2. Total commitment to quality.
3. Universal participation and teamwork.
4. Adopt a continuous process improvement methodology.

1. Customer Needs and Expectations

In TPC, the customer is most likely not the individual or company buying the end product. A typical production area is composed of many different people and processes. Each receives raw material, sub-assemblies, or components from somewhere else. Each are customers for previous steps in the process, and vendors for the next steps. In other words, we are all somebody's customer and somebody's vendor.

In particular, process and production engineers have a variety of customers, each with his own needs and expectations. In most cases, these customers are the people on the production line. To properly serve these customers, the engineer must fully understand not only what the boss wants, but what the production people need and expect from engineering. This cannot be learned from books, reading reports, attending meetings, or sitting at a desk. The primary key to our success in implementing TPC was a move made several years ago, when all of the engineers were directed to spend at least half of their time physically in the production area.

Since these engineers had no direct assignment or project which kept them occupied while on the production floor, they learned to communicate with the production people, eventually developing a strong rapport with them. Out of this grew an understanding of all of the petty problems that impact production.

CH2186-0/85/0000-0475$1.00©1985IEEE

2. Commitment to quality

The definition of quality appropriate to a given area or product is a management decision. Quality can be defined in terms of conformance to requirements [9]. In this case, if a product does not conform to specifications, defined by management and/or engineering, there is no authority to ship.

Quality can also be defined in terms of fitness for use. Here, management must be fully aware of the end customer's needs and expectations, and define product quality in a manner that fulfills these needs and expectations.

Once the level of quality is defined, quality becomes a matter of doing the right thing, and doing it right the first time, every time.

3. Universal Participation

Our first experiments with universal participation involved Quality Circles. This format, which is by definition voluntary, unstructured, and led by a member of the circle, proved unsuitable for a high-technology production area.

Further experimentation led eventually to action teams composed of all of the production people in the production area, the engineers supporting that area, and the management team that directs that area, functioning as a problem solving team. The problems are defined by management. The team always includes all of the necessary resources to define the problem, determine the appropriate solution, and implement that solution.

It is in the action teams that the good communication and rapport developed between the line people and engineering proved to be so valuable. The level of trust and concern shown by both broke down any barriers that might have hindered potential process improvements. The usual finger pointing was avoided.

4. Continuous Process Improvement Methodology

The above define components of a TPC program which we found to be necessary but not sufficient conditions upon which to base a system of total process control. Figure 1 shows a flow chart of the process we use.

The first step is to determine what the project is to be, i.e., what process will be improved? Knowledge of the next customer's needs and expectations are critical at this point. What needs improvement? How serious is the problem? Who sets the priorities?

The next step is to determine the current, actual flow diagram of the process. In our experience, no matter how well management and engineering thought they understood the process, the initial flow chart had to be significantly modified after checking with the people who actually do the work. Full knowledge of the real-world process alone can sometimes show where significant improvement might be made.

Next, in order to measure the degree of success that the project has, we need some form of measurement of success, a reduction in labor hours, increased yield, reduction in scrap and rework, etc. This is the process performance measure (PPM). It is primarily a statement of the end result desired and the method by which we know we've achieved the desired end result.

At this point, brainstorming is appropriate to try to determine all of the possible problems which might be causing the results we currently have. Once these are determined, a cause-effect diagram (also known as a fish-bone or Ishikawa [7] diagram) can be drawn. Although at first glance this appears to be a trivial exercise, because as good engineers we know all of the answers, the insights from everybody involved in the action team must be taken into account. It is amazing how much the production people really know about the process. Ignoring this valuable resource may lead to solving the wrong problem, or wasting time that might be better spent to solve yet another problem.

Now that the list of possible causes has been determined and the inter-relationships of the various causes, it is time to gather data. A fact without data is just another opinion. This usually becomes the most time consuming part of the process improvement methodology. Again, although we may think we know where the problem lies, in almost every case we've examined, the original cause postulated turned out to differ significantly from the final cause(s) determined through proper analysis of the data gathered. Since the production people are a part of the problem solving team, an appropriate action is to have the data taken by them as part of their normal work function. The challenge to management and engineering is to make the data taking as efficient and meaningful as possible. A few good pieces of data are worth far more than a filing cabinet full of data no one ever looks at.

Data analysis becomes the next step in the process. Many analytic tools can be used. Most are far too complicated for production people to master successfully. However, there are some that are relatively easy to understand and still quite powerful. The Pareto chart identifies major causes of a problem and prevents wasting time on trivial causes. The histogram plots measurement distribution which may give clues about process stability. Scattergrams, where measured values of one parameter is plotted against the measured values of another can identify functional relationships between various parameters. Time-line plots show cycles and trends in the production process as a function of time. A good reference for data gathering techniques and simple analytic methods is Ishikawa [7].

Control charts [7,8] give information as to whether a particular production process is in statistical control. By definition, a process in statistical control maximizes the productivity of that particular process. This does not mean that the process is giving desirable results. It indicates that the process will give consistent results, time after time. To improve results obtained or to meet specifications requires engineering/management intervention to change the process. Usually however, a control chart at this point will indicate a process that is totally out of control.

The combination of the various analytic tools shown usually result in a determination of the major cause(s) of the problem. Once the cause is identified, we enter the "Deming" loop shown at the bottom of Figure 1. The four steps are:

a. Plan the process improvement.

b. Implement the plan.

c. Monitor the results through control charting.

d. If successful, document the solution, and institutionalize the improvement. If not, go back to step a.

The above is usually sufficient to solve most production problems. It has also the outstanding attribute of involving everyone in the definition and solution of the problem, 100% buy-in by everyone.

The Results

Experience gained in many areas in manufacturing, accounting, data processing, and marketing indicate that the type of improvement seen usually is not just a 10-20% improvement, but usually improvements by factors of two, three, or more.

As an example, Figure 2 shows the results seen during the past year and a half on repair times for the 2813B quartz pressure gauge. At the start of the project, we determined that the major external customer complaint was the very long turn-around time on probe repairs. Average repair time for years had been about three months, with a maximum time near nine months. At this point, we set a goal of one month in-factory repair time. The solutions required restructuring the entire repair function. The current results are an average repair time of 2.1 days, with a maximum of 9 days. In the process, the manufacturing repair cost was also reduced significantly, resulting in additional profit.

Figure 3 shows a similar experience with oscillator factory repairs. This project was started about 8 months ago. Results currently show a reduction in average repair time from 100 days to 15 days, a greater than six-fold reduction in repair time with no change in labor requirements or factory repair costs.

Of more importance to this paper are the results shown in the manufacturing of high precision, 10 MHz, 3rd overtone, SC-cut quartz crystal resonators.

At the start of the project, crystal yields out of X-ray angle correct were at 99%. As a result, no effort was made to improve this.

Initial yield during surface contour and polish was 75%. After analysis, the primary causes of failure were identified as edge chips and edge fractures. The holding buttons were re-designed to provide better protection during beveling. The beveling process was also changed to give more consistent results. All other processes in the area were also reviewed and rewritten for more clarity where necessary. Current yield in this area is 95%.

Clean room operations provided the major source of yield loss at the start of the project. Average yield ran at 57%. Significant losses occurred during edge plating, base plating, and frequency plating, usually resulting in poor mechanical brazing to the header and excessive resistance losses. Involving all of the production people in the problem permitted a thorough review of all production processes and identified those procedures which were either poorly written or otherwise unclear to the production people. Involvement also created a new expectancy level regarding quality among the production people. Equipment used was also reviewed, complaints noted, and where justified and necessary, equipment changes and modifications were made. Crystal yields at various points in the process were charted and posted on a daily basis, making all problems visible to everyone. Current yield is 90%.

At project start, test yields were running at 82%. Current yield is 89%. Although no process changes were made at test, control charting was instituted to track total process stability. Of interest are the control charts on measured series resistance shown in Figure 4, and the charts on 1st order temperature coefficient shown in Figure 5. The series resistance charts are particularly sensitive to clean room processing problems and material problems. Much the same sensitivity can be seen control charting motional capacitance. The control charts on 1st order temperature coefficient are a direct measure of angle correction accuracy.

The major project started 3 years ago. During this time, the overall yield has increased from 35% to 75%. This was the primary goal of the project. All of this was accomplished without heavy investment in equipment or automation. There were no "home runs" hit during the project. Every gain was another small step, a minor error corrected, a little more attention to detail, up-dating of documentation, a new design on a production fixture, etc. The project is still continuing, looking at any further gains that might be made.

Of even more importance are the side effects. Several months after project start, normal production fire-fighting was no longer required. This has continued with no major problems and no production stoppage for the last two years.

During the same period, with the productivity improvements seen, the factory cost today of this crystal is less than 40% of that at project start. This is due to yield improvements and individual productivity improvements which more than offset the rising cost of materials and labor.

Other unexpected improvements were an improvement in acceleration sensitivity and oscillator aging.

Conclusions

Proper implementation of a total process control methodology, utilizing techniques from a variety of sources, can produce multi-fold improvements in quality and productivity with minimal investment in capital equipment improvements.

Acknowledgment

Much of the material and concepts presented here were originally developed by Jack Herris and others of the Santa Clara Division Quality Assurance group.

The authors wish to acknowledge the efforts of the Crystal Products Engineering group, Jim Collin, Doug Dull, and Gwen Kaitz, and all of the production people in the crystal area, whose help, interest, and concern made the project possible.

References

[1] Schonberger, Japanese Manufacturing Techniques, New York: Free Press, 1982.

[2] Kantor, The Change Masters, New York: Simon and Schuster, 1983.

[3] Gruska, Theory D - Deming, Dearborn Heights, MI: Multiface Publishing, 1981.

[4] Deming, Quality, Productivity and Competitive Position, Cambridge Massachusetts: MIT Center for Advanced Engineering Study, 1982

[5] Juran, Quality Control Handbook, New York: McGraw-Hill, 1974.

[6] Kepner & Tregoe, The New Rational Manager, New York: McGraw-Hill, 1981.

[7] Ishikawa, Guide to Quality Control, Tokyo: Asian Productivity Organization, 1982.

[8] Statistical Quality Control Handbook, Indianapolis: Western Electric Co., 1958.

[9] Crosby, Quality is Free, New York: New American Library, 1980.

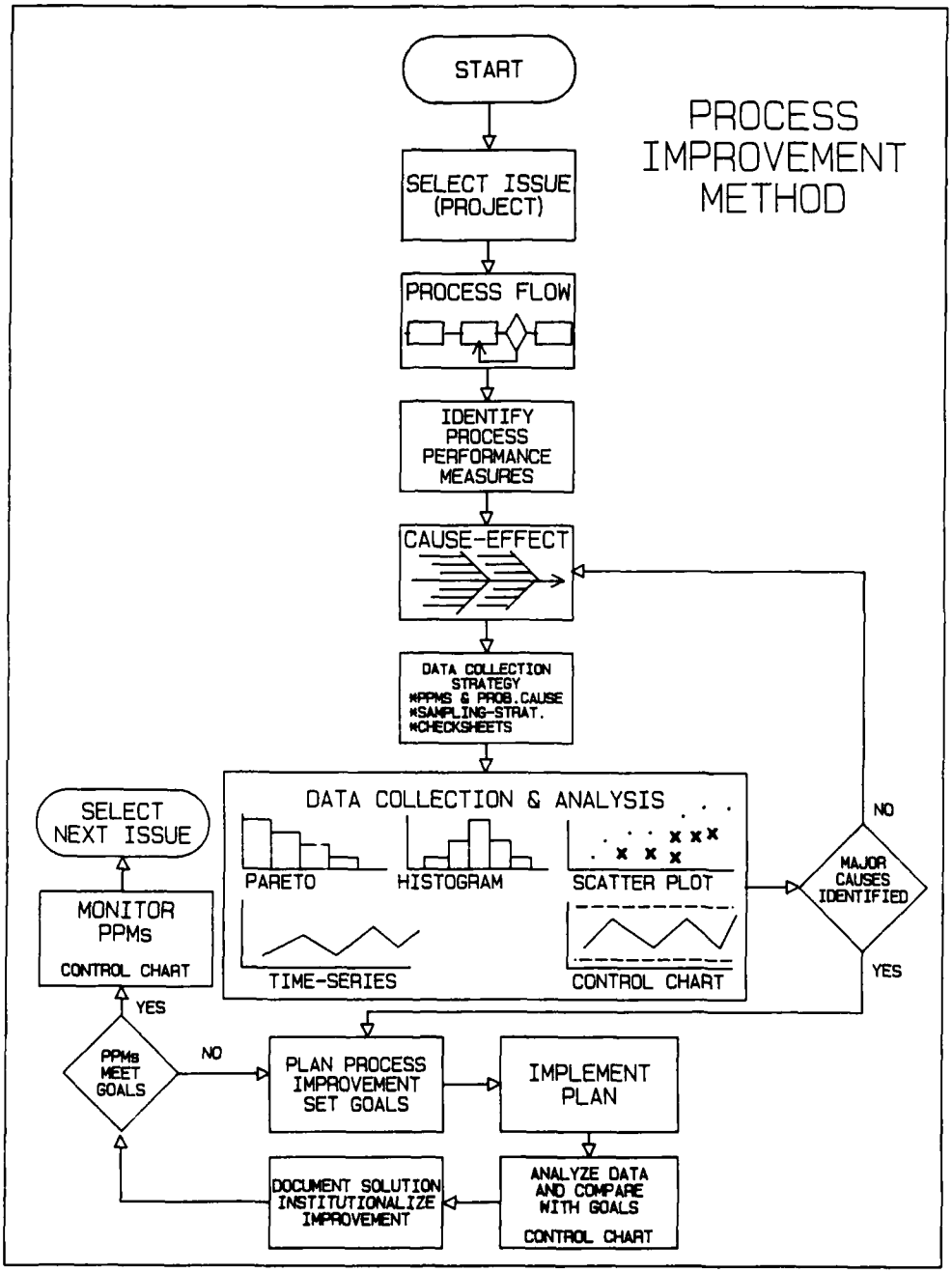


FIGURE 1. PROCESS IMPROVEMENT METHOD

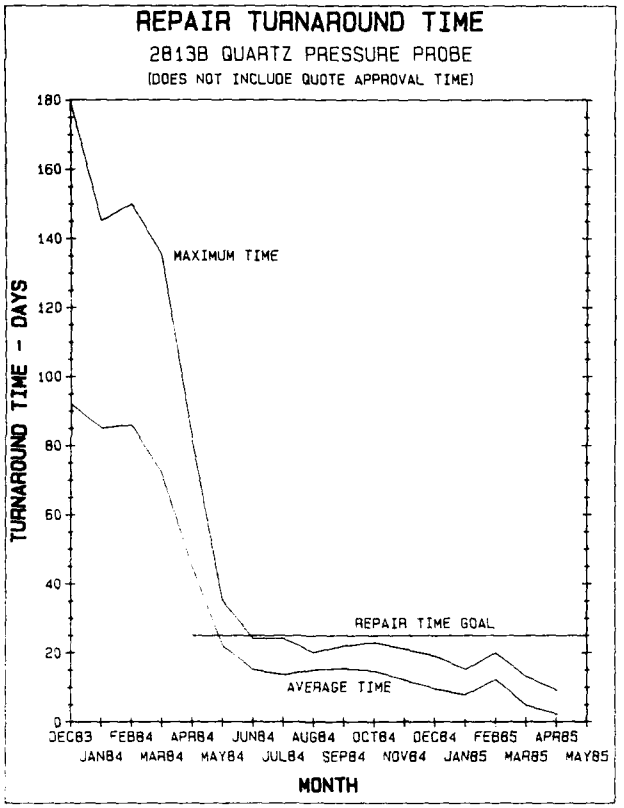


FIGURE 2.

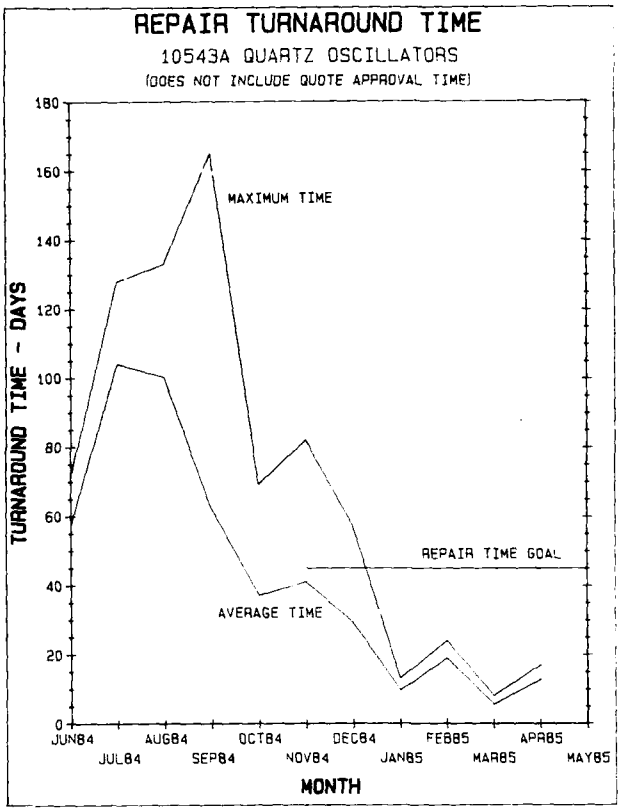


FIGURE 3.

CONTROL CHART - SERIES IMPEDANCE

SC-cut CRYSTALS - SPECIFICATION - 40 TO 70 OHMS

4/18/85 THROUGH 5/3/85

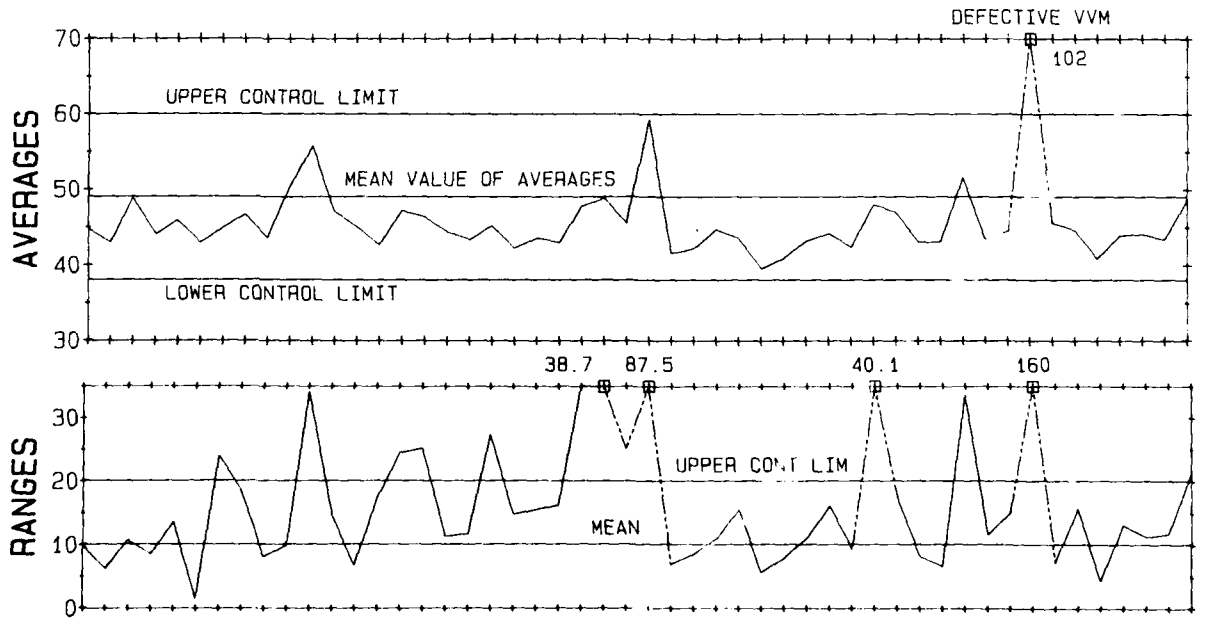


FIGURE 4.
CONTROL CHART SHOWING EFFECTS OF EQUIPMENT MALFUNCTION AND MATERIAL VARIATIONS

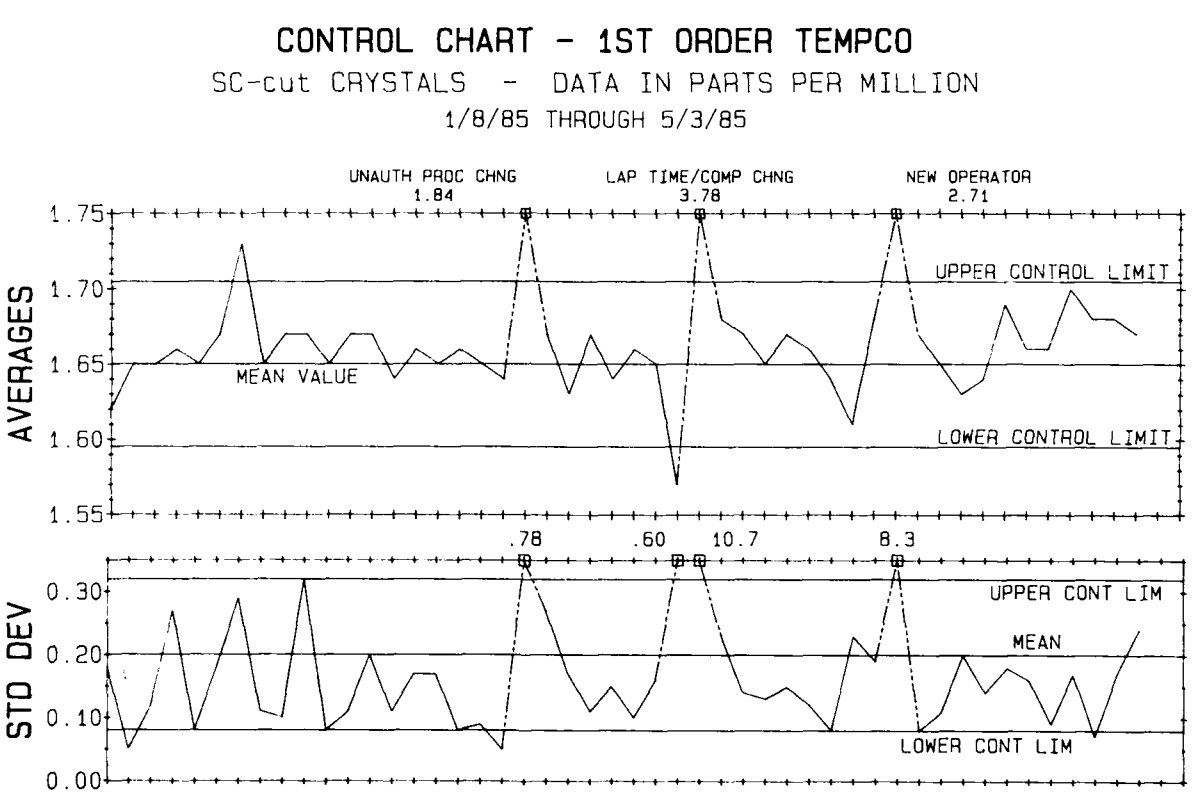


FIGURE 5.
CONTROL CHART SHOWING EFFECTS OF NOT FOLLOWING PROCESS AND ENGINEERING INTERVENTION

VHF MONOLITHIC CRYSTAL FILTERS FABRICATED BY CHEMICAL MILLING

R.C. Smythe, M.D. Howard, & J.R. Hunt
Piezo Technology, Inc.
Orlando, Florida

Abstract

The fundamental frequency limitations imposed by conventional quartz wafer machining can be overcome by chemical milling. Using this process, two-pole, fundamental mode AT-cut monolithic filters have been fabricated at frequencies from 70 to 250 MHz. Four- and six-pole tandem monolithic filters have also been constructed. Chemical milling allows the realization in an inductorless structure of filter bandwidths previously unattainable at VHF. Capabilities and limitations are discussed.

Introduction

The introduction of dual-mode resonators in the '60's greatly advanced the VHF crystal filter art by simplifications resulting from the reduced number of components as compared with discrete-resonator filters. Most importantly, it afforded improvements by eliminating the need for balanced transformers, which increase in difficulty with increasing frequency. Further restrictions remain, however, associated with the maximum fundamental frequency obtainable using conventional wafer lapping and polishing techniques. Since, at a given frequency and bandwidth, filter impedance is proportional to the third power of the overtone, increasing the realizable fundamental frequency greatly extends the capabilities of crystal filters.

Vig, et al [1, 2] developed the use of chemical etching for polishing quartz, and suggested that etching might be used to fabricate resonators having the ring-supported wafer structure proposed by Guttwein, Ballato, and Lukaszek [3] and others. The development of such chemically etched, or milled, VHF and UHF ring-supported resonators has been carried out in our organization [4]. The same process techniques have also been applied to the fabrication of monolithic two-pole filters at fundamental frequencies up to 250 MHz. Four- and six-pole tandem monolithic filters have also been produced. This paper describes some of the filters which have been developed and discusses the capabilities and limitations of this approach.

Two-Pole Fabrication

Figure 1 shows the ring-supported wafer structure of a chemically milled two-pole monolithic filter. Aluminum electrodes were deposited using photofabricated aperture masks and electron-beam-gun evaporation. The electrode configuration shown is similar to that used by us for conventional VHF two-poles. An alternative arrangement, discussed below, was also employed. The two-pole devices were packaged in standard 3-lead holders dimensionally equivalent to type HC-18/U except for a reduced height of 0.450 inch (11.4 mm).

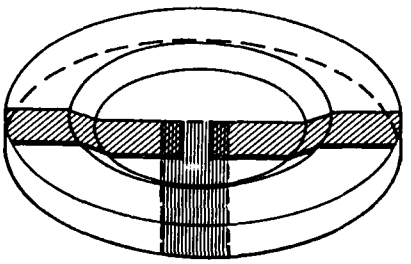


FIG. 1. RING-SUPPORTED MONOLITHIC FILTER WAFER

As frequency and bandwidth increase, required electrode dimensions decrease. The use of conventional aperture masks to define electrode patterns is limited by aperture dimensional tolerances and by front-to-back mask registration errors. To alleviate the former to some degree and to essentially eliminate the effect of mis-registration, the novel electrode configuration shown in figure 2 was devised and used to fabricate the 252 MHz two-pole devices shown in the next section. As can be seen, the electrode dimensions for each resonator are independent of lateral mis-registration. The effects of electrode connecting tabs in these configurations can be approximated analytically and have been taken account of in design. Qualitative effects of tabs have been discussed previously [5].

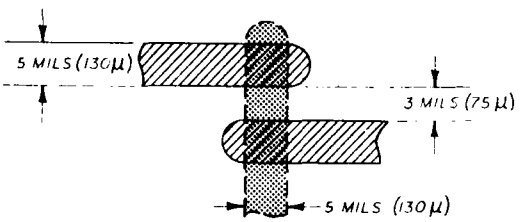


FIG. 2. NOVEL ELECTRODE CONFIGURATION USED FOR 252 MHz TWO-POLE FILTER

Results

Table 1 summarizes the primary characteristics of seven representative filters. All seven use fundamental mode, AT-cut two-poles. Figures 3-12 show attenuation and delay responses. The responses are characterized by low insertion loss, reasonable unwanted mode response, and high ultimate attenuation. Natural impedances run from 470 to 1400 ohms.

CH2186-0/85/0000-0481$1.00©1985IEEE

TABLE 1. Summary Of Filter Characteristics

Freq. (MHz)	Number Of Poles	Passband Atten. (dB)	Passband BW (kHz)	Stopband Atten. (dB)	Stopband BW (kHz)	Ultimate Atten. (dB)	Natural Impedance (Ohms)	Flat Loss (dB)	Figure(s)
70	2	3	80	20	235	50	1000	.85	3
72.5	4	3	78	45	252	80	1300	2.9*	4, 5
72.5	6	6	81	60	195	90	1400	2.7*	6
125.16	2	3	102	20	320	50	1000	2.4*	7
125.40	4	3	95	60	462	100	750	2.3*	8, 9
125	6	3	101	60 100	251 420	110	750	2.6*	10, 11
252	2	3	135	20	425	45	470	2.0	12

* Includes input/output 50 ohm matching loss.

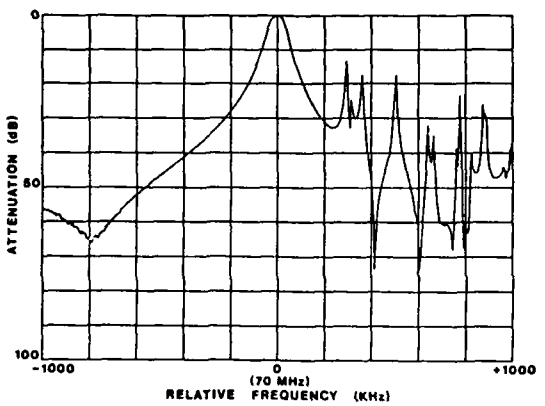


FIG. 3. ATTENUATION CHARACTERISTIC, TWO-POLE, 70 MHz FUNDAMENTAL MODE MONOLITHIC FILTER; 3dB BW IS 80 kHz; 20dB BW IS 235 kHz.

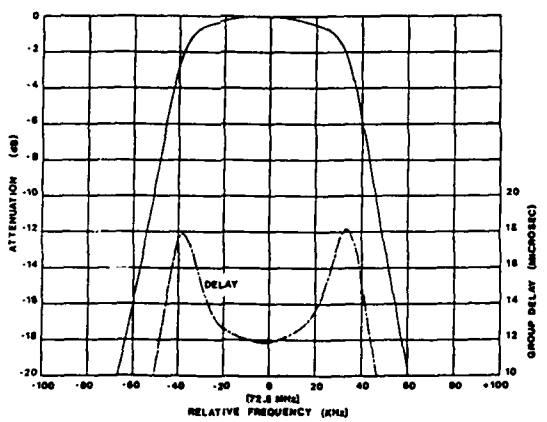


FIG. 4. PASSBAND RESPONSE, FOUR-POLE, 72.5 MHz FUNDAMENTAL MODE TANDEM MONOLITHIC FILTER; 3dB BW IS 78 kHz.

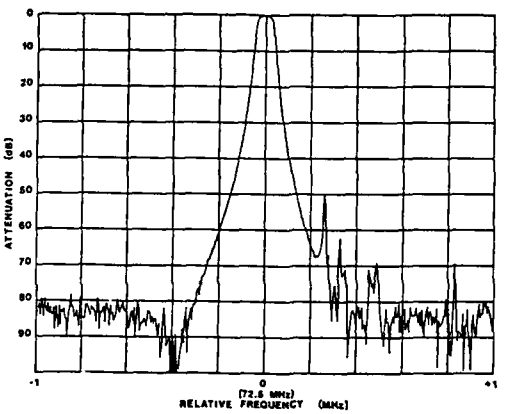


FIG. 5. STOPBAND RESPONSE, FOUR-POLE, 72.5 MHz FUNDAMENTAL MODE TANDEM MONOLITHIC FILTER, 45dB BW IS 252 kHz

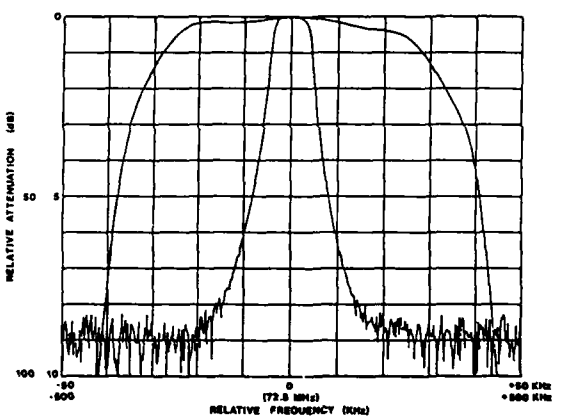


FIG. 6. PASSBAND AND STOPBAND RESPONSE, SIX-POLE 72.5 MHz, FUNDAMENTAL MODE TANDEM MONOLITHIC CRYSTAL FILTER; 6dB BW IS 81 kHz; 60dB BW IS 195 kHz.

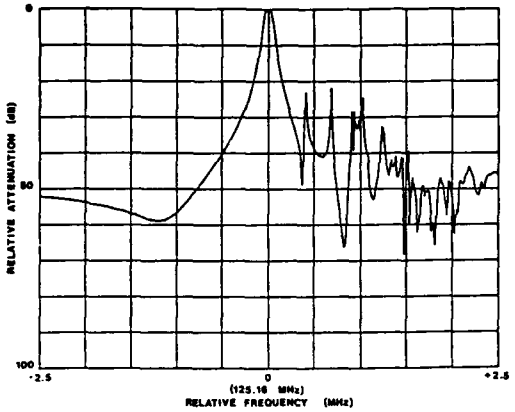


FIG. 7. ATTENUATION CHARACTERISTICS, TWO-POLE 125.16 MHz FUNDAMENTAL MODE MONOLITHIC FILTER; 3dB BW IS 102 kHz; 20dB BW IS 320 kHz.

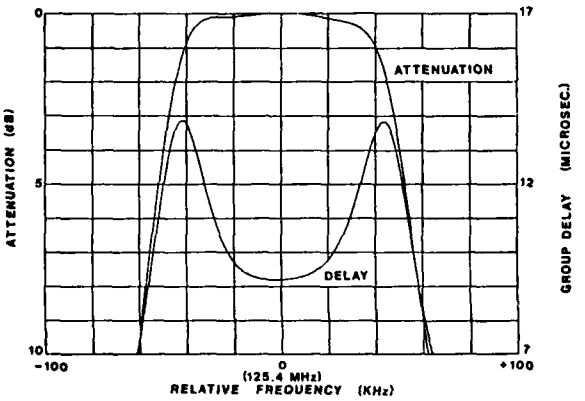


FIG. 8. PASSBAND RESPONSE, FOUR-POLE, 125.4 MHz FUNDAMENTAL MODE TANDEM MONOLITHIC FILTER; 3dB BW IS 95 kHz.

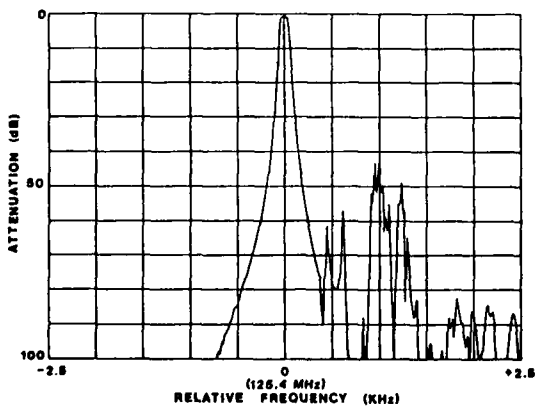


FIG. 9. STOPBAND RESPONSE, FOUR-POLE, 125 MHz; FUNDAMENTAL MODE TANDEM MONOLITHIC FILTER; 60dB BW IS 462 kHz.

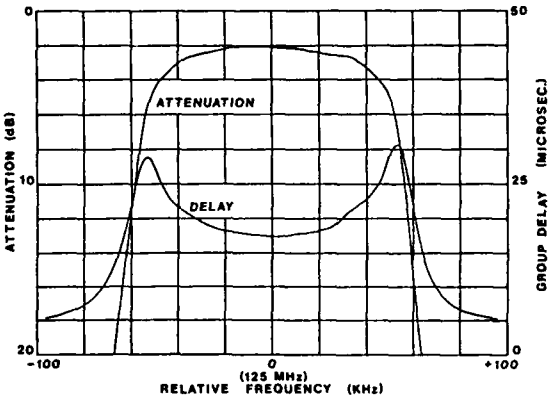


FIG. 10. PASSBAND RESPONSE, SIX-POLE, 125 MHz FUNDAMENTAL MODE TANDEM MONOLITHIC FILTER; 3dB BW IS 101 kHz.

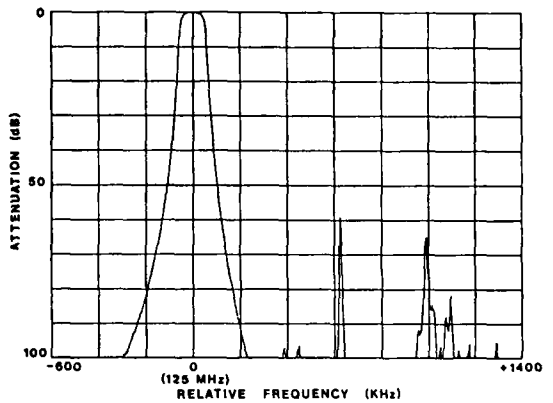


FIG. 11. STOPBAND RESPONSE, SIX-POLE, 125 MHz FUNDAMENTAL MODE TANDEM MONOLITHIC FILTER; 100dB BW IS 420 kHz.

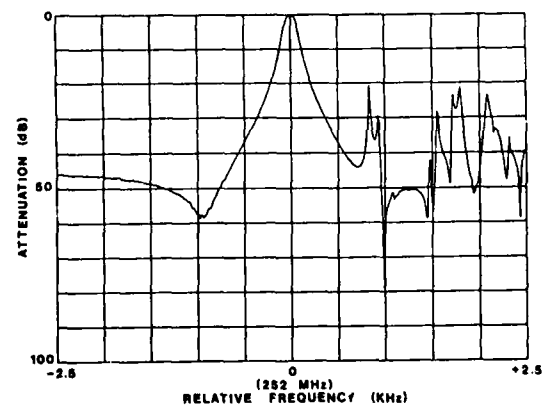


FIG. 12. ATTENUATION CHARACTERISTIC, TWO-POLE, 252 MHz, FUNDAMENTAL MODE MONOLITHIC FILTER; 3dB BW IS 135 kHz; 20dB BW IS 425 kHz.

Bandwidth & Frequency Capabilities

The use of VHF fundamental mode devices greatly extends the bandwidth and frequency range of monolithic and tandem monolithic crystal filters. Of particular interest for tandem monolithic filters is the maximum inductorless bandwidth [6, 7]. While it should be emphasized that practical filters are realizable at bandwidths well beyond the inductorless limit, the "spreading" coils required in that region add to filter size and cost, increase flat loss, and degrade passband symmetry. Because the maximum inductorless bandwidth is roughly proportional to f/N^2, overtone tandem monolithic filters are only rarely inductorless. Thus, the availability of VHF fundamental mode monolithics makes inductorless tandem monolithic filters possible for the first time at frequencies up to at least 200 MHz. Because of holder capacitance and stray capacitances the inductorless limit as a fraction of center frequency decreases with increasing frequency, from about 0.3% at 10 MHz to 0.085% or so at 200 MHz, using conventional crystal enclosures. At 70 MHz the inductorless limit is about 100 kHz; at 150 MHz, about 150 kHz using conventional enclosures. By eliminating just 0.5 pF of nodal capacitance through improved packaging, these limits could be increased by up to 50%.

Limits on monolithic filter realization are imposed by lateral dimensions and film thickness of the electrode array. These are determined from the acoustical design based on the theory of Tiersten [8] in the usual way [7]. For uniform design criteria, electrode lateral dimensions are inversely proportional to center frequency and the square root of bandwidth.

$$l_e \propto 1/(f_o \cdot BW)^{1/2}$$

Where l_e represents any lateral electrode dimension. There is a weak dependence on overtone as well [7, 8].

The 5 mil (130 micrometer) electrodes of the 252 MHz and 135 kHz two-pole monolithic, figures 2 and 12, approach the lower limit of conventional shadow mask technology. Using photolithography [9] smaller and more accurately defined electrode arrays should be possible.

Electrode film thickness, t_e, varies with overtone, bandwidth and frequency according to

$$t_e \propto N \cdot BW/f_o^2$$

For the 252 MHz example, the thickness of the aluminum electrode is about 300 Angstroms, near the lower practical limit. Extension beyond this limit would require the use of recessed electrodes [10] or novel electrode structures.

Ruggedness

The ring-supported wafers used in these devices were chemically milled from 30 MHz (55 micrometer) carefully polished AT-cut blanks, 6.35 mm. in diameter. After electroding they were mounted in conventional HC-18 series holders. Epoxy-silver cement was used. Ribbon mounts were of a standard hook configuration.

A series of shock tests were performed on a randomly chosen 125 MHz unit. The unit was subject to six shocks at 1000 g's, then at 2000 g's, and then at 3000 g's. Duration of the shock pulse was 0.3 ms. Figure 13 shows the measured filter response at the start of the test and after each series of drops. Failure occurred at 5000 g's. This is at least comparable to the performance of conventional units using this holder and mount. Still greater shock resistance could be obtained by obvious design changes-reduced blank diameter, use of a TO-5 holder - as well as by refinement of the mount design. A further possibility is chemical milling of the support ring.

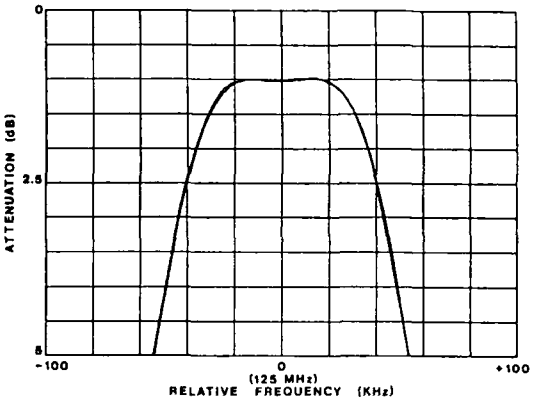


FIG. 13. PASSBAND RESPONSE OF TWO-POLE, 125 MHz MONOLITHIC FILTER BEFORE AND AFTER SHOCK TESTING.

Conclusions

Using chemically-milled ring-supported AT-cut quartz wafers, low-loss fundamental mode two-pole monolithic and four- and six-pole inductorless tandem monolithic filters have been fabricated at frequencies up to 250 MHz, and bandwidths up to 0.1% of center frequency. For the examples given, flat loss was below 3 dB in all cases. Shock testing has shown the ruggedness of these devices. The use of chemical milling has opened up a range of frequency and bandwidth not covered by other filter technologies.

References

[1] Vig, J.R., J.W. LeBus, & R.L. Filler, "Chemically Polished Quartz," Proc. 31st Annual Symposium On Frequency Control (AFCS), pp. 131-143; 1977.

[2] Vig, J.R., R.J. Brandmayr & R.L. Filler, "Etching Studies on Singly- & Doubly-Rotated Quartz Plates," Proc. 33rd AFCS, pp. 351-358; 1979.

[3] Guttwein, G.K., A.D. Ballato, & T.J. Lukaszek, U.S. Patent 3,694,677; 26 Sept. 1972.

[4] Hunt, J.R., & R.C. Smythe, "Chemically Milled VHF & UHF Resonators," these Proceedings.

[5] Horton, W.H., & R.C. Smythe, "Theory of Thickness-Shear Vibrators with Extensions & Applications to VHF Acoustically Coupled-Resonator Filters," Proc. 21st AFCS, pp. 160-178; 1967.

[6] Smythe, R.C., "Communications Systems Benefit From Monolithic Crystal Filters," Electronics, v. 45, no. 3, pp. 48-51; 31 Jan. 1972.

[7] Smythe, R.C., "Bulk Acoustic Wave Filters," in Precision Frequency Control, (E.A. Gerber & A.D. Ballato, ed.s), pp. 188-228; Academic Press, New York; 1985.

[8] Tiersten, H.F., "An Analysis of Overtone Modes in Monolithic Crystal Filters," Proc. 30th AFCS, pp. 103-108; 1976.

[9] Shanley, C.W., L.N. Dworsky, J.A. Whalin, G.G. Clifford, & M.N. Scansaroli, "Processing of a Five-Resonator VHF Crystal Device," these Proceedings.

[10] Curran, D.R., & D.J. Koneval, "Factors in the Design of VHF Filter Crystals," Proc. 19th AFCS, pp. 213-268; 1965.

THE MOTOROLA MULTI-POLE MONOLITHIC FILTER PROJECT

L. N. Dworsky
Motorola, Inc.
1301 Algonquin Rd.
Schaumburg, Ill. 60196

C. S. Shanley
Motorola, Inc.
8000 W. Sunrise Blvd.
Plantation, Fla. 33022

Abstract

A 4-pole VHF monolithic crystal filter and a crystal resonator on a single crystal blank have been designed and are being manufactured for use in a radio pager. This paper reviews the design of the device, emphasizing how the various design specifications led to a set of manufacturing tolerances which in turn influenced the design of the device and the fabrication processes. These fabrication processes are detailed in a companion paper [1].

Introduction

The quartz crystal device to be discussed was designed as part of a larger design effort to produce the VHF radio pager whose block diagram is shown in Figure 1. All RF selectivity in this pager is due to the 4-pole filter at a nominal signal frequency of 152 MHz. This design approach yields a very simple high performance radio receiver. Since the radio intermediate frequency is 35 KHz, the local oscillator frequency is only 35 KHz below (or above) the signal frequency. The relatively small difference between the signal and the local oscillator frequencies suggests that the local oscillator crystal resonator could be fabricated on the same crystal blank as the filter.

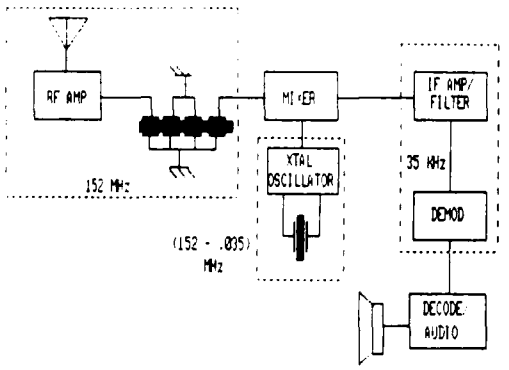


Figure 1. VHF Radio Pager Block Diagram

Filter Specifications and Design

The performance and durability specifications for the device were determined both by the electrical and mechanical (environmental) requirements of a VHF radio pager. Common to both the filter and single resonator are requirements for temperature stability within ±15 ppm from 0 to 65 degrees Centigrade, the ability to withstand in excess of 5000 g shocks, and less than 1 ppm/year aging. The temperature stability taken alone appears well within the standard capabilities of AT quartz thickness-shear resonators, but maintaining this stability in a high (for VHF quartz devices) shock environment requires that careful attention be paid to the mounting of the device - as shall be discussed below.

The 4-pole filter's 3 dB bandwidth is 12 KHz, and the maximum center frequency insertion loss is 5 dB. The filter matches into parallel R-L terminations. The inductive components of these terminations are solely to resonate with the shunt capacitances of the input and output resonators. The resistive components are both 2000 Ohms. A minimum insertion loss filter design realizing all of the above at 152 MHz requires unloaded resonator Q's of greater than 65000. Such a design in a 4-pole filter meets the radio system requirements of filter skirt fall-off at adjacent paging channel separation of 25 KHz. The stop-band rejection of the filter must be greater than 70 dB. It should be pointed out that there is no inherent necessity for a minimum insertion loss (as opposed to a Butterworth, Chebychev, or any other) filter response in the design or manufacture of the filter.

As discussed above, the local oscillator frequency is 35 KHz below the filter center frequency. The local oscillator motional resistance must be less than 200 Ohms. Due to the unusually low intermediate frequency of the radio, spurious responses of this resonator fall on or very close to the radio signal frequency. This results in undesirable noise degradation of the radio performance. The resonator spurious mode motional resistances must therefore be as high as possible.

Figure 2 shows the insertion loss of the filter designed to the above specifications, using a conventional coupled L-C resonator model for the filter [2]. This response is idealized in that spurious modes and subtle deviations from the L-C model do not appear.

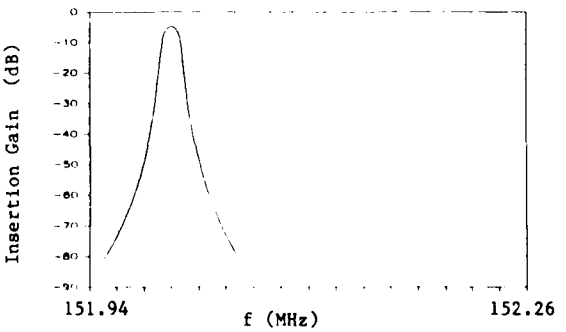


Figure 2. Circuit Model Filter Response

In order to translate the electrical model into a physical device, it is first necessary to decide on an "overtone" of crystal operation. The table below shows approximate crystal blank thickness and filter electrode dimensions as a function of the overtone of operation. The electrodes are assumed to be square for convenience here and maximum electrode thickness is that thickness of an aluminum electrode which would

CH2186-0/85/0000-0486$1.00©1985IEEE

allow for no trapped (single resonator) spurious modes for the parameters given. These calculations were performed using the resonator model presented by Tiersten [3].

N	Electrode Size	Blank Thickness	Max Electrode Thickness
1	.04 X .04 (mm)	.011 (mm)	40,000 (Ang)
3	.19 X .19	.034	2,700
5	.39 X .39	.055	1,000
7	.65 X .65	.078	450

N = 1 and N = 3 were rejected because the blank is simply too thin to be practicable - both in terms of blank fabrication and resulting device sturdiness. At the other extreme, N = 7 was rejected both because the electrodes are so large that the overall device size would be unacceptable and the allowable electrode thickness is so small that the thin film deposition process would be difficult. At N = 5 the blank is still quite thin, but it was judged that this problem was manageable. The required 4 resonator filter and single resonator easily fit onto a 1/4 inch (6.4 mm) blank.

Once the overtone is chosen, in addition to the information above, the electrical design also leads to the physical separation (gap) between electrodes [4,5]. Figure 3 shows a simulation of the resulting filter performance, using a discrete element model which was presented at a previous symposium [6]. As may be seen, the pass-band and skirt response is in overall agreement with the electrical model. The simulation, of course, also predicts filter spurious response - approximately 40 dB down and 180 KHz away from the filter pass-band.

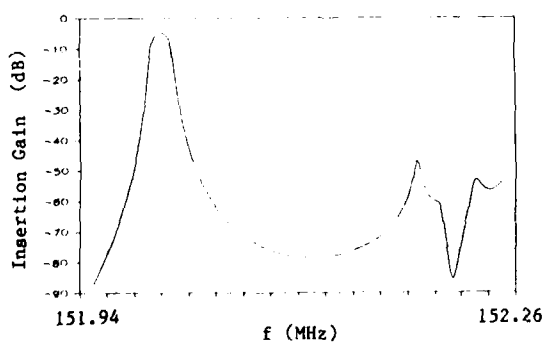


Figure 3. Simulation Filter Response

Figure 4 shows the simulation of the filter pass-band in greater detail than did Figure 3. Here it can be seen that while the agreement with the electrical element model is quite good close to the center frequency, at 25 KHz above the center frequency the insertion loss is approximately 6 dB less than it is at 25 KHz below the center frequency. This predicted "skewing" of the skirt response agrees with actual device results. It may be explained by noting that the coupling in a monolithic crystal filter is better described by a cut-off waveguide operating near its cut-off frequency than by the mutual inductance of the electrical model [7].

Figure 5 shows the physical layout of the filter on the crystal blank. The solid areas represent metal

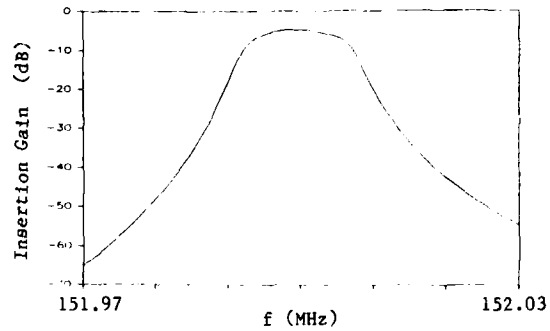


Figure 4. Simulation Filter Response (Passband)

on the top-side of the blank. Each of the four electrodes is connected to a bonding pad at the periphery of the blank. The four unfilled lines represent metal on the underside of the blank, joining together at a central metal "pad." This pad is attached using metalized epoxy to a pedestal in the crystal package and acts as both the mechanical mount and the filter ground return. The bonding pads are connected, using ultrasonically bonded aluminum wires, to the package. This somewhat unusual mounting and attachment scheme, although wasteful of quartz surface area, is very rugged and resolves the problem of the fragility of the thin quartz blanks [1].

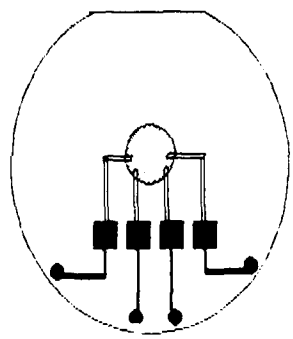


Figure 5. Physical Layout of Filter

Filter Parameter Tolerances

In order to design a manufacturing system for these devices, it is first necessary to understand the effects of the various possible physical parameter variations upon the filter performance. This is a two step process - first the effects of an electrical parameter change on the filter performance must be examined and then these effects must be correlated to the physical device parameters which will cause this electrical parameter change. Since in actual manufacturing processes, all variations are happening at once, a Monte-Carlo analysis is necessary to predict the expected range of filter variations. For the purposes of this discussion, the simpler case of varying one parameter at a time will be utilized. While not as accurate as a Monte-Carlo simulation, this latter procedure is much easier to present and leads to very reasonable conclusions.

Variations in electrode size (lateral dimensions) will cause changes in both motional parameters of the individual resonators and in the inter-resonator coupling. The effect of motional parameter changes on

the filter response due to a change in electrode size is minor when compared to the effect of the accompanying inter-resonator coupling change. Figure 6 shows the effect of changing the coupling, K, on the 3 and 6 dB bandwidths of the filter. As may be seen, a 3 dB bandwidth tolerance of +.25 kHZ requires a coupling tolerance of approximately ±2%.

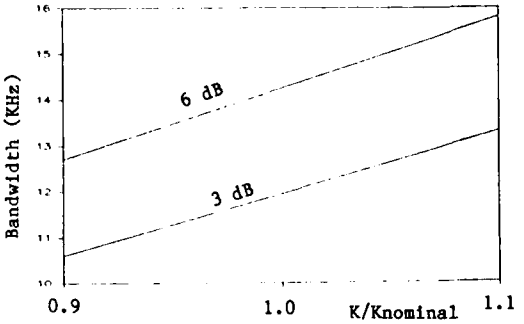


Figure 6. Bandwidth Variation vs. K

Figure 7 shows the variation of K with electrode size. This figure assumes that electrode center-to-center spacing is held constant while the electrode size varies. This would be the case due to aging masks, processing errors, etc. Here, a larger electrode means a smaller edge-to-edge electrode gap, and hence increased coupling due to increased electrode size. A coupling variation of 2% is caused by a size error of approximately .002 mm (.1 mils). Maintaining electrode size within .002 mm is not within the capability of shadow mask deposition techniques, and hence it was necessary to develop a two-sided photolithographic capability in order to achieve the required tolerances. It may be shown that a front-to-back alignment capability comparable to the electrode size tolerance capability is also necessary.

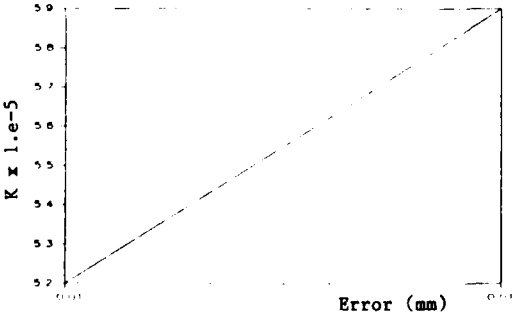


Figure 7. K vs. Electrode Size Error

Figure 8 shows the variation of K with electrode thickness. In the manufacture of a thickness shear resonater, the electrode thickness (actually, the mass) is adjusted so that the resonance frequency is correct. Figure 8 implies that if the crystal blank is not very close to the correct frequency (thickness) then the process of tuning the four resonators that comprise the filter to their correct frequencies will cause it to have an unacceptable bandwidth. This means that the crystal blanks to be used must be prepared to a very tight tolerance. The fact that the filter is comprised of four resonators spaced some distance from each other on the crystal blank further complicates the situation, in that the parallelism of the blank becomes an important issue. A procedure for examining the frequency of the blank as a function of location on the surface had to be developed [8], and the blanks are specified in terms of surface profile as well as average or maximum thickness.

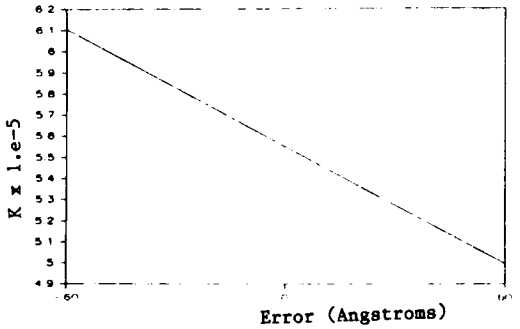


Figure 8. K vs. Elecrode Thickness Error

Figure 9 shows the insertion loss response of a typical filter. As may be seen, the response is essentially that predicted by the simulation. The stop-band rejection of the actual filters is primarily determined by inter-electrode and stray capacitances, and is therefore not predicted by the simulation.

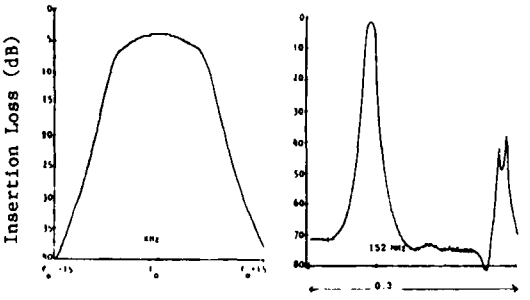


Figure 9. Typical Filter Response

Multi-Frequency Filter Design

In principal, it is necessary to go through the entire design process described above for every center frequency desired. As explained above, merely changing the blank frequency drastically affects the coupling, K, and hence the filter bandwidth. Figure 10 shows the 3 dB bandwidth versus filter center frequency due to simply changing the blank frequency about its nominal value.

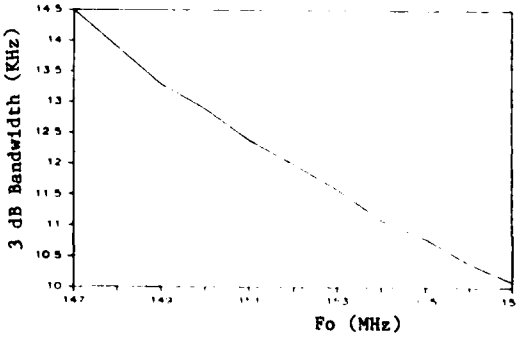


Figure 10. Uncorrected Bandwidth vs Center Frequency

In an attempt to correct the filter response without redesigning the filter, the coupling, K, was adjusted so as to return the 3 dB bandwidth to its nominal value. Figure 11 shows the required value of K versus the filter center frequency. In practice, this is accomplished by offsetting the electrode thickness by the proper amount, and adjusting the blank thickness so that the desired center frequency is obtained. While filters with the correct 3 dB bandwidth are obtained in this manner, the overall insertion loss response is not exactly the same as that of the original design.

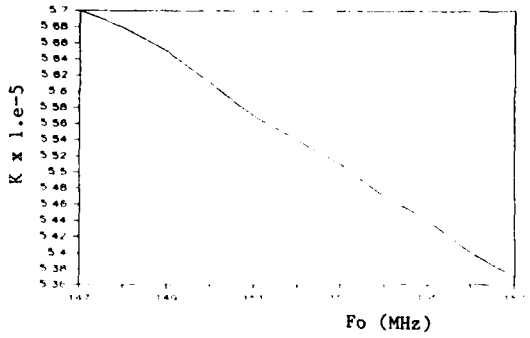


Figure 11. K to Correct 3 dB Bandwidth

Figure 12 shows the insertion loss response error for filters built using the nominal design, but with blanks and electrode thicknesses adjusted so that the center frequencies will be 3 MHz above and 3 Mhz below the nominal frequency, and with the filters having the correct 3 dB bandwidth. These responses are therefore essentially identical to the nominal response for frequencies approximately ±12 KHz about the center frequency, whereas at 25 KHz below center frequency there is about a 1 dB error. In practice, this means that a nominal design (i.e. mask set) may be used for approximately a 5 MHz range of center frequencies without serious degradation of filter performance.

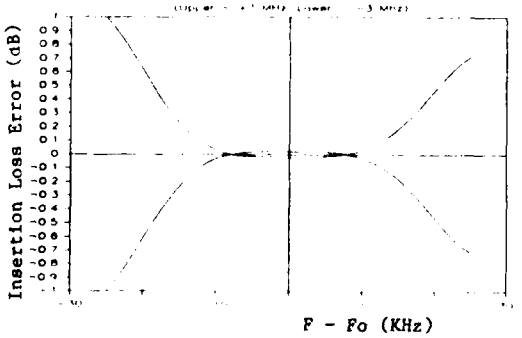


Figure 12. Corrected Filter Response Error

Single Resonator Design

The required single resonator motional resistance and the experimentally obtainable device Q together predict a resonator surface area of approximately one square millimeter. The resonance frequency of this resonator, assuming the same metal thickness as that of the filter on the same crystal blank, is shown in Figure 13 versus the aspect ratio of the (rectangular) electrode. For convenience, this resonance frequency is shown referenced to the filter center frequency. As may be seen, the required 35 KHz frequency separation is obtained for an resonator with approximately a 9:1 length to width (or vice-versa) aspect ratio. While resonators built with this aspect ratio operated properly, their spurious responses were deemed unacceptably large for the intended application.

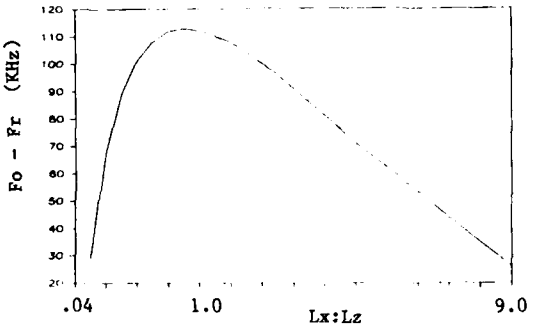


Figure 13. Fo - Fr vs Lx:Lz for a Fixed Area Resonator

In order to circumvent the spurious mode problem, it was necessary to design a resonator whose largest dimension would be small enough so as not to cause spurious mode problems. Figure 14 shows the evolution of such a design. A square resonator (having the correct surface area) would have a low enough motional resistance and acceptable spurious mode performance, but would be *too* low in frequency. If this resonator were sliced in half, each of the halves would would have approximately twice the desired motional resistance, acceptable spurious mode performance, and be approximately 15 KHz higher in frequency than desired. If the two halves were electrically connected in parallel and then brought close enough *to each other* to couple, they would exhibit only the even coupled mode resonance, have the correct motional resistance, acceptable spurious mode performance, and a resonance frequency somewhere between the 15 KHz too high and 100 KHz too low - depending upon the gap between them. Figure 15 shows the frequency separation as a function of this gap, with a gap of approximately .08 mm producing the desired result.

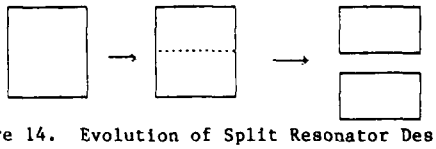


Figure 14. Evolution of Split Resonator Design

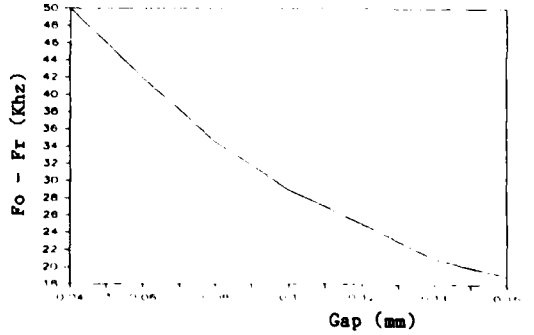


Figure 15. Fo - Fr vs Gap of Split Resonator

Summary

Figure 16 shows the complete filter - resonator device. The original underside pedestal has been electrically split so that the single resonator is not connected to the filter ground return. The 4 dark spots shown on the pedestal sections are the areas where the crystal is actually attached to four raised spots on the package substrate. The spots are chosen so as to be centered along the axes of 0 frequency - stress sensitivity [9]. This choice of mounting locations allows the frequency - temperature response of both the filter and the single resonator to be manufacturably repeatable and indistinguishable from the frequency - temperature responses obtained by conventional mounting systems.

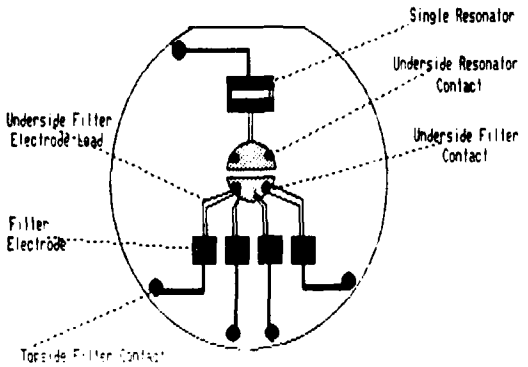


Figure 16. Final Device Layout

The device design described above has proven to be economically produceable with a repeatability of characteristics better than is typical of RF quartz devices. With many thousands of units in the field for several years, it has been demonstrated that the ruggedness and overall durability of the devices is excellent. Since the original design was executed, devices have been designed and are in production from 35 to 90 MHz (3rd overtone) and 90 to 170 MHz (5th overtone). The heavy initial emphases on extensive modeling and tight production tolerance control have resulted in a series of devices whose designs are easily customized, reliably predicted, and reproducibly manufactured. The mechanical design along with the production control has resulted in devices which have exhibited and continue to exhibit excellent field reliability.

References

1. C. Shanley, L. Dworsky, J. Whalin, G. Clifford & M. Scansaroli, "Processing of a Five Resonator Crystal Device." Proc. 39th AFCS, 5/85.

2. R. Sykes & W. Beaver, "High Frequency Monolithic Crystal Filters with Application to Single Frequency and Single Sideband Use." Proc. 20th AFCS, 4/66.

3. H. Tiersten, "Analysis of Trapped Energy Resonators Operating in Coupled Modes of Thickness Shear and Thickness Twist." J. Acoust. Soc. Am., Vol 59, #4, 4/76.

4. H. Tiersten, "An Analysis of Overtone Modes in Monolithic Crystal Filters." Proc. 30th AFCS, 5/76.

5. L. Dworsky, "Monolithic Crystal Filter Design Using a Varitional Coupling Approximation." IEEE Trans. on Sonics & Ultrasonics, Vol SU-28, #4, 7/81.

6. L. Dworsky, "Discrete Element Modeling of AT - Quartz Devices." Proc. 32nd AFCS, 5/78.

7. L. Dworsky, "An Improved Circuit Model for Monolithic Crystal Filters." IEEE Trans. on Sonics & Ultrasonics, Vol SU-28, #4, 7/81.

8. L Dworsky & G. Kennedy, "Air Gap Probe Evaluation of Thin Quartz Plates." Proc. 35th AFCS, 5/81.

9. J. Ratajski, "The Force Sensitivity of AT - Cut Quartz Crystals." Proc. 20th AFCS, 4/66.

MONOLITHIC CRYSTAL FILTERS HAVING IMPROVED INTERMODULATION & POWER HANDLING CAPABILITY

M.D. Howard, R.C. Smythe & P.E. Morley
Piezo Technology, Inc.

Abstract

Recent studies of elastic non-linearity in quartz resonators suggest that power handling capabilities and in-band intermodulation performance could be improved by using BT- or SC-cut resonators in filter realization. These improvements have potential advantages in spectrum cleanup applications, where increased power handling capability may offer better signal to noise ratios, and in communications system applications where receiver and transmitter exciter characteristics are limited by existing filter intermodulation performance. Improvements are not obtained without penalty. Increases in impedance level, C_0/C_1 ratio and frequency change with temperature cause degradation of other performance parameters. These design trade-off's are discussed.

Results on SC- and BT-cut monolithic crystal filters in the frequency range 25 MHz to 125 MHz are presented. Data include measurements of conventional filter parameters as well as those relating to elastic non-linearities. Measurements of the motional and coupling parameters of SC- and BT-cut two-pole monolithic devices are compared with theoretical predictions.

Introduction

Over the past decade or so advances in semiconductor technology have lead to the development of active devices with improved linearity and power handling capabilities. As active circuits have improved, the quartz crystal filter has become one of the limiting factors when low levels of third order intermodulation product (IM_3) are required. In the past, improvements in IM_3 generated in AT-cut quartz crystal filters have been achieved by improving processing techniques [1] and by the development of theory to assist in analyzing the anelastic properties of AT-cut quartz and its contribution to IM_3 generation [2,3,4], which in turn allows the optimization of design variables. However, the next generation of military H.F. radios is expected to demand even more stringent IM_3 performance from crystal filters. Also, in order to improve exciter signal-to-noise ratios it is desirable to have as much signal amplification as possible before filtering. To achieve these objectives filters capable of handling higher power levels while still operating in a linear manner are needed.

Concurrent with these pressing requirements are the demands of next generation instrumentation. As quieter, more linear measurement systems are required for the development of electronic warfare and global positioning systems the demands on spectrum clean-up filters likewise become more stringent.

Non-linearities in resonators can be considered as being of four types:

1) Non-linearity of effective series resistance, an effect which is often related to surface condition and surface contamination. This effect is largely process related and, providing units are fabricated in a similar manner and have similar surface motions, is expected to be similar for all cuts. The filter parameters most affected are out-of-band intermodulation performance, and variation in loss at low signal levels.

2) Non-linearity of motional parameters, an effect related to the anelastic properties of quartz and most apparent at high drive levels. Filter parameters most affected will be power handling capability and inband intermodulation products. This type of non-linearity will be dependent upon choice of cut and resonator design.

3) Thermal effects. Dissipated power is distributed non-uniformly within the resonator in accordance with the squared amplitude of vibration, giving rise to thermal gradients which induce non-linear behavior.

4) Non-linear coupling to unwanted modes of vibration, which may affect both motional parameters and effective series resistance. While this effect is somewhat dependent on cut and is thought to be less prevalent in SC-cuts than in AT or BT it can often be eliminated by proper resonator design. Non-linear coupling is most prevalent at high drive levels and affects inband IM_3 performance as well as filter amplitude and phase response.

Non-linear resonance measurements suggested that the anelastic properties of SC- and BT-cut quartz resonators might be superior to AT-cut; hence, in-band IM and power handling might be improved. Accordingly, a program to measure the relevant effective non-linear elastic constant for these cuts was instituted. Early results from that program are reported in a companion paper [5]. At the same time, a second program, reported here, was undertaken to design and build SC- and BT-cut monolithic filters and to measure their intermodulation characteristics and power handling capabilities.

The Design of Coupled Resonator Structure on SC- & BT-Cut Monolithic Dual Resonators

Monolithic coupled resonator elements have replaced discrete resonators in most filter applications utilizing AT cuts. One of the major driving forces, along with economy and size, factors of prime concern to any user, is the effect of the hybrid coils required to construct filters using discrete resonators. These coils contribute excessive loss in the passband and will not give proper stopband attenuation unless manufactured with extreme care. In addition, especially at high impedance levels, there

CH2186-0/85/0000-0491$1.00©1985IEEE

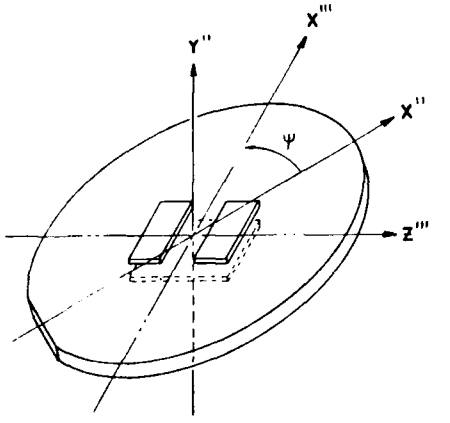
can be severe problems in obtaining a proper balance of lattice arms, which will also degrade stopband performance.

SC- and BT-cuts have motional inductances of approximately 3.5 and 4.0 times greater than that of AT-cuts of the same electrode area and overtone. Consequently, parasitic effects of hybrid coils will be even more pronounced than in filters made with AT-cuts. To produce filters which offer an optimum combination of properties it was concluded that coupled resonators are an essential building block.

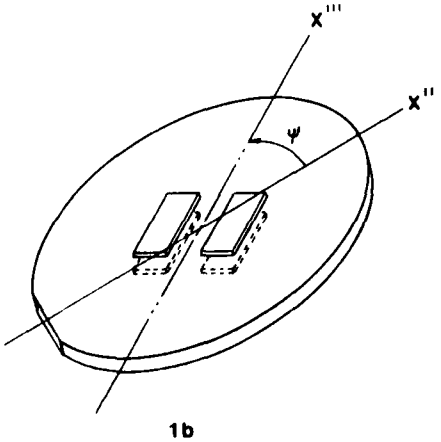
A three-dimensional, two-pole AT-cut coupled resonator analysis has been given by Tiersten [6]. The extension to the BT-cut is immediate. For the SC-cut, the effective elastic constants M_n and C_{55}, of equation 2-9 [6] should be replaced with M_n' and P_n', as given by Stevens [7]. The angular orientation Ψ, of the electrodes (fig. 1a) for wave propagation along the electrode axes, which is zero for the singly-rotated AT- and BT-cuts, is in general non-zero for doubly-rotated quartz cuts. For the SC-cut its value is dependent on overtone as listed in table 1.

TABLE 1.

Overtone (n)	Ψ
1	4.3°
3	25.5°
5	34.1°
7	12.0°



1a



1b

FIG. 1.

The fundamental third, and fifth overtones of the SC-cut are quite suitable for the design of coupled resonators due to their closeness to being isotropic, although the resulting motional capacitance is lower. The BT-cut, however, is strongly anisotropic, especially for overtones. For an equi-modal resonator the ratio of electrode x- and z- dimensions is about 2 for the fundamental, 3 for the third overtone, and 6 for the fifth.

Typical electrode geometries for all cuts are shown in figs. 1a & 1b. For the former, the analysis has been modified to include the mass loading in the inter-resonator gap. Particularly at VHF, attention must also be paid to the effect of electrode tabs [8].

Results On Two-Pole Monolithic Resonators

Two-pole monolithic structures have been manufactured on both SC- and BT-cut fundamental mode blanks at 25 MHz. SC-cut monolithics have also been manufactured at 97.8 MHz operating on the third overtone and at 125 MHz on the fifth overtone.

As noted above, for overtone operation BT-cuts exhibit extreme anisotropy of propagation. In practice this leads to designs having electrode configurations which are not suitable for manufacture with existing techniques. However, some design data for BT-cuts is included. Also, for the purposes of comparison, some AT-cut data will be presented.

Figures 2a through 2d illustrate the electrode configurations used. Fig. 2a also illustrates the method of defining electrode dimensions. All electrode configurations were obtained with existing masks and are not necessarily optimum. Two of these configurations, 2b and 2c, were initially designed for applications where large tab leads were thought to be advantageous. The effect of these large tabs was not included in the design analysis, which may account for the discrepancy between some of the predicted and measured mode spectra.

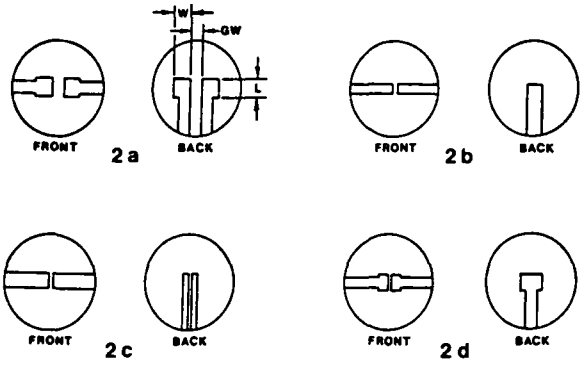


FIG. 2.

Tables 2a, 3a and 4a give electrode dimensions, electrode configuration, mass loading in kHz and coupling axis for fundamental, third and fifth overtone designs. For the SC-cuts the Z''' and X''' are for electrodes oriented according to Table 1, as indicated in fig. 1a and 1b. Table 2b, 3b and 4b give the corresponding calculated and measured values of motional inductance, mode spacing and spurious response locations.

In general a good correlation is seen between the measured mode spectra and calculated mode spectra. However, measured motional inductances are less than calculated values in all cases.

All two pole devices were designed to give filters having 3 dB bandwidths close to 20 kHz with some variations imposed by the availability of masks for electrode deposition. Electrode areas are comparable on all fundamental units and also on the third overtone units. However, the fifth overtone SC-cut units were designed to have substantially larger electrodes than the comparable AT-cut units in order to obtain practical impedance levels.

TABLE 2a. FUNDAMENTAL TWO-POLE DESIGN DATA

Cut	Electrode Dimensions w (mils)	h (mils)	gw (mils)	Electrode Type	Mass Loading (kHz)	Coupling Axis
SC	45	55	12	a	180	Z'''
BT	40	45	8	d	200	Z''
AT	50	50	15	a	200	X''

TABLE 2b. FUNDAMENTAL TWO-POLE MODE RESPONSE DATA

Cut	Motional Inductance (mH)		Mode Spacing (kHz)		Unwanted Modes					
					Mode Number		Freq. Ref. f_s (kHz) Symmetric		Antisymmetric	
	Calc.	Meas.	Calc.	Meas.	Coupled	Cross	Calc.	Meas.	Calc.	Meas.
SC	28.1	24.7	12.4	12.2	1	0	101	84	-	-
BT	44.4	39.5	10.9	12.9	1	0	70	55	-	-
AT	8.1	6.95	18.3	19.3	0	1	122	122	140	139
					1	0	148	148	-	-

TABLE 3a. THIRD OVERTONE DESIGN DATA

Cut	Electrode Dimensions w (mils)	h (mils)	gw (mils)	Electrode Type	Mass Loading (kHz)	Coupling Axis
SC	24	40	7	d	150	X'''
AT	25	40	9	c	100	X''

TABLE 3b. THIRD OVERTONE MODE RESPONSE DATA

Cut	Motional Inductance (mH)		Mode Spacing (kHz)		Unwanted Modes					
					Mode Number		Freq. Ref. f_s (kHz) Symmetric		Antisymmetric	
	Calc.	Meas.	Calc.	Meas.	Coupled	Cross	Calc.	Meas.	Calc.	Meas.
SC	33.1	32.3	15.4	14.8	0	1	50.3	49.5	66.3	66.5
					1	0	82.1	72.5		
AT	8.5	7.5	12.8	12.0	0	1	52.1	50	65	66.7

TABLE 4a. FIFTH OVERTONE DESIGN DATA

Cut	Electrode Dimensions w (mils)	h (mils)	gw (mils)	Electrode Type	Mass Loading (kHz)	Coupling Axis
SC	27	42	7	d	180	X'''
AT	16	25	10	b	150	Z''

TABLE 4b. FIFTH OVERTONE MODE RESPONSE DATA

Cut	Motional Inductance (mH)		Mode Spacing (kHz)		Unwanted Modes					
					Mode Number		Freq. Ref. f_s (kHz) Symmetric		Antisymmetric	
	Calc.	Meas.	Calc.	Meas.	Coupled	Cross	Calc.	Meas.	Calc.	Meas.
SC	61.4	48.6	12.6	12.0	0	1	49.3	51.1	61.9	-
					0	2	-124	123.5	136	140.5
					1	0	69.6	69	109.3	107
AT	62.3	43.3	16.5	17.5	None Trapped					

Results On Filters

In total eleven different filters have been tested. Table 5 lists cut, center frequncy, number of poles and serves as an index for figure numbers and tables. All filters were manufactured from the coupled two pole resonators described as above .

TABLE 5. SUMMARY OF RESULTS

Cut	Center Freq. (MHz)	Over-tone	# Of Poles	Linear Transfer Properties		Inband IM_3	Power Handling
				Table #	Figure #	Table #	Figure #
SC	25	1	2	6	3 & 4	9	25
BT	25	1	2	6	5	9	26
AT	25	1	2	6	6	9	24
SC	97.8	3	2	7	7	9	23
AT	97.8	3	2	7	8	9	22
SC	97.8	3	4	7	9	9	—
AT	97.8	3	4	7	10	9	—
SC	125	5	2	8	11	9	21
AT	125	5	2	8	12	9	20
SC	125	5	4	8	13	9	—
AT	125	5	4	8	14	9	—

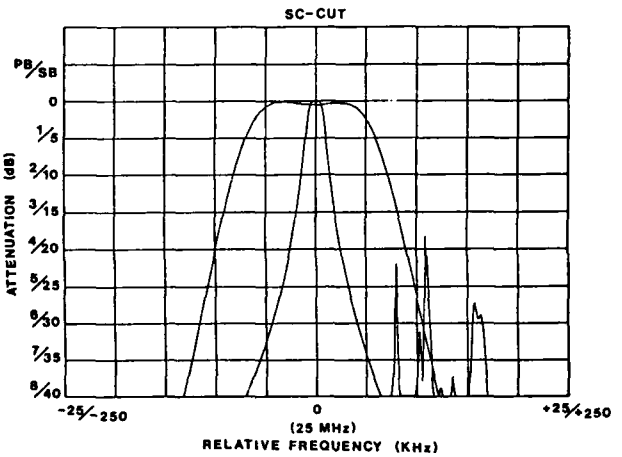


FIG. 3.

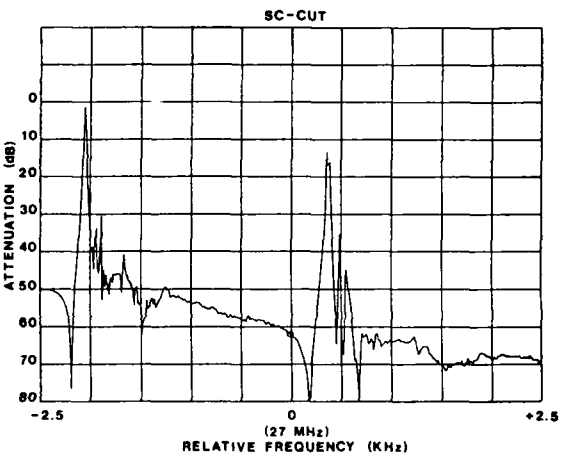


FIG. 4.

The results are grouped in three sections. The first presents measured results on linear transfer properties in both passband and stopband. The second presents results on inband IM_3 products, and the third section presents some qualitative data on power handling capabilities.

Linear Transfer Properties

For fundamental devices all electrode areas were in the region of 2400 sq mils. The SC- and BT-cut devices have a higher impedance level but the coils availabe for impedance matching networks at this frequency are such that there is no appreciable penalty in insertion loss. The AT- and SC-cuts have comparable spurious responses, but the BT-cut showed a strongly excited mode at approximately +50 kHz, returning to 3 dB while AT- and SC-cut had worst case spurs in the 15 - 20 dB range. The B-mode of the SC-cut returns to 10 - 15 dB (fig. 4).

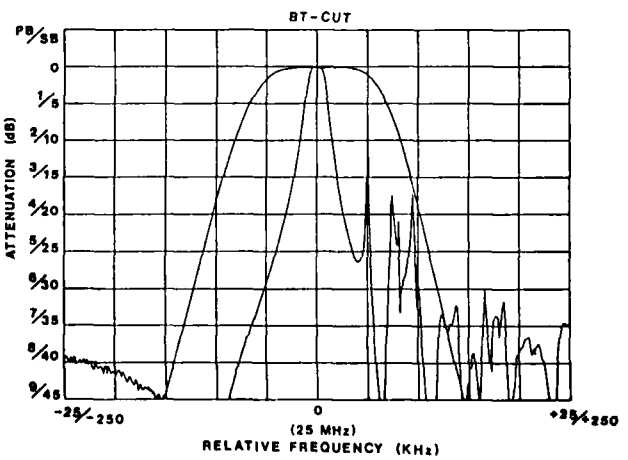


FIG. 5.

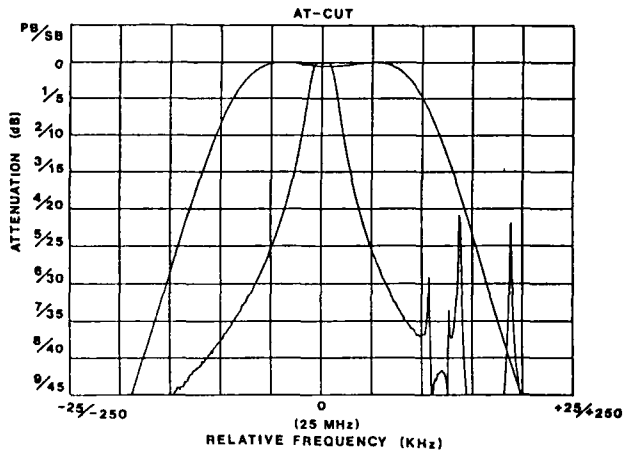


FIG. 6.

TABLE 6. TYPICAL FUNDAMENTAL FILTER RESPONSE DATA

	CUT		
	SC	BT	AT
Number Of Poles	2	2	2
Source/Load Impedance (Ohms)	1800	3300	720
Flat Loss (dB)	1.1	1.0	0.8
Ripple (dB)	0.0	0	0.1
BW_3 (kHz)	17.3	18.8	21.5
BW_{15} (kHz)	38	45	48
Min. Spur (dB)	19.8	3.3	15
Flyback (dB)	51	35	50

Third-overtone devices were made with electrode areas of approximately 2000 sq mils. At this frequency the SC-cuts are seen to have a higher flat loss than the AT-cut units, due to greater loss in the matching networks at higher impedance levels. The SC-cuts are also seen to give a somewhat worse spurious response spectra and less ultimate attenuation than the AT units.

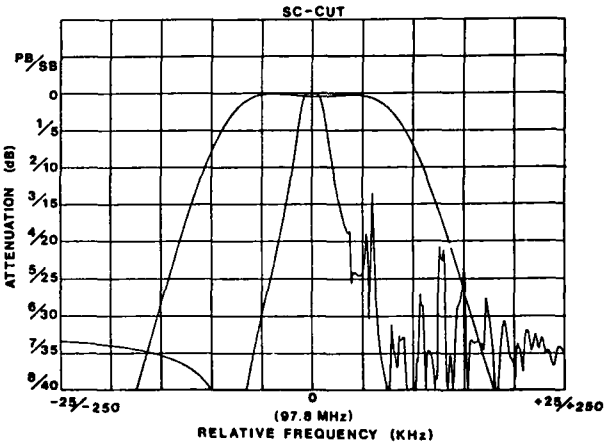


FIG. 7.

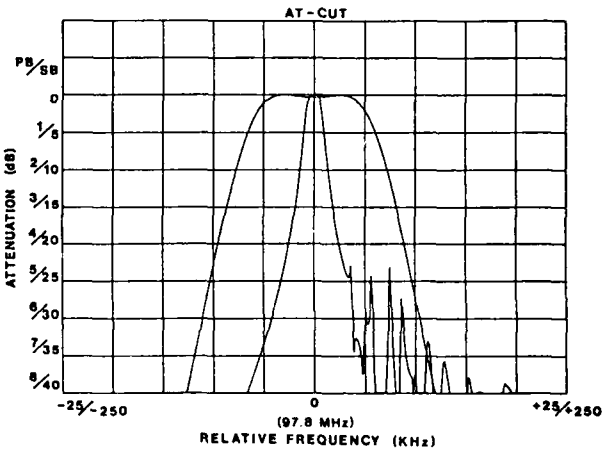


FIG. 8.

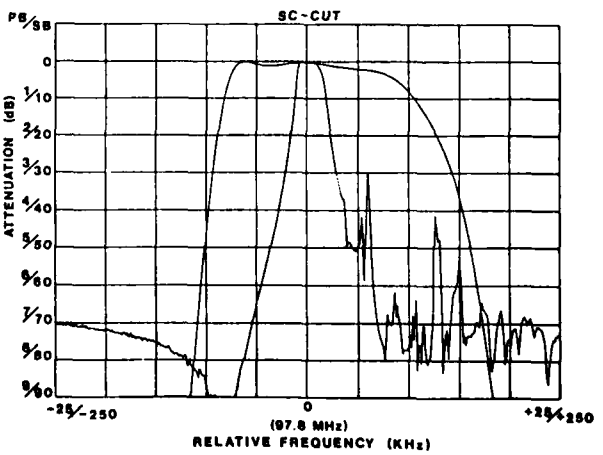


FIG. 9.

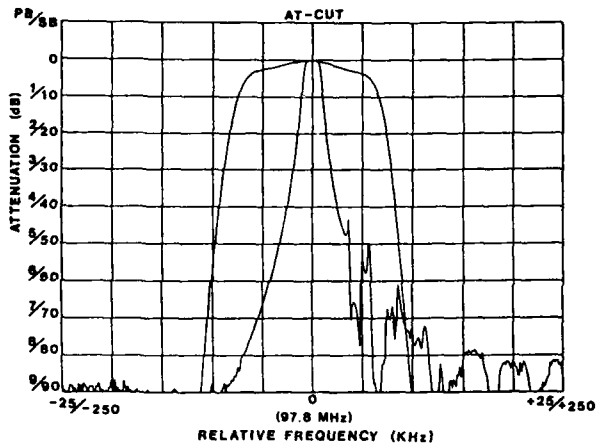


FIG. 10.

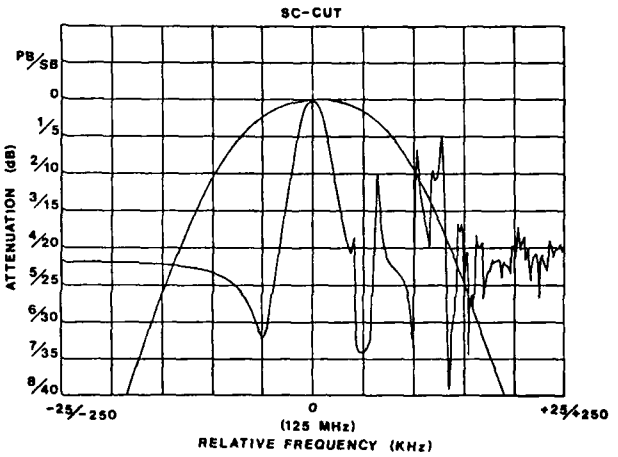


FIG. 11.

TABLE 7. TYPICAL THIRD OVERTONE FILTER RESPONSE DATA

	CUT			
	SC	SC	AT	AT
Number Of Poles	2	4	2	4
Source/Load Impedance (Ohms)	3000	4500	700	1000
Flat Loss (dB)	2.0	4.1	1.75	3.0
Ripple (dB)	0.1	0.1	0.1	0.0
BW_3 (kHz)	24.4	23.8	20.0	19.5
BW_{15} (kHz)	53.3	-	44.2	-
BW_{60} (kHz)	-	114	-	96.4
Min. Spur (dB)	13.6	26	23	43
Flyback (dB)	31	52	42	90

Fifth-overtone devices were made with differing electrode dimensions. Using the same electrode area as its AT-cut counterpart, the SC-cut filter would have had an impedance level in the region of 25k ohms, an unacceptably high value at 125 MHz. The electrode dimensions of the SC-cuts were therefore increased to maintain a similar impedance level to the AT-cuts; consequently,the flat loss of SC- and AT-cut devices is comparable. The spurious response spectrum of the SC-cut is somewhat worse than that of the AT but, considering the increase in electrode dimensions, is better than might be expected. The ultimate attenuation of the AT-cuts is again better than that of the SC-cuts.

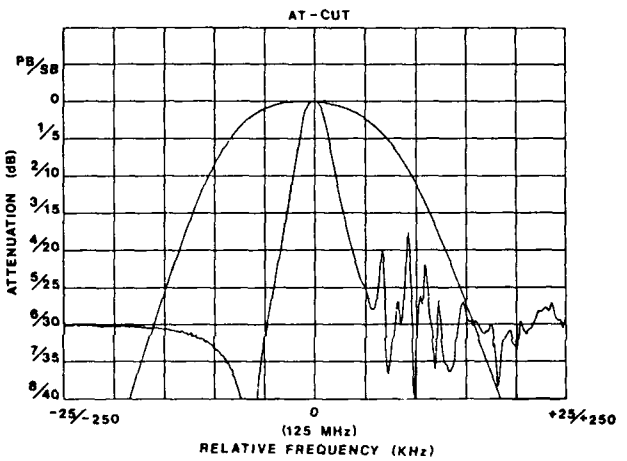


FIG. 12.

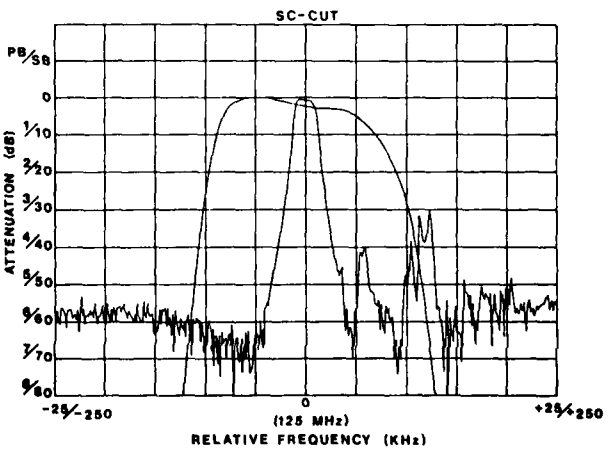


FIG. 13.

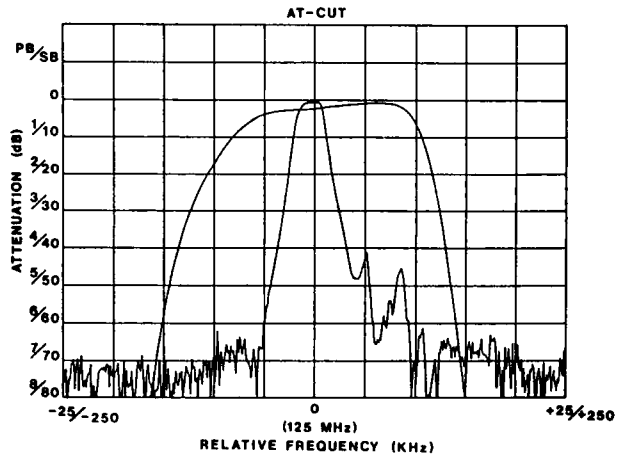


FIG. 14.

TABLE 8. TYPICAL FIFTH OVERTONE FILTER RESPONSE DATA

	CUT			
	SC	SC	AT	AT
Number Of Poles	2	4	2	4
Source/Load Impedance (Ohms)	4500	7000	4500	7000
Flat Loss (dB)	3.8	4.5	3.5	4.4
Ripple (dB)	0	0	0	0
BW_3 (kHz)	22	20.4	27	25
BW_{15} (kHz)	47	-	55	-
BW_{60} (kHz)	-	80	-	109
Min. Spur (dB)	5	31	17.5	40
Flyback (dB)	22	50	30	65

Inband IM_3 Products

An analysis of IM_3 due to anelasticity for a single resonator has been carried out by Tiersten [2, 3] for the AT-cut. An equivalent circuit representative is discussed below. For test tones of equal amplitude (the usual case) sufficiently close to resonance frequency and for a typical degree of trapping, Tiersten's result may usefully be approximated by [9, 10].

$$IM_3 = 130 + 20 \log_{10}(A_E N(BW_3/F_o)^2/P_{TT}) dB \quad (1)$$

where:

IM_3 = Ratio of P_{TT} to IM_3 product (dB)
A_E = Electrode area (sq mils)
N = Overtone
BW_3 = 3 dB bandwidth (Hz)
F_o = Filter center frequency (Hz)
P_{TT} = Power per tone (mW)

While Tiersten's analysis was restricted to single resonators, by ignoring the change in mode shape it may be applied with resonable accuracy to the two-pole case.

Equation (1) can be rewritten:

$$IM_3 = K_c + G(f) \quad (2)$$

where K_c is a constant dependent on cut and

$$G(f) = 20 \log_{10}(A_E N(BW_3/F_o)^2/P_{TT})$$

As all parameters in G(f) are either known or can be measured, G(f) can be calculated for any individual filter. The IM_3 can be measured and therefore K_c can be estimated for each cut.

Table (9) shows results estimated for K_c on all three cuts. The AT results on fundamental and third overtone are very close to the previously implied value of 130. However, the fifth overtone results on AT-cuts erroneously suggest a value of 140 for K_{AT}.

The SC-cut results are fairly uniform about a value of 136 for K_{SC}, and the one result on BT-cuts suggests a value of 134 for K_{BT}.

TABLE 9. CALCULATION OF K_{AT}, K_{SC} & K_{BT} FROM MEASURED DATA

Cut	Freq.	G(f) (Calc.)	IM_3 dB (Meas.)	K_c (IM_3 - G(f))
AT	25	-54.5	73.7	128.2
AT	97.8	-76.5	53.9	130.4
AT	125.0	-80.6	60.0	140.6
SC	25	-58.5	77.3	135.8
SC	97.8	-75.4	59.6	135.0
SC	125	-75.4	61.0	136.4
BT	25	-61.5	72.6	134.1

$$K_{AT} \doteq 130$$

$$K_{SC} \doteq 136$$

$$K_{BT} \doteq 134$$

An Analysis Of Inband IM_3 Products. An equivalent circuit representative of intermodulation in a quartz resonator is shown in fig. 15. For test tones at f_1 and f_2 the third-order IM product at $f_{IM1} = (2f_1 - f_2)$ is represented by a controlled voltage source whose rms amplitude is given by

$$V_{im} = j \alpha I_1^2 I_2^* \quad (3)$$

where:

I_1 is the current at f_1

I_2 is the current at f_2

and α is a constant dependent on resonator cut and design details.

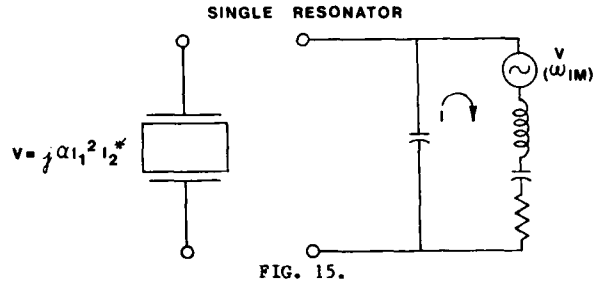


FIG. 15.

This principle can be extended to a coupled two-pole resonator structure, fig. (16). In this structure we have introduced two current-controlled voltage sources, one in each series resonant arm, the voltages generated are defined in a similar manner to the single resonator as:

$$V_{21} = j\ \alpha\ I_{12}^2\ I_{22}^* \qquad (4)$$

$$V_{41} = j\ \alpha\ I_{14}^2\ I_{24}^*$$

where:

I_{12} is the current in branch 2 f_1

I_{22} is the current in branch 2 f_2

I_{14} is the current in branch 4 f_1

I_{24} is the current in branch 4 f_2

and V_{21} and V_{41} have a frequency f_{IM1}

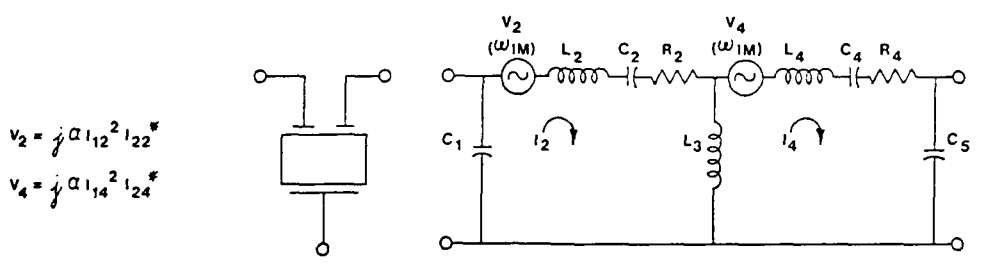


FIG. 16.

If the two-pole circuit is properly terminated and excited with a test tone P_1, at the desired power level and frequency f_1, I_{12} and I_{14} can be calculated, as may I_{22} and I_{24} for excitation by a test tone P_2 at frequency f_2. The voltages V_{21} and V_{41} at f_{IM1} are calculated from Eq. (4) and the mesh equations for the network solved at frequency f_{IM1} to give an IM_3 output level.

This procedure can be repeated for the second IM_3 product at a frequency

$$f_{IM2} = 2\ f_2 - f_1$$

where:

$$V_{22} = j\ \alpha\ I_{22}^2\ I_{12}^*$$

$$V_{42} = j\ \alpha\ I_{24}^2\ I_{14}^*$$

Using this analysis method the IM_3 generated as the test tones are swept through the filter passband can be calculated. Fig. 17 shows the predicted variation in IM_3 generated as the test tones are swept across the passbands of two-pole Butterworth and 0.5dB Tchebycheff filters. Frequency is normalized to the filter half-bandwidth and the absolute value of IM_3 generated is determined by the value of α. The peak-to-peak variation across the passband is independant of α but does show some dependence on the tone separation. The results shown are for a tone separation of 5% of the 3dB bandwidth.

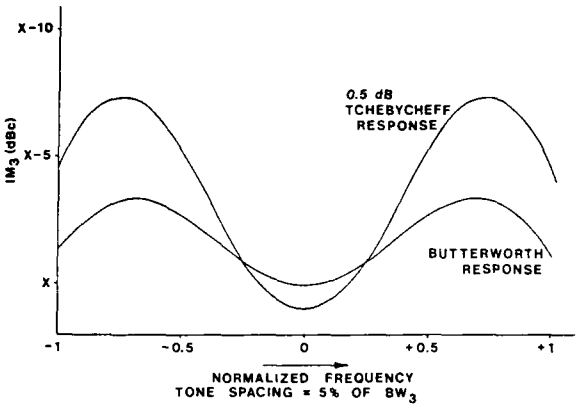


FIG. 17.

These results indicate that if measurements are made at filter center frequency then the IM_3 generated will be only slightly dependent on filter response type. For the Butterworth case the mid-band value will be the same as the single-resonator value [9, 10].

The results of this analysis are compared in figs. 18 & 19 with measured values of IM_3 for the 125 MHz AT- & SC-cut, fifth overtone units. The SC-cut unit has a good correlation with predicted performance, especially over the upper and center portions of the filter passband. In the lower passband measured results deviate slightly from those predicted.

The AT-cut units, however, show very little agreement with predicted performance except in overall response shape. The measured response is discontinuous in at least one section of the passband and possibly more.

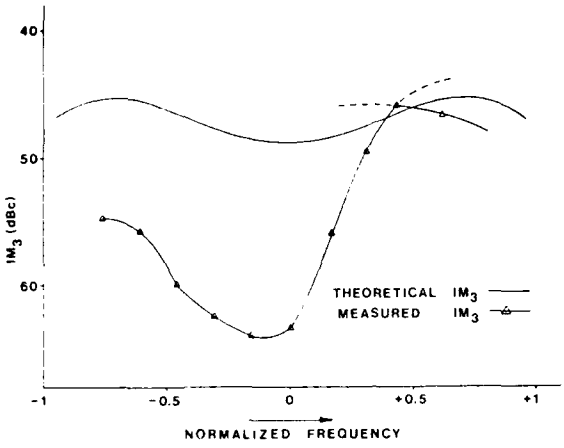


FIG. 18.

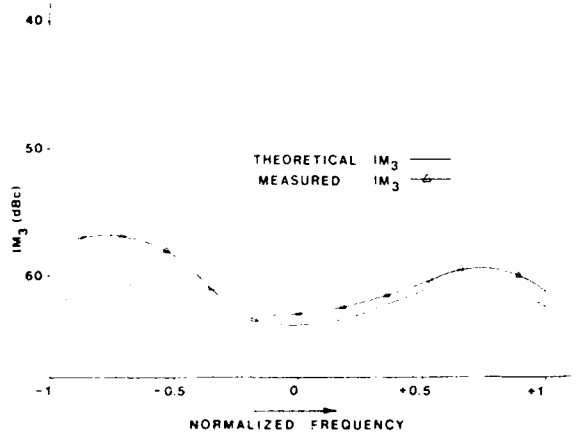


FIG. 19.

It is thought that the high value of K_{AT} obtained on the fifth overtone devices was due to non-linearly coupled modes, discussed below, which may be considered as introducing a second current-controlled voltage source at the frequency of IM_3 generation. As the phases of the two sources are presumably independent the situation can arise where the two are out of phase and a cancellation of generated IM_3 occurs. This phenomenon is probably not controllable and therefore not useful. The validity of this hypothesis is strengthened by the good agreement of the SC-cut device with predicted performance and the fact that the only deviations of the SC-cut results over the passband frequency range are minor and are in the vicinity of a weakly coupled mode.

Power Handling Capabilities

Two-pole filter responses are plotted at drive levels between 0dBm and 25dBm. The filter response plots are overlaid to show the shift in frequency with increased drive and deterioration of response shape as coupled modes are introduced at the higher drive levels.

Figs. 20 & 21 show the 125 MHz AT- and SC-cut fifth overtone results. The AT-cut filters shows evidence of at least three strongly coupled modes compared with one mildly excited mode in the case of the SC-cut unit. For these AT's the coupled modes may explain the IM_3 performance shown in fig. 18. Note that at 15dBm the AT-cut unit has shifted frequency to the point of being barely useful while the SC continues to handle power up to +20dBm without serious degradation.

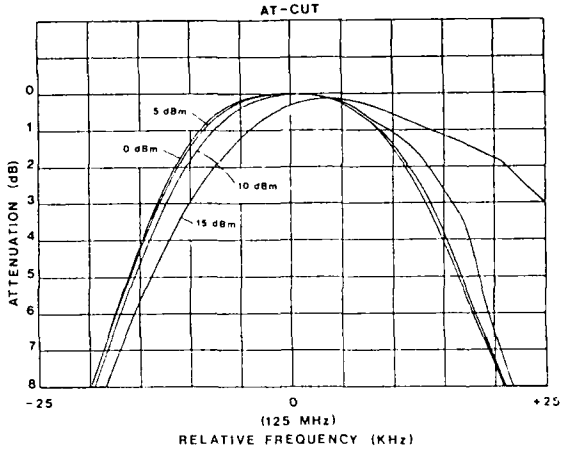


FIG. 20.

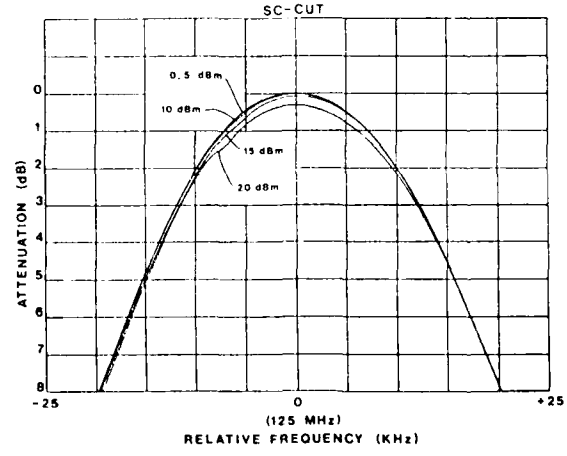


FIG. 21.

These results are repeated on the 97.8 MHz third overtone units, figs. 22 & 23. At 15dBm the response of the AT-cut filter is grossly distorted while the SC-cut remains eminently useable up to 25dBm.

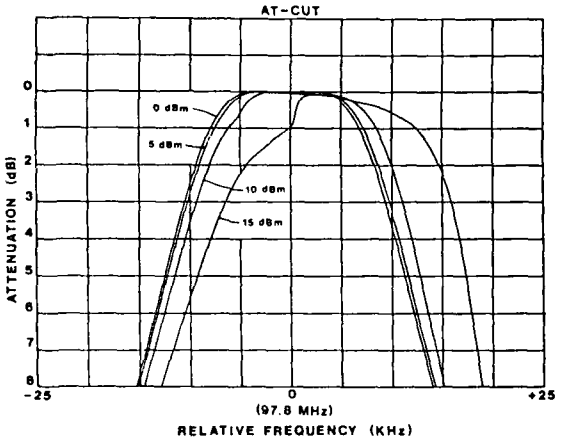


FIG. 22.

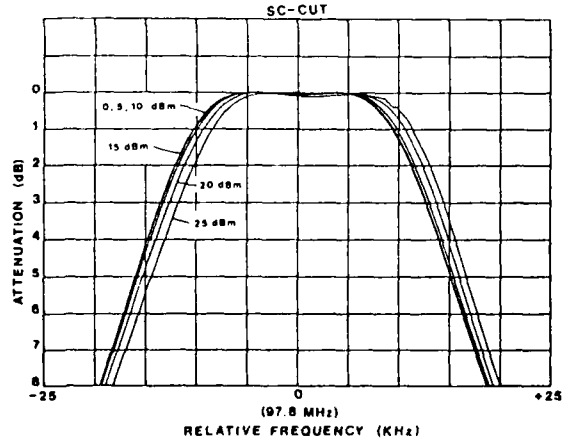


FIG. 23.

On the fundamental, since percentage bandwidths are now significantly increased, all units are handling increased power levels more effectively, figs. 24, 25 & 26. However, the AT-cut again shows evidence of three coupled modes at 25dBm which significantly distort the passband response. The BT-cut unit shows excellent power handling up to 25dBm with only a slight downward shift in frequency while the 6 responses of the SC-cut (0, 5 ... 25dBm) almost exactly overlay.

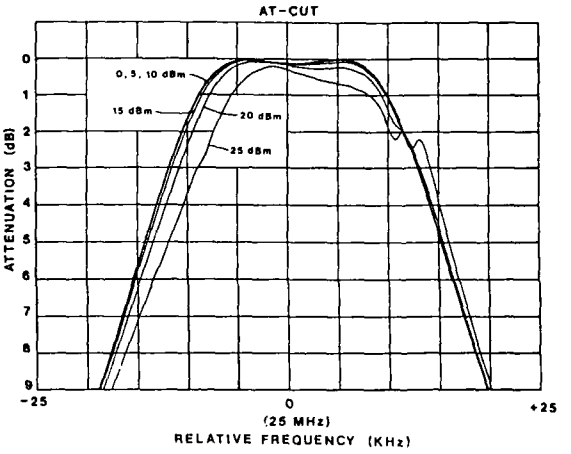


FIG. 24.

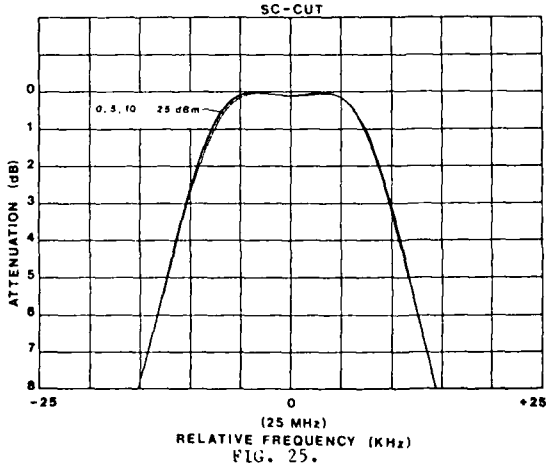


FIG. 25.

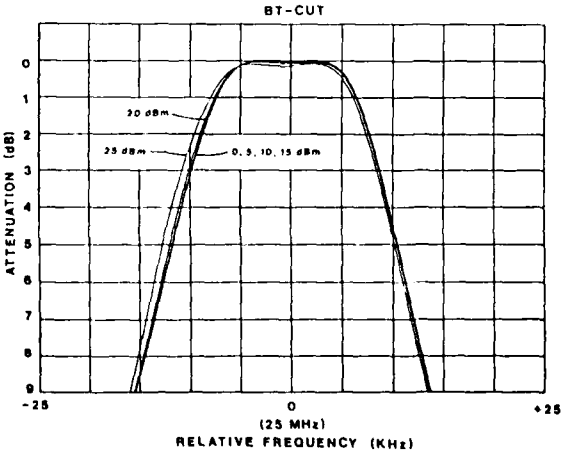


FIG. 26.

Design Trade-Offs

The improved linearity of BT- and SC-cut devices must be traded against less favorable frequency-temperature characteristics, impedance level, and capacitance ratio. These are now briefly discussed.

Frequency-Temperature Performance

Table 10 illustrates typical frequency temperature deviations that may be expected for AT-, SC- and BT-cuts over a military temperature range of -55°C to +95°C. The AT-cut offers superior performance, though the SC-cut, by proper selection of turn over temperature, offers reasonable performance. By the use of an oven or heater to limit cold end temperatures to -25°C the SC-cut performance could be improved by a factor of 2.

TABLE 10. FREQUENCY-TEMPERATURE PERFORMANCE

CUT	Δf (ppm)[1]	
	-55°C +95°	-25°C +95°
AT	±25	±15
SC	+7/-85	+0/-50
BT	+10/-260	+0/-160

Note: [1] For an angular tolerance of $\pm 2'$ on theta

Motional Inductance & Capacitance Ratios

Table 11 lists capacitance ratios (uniform mode) for the three cuts and inductance ratios referenced to an AT-cut of the same electrode area. These are used to compare maximum inductorless bandwidths for the three cuts on fundamental. The advantage of the AT-cut is quite apparent.

TABLE 11. MOTIONAL VALUES

CUT	CAPACITANCE RATIO	INDUCTANCE (REF AT)	MAX. INDUCTORLESS BW (%)
AT	159	1	0.32
SC	496	3.37	0.10
BT	390	3.74	0.14

Attenuation Levels

SC- and BT-cuts are both thicker wafers than AT-cuts and therefore have somewhat more bridging capacitance. More importantly, these two cuts have higher impedance levels. Both factors result in less ultimate attenuation for SC- and BT-cuts than for AT's.

Flat Loss and Spurious Response Trade-Offs

Particularly at higher frequencies a critical trade-off exists between flat loss and spurious response spectra. In order to reduce flat loss introduced by matching circuitry, larger electrodes are required to reduce impedance levels. However, as electrode dimensions are increased not only do unwanted modes become more closely spaced while more unwanted modes become trapped but also the closest modes become strongly trapped and severely degrade filter stopband performance. This trade-off will assume an even more delicate balance in the case of SC- and BT-cuts where, for similar mode spectra, the flat loss will be higher than that of the AT-cut due to increased impedance levels.

One other factor in this trade-off is the physical volume available. As volume is increased better coils can be introduced and the flat loss incurred to obtain improved spurious response performance can, at least in part, be recovered.

Conclusions

To the best of the authors' knowledge this is the first presentation of results on monolithic filters manufactured on SC- and BT-cut wafers. The design theory of coupled two-pole devices has been extended to cover these cuts with, in general, a good correlation between measured and calculated mode spectra.

A technique has been proposed that will assist in the analysis of IM_3 generated by multi-pole filters. While limited to two-pole filters in this paper the method is readily adaptable to any number of poles. Using this method the results of sweeping test tones though a filter passband were predicted and confirmed by measured results.

The IM_3 generated by a two-pole filter at filter center frequency has been shown to be represented by a simple equation. This equation is broken in two parts, a constant dependent on cut and a second term which is calculated from filter and resonator design considerations. The constant has been found experimentally for the three cuts. The results suggest that for the same frequency, bandwidth, overtone and electrode area BT-cuts give a 4dB improvement over AT-cuts and SC-cuts a 6dB IM_3 improvement.

Finally, SC- and BT-cuts have both been shown to have a significantly greater power handling capability than AT-cuts.

References

[1] Horton, W.H. & Smythe, R.C., "Experimental Investigations of Intermodulation in Monolithic Crystal Filters," Proc. 27th Annual Symposium on Frequency Control (AFCS), pp. 243-245; 1973.

[2] Tiersten, H.F., "Analysis of Intermodulation in Rotated Y-Cut Quartz Thickness-Shear Resonators," Proc. 28th AFCS, pp. 1-4; 1974.

[3] Tiersten, H.F., "Analysis of Intermodulation in Thickness-Shear and Trapped-Energy Resonators," Acoust. Soc. Am., v. 57, no. 3, pp. 667-681;

Mar. 1975.

[4] Smythe, R.C., "Intermodulation in Thickness-Shear Resonators," Proc. 28th AFCS, pp. 5-7; 1974.

[5] Smythe, R.C. & Morley, P.E., "Experimental Evaluation of the Effective Non-Linear Elastic Constant for Trapped-Energy and Contoured Resonators," these Proceedings.

[6] Tiersten, H.F., "An Analysis of Overtone Modes in Monolithic Crystal Filters," Proc. 30th AFCS, pp. 103-108; 1976.

[7] Stevens, D.S., "I. An Analysis of Arbitrarily Anisotropic Contoured Quartz Crystal Resonators. II. The Evaluation of the Temperature Dependence of Some Harmonic Frequencies," Ph.D. Thesis, Rensselaer Polytechnic Institute, Troy, N.Y.; 1984.

[8] Horton, W.H., & Smythe, R.C., "Theory of Thickness-Shear Vibrators With Extensions and Applications to VHF Acoustically-Coupled-Resonator Filters," Proc. 21st AFCS, pp. 160-178; 1967.

[9] Smythe, R.C., "VHF Monolithic Filters," Rpt. No. ECOM-72-0025-F, U.S. Army Electronics Command, Ft. Monmouth, N.J.; Feb., 1976.

[10] Smythe, R.C., "Bulk Acoustic Wave Filters," in Precision Frequency Control (E.A. Gerber and A.D. Ballato, ed.s), pp. 188-228; Academic Press, New York; 1985.

MONOLITHIC FILTERS USING ION ETCHED FUNDAMENTAL MODE RESONATORS BETWEEN 60 AND 100 MHz

By J. BRAUGE, M. FRAGNEAU AND JP. AUBRY

C E P E , COMPAGNIE D'ELECTRONIQUE ET DE PIÉZO ELECTRICITÉ

95 100 - ARGENTEUIL FRANCE

Ion beam milling technique which was introduced some years ago, allows to manufacture high frequency crystal resonators on fundamental mode. Filters using discrete VHF ion etched resonators are now currently delivered between 60 and 200 MHz. Typical bandwidths due to the C1 / Co ratio of the fundamental mode are reached.

In order to reduce size, cost and to increase reliability, CEPE has studied new filters using monolithic cells based on ion etched resonators.

Details on the calculation and on the manufacturing process of the experimental cells are shown between 60 and 100 MHz. Results obtained with filters using these monolithic cells are presented. Typical widest bandwidth which is reached is 140 KHz at 70 MHz, with very good shape factor, high spurious suppression and high stopband attenuation. Typical results with 2, 4, 6 and 8 poles (1, 2, 3 or 4 cells) are presented.

1. INTRODUCTION

Progresses on UHF and VHF telecommunications call for frequency generation and filtering at high frequencies. Since 1977, some VHF oscillators have been manufactured on direct frequency generation between 500 and 1000 MHz using quartz or $LiTaO_3$ ion etched resonators working on overtones.

On the same way, VHF telecommunications can take advantages of intermediate frequencies (I F) higher than the classical today's range. I F are now required between 70 and 300 MHz Other specific applications such as rythm frequency restoring require also filters in this range or higher in frequency.

Typical specifications of such filters call for wide bandwidths (between 0,01 and 2 %), low insertion losses, high stopband attenuation, low level of spuriousses, high selectivity, low shape factor, small size, low cost..

Those filters mostly require fundamental mode piezoelectric resonators, on quartz or Lithium Tantalate.

Filters using SAW, chemical etched, ion etched and, in the next years, composite resonators are main competitors in this frequency range.

The introduction of coupled resonators in the filter design will allow a great reduction of size and cost of the complete filter. It will also increase reliability and avoid some additional components.

A computer assisted plating equipment devoted to mass production of monolithic cells have been developped. It allows adjustment of both frequencies and bandwidth within only one automatic operation.

Preliminary results obtained with monolithic cells at 70 and 100 MHz are presented.

Typical results of 4, 6 and 8 poles 70 MHz band pass filter using these coupled resona-

CH2186-0/85/0000-0504$1.00©1985IEEE

tors are shown.

2. VHF RESONATORS

Since 1977, two main techniques have been in competition for manufacturing high frequency fundamental mode crystals.

Chemical etching have been intensively studied [1-3]. It seems that there are still some limitations on frequency (depending on the cuts) [3,4]. Furthermore, raw material must be carefully selected and the manufacturing process is still expensive because of visual inspections required for rejection of etch defects and because of a low industrial yield.

Moreover, the chemical etching process requires stringent safety protection due to the use of hydrofluonic acid (pure ammonium bifluoride solutions do not produce deep etched good enough surfaces).

Thanks to Marc BERTE [5], the ion beam milling technique has allowed, since 1977, to get VHF fundamental mode resonators of various cuts in various materials.

Despite the cost of the initial investment, manufacturing cost of ion etched resonators can reached reasonable values. Ion etching allows to mill simultaneously great quantities of blanks and is an effective time less process. Good yields are also an efficient money saving mean.

.2.1. ION BEAM MILLING

The ion beam etching is now a well known process. 5 mm diameter - 50 μm thick blanks are first prepared by conventional mechanical polishing. A glass mask is used which leaves visible the central area of the quartz blank only. This sandwich is exposed to an Argon ion flow. The central part of the quartz disk is then milled, down to thickness from 30 to 5 μm. (2.8 μ i.e. 600 MHz fundamental AT was obtained in Laboratory [9]). The ion milling rate is about 8 μm / hour for quartz and 5 μm per hour for $LiTaO_3$. The ion etch machine can work quite 24 hours a day because of the automatic process control.

Ion beam milling technique offers many advantages compared to chemical etch or others VHF resonators :

- quality of etched surface is not different from cuts to cuts (AT, BT, SC ...) in quartz.

- etching rate is very reproducible for a given cut

- materials such as quartz or $LiTaO_3$ are identically etched (etch rates lightly different) (up to now, there is no significant chemical etchant for $LiTaO_3$).

- size of milled area is large enough for a pair of electrodes

- flatness and parallelism are good enough under the area covered by electrodes

- high QF product on overtone (BT cut $QF > 35.10^{12}$)

Reactive ion beam etching was shown last year [7] to be a promising improvement of ion etch. Surface quality seems to be equivalent to classical ion etch one, while the etch rate is significantly increased.

.2.2. ION ETCHED RESONATORS PERFORMANCES

Since the first production of ion etched crystals in 1979, CEPE has manufactured a lot of VHF discrete resonators between 70 and 300 MHz on fundamental mode.

Typical electrical results which have been observed on hundreds or thousands of pieces of each type are summarized on Table I.

At frequencies higher than 200 MHz, the S parameter method [9] is used to perform electrical measurement.

This method allows the measurement of two stray capacitances (C2, C3) between ground and both ends of classical equivalent network (crystal is then seen as 4 ports network). The Co values in table I do not take into account these stray capacitances.

For 600 MHz fundamental mode resonator, the ratio Co / C1 = 855 is still high compared to the theoretical value. C2 and C3 do not take into account all parasitic effects. Around 600 MHz, the equivalent 4 ports network does not describe completely the resonator (or the resonator is degraded at such a fundamental frequency).

For 200 MHz SC cut fundamental and 1060 MHz BT cut fifth overtone, the Co and L measured by S parameters do agree very well with theoretical values (experimental Co / C1 : SC cut = 900 ; BT cut $\simeq$ 300 x n^2).

The stray capacitances C2 and C3 seem to be due to the kind of can which have been used. Reduction of these capacitances is required in most of the VHF filter applications.

This has been achieved by a new can with internal connection lengths as low as possible. Stray capacitances have been reduced by a factor 4.

.2.3. COUPLED RESONATORS DESIGN

The Tchebyscheff filter synthesis gives the resonant frequencies (symmetric and antisymmetric) to be reached and the electrical parameters required. Through classical formulae based on investigations by Sykes, Beaver, Sheahan, Muir et al. (11.14) one can reached the resonators design parameters :

$$\frac{Fa - Fs}{Fo} = \frac{C}{Fo^2 \phi^2} \exp (- \xi d)$$

where :

- Fa and Fs are the antisymmetric and symmetric frequencies,
- Fo is the unplated quartz frequency, d is the electrode separation
- $\xi = A \Delta^{-1/2} / t$ is the propagation constant for thickness shear, t is the thickness of crystal slab, Δ is the plateback
- $\phi = d + 2 \ell + \frac{Bt}{\Delta^{1/2}}$ is the effective length of motional volume, L is the length of the plate along X (or Z') - crystallographic axis and A, B and C are constants which depend on elastic coefficients Cij.

The crystal bandwidths logarithm versus plateback are shown on Fig. 4.There is no large difference between experimental and theoretical results (except for slope of the experimental curve which is more important than the theoretical one).

.2.4. AUTOMATIC PLATING

To reach as soon as possible industrial quantities, an automatic plating equipment have been designed. This equipment shown on figures 2 and 3 is mainly based on :

- vacuum pumps
- plating and masking (controlled by relay)
- visualisation
- frequency measurements
- micro computer

This original technique cancels first metallization and manual finishing adjustment. The automatic plating equipment devoted to mass production of monolithic cells have been developped to allow adjustment of symmetric and antisymmetric frequencies and bandwidth in a very short period of time.

The symmetric frequency is adjusted by loading the common electrode and the bandwidth by loading the coupled side. The equilibrium between the two half-cell frequencies is obtained by loading only one of coupled side plated areas.

For the new 70 MHz ion etching monolithic resonators, the automatic plating system was transformed by adding a high frequency tra-

cking oscillator (up to 300 MHz).

.2.5. COUPLED RESONATORS RESULTS

The main parameters we can use to characterize a monolithic cell are :

- frequency adjustment (symmetric resonance and bandwidth)
- electrical parameters
- Q factor
- Spuriousses
- f (T) behaviour of both symmetric and antisymmetric frequencies.

Three different types of cells were performed to build 70 MHz filters :

- 1) Monolithic cells for 8 poles filter (140 KHz BW - Fo = 70 MHz)

* 1A) External cell

F_{sym} : 69 904 000 Hz
B W : 103 000 Hz
R : 11 ohms
L : 1,7 mH
Co : 4,10 pF
Co/C1 : 5 300
Q : 67 500
Spur Frec : 300 KHz

* 1B) Internal cell

F_{sym} : 69 930 000 Hz
BW : 70 000 Hz
R : 10 ohms
L : 1,6 mH
Co : 3,90 pF
Co/C1 : 4 700
Q : 69 000
Spur Frec : 400 KHz

- 2) Identical cells for 4 poles filters (140 KHz BW - Fo = 70 MHz)

F_{sym} : 69 912 000 Hz
BW : 93 000 Hz
R : 11 ohms
L : 1,7 mH
Co : 3,80 pF
Co / C1 : 5 200
Q : 71 000
Spur Frec : 420 KHz

Fig. 5 gives the transmission response of such cells. One can notice the low level of spuriousses and the cleanliness of the spectrum close to the carrier.

Fig. 6 gives the f (T) behaviour of both frequencies (f sym and f antisym) of a monolithic cell.

One can see that very flat response close to the " o " angle are obtained. In a batch of such cells, the dispersion of frequencies and parameters are :

$\Delta F/F < 9.10^{-6}$
$\Delta L/L < 3\%$
$\Delta R/R < 7\%$

Fig. 7 gives the impedance level variation of the highest frequency of two cells as a function of drive level. One can observed a more important variation for the vacuum encapsulation resonator than for the nitrogen atmosphere resonator.

3. MONOLITHIC CELLS FILTERS RESULTS

Coupled resonators have been used to build up filters with 70 MHz as central frequency. Two, four, six and eight poles have been tried. We have already discussed about the stray capacitance, C2 and C3 of VHF resonators or cells.

The filters we have manufactured have been degraded by these capacitances, mostly on the upper frequency side. In prototypes filters, this effect has been balanced by coils. Low stray capacitance cans will avoid such coils.

* TWO POLES

Fig. 8 gives transmission characteristic of a two poles filter, built around one coupled cell.

The insertion loss is very low : 1.5 dB and the 3 dB bandwidth (+ 68 KHz / - 70 KHz around Fo = 70 MHz) is close to the expected one.

Nevertheless, this filter has a too rich spurious spectrum.

* FOUR POLES

This filter has been built around two identical cells. Filter response is shown on fig. 9.

The insertion losses are about 2 dB and the 3 dB points are located at + 60 KHz and - 66 Khz around Fo.

40 dB points are - 170 and + 140 KHz, 60 dB one are - 280 and + 202 KHz away from Fo.

In this filter, the upper side bandwidth is also affected by stray capacitance and balancing coils were used. Filter bandwidth is quite symmetric around Fo.

Spuricusses are down to 40 dB. Stopband attenuation reaches 70 dB.

* SIX POLES

Transmission characteristic of a 6 poles monolithic filter is shown on fig. 10. The 3 dB bandwidth is given + 67 and - 66 KHz versus 70 MHz. Coils have been used to balance stray capacitances.

Stopband attenuation is about 80 dB, spuriousses are down to less than - 60 dB. 40 dB points are at + 112 and - 113 KHz, 60 dB are at + 153 and - 160 KHz away from Fo. Insertion loss is lower than 2.6 dB.

* EIGHT POLES

A eight poles filter has been manufactured with 2 x 2 identical cells.

Fig. 11 shows the actual design of this filter, including the coils.

Fig. 12 gives the frequency response of this filter. The 3 dB points are at + 68 and - 66 KHz, the 40 dB ones at + 94 and - 91 KHz, the 60 dB ones are + 114 and - 115 KHz.

The insertion loss is only 3.1 dB. The filter becomes more symmetric as far as the number of poles increases.

Eight poles filter has a dynamic range higher than 80 dB and is very clear from spuriousses. The main one is 400 KHz away from Fo and is lower than 70 dB versus passband level.

Other filters, mainly around 100 MHz have been built up and tested. They show the same behaviour that the 70 MHz ones.

Up to now, we dealt mainly with wide band filters. Typical bandwidth around 0,2 % (compared to $\frac{C1}{Co} \simeq 0,3$ %) were obtained.

On the other hand, by adding coupling components between cells, it is possible to reduce bandwidth. The lower limit is given by the insertion loss which is allowed and by the Q factor of resonators. For example, if one allows 6 dB IL, the Q factor being 120 000, bandwidths as low as 0,07 % (± 25 KHz) or lower can be achieved.

CONCLUSION

Acoustically coupled monolithic cells have been fabricated on VHF ion etched quartz crystal blanks. The ion etched surface quality leads to very high Q resonators (Q = 120000 at 70 MHz).

An automatic plating equipment which allows to adjust simultaneously frequencies and bandwidth is now currently used to manufacture coupled cells.

The main trouble we have met in the course of this study is due to the stray parasitic capacitances of the crystal can which induces rounded upper side corner of the passband response.

This problem is solved by the use of additional coils. Low stray capacitance holders will avoid these coils.

High performances 2, 4, 6 and 8 poles filters using such cells have been tested. Insertion losses between 2 and 3 dB, shape factors between 3.8 and 1.7 were obtained (4 - 8 poles). Stopband attenuation higher than 80 dB is reached with 6 or 8 poles.

The use of monolithic cells allows to reduce greatly size, cost and to increase performance and reliability of new wide band VHF filters.

REFERENCES

1. **J. R. VIG, J.W. LEBUS and R.L. FILLER**
   Proc. 31st Ann. Freq. Control Symp. 1977
   p. 13

2. **J.R. VIG, R.J. BRAUDMAYR and R.L. FILLER**
   Proc. 33rd Ann. Freq. Control Symp. 1979
   p. 351

3. **W.P. HANSON**
   Proc. 37Th Ann. Freq. Control Symp. 1983
   p. 261

4. **P.J. KAVOLIS and W.P. HANSON**
   Proc. 38TH Ann. Freq. Control Symp. 1984
   p. 251

5. **M. BERTE**
   Electronics Letters Vol. 13 1977 p. 245
   Proc. 31st Ann. Freq. Control Symp. 1977
   p. 122

6. **J.P. AUBRY**
   Proc. of IEEE Ultrasonic Symposium 1983
   p. 478.

7. **J.S. WANG, S.K. WATSON and K.F. LAU**
   proc. 38Th Ann. Freq. Control Symp. 1984
   p. 101

8. **B. D'ALBARET and P. SIFFERT**
   Proc. 36Th Ann. Freq. Control Symp. 1982
   p. 405

9. **J.P. AUBRY, E. GERARD and S. LECHOPIER**
   Proc. 37Th Ann. Freq. Control Symp. 1983
   P. 306.

10. **L. BIDART and J. CHAUVIN**
    Proc. 35Th Ann. Freq. Control Symp. 1983
    P. 365

11. **A.J.L. MUIR**
    Systems Technology 1973 No 17 p. 16

12. **D.S. SAEANAN**
    Proc. IEEE 1970 P.260

13. **R.A. SYKES and W.D. BEAVER**
    Proc. 20Th Ann. Freq. Control Symp. 1966
    P. 288

14. **W.D. BEAVER**
    Proc. 21st Ann. Freq. Control Symp. 1967
    p. 179.

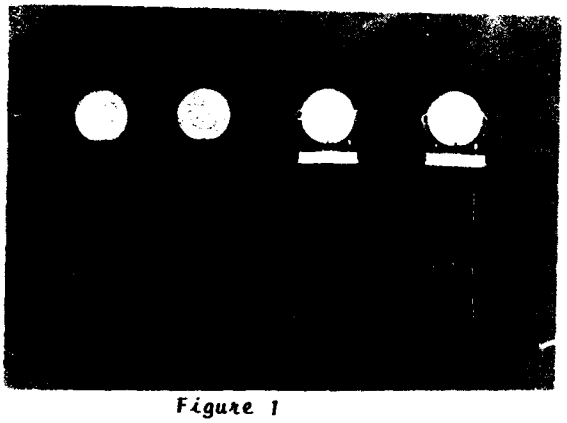


Figure 1

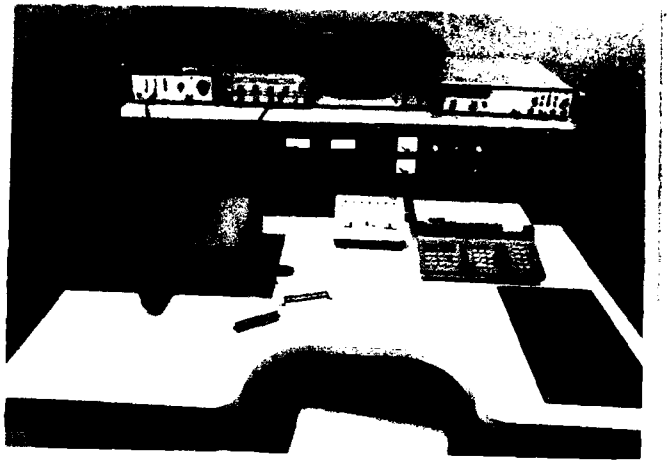
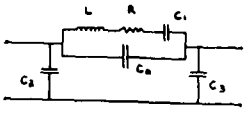


Figure 2

VHF ION ETCHED RESONATORS
STRAY CAPACITANCES



Fo	CUT/OVERTONE	Co	C2,C3	C1/Γ	EXACT C1/Co	$\frac{C1}{Co} = 8\frac{k^2}{\pi^2}$
70 MHz	AT FUND	0.63	0.4/0.4	1/500	1/340	1/160
200 MHz	SC FUND	0.9	0.6/0.6	1/1300	1/900	1/640
600 MHz	AT FUND	0.4	1/1	1/2000	1/800	1/160
1.06 GHz	BT 5TH	0.34	2/2	$1/1100.N^2$	$1/300.N^2$	$1/300.N^2$

STRAY CAPACITANCES ARE OF GREAT INFLUENCE IN HIGH FREQUENCIES CRYSTAL RESONATORS.

Table 1

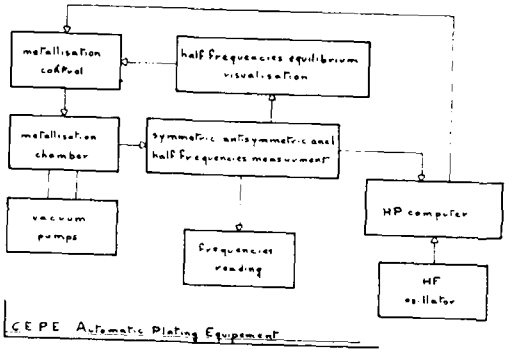


Figure 3

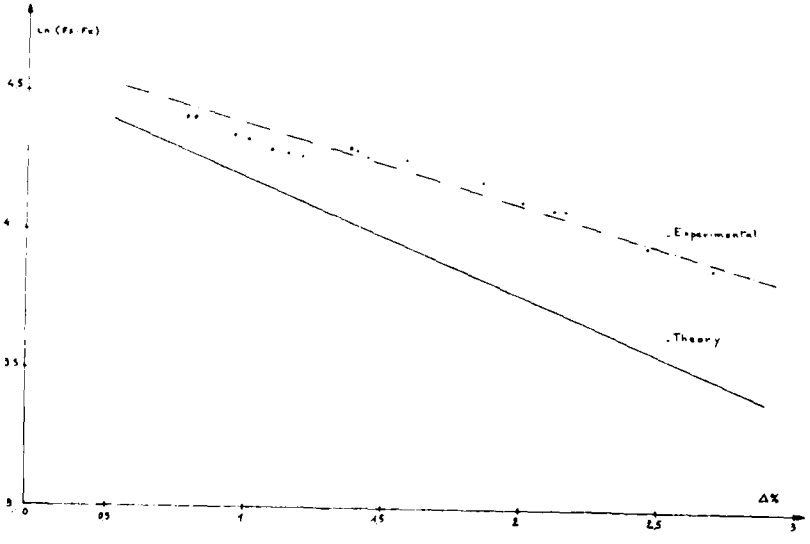


Figure 4

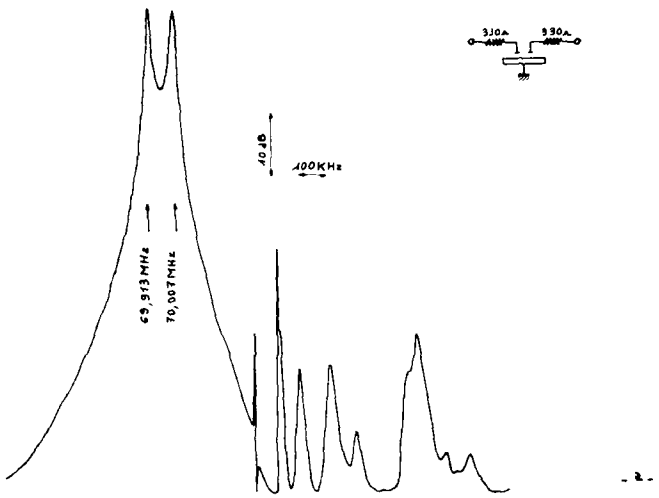
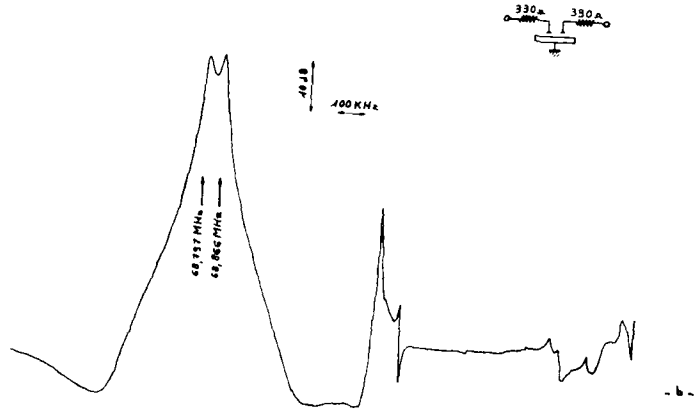


Figure 5
Monolithic cells transmission response



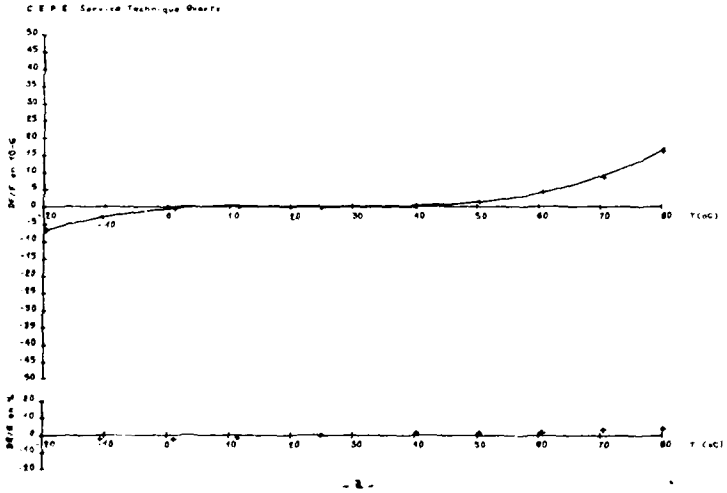


Figure 6 a)
δ (T) behaviour symmetric mode.

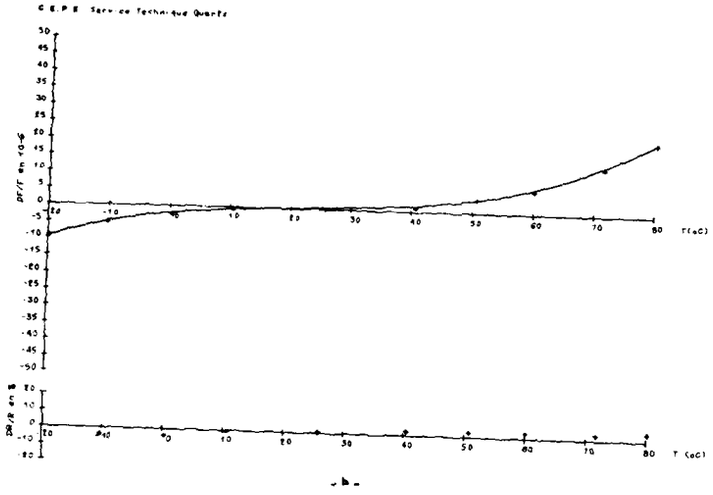


Figure 6 b)

δ (T) behaviour antisymmetric mode

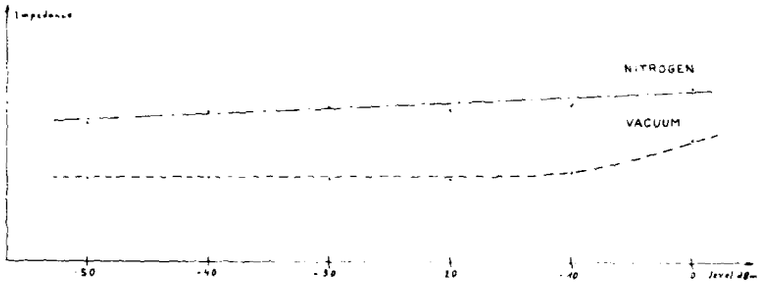


Figure 7

Impedance versus drive level

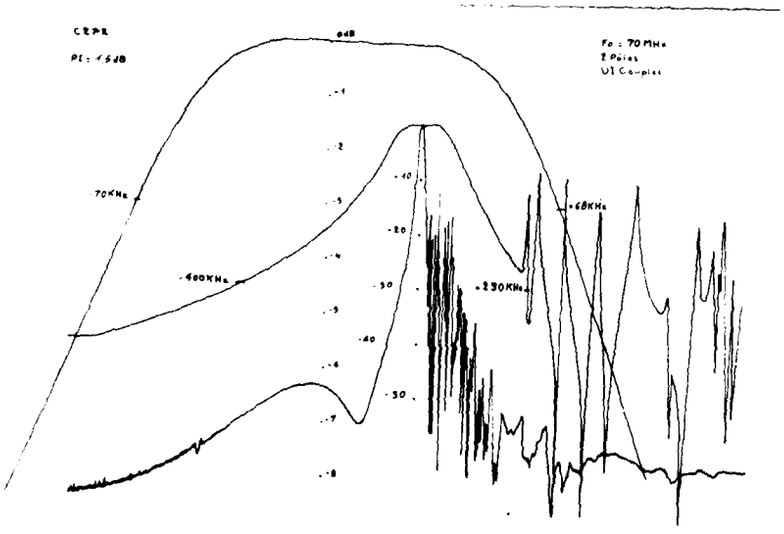


Figure 8

2 Poles filter frequency response

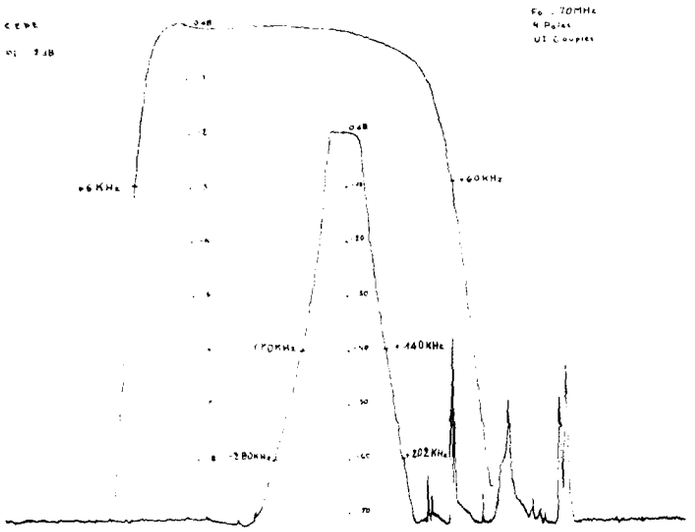


Figure 9
Four poles transmission response

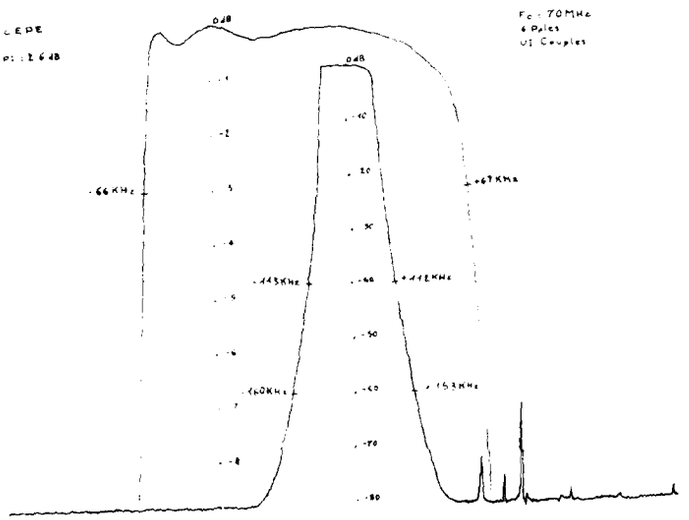
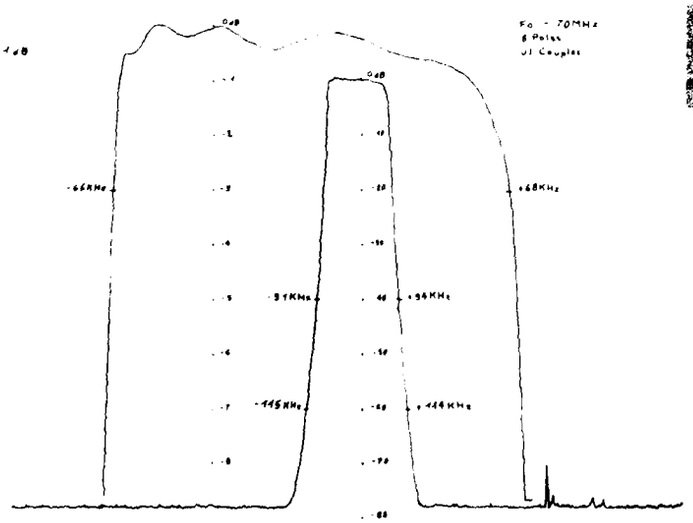


Figure 10
Six poles transmission response



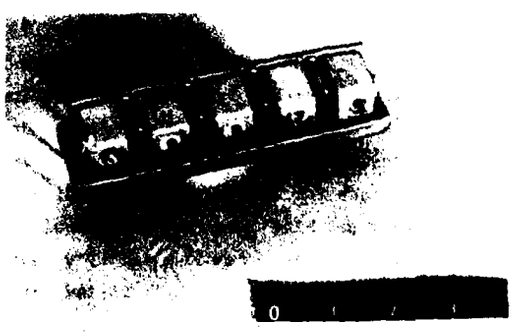


Figure 11
Eight poles filter design

Figure 12
Eight poles response

SCATTERING PARAMETERS OF ELECTRICALLY LOADED S.A.W. GROUP TYPE UNIDIRECTIONAL TRANSDUCERS

A. R. Reddy

Indian Telephone Industries, R & D
Bangalore 560016, India

S. K. Lahiri

Department of Electronics and E C E
Indian Institute of Technology
Kharagpur 721302, India

The paper presents a simple analysis of the scattering parameters of a group type unidirectional transducer (GUDT). Simple analytical expressions are derived for the different scattering parameters as a function of a directivity parameter defined as the ratio of acoustic wave amplitudes launched in the backward and forward directions when the GUDT is properly phased. The theoretical plots of the scattering parameters vs. load resistance at the resonance frequency are compared with experimental results on a 21·8 MHz GUDT on YZ-$LiNbO_3$ with 4 % fractional bandwidth. There is a satisfactory quantitative agreement between theoretical and experimental results excepting for the acoustic reflection and transmission coefficients. The discrepancy in the results of reflection and transmission coefficients is attributed to second-order reflection phenomena which are not considered in the present analysis. The analysis may prove useful for the design of a low loss filter by providing easy means of quick computation of the important scattering parameters of GUDTs.

Introduction

Group type unidirectional transducers (GUDT) driven by quadrature phase signals / 1 / have been found to be much useful for the realization of SAW low loss filters in recent years. GUDT with a meander ground busbar does not involve any complex multi-layer cross over technique for fabrication and requires simpler phasing-cum-matching networks as compared to three phase unidirectional transducer (TPUDT), while its pass band response is comparable to that of TPUDT. These transducers have found practical use since long, but little attention has so far been paid to analyze their scattering properties. Scattering parameters of ordinary bidirectional interdigital transducers (IDTs) have been studied extensively by various workers / 2 - 4 /. These parameters are very useful in the design of different SAW devices. Farnell, et al. / 5 / made a crossed-field model analysis of unidirectional transducers and extended their analysis to study scattering parameters. The analysis covered broadly the different UDTs with multi-phase drive, but did not consider the specific case of GUDT. It yielded fairly accurate results obtained by lengthy computations. The present authors / 6 / studied the frequency response of various scattering parameters of both IDT and GUDT under phased and matched condition by deriving simple analytical expressions based on crossed-field model equivalent circuit. In this paper the authors present a simple analysis of the scattering parameters of GUDT to study their dependence on external load. Experimental results are also presented and compared with theoretical results.

Theory

A typical GUDT with the minimum component phasing-cum-matching circuit / 7 / for series tuning configuration is shown in Fig. 1. Under phased condition

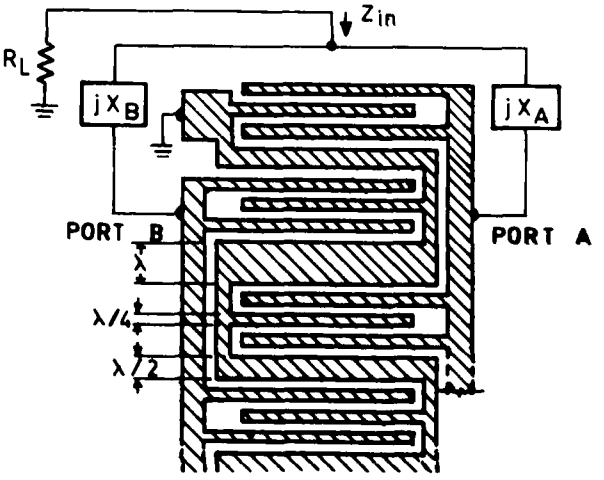


Fig. 1. The schematic of a GUDT with the minimum component phasing-cum-matching circuit.

CH2186-0/85/0000-0514$1.00©1985IEEE

the currents entering the two ports (A and B) are of equal magnitude and have a phase difference of 90^{o}. For perfect phasing at the resonance frequency f_o, the two external series reactances, X_A and X_B, are to be made equal to / 7 /

$$X_A = a(\sin\theta + \cos\theta) \qquad \cdots (1)$$
$$X_B = a(\sin\theta - \cos\theta)$$

where $a = [G_o^2 + (2\pi f_o C_T)^2]^{-\frac{1}{2}}$

$\theta = \tan^{-1}(2\pi f_o C_T / G_o)$

G_o is the radiation conductance of each port at resonance and C_T is the static capacitance of each port. In most cases, one of the reactances is positive and the other is negative thus representing inductance and capacitance respectively.

When the above phasing condition is satisfied at the resonance frequency, f_o by proper choice of components, the input impedance as seen by the external load (R_L) is

$$Z_{in} = a\cos\theta = R_o \qquad \cdots (2)$$

which is a pure resistance. R_o is the radiation resistance of each port at resonance.

If $R_L = R_o$, the transducer is said to be properly phased and matched at the resonance frequency. On the other hand, if $R_L \neq R_o$, the transducer is phased but mismatched.

As shown in Fig. 1, the GUDT may be considered as a three-port network with two acoustic ports radiating surface acoustic waves in the forward and backward directions, and one electrical port across which the load (R_L) is connected. Denoting the acoustic ports by 1 and 2 and the electrical port by 3, one may express the scattering matrix, S of the GUDT as

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{12} & S_{22} & S_{23} \\ S_{13} & S_{23} & S_{33} \end{bmatrix} \qquad \cdots (3)$$

Due to the reciprocity between the three ports, $S_{12} = S_{21} = S_{12}$, $S_{23} = S_{32} = S_{23}$ and $S_{13} = S_{31} = S_{13}$.

If the GUDT is lossless, the scattering matrix, S is unitary / 8 / which follows from energy conservation. Thus

$$\sum_{j=1}^{3} S_{kj} S_{mj}^{*} = \delta_{km} \; ; \; k,m = 1,2,3 \qquad \cdots (4)$$

where δ_{km} is the Kronecker delta function, and the asterisk ($*$) denotes complex conjugate.

The above condition yields

$$|S_{11}|^2 + |S_{12}|^2 + |S_{13}|^2 = 1 \qquad \cdots (5a)$$
$$|S_{12}|^2 + |S_{22}|^2 + |S_{23}|^2 = 1 \qquad \cdots (5b)$$
$$|S_{13}|^2 + |S_{23}|^2 + |S_{33}|^2 = 1 \qquad \cdots (5c)$$

S_{33} represents the electrical reflection coefficient at port-3. As discussed above, at the resonance frequency, $Z_{in} = R_o$. This reflection occurs when the port is mismatched, i.e., $R_L \neq R_o$. Obviously, at the resonance frequency, f_o,

$$S_{33} = (1 - R_L/R_o)/(1 + R_L/R_o) \qquad \cdots (6)$$

S_{13} and S_{23} give the acousto-electric conversion from port-1 to port-3 and from port-2 to port-3 respectively. We may define a directivity parameter, d as

$$d = |S_{23}|/|S_{13}| = |S_{32}|/|S_{31}| \qquad \cdots (7)$$

which physically represents the ratio of acoustic wave amplitude from port-2 to that from port-1, when RF electrical signal at the resonance frequency is applied at port-3.

Substituting (7) in (5c),

$$S_{13} = \sqrt{\frac{1 - |S_{33}|^2}{1 + d^2}} \qquad \cdots (8)$$

$$S_{23} = \sqrt{\frac{1 - |S_{33}|^2}{1 + d^2}} \cdot d \qquad \cdots (9)$$

Eqns. (6), (8) and (9) give the dependence of S_{33}, S_{13} and S_{23} on the external load, R_L.

In order to determine the load dependence of other independent scattering parameters, viz., S_{11}, S_{22} S_{12}, we proceed as follows. If an acoustic signal of magnitude u_1 is incident on the port-1 of a phased and matched GUDT, then the induced electrical signal ($u_3 = |S_{13}|u_1$) at port-3 will be completely absorbed in the load. For a mismatched transducer, however, a part of u_3 of magnitude equal to $|S_{33}|u_3$ will be reflected back from the load, which in turn will regenerate acoustic waves. Assuming the transducer to be lossless, the magnitude of the regenerated signal at port-1 will be $|S_{13}|\cdot|S_{33}|\, u_3$. The acoustic

reflection coefficient at port-1 (S_{11}) is the ratio of the magnitude of the regenerated signal to that of the input acoustic signal. Hence

$$S_{11} = |S_{13}|^2 \cdot |S_{33}| \quad \cdots (10)$$

Substituting (10) in (5a),

$$S_{12} = \sqrt{1 - |S_{11}|^2 - |S_{13}|^2} \quad \cdots (11)$$

and from (5b),

$$S_{22} = \sqrt{1 - |S_{12}|^2 - |S_{23}|^2} \quad \cdots (12)$$

Expressions (6) - (12) for the different scattering parameters are valid for the phased condition at the resonance frequency. The dependence of the scattering parameters on the normalised load resistance (R_L/R_o) for a directivity of -10 dB as obtained from the above expressions is shown in Figs. 2 - 6.

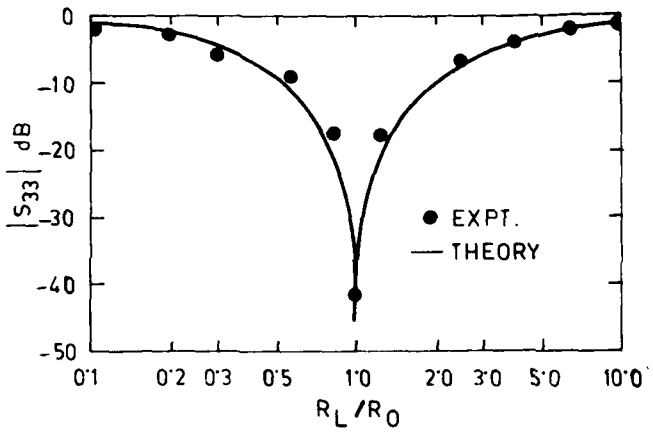


Fig. 2. Dependence of the electrical reflection coefficient (S_{33}) on load resistance.

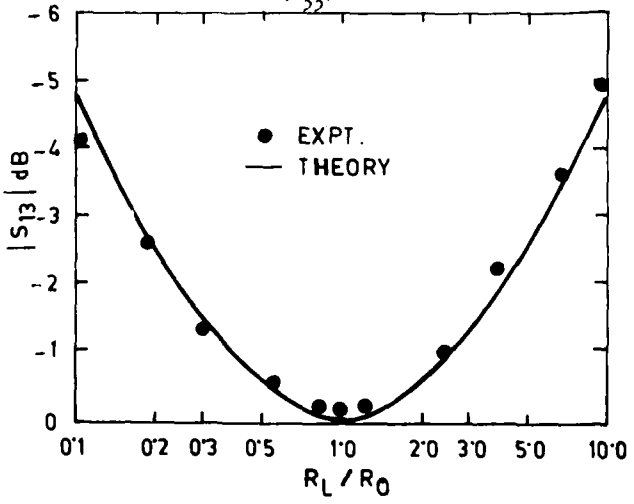


Fig. 3. Dependence of S_{13} on load resistance. Directivity -10 dB.

Experiments

For the measurement of the different scattering parameters, a GUDT with a center frequency of 21·8MHz consisting of 12 groups per port and 2 finger pairs per group, was used. The electrode pattern was defined by photolithography on YZ- $LiNbO_3$ substrate and encapsulated in a dust free metal package with RF shield. The transducer exhibited a fractional bandwidth of 4 % and a directivity of -10 dB under the best phasing condition. More precise phasing arrangement and narrower bandwidth might improve

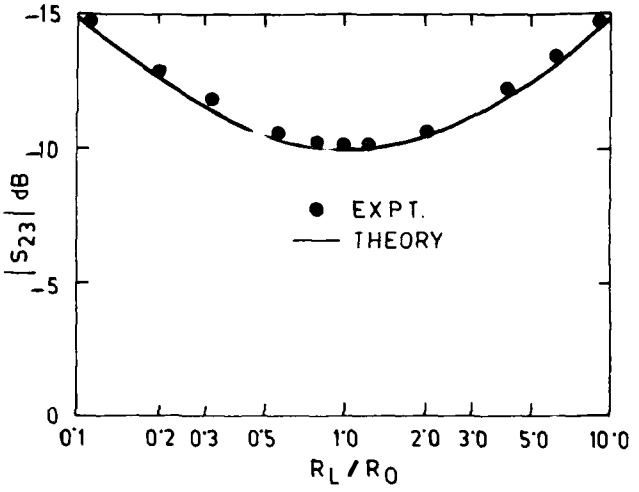


Fig. 4. Dependence of S_{23} on load resistance. Directivity -10 dB.

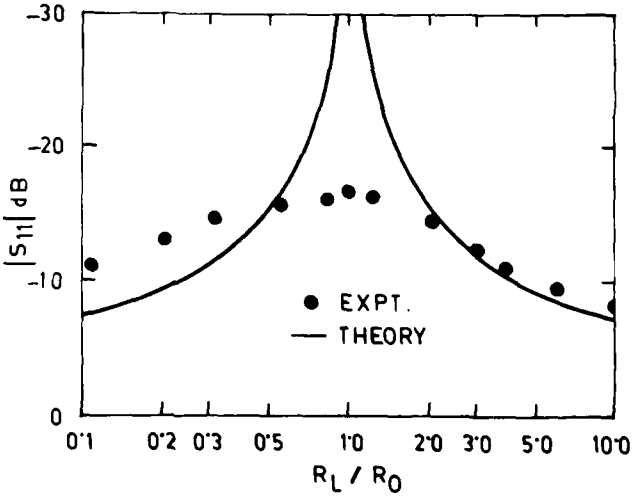


Fig. 5. Dependence of acoustic reflection coefficient, S_{11} on load resistance. Directivity -10 dB.

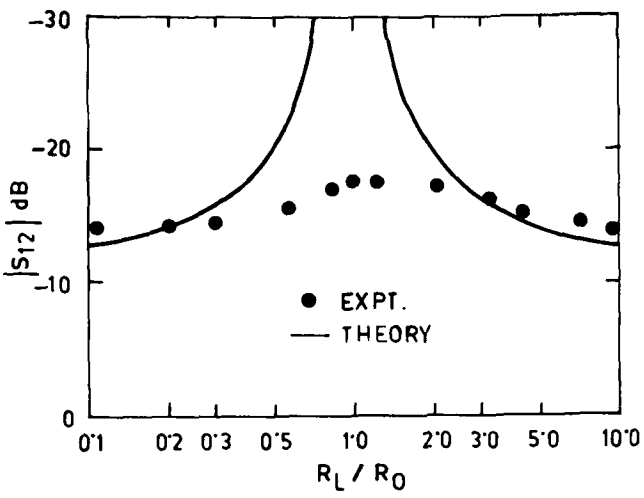


Fig. 6. Dependence of acoustic transmission coefficient, S_{12} on load resistance. Directivity -10 dB.

the directivity. The scattering parameters (excepting S_{33}) were measured by the RF burst technique/ 2 /. The latter uses an efficient input transducer at one end of the substrate to launch surface acoustic waves which travel towards the test GUDT placed on the propagation path. Two wide band IDTs are put at equal distances on either side of the test GUDT. The IDT located between the input transducer and the GUDT measures the reflected wave from the GUDT, and the other IDT measures the transmitted wave. When RF bursts at the resonance frequency (21·8 MHz) are applied at the input transducer, different delayed RF pulses are observed at the output of the test GUDT and at the two wide band IDTs. Their relative magnitudes were used to determine the values of the reflection (S_{11}, S_{22}) , transmission (S_{12}) and acousto-electric conversion (S_{13}, S_{23}) coefficients for the forward and reverse phasing conditions of the GUDT and for the various values of the load resistance. The experimental results are shown along with the theoretical curves in Figs. 3 - 6.

Since the direct measurement of S_{33} was difficult at the lower operating frequency, it was estimated indirectly with the help of the following equation obtained from (8) :

$$|S_{33}| = \sqrt{(1 - d^2)(1 - |S_{13}|^2)} \quad \cdots \quad (13)$$

The experimental values of S_{33} are shown in Fig. 2.

Discussion of Results

It may be noted that the experimental results match closely with the theoretical results for the scattering parameters representing acousto-electric conversion (S_{13}, S_{23}) and electrical reflection (S_{33}). There is, however, lot of discrepancy between experimental and theoretical results for acoustic reflection and transmission coefficients (S_{11} , S_{12}). According to the present analysis, both S_{11} and S_{12} should approach zero value (or negative infinity in dB scale) for phased and matched condition at resonance (R_L/R_o = 1), so that the entire acoustic power incident at port-1 is converted into electrical power and absorbed in the load (S_{13} = 1 or 0 dB). This is also expected physically for an ideal unidirectional transducer. The experimental results, however, show that both S_{11} and S_{12} decrease as R_L approaches R_o and attain minima (-16 to-17 dB) at $R_L = R_o$. In separate experiments / 6 / we have seen that the acoustic reflection and transmission cannot be minimised below certain value (-20 dB) even if the two ports are shortcircuited. It establishes the fact that S_{11} and S_{12} are determined not only by regeneration but also by reflections caused by piezoelectric field shorting under electrodes and mass loading. The present analysis ignores the second-order effects. The large discrepancy between the theoretical and experimental results for S_{11},S_{12} , particularly near the matching condition ($R_L = R_o$), may therefore be attributed to second-order reflection effects.

Conclusions

The simple analysis presented in this paper enables one to compute quickly the scattering parameters of a phased GUDT with a known value of directivity for different external loads. The knowledge of scattering parameters is useful in the design of low insertion loss filters made of GUDTs. For example, S_{11} determines the triple transit echo and pass band ripple. The experimental results on S_{11} vs. R_L/R_o as reported above show that S_{11} does not fall below a certain minimum (-16·5 dB) when the transducer

is phased as well as matched. Thus it is expected that the pass band ripple will be suppressed to a minimum value when the transducer is tightly matched unlike the case of an ordinary IDT which gives the minimum reflection for short circuit load. Bray, et al. /9 / observed experimentally similar characteristics of a filter made of GUDTs, which is established firmly in the present study. It is therefore concluded that GUDTs with high directivity produce very small pass band ripples in SAW filters with low insertion loss when the transducers are accurately phased and tightly matched.

References

1. K. Yamanouchi, F. M. Nyffeler and K. Shibayama, " Low insertion loss acoustic surface wave filter using group type unidirectional transducers," Proceedings of IEEE Ultrasonics Symposium, 1975, pp 317 - 321.
2. S. G. Joshi and P. Sudhakar, " Scattering parameters of interdigital surface acoustic wave transducers," IEEE Trans., vol. SU-24, pp 201 - 206, 1977.
3. R. L. Rosenberg, " Wave-scattering properties of interdigital SAW transducers," IEEE Trans., vol. SU-28, pp 26 - 41, 1981.
4. C. M. Panasik and B. J. Hunsinger, " Scattering matrix analysis of surface acoustic wave reflectors and transducers," IEEE Trans., vol. SU-28, pp 79 - 91, 1981.
5. G. W. Farnell, C. K. Jen and E. L. Cambiaggio, " Multiphase interdigital transducers and crossed-field model," IEEE Trans., vol. SU-27, pp 184 - 195, 1980.
6. A. R. Reddy and S. K. Lahiri, " Scattering parameters of interdigital and group type unidirectional SAW transducers," Journal of the Institution of Electronics and Telecom. Engineers, vol. 31, pp 46 - 51, 1985(India)
7. D. C. Malocha and B. J. Hunsinger, " Tuning of group type unidirectional transducers," IEEE Trans., vol. SU-26, pp 243 - 245, 1979.
8. H. J. Carlin and A. R. Giordano, " Network Theory : ....," Prentice-Hall, 1964, pp 222 - 281.
9. R. C. Bray, et al., " SAW filters using GUDTs : sources of problems," Proceedings IEEE Ultrasonics Symposium, 1982, pp 227-232.

39th Annual Frequency Control Symposium - 1985

A NEW ALL QUARTZ PACKAGE FOR SAW DEVICES

T.E. Parker, J. Callerame and G.K. Montress

Raytheon Research Division
131 Spring Street
Lexington, MA 02173

ABSTRACT

A new all quartz package has been developed for SAW devices. Over 35 devices covering the frequency range of 187 MHz to 425 MHz have been fabricated using this new packaging technique. With the all quartz package, the SAW substrate serves as the bottom of the package, and a matching piece of single crystal quartz is used as the cover. A glass frit provides the seal between the two quartz plates and serves as a spacer between the two inside surfaces. In all respects the device performance has equaled or exceeded the performance of similar devices in conventional TO-8 cold-weld enclosures. Aging data on six resonators shows a long term stability in the range of 0.1 to 0.5 PPM/year and a short term stability in the range of 2 to 6 x 10^{-11} for a 0.1 second gate time has also been observed on a number of devices. The vibration sensitivity is close to 1 x 10^{-9} per g. The all quartz package allows considerable flexibility in the choice of substrate mounting materials since the active surface is protected. The greatest potential of the all quartz package is to make possible very small SAW oscillators using hybrid or monolithic circuits that can deliver state-of-the-art frequency stability.

INTRODUCTION

A major obstacle in the design and fabrication of compact, low vibration sensitivity SAW oscillators has been the lack of a SAW substrate mounting technique which is compatible with achieving good long-term frequency stability. Also, in order to minimize vibration sensitivity, the back of the SAW substrate should be uniformly supported, preferably with a soft material. Unfortunately, nearly all such supporting materials outgas excessively and are thus not appropriate for achieving good long-term frequency stability. Similar problems are encountered in the fabrication of high performance hybrid circuit SAW oscillators as a means of reducing oscillator size. The high temperatures and degree of cleanliness necessary to assure good long-term frequency stability are incompatible with standard hybrid circuit components, assembly and packaging techniques. The use of a separate conventional cold-weld sealed package for the SAW device will result in good long-term frequency stability, but only at the expense of increased oscillator size. In order to address these problems a new all quartz package (AQP) has been developed which affords state-of-the-art long-term frequency stability and vibration sensitivity performance for SAW resonator based oscillators.

More than thirty five one- and two-port SAW resonators in the frequency range of 187 to 425 MHz have been fabricated using this all quartz package design. The unloaded Q (Q_u) for all devices ranged from 70 to 90% of the material limited Q. No degradation in device performance (e.g., insertion loss, loaded Q, etc.) has been noted for all quartz packaged devices, when compared to similar devices sealed in cold-weld TO-8 type packages. A number of devices have been frequency trimmed using RIE techniques prior to all quartz package sealing, and six all quartz packaged SAW resonator devices are currently undergoing oscillator aging tests. Several devices have been temperature cycled and also the short term frequency stability has been measured on a number of devices. Preliminary measurements of vibration sensitivity have also been made, and in all cases, the observed results have been as good as, if not better than the results on similar devices sealed in conventional cold-weld sealed TO-8 packages.

CH2186-0/85/0000-0519$1.00©1985IEEE

The all quartz package overcomes many of the problems encountered in obtaining exceptionally good long-term frequency stability in a miniature hybrid circuit SAW oscillator. Also, the all quartz package affords considerable latitude in the choice of SAW substrate mounting materials and techniques. With the incorporation of monolithic electronic circuits, the cover plate of the all quartz package could serve as a carrier for the electronic circuitry. Then the entire oscillator would not be appreciably larger than just the all quartz packaged SAW device itself. The potential for further reducing vibration sensitivity using the all quartz packaging technique is readily apparent. Three present areas of considerable interest involve: (1) investigating new mounting techniques and materials for all quartz packaged devices, with a goal of achieving vibration sensitivities approaching 1 x 10^{-10} ($\Delta f/f$ per g); (2) developing techniques for frequency trimming, with a goal of final frequency settability to +/- 1 ppm; and (3) extending the operating frequency range for all quartz packaged devices to 1 GHz.

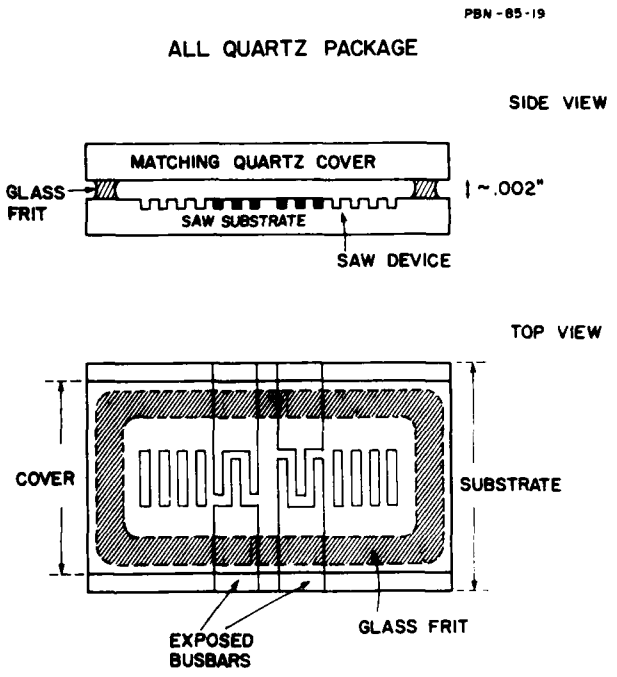


Figure 1. Diagram of the All Quartz Package (AQP) for SAW devices.

A summary of experimental results to date involving the all quartz packaging technique, along with a discussion of potential package enhancements follows.

AQP FABRICATION

Although several earlier attempts to encapsulate SAW devices using an "all quartz packaging" concept have been reported, these efforts have met with very limited success.[1,2,3,4] To the best of our knowledge, the aging performance achieved either did not compare at all favorably with that obtainable using standard cold-weld package sealing techniques, or else they were very difficult to fabricate. Our own effort to develop a two piece all quartz package (AQP) technique which would be compatible with high vacuum sealing has lead to the approach shown schematically in Fig. 1. The SAW substrate is typically 0.6 inches long, 0.5 inches wide and 0.035 inches thick. The identically oriented quartz cover plate is generally of the same length and thickness as the SAW substrate, but somewhat narrower in order to allow electrical connections to the exposed busbars connected to the SAW device transducers. A glass frit is then used to seal the two quartz plates together in a vacuum chamber. Both vitrifying and devitrifying types have been used. Figure 2 contains a photograph of a completed all quartz packaged 400 MHz SAW resonator device. The technique may be used to package either SAW resonators or delay lines. A number of sealed packages have been subjected to both coarse leak testing (bubble tester) and fine leak testing (helium) with an ultimate instrument detection limit of 4.1 x 10^{-9} cc/min with no measurable leaks observed. The transparency of the package cover assists in the observation of coarse leaks due to either inadequate frit coverage or bubbles developing within the frit during the sealing process.

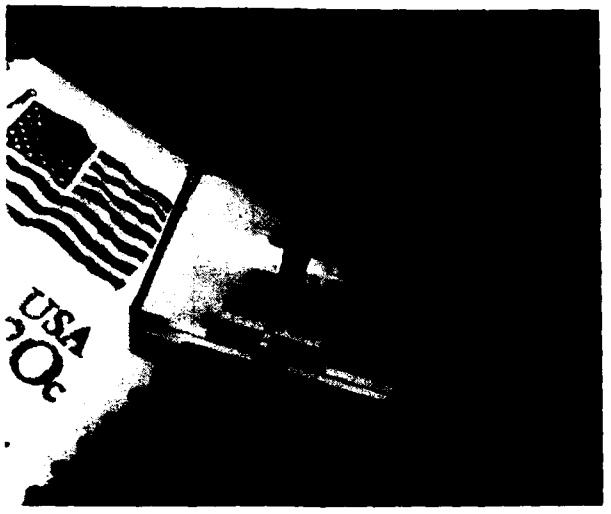


Figure 2. 400 MHz SAW resonator in an All Quartz Package.

Several all quartz packaged devices have been repeatedly subjected to temperature cycling over the -40°C to +80°C range without any leaks developing as a consequence of these temperature excursions. Figure 3 shows

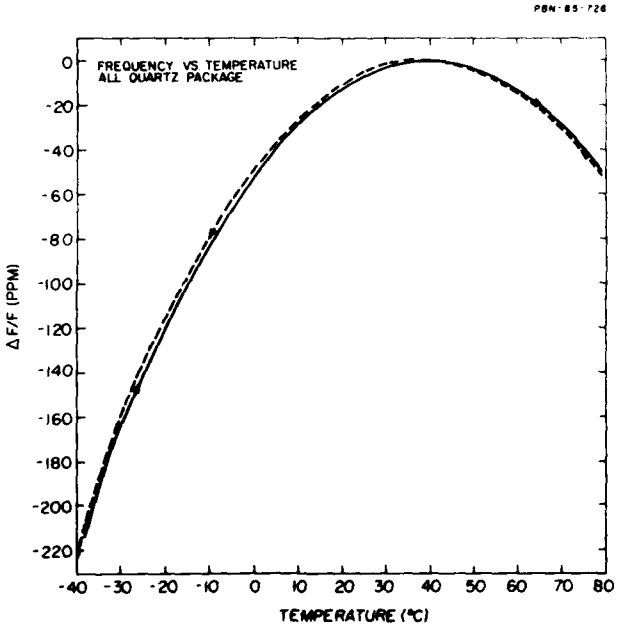


Figure 3. Resonant frequency versus temperature for a 310 MHz SAW resonator in an All Quartz Package.

a typical resonant frequency versus temperature characteristic for a 310 MHz AQP SAW resonator device. The maximum hysteresis between heating and cooling runs is only +/- 1 ppm, typical for the 1°C/minute heating and cooling rates used for the measurement. Similar results have been observed for devices packaged in cold-weld sealed TO-8 enclosures and subjected to the same temperature cycling.

A frequency shift of typically +35 ppm (e.g., ~ +12 KHz for a 400 MHz SAW device) has been observed to result from the all quartz package sealing process when applied to a SAW resonator. Table 1 summarizes our

Table 1

AQP Sealing Induced Resonant Frequency Shift

	187 MHz 2 Devices		353 MHz to 397 MHz 20 Devices		401 MHz 3 Devices		425 MHz 3 Devices	
	Avg. (KHz)	Range (KHz)	Avg. (KHz)	Range (KHz)	Avg. (KHz)	Range (KHz)	Avg. (KHz)	Range (KHz)
Frequency Shift	+25	+20 to +30	+12	+	+35	+	+35	+

measured package sealing induced frequency shifts for a number of SAW resonators designed for operation in the 187 MHz to 425 MHz frequency range. This frequency shift is somewhat less than that typically found for devices packaged in cold-weld sealed TO-8 type enclosures, and eliminates completely those frequency shifts associated with device mounting and wire bonding techniques in the cold-weld sealed TO-8 type package as well.

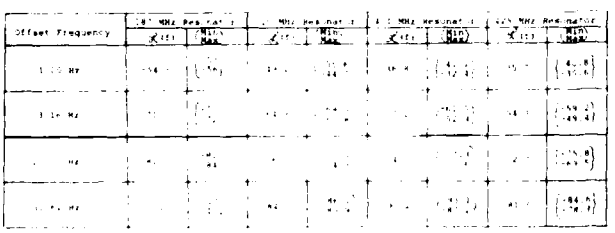
AQP SAW OSCILLATOR PERFORMANCE

A total of more than 35 one- and two-port SAW resonators in the frequency range of 187 MHz to 425 MHz have now been fabricated using the all quartz packaging technique. Many of these devices have been trimmed to final frequency using standard RIE techniques prior to AQP sealing.[5,6] No degradation in device performance has been observed, e.g.,

insertion loss, loaded Q, etc., and in fact overall device performance has in several respects been better when compared to similar devices packaged using cold-weld sealed TO-8 type enclosures.

Short-Term Stability (Noise)

Table 2 summarizes measured single sideband (SSB) phase noise to carrier levels on a number of SAW resonator oscillators, for several values of carrier offset frequency. In general, we have found less device to device variation and somewhat better performance for AQP sealed SAW resonator devices than for cold-weld sealed devices. At the present time, we are continuing to design, fabricate and test additional devices operating over a wide frequency range in order to establish the source of this apparent improvement in oscillator performance, which is perhaps due in part to the all quartz packaging technique. Table 3 summarizes measured SAW resonator oscillator fractional frequency stability for a number of gating times. Once again, these results are comparable to those found for similar devices packaged in TO-8 type cold-weld sealed enclosures, except that the scatter is somewhat reduced.



Long-term Stability (Aging)

Six AQP SAW resonator oscillators are presently undergoing oscillator aging tests. They have now been operating for periods of time ranging from twenty-seven to eighty weeks. The measured or extrapolated aging rates for all six devices are less than 1 ppm/year, typically falling in the range of

Table 3

Oscillator Fractional Frequency Stability

($\sigma_y(\tau)$, $\Delta f/f \times 10^{-11}$)

Gating Time	187 MHz Resonator		310 MHz Resonator		403 MHz Resonator		425 MHz Resonator	
	$\sigma_y(\tau)$	(Min Max)	$\sigma_y(\tau)$	(Min Max)	$\sigma_y(\tau)$	(Min Max)	$\sigma_y(\tau)$	(Min Max)
3.16 msec	5.6	(5.2 6.0)	24.4	(18.7 34.6)	4.8	(4.1 5.7)	17.5	(16.0 18.9)
10.00 msec	2.4	(2.3 2.4)	8.9	(7.1 12.1)	3.8	(3.5 4.1)	9.4	(5.2 13.5)
31.62 msec	2.4	(2.2 2.6)	3.8	(2.6 5.3)	3.6	(2.0 6.1)	4.9	(3.2 6.6)
100.00 msec	2.2	(2.1 2.2)	2.7	(1.4 4.2)	4.4	(2.1 7.1)	5.5	(3.3 7.7)

0.1 to 0.5 ppm/year. Figure 4 shows the aging characteristics for three 310 MHz AQP SAW resonator oscillators. These results, which represent current state-of-the-art performance for SAW oscillators, are comparable to results obtained using cold-weld sealed TO-8 type packages. Figure 5 illustrates equally good results obtained on AQP SAW resonator oscillators operating at 425 MHz (two) and 187 MHz (one). It should be emphasized that these results are for the first six AQP SAW resonator oscillators which were completed and put on test, they are not the results for selected devices. These experiments have been carried out with the devices housed in ovens designed to stabilize

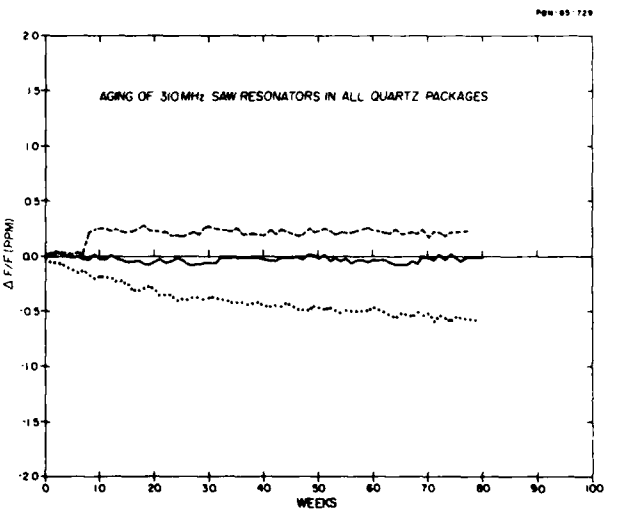


Figure 4. Aging characteristics of three 310 MHz resonators in All Quartz Packages.

the oscillators at $T \simeq 35^{o}C$, their nominal turn-over temperature. Thermocouple temperature readings, in conjunction with the measured frequency versus temperature characteristics for each oscillator, are used to correct the measured frequency data for temperature induced changes.

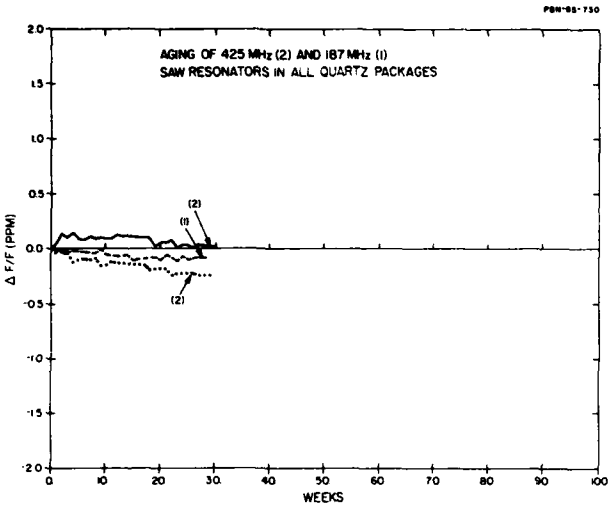


Figure 5. Aging characteristics of two 425 MHz (2) and one 187 MHz (1) SAW resonators in All Quartz Packages.

Vibration Sensitivity

The all quartz package affords considerable latitude in the choice of SAW substrate mounting materials and techniques. It is this aspect of the AQP design that holds out the prospect of developing a SAW resonator oscillator with reduced vibration sensitivity. Previously reported data for SAW oscillators have not demonstrated performance better than 1-2 x 10^{-9} $\Delta f/f$ per g).[7,8] While we are still actively investigating mounting techniques which will result in perhaps as much as an order of magnitude reduction in oscillator vibration sensitivity, we have rather readily been able to achieve comparable performance for AQP SAW resonator oscillators using a variety of mounting techniques. Figure 6 summarizes our preliminary vibration sensitivity measurements performed, on a 400 MHz SAW resonator oscillator using three different mounting techniques, namely: double sided tape, vacuum grease, and RTV.

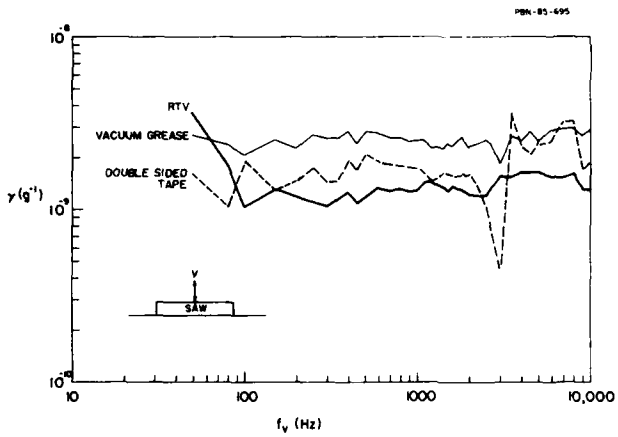


Figure 6. Vibration sensitivity of an AQP 400 MHz SAW resonator for three different mounting materials.

SUMMARY AND CONCLUSIONS

The all quartz package overcomes many of the problems encountered in achieving exceptionally good long term stability in a miniature hybrid circuit SAW oscillator. The all quartz packaging technique just described achieves state-of-the-art oscillator aging performance, while also possessing good short-term stability (noise) properties. Additionally, considerable flexibility is available in the choice of substrate mounting techniques, since the AQP SAW device is fully protected from any external sources of contamination. Figure 7 illustrates these points for a possible AQP SAW resonator oscillator design approach. The incorporation of either hybrid or monolithic electronic circuits attached to the cover plate of the all quartz package opens up new areas for SAW oscillator application, since then the entire oscillator would not be appreciably larger

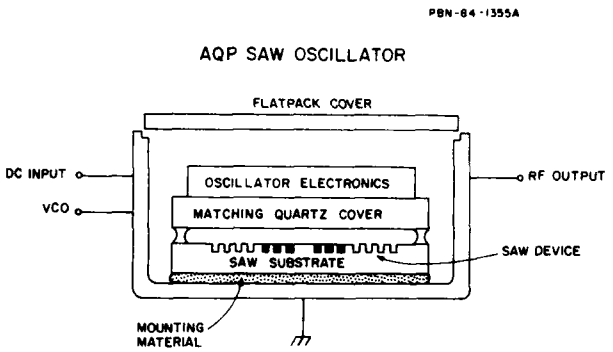


Figure 7. Illustration of a small SAW oscillator using the All Quartz Package and hybrid or monolithic electronics.

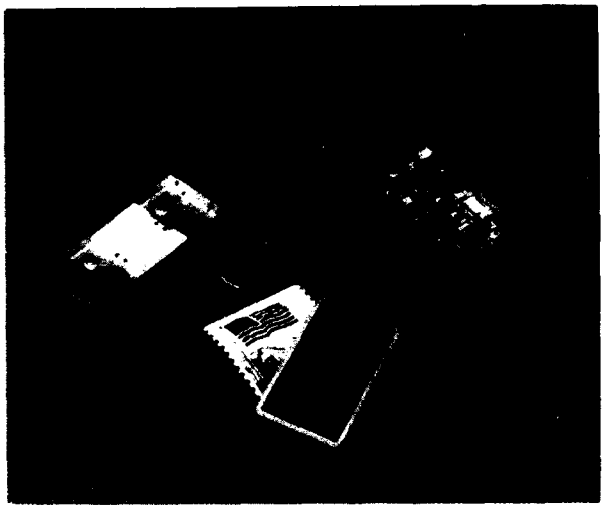


Figure 8. A hybrid SAW oscillator using the All Quartz Package.

than just the all quartz packaged SAW device itself. Figure 8 illustrates an intermediate step in this oscillator development which demonstrates the potential for reduced oscillator size and minimum dc power dissipation, while still achieving state-of-the-art oscillator performance. The 400 MHz SAW resonator oscillators entire volume is only about 0.25 cubic inches, and yet typically we find that $\mathcal{L}(f_m)$ = -65 to -75 dBc/Hz at f_m = 10 Hz, while $\sigma_y(\tau)$ = 3 to 5 x 10^{-11} for $0.01<\tau<1$ second. Total oscillator dc power dissipation is only about 20 milliwatts, for $P_{out} \simeq$ + 3 dBm. At the present time we are continuing to focus on three specific areas of interest: (1) investigating new mounting techniques and materials in order to reduce oscillator vibration sensitivity; (2) developing alternative techniques for frequency trimming; and (3) extending the operating frequency range for all quartz packaged devices to 1 GHz.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the valuable assistance of Fred Borchelt in developing the glass frit sealing technique and J. Lang, E. Sabatino and J. Columbus in fabricating devices.

REFERENCES

1. Gilden, M., G.K. Montress, D.E. Cullen, R.A. Wagner and T.M. Reeder, "Design of Stable SAW Oscillators Operating above 1 GHz," 1978 Ultrasonics Symposium Proceedings, IEEE Cat. No. 78CH1344-ISU, pp. 452-457, September 25-27, 1978.

2. Montress, G.K., R.A. Wagner and M. Gilden, "Stable Microwave SAW Oscillators for Aging Studies," 1979 IEEE Ultrasonics Symposium Proceedings, IEEE Cat. No. CH1482-9, pp. 886-889.

3. Weirauch S.F., R.J. Schwartz and R.C. Bennett, "SAW Resonator Frit-Bonded Pressure Transducers," 1979 IEEE Ultrasonics Symposium Proceedings, IEEE Cat. No. CH1482-9, pp. 874-877.

4. Parker, T.E, "Analysis of Aging Data on SAW Oscillators," 34th Annual Frequency Control Symposium - 1980, U.S. Army Electronics Command, Fort Monmouth, New Jersey, pp. 292-301. Copies available from Electronics Industries Association, 2001 Eye Street, N.W., Washington, D.C., 20006.

5. Tanski, W.J., "Surface Acoustic Wave Frequency Trimming of Resonant and Traveling Wave Devices on Quartz," Applied Physics Letters, Vol. 39, No. 1, pp. 40-42, July 1, 1981.

6. Cross, P.S., and W.R. Shreve, "Frequency Trimming of SAW Resonators," IEEE Transaction on Sonics and Ultrasonics, Vol. SU-29, No. 4, pp. 231-234, July 1982.

7. Parker, T.E., and J. Callerame, "Sensitivity of SAW Delay Lines and Resonators to Vibration," 1981 IEEE Ultrasonics Symposium Proceedings, IEEE Cat. No. 81CH1689-9, Vol. 1, pp. 129-134, October 14-16, 1981.

8. Parker, T.E., "Development of Precision SAW Oscillators for Military Applications," 36th Annual Frequency Control Symposium - 1982, U.S. Army Electronics Command, Ft. Monmouth, N.J., pp. 453-458, (1982). Systematics General Corporation, Airport Plaza, Suite 902, 2711 Jefferson Davis Highway, Arlington, VA. 22202.

A SYSTEM FOR PRECISION PARAMETER MEASUREMENTS ON QUARTZ CRYSTAL RESONATORS AND BIPOLES

R C Peach and S E Morris

GEC Research Limited
GEC Hirst Research Centre
East Lane
Wembley Middlesex HA9 7PP UK

Abstract

Supported by the UK Ministry of Defence, a comprehensive system for precision parameter measurements on crystal resonators and bipoles has been developed.

The system has been based on automatic network analysers, with calibration by APC-7 or APC-3.5 coaxial standards. Much of the initial work was performed on a Rohde and Schwarz vector analyser, but the system is now being configured for a HP 3577A network analyser. Special test jigs have been developed to allow the crystal units to be mounted at the same APC-7 (or APC-3.5) interfaces as the calibration components, thus minimising potential errors. Frequency range is restricted only by the particular network analyser employed, and not by any intrinsic limitation in the method.

The values of the equivalent circuit parameters are inferred from the S parameter data by an optimal fitting procedure. The system will perform transmission and reflection measurements on single resonators, and transmission measurements on bipole resonators. The fitting procedure uses four variable parameters for a single resonator measurement (R,L,C,C_0), and eight for a bipole. The iterative algorithm used in this method has been chosen to guarantee stability under adverse conditions.

Introduction

For a decade or more the IEC-444 pi-network [1] has been the mainstay of fundamental crystal measurements. However, it is widely appreciated that the pi-network is unsatisfactory for use above 100 MHz, and attempts to extend the range by compensation of the crystal static capacitance are less than convincing. In addition, the pi-network requires its own set of calibrated resistors and load capacitors, and the actual calibration of these has proved to be a chronic problem which is unlikely to be resolved.

For these reasons, a few years ago the UK Ministry of Defence raised a requirement for a new standard system for crystal resonator measurement, which would supersede the IEC pi-network. Considering available R.F. measurement techniques, a network analyser based system was found to offer the best combination of frequency range, availability, accuracy, and traceability. A crude demonstration system was then constructed [2], which used a generalised fitting procedure to infer the equivalent circuit parameters from the network analyser S parameter data. The MOD subsequently provided funding to develop this into a full standard reference system.

Automatic network analysers can perform measurements over very wide frequency ranges, with an accuracy which is directly related to coaxial standards of impedance. The advantages of such an approach have been realised for some time; it was first applied to crystal resonators by Pustarfi and Smith in 1973 [3], and now, with the rapidly decreasing price of suitable analysers, it is gaining wide acceptance throughout the crystal industry. Off the shelf systems do not, however, provide a direct method of determining crystal equivalent circuit parameters, and there is a considerable software problem in deducing these quantities from the basic S parameter data. In addition, to facilitate crystal measurements and allow accurate instrument calibration, special test jigs are required.

The prime objective of the project has been to develop a system with maximum accuracy and traceability. At the outset the secondary objectives were somewhat confused, and some rationalisation has been necessary. The idea of applying the method to scalar instruments, albeit with some loss of accuracy, has been abandoned in recent implementations of the software. Originally it was also intended to write all software in BASIC to facilitate program transfer between different instrument controllers. However much of this compatibility is illusory, and the present system operates specifically on Hewlett-Packard series 200 computers (models 9816-9836), with all mathematical routines in compiled Pascal code; although this somewhat restricts transferability, the improvement in performance is enormous. In some respects the scope of the original project has been extended. It was realised that, with the chosen measurement method, an extension to the important case of bipole resonators was relatively straightforward, and this type of measurement is supported in recent versions of the system.

The basic principles of the system are similar to those of the original demonstrator, although the details differ enormously. Automatic error correction is used to eliminate systematic errors from the data, and the circuit parameters of the given device are then all deduced simultaneously by an optimal fitting procedure designed to make best use of the measured data. Much of the development work has been performed on a Rohde and Schwarz ZPV vector analyser, which has a frequency range extending up to 2 GHz. The system is currently being reconfigured to operate with the new Hewlett-Packard 3577A network analyser.

Measurement jigs and system calibration

One of the principal attractions of vector network analysers for precision R.F. measurements is the possibility of using comprehensive error correction algorithms. The underlying idea of this is quite simple. The system is described by a hypothetical model defined by a number of 'error' parameters; the values of these parameters are determined by measurements on known standard components, and once determined they may be used to compute corrected values for the S parameters of an arbitrary device. For a two port analyser a 12-term error model is the most widely used. A full discussion of the choice of such models would not be appropriate in this paper, and is anyway well covered in the literature, [4,5]; however, for reference, it is worth stating the defining equations.

CH2186-0/85/0000-0527$1.00©1985IEEE

$S^M_{11}\ S^M_{12}\ S^M_{21}\ S^M_{22}$ measured S parameters

$S^A_{11}\ S^A_{12}\ S^A_{21}\ S^A_{22}$ corrected S parameters

$e_{00}, e_{01}, e_{11}, e_{30}, e_{32}, e_{22},$

$e'_{33}, e'_{32}, e'_{22}, e'_{03}, e'_{01}, e'_{11}$ error parameters

If we define matrices R, Q, E, E', A and A' by

$$R = \begin{pmatrix} (S^M_{11} - e_{00})/e_{01} & (S^M_{12} - e'_{03})/e'_{01} \\ (S^M_{21} - e_{30})/e_{32} & (S^M_{22} - e'_{33})/e'_{32} \end{pmatrix} \quad (1)$$

$$Q = \begin{pmatrix} 1 + e_{11} R_{11} & e'_{11} R_{12} \\ e_{22} R_{21} & 1 + e'_{22} R_{22} \end{pmatrix} \quad (2)$$

$$E = \begin{pmatrix} e_{11} & 0 \\ 0 & e_{22} \end{pmatrix} \qquad E' = \begin{pmatrix} e'_{11} & 0 \\ 0 & e'_{22} \end{pmatrix} \quad (3)$$

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad A' = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad (4)$$

Then S^M may be expressed in terms of S^A by

$$R = [(S^A)^{-1} - E]^{-1}A + [(S^A)^{-1} - E']^{-1}A' \quad (5)$$

and the inverse equation is

$$S^A = R.Q^{-1} \quad (6)$$

The notation for the error parameters may appear a little bizarre, but it follows conventional usage. If a number of calibrated standards are available, that is devices for which S^A is known, then by measuring their apparent S parameters (S^M), and using equation (5), the values of the error parameters can be deduced. A number of procedures are possible but, for the present system, calibration is performed by measuring three different reflection standards at each of the two ports, by measuring the transmission coefficients (S_{21} and S_{12}) with each port terminated by a matched load, and by measuring all four S parameters with the ports connected by a line of known characteristics. These 12 measurements are sufficient to determine all of the error parameters. The reflection standards can be selected arbitrarily but at the frequencies normally encountered in crystal measurements, a short circuit, a matched load, and a shielded open are the most reasonable choice. The error model supposes that the network analyser can be described by a linear network, and obviously some small non-linear errors will not be corrected by this procedure; in addition, for most analysers, there are some implicit assumptions about the reproducibility of the R.F. switches. These effects are however very small and the network analyser offers a combination of speed and accuracy which is not matched by any other approach. The error parameters themselves are functions of frequency, and a calibration at all possible measurement frequencies would require a prohibitive amount of storage; the calibration is therefore performed at selected points, maybe 100 in total, and the software then provides a system of interpolation.

The error model is referred to a reference plane at the connector interfaces, and if it is necessary to interpose a mounting jig between the connectors and the component to be measured, then some system of calibration for the jig itself is necessary. This is potentially a most unsatisfactory situation, and for this reason, special jigs have been designed which have a common plane of measurement and calibration. Two jigs have been found to be sufficient to accommodate all the commonly used types of crystal enclosure. The larger of the two jigs is shown in Figures 1 and 2;

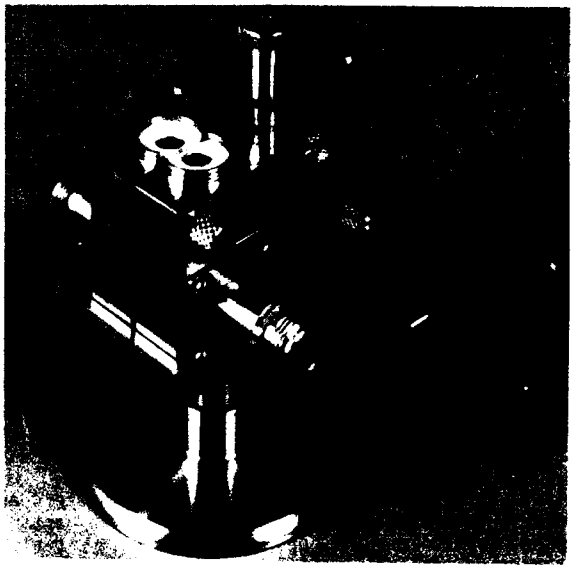


Figure 1 : APC-7 JIG.

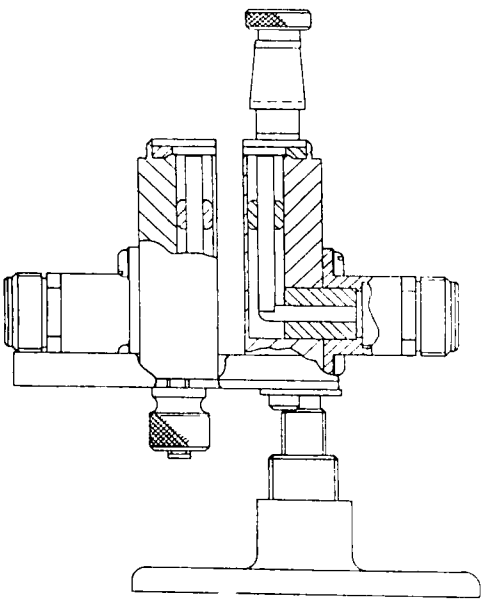


Figure 2 : APC-7 jig cross section.

coupling to the analyser is via precision N type connectors and the jig provides a 50 Ω connection to the measurement point, which is in the form of an APC-7 connector interface to allow direct attachment of calibration components. Hollow interchangeable centre pins are provided for the APC-7 connectors so that a variety of cans with different lead diameters may be accommodated. Two different types of measurement may be performed; for a transmission measurement, where one of the resonator terminals is connected to each of the

analyser ports, the jig is as shown, but for a reflection measurement, where one of the terminals is grounded, the sliding body of the jig may be replaced by an earthed pin clamp. For small crystal cans the APC-7 jig is too large for use in transmission, therefore, a similar version based on APC-3.5 connectors has also been developed, (Figure 3). This

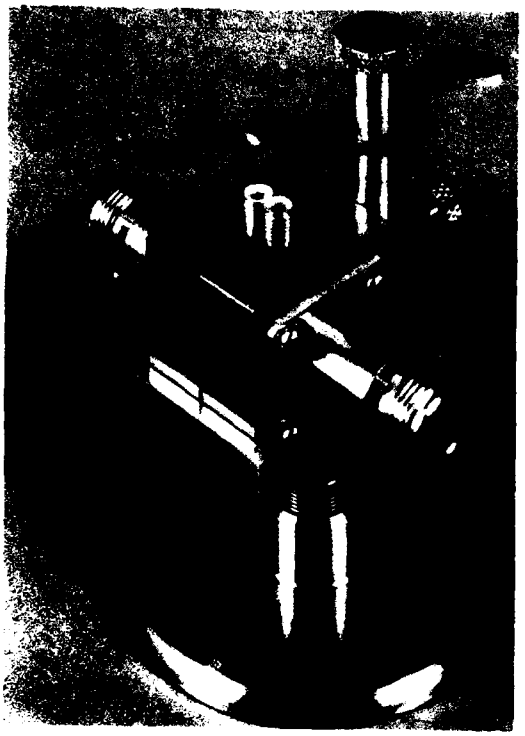


Figure 3 : APC-3.5 JIG.

jig is similar to the larger one, but no option for reflection measurements is included; for reflection the APC-7 jig is suitable in all cases. Both jigs are provided with a spring loaded earthing clamp so that metal crystal cans can be grounded. The frequency range over which measurements can be performed is dictated by the characteristics of the network analyser itself, and not by any feature of the test jigs or of the measurement method.

The jigs themselves are machined entirely out of stainless steel for maximum durability and resistance to corrosion. To avoid any possible confusion over the structure of the mounting jigs full engineering drawings and specifications have been prepared.

System software

The system consists of two elements, a calibration programme, and a measurement programme; the overall structure is shown in Figure 4. The calibration programme consists of a control routine, which is responsible for instrument control and for interfacing with the operator, and a data handling routine, which processes the S parameter data and computes the error parameter interpolation coefficients. The control routine is, of course, highly dependent on the chosen network analyser, but it involves relatively little computation; the data handling routine is quite independent of any particular instrument, and performs most of the numerical operations. To enhance the efficiency of the system, the data routine is written

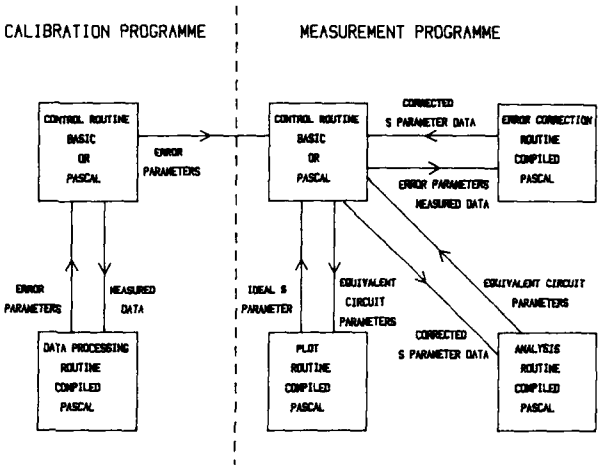


Figure 4 : Software system.

in Pascal and is used in the form of a compiled module, which, on an HP200 series computer, can be called from a Pascal or BASIC control programme. This approach has the advantage that it is relatively easy to reconfigure the system for new instruments, while at the same time maintaining a high level of computational efficiency.

The error parameters are deduced from various S parameter measurements by the method outlined in Section 2. The output from the programme is not the error parameters themselves, but interpolation coefficients which may be used to compute values at arbitrary frequencies. In the present system the method of interpolation by cubic splines has been chosen [6]. In an interval between two calibration points, each parameter is represented by a cubic polynomial in frequency; at a boundary between successive intervals the polynomials are chosen to match in function value, first derivative and second derivative. The method of cubic splines is optimal in the sense that it draws the smoothest possible curve through a given set of points and is therefore one of the safest available interpolation procedures. To determine the value of a parameter at a given frequency requires the evaluation of a cubic polynomial, which involves a minimum of three multiplications and three additions. The use of linear splines (straight line interpolation between successive points) would be simpler, but noticeable discrepancies can arise unless a high number of calibration points are employed. For this reason the increased complexity of cubic splines is more than offset by the greater robustness and reliability which they offer. The 12 error parameters each have a real and an imaginary part, and a cubic polynomial is defined by four coefficients. Therefore, if there are N calibration points, 12x2x4x(N-1) = 96(N-1) real polynomial coefficients must be stored; if N = 101, and 8 bytes of storage per number are required, then this would occupy a total of 75 Kbytes of memory. In practice, calibration is only performed infrequently, perhaps once a week, and the calibration data is stored on floppy disks for access by the measurement programme.

The structure of the measurement programme is similar to that of the calibration programme. A control routine organises instrument and operator interfaces, while compiled subroutines perform all the computation intensive procedures. The modular structure is designed so that engineers can configure the control parts of the programme to their individual requirements, and simply call the 'black box' compiled

routines to deal with the data interpretation. The engineer does not require a detailed knowledge of the operation of the compiled routines, and attempts have been made to make these as robust as possible, so that predictable results can be obtained under all reasonable conditions.

The error correction routine is quite simple; it merely reconstitutes the error parameters from the interpolation data at the specified frequency, and calculates the corrected S parameters using equation (6). The plot routine is equally simple; it calculates a suitable S parameter from the equivalent circuit parameters to provide a visual comparison of the measured and theoretical responses.

Most of the programme complexity is contained in the analysis subroutine; this takes the corrected S parameter data and returns the optimal values of the equivalent circuit parameters. The optimisation procedure is fully automatic, and very reliable, although a sophisticated user may 'tune' the performance to some degree by adjusting certain variables in the subroutine parameter list. At present three distinct types of measurement are supported. The first, a reflection measurement, assumes that the crystal may be represented by the two terminal equivalent circuit of Figure 5. The second, a transmission measurement, assumes the equivalent circuit of Figure 6. The third, also a transmission measurement, is intended for bipole resonators

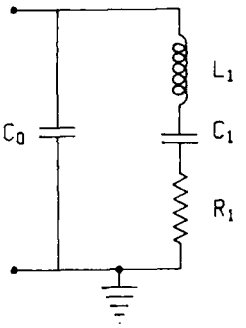


Figure 5 Single resonator two terminal equivalent circuit.

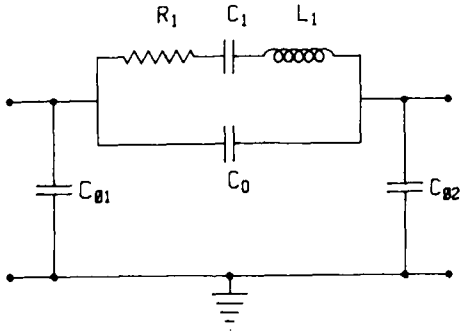


Figure 6 : Single resonator three terminal equivalent circuit.

and assumes the equivalent circuit of Figure 7. There appears to be less general agreement on equivalent circuits for bipoles than for single resonators. Using the normal mode expansion of Lloyd [7], the circuit shown in Figure 7 has a similar justification to that of Figure 5 for a single resonator, and is therefore regarded by the authors as being the most reasonable simple representation. For the case of a physically symmetric bipole, the turns

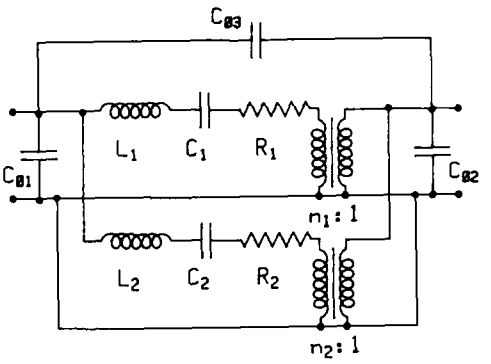


Figure 7 : Bipole equivalent circuit.

ratios of the transformers are 1:1 and 1:-1, and the circuit may be redrawn in standard half lattice form; in all cases the T equivalent circuit of Figure 8 may be used.

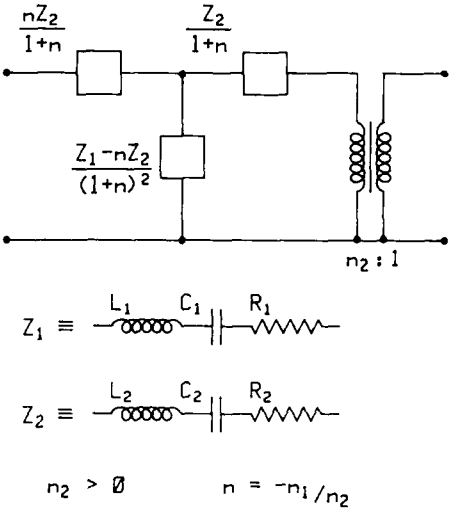


Figure 8 : Alternative bipole equivalent circuit.

In these circuits, not all the circuit elements are used as adjustable parameters in the final fitting procedure. For a transmission measurement on a single resonator (Figure 6), C_0, R_1, L_1, C_1 are used as variables in the fitting routine. Similarly, for the bipole, C_{01}, C_{02} and C_{03} are evaluated away from resonance and are assigned fixed values in the fitting routine.

For a single resonator a quantity Y may be defined by

$$Y = j\omega C_0 + \left(R_1 + j\omega L_1 + \frac{1}{j\omega C_1}\right)^{-1} \qquad (7)$$

In both reflection and transmission measured values of Y are easily obtained from the S parameters

$$R_0 Y = \frac{1-S_{11}}{1+S_{11}} \qquad \text{in reflection} \qquad (8)$$

$$R_0Y = \frac{2S_{12}}{(1+S_{11})(1+S_{22})-S_{12}S_{21}} \quad \text{in transmission} \qquad (9)$$

$$(R_0 = 50\ \Omega)$$

If Y_i denotes the value of Y, regarded as a function of the four adjustable parameters, evaluated at ω_i, and Y_i^A denotes the corresponding measured quantity, then an error function is defined by

$$E = \sum_i \frac{1}{|Y_i^A|} \cdot |Y_i - Y_i^A|^2$$

E involves a summation over all measurement points, and represents a measure of the difference between theoretical and experimental responses; the optimal values of C_0, R_1, L_1, C_1 are defined as those which minimise the value of E. There are many ways in which E could be defined, and the present choice is a compromise between a number of factors. In the past, several workers have chosen to deduce values of R_1, L_1, C_1 by fitting to $[ReY]^{-1}$ [3,8]. As

$$[ReY]^{-1} \approx R_1 + 4(\omega-\omega_0)^2 \frac{L_1^2}{R_1} \qquad (11)$$

this is a very simple procedure, requiring only a parabolic fit, but apart from analytic simplicity it has few advantages.

$$Y = j\omega C_0 + \frac{R - j\left(\omega L_1 - \frac{1}{\omega C_1}\right)}{R^2 + \left(\omega L_1 - \frac{1}{\omega C_1}\right)^2} \qquad (12)$$

The real part of Y has a maximum at the resonant frequency (ω_0), while the imaginary part, neglecting C_0, has a zero. For small deviations from ω_0, variations in ReY are second order in $\delta\omega$ whereas those of ImY are first order, and it can therefore be argued that ImY is the more sensitive indicator of frequency. Indeed phase zero methods of determining resonant frequency rely on finding the zero of ImY. In addition to this, reliable values of ReY can only be measured very close to the resonant frequency, and when a crystal response is observed on a network analyser ImY dictates the response over most of the frequency range. Aside from the question of analytic simplicity, it is therefore highly advantageous to make use of ImY, and it was found to be impractical to make a robust analysis routine using a fit to ReY above. Re-writing E as

$$E = \sum_i |Y_i^A| \cdot \left|1 - \frac{Y_i}{Y_i^A}\right|^2 \qquad (13)$$

it can be seen that relative errors at the point i are weighted by $|Y_i^A|$, thus emphasising the region around resonance. E could have been chosen as $|Y_i - Y_i^A|^2$, but this tends to produce poor agreement in the vicinity of the anti-resonance. The form of the error function that has been adopted is therefore a reasonable compromise between simplicity and reliability.

Similar considerations apply to the choice of an error function for bipoles. The admittance matrix for the circuit shown in Figure 7 is particularly simple;

$$Y = Y_0 + Y_1 + Y_2$$

$$Y_0 = j\omega \begin{pmatrix} C_{01}+C_{03} & -C_{03} \\ -C_{03} & C_{02}+C_{03} \end{pmatrix}$$

$$Y_1 = \begin{pmatrix} 1 & -n_1 \\ -n_1 & n_1^2 \end{pmatrix} \left(R_1 + j\omega L_1 + \frac{1}{j\omega C_1}\right)^{-1} \qquad (14)$$

$$Y_2 = \begin{pmatrix} 1 & -n_2 \\ -n_2 & n_2^2 \end{pmatrix} \left(R_2 + j\omega L_2 + \frac{1}{j\omega C_2}\right)^{-1}$$

so it is desirable to use Y as the basis of the function. Only the eight variables, R_1, L_1, C_1, n_1, R_2, L_2, C_2, n_2 are used in the fitting procedure; it is assumed that C_{01}, C_{02} and C_{03} have been determined away from resonance, and they are assigned fixed values. The admittance matrix Y is related to the scattering matrix S by

$$R_0Y = (1+S)^{-1}(1-S) \qquad (R_0=50\ \Omega) \qquad (15)$$

in matrix notation. Denoting the theoretical value of the admittance matrix at ω_i by Y_i and the experimental value by Y_i^A, a suitable choice for E is

$$E = \sum_i Tr\left[(Y_i - Y_i^A)^+ (Y_i - Y_i^A)\right] \qquad (16)$$

($^+$ denotes Hermitian adjoint, Tr denotes trace).

Again this choice is somewhat arbitrary, designed to keep the analysis fairly simple, but it has been found to work extremely well in practice.

Having selected expressions for the error function E, the basic operation of the system is fully defined, but the important technical problem of finding the minimum of E remains. Fortunately this type of problem has been much studied, and a wide variety of methods are available. Except in a few very special cases all of these rely on an iterative procedure which will converge to the solution in an unspecified number of iterations; in general it is usually possible to determine the asymptotic rate of convergence, but little else. Some methods only require a knowledge of the function values at each iteration, whereas others may require values of E and its first and second derivatives. Not surprisingly, the more derivative information which can be supplied, the more powerful the algorithm that can be employed. Fortunately, the derivatives of the chosen functions can be computed very easily, and it is possible to use a full Newton method to compute the minima. For the single resonator E depends on four parameters, which are chosen to be

$$x_1 = C_0,\ x_2 = R_1,\ x_3 = L_1,\ x_4 = C_1^{-1}$$

$$\frac{\partial E}{\partial x_j} = \sum_i \left|\frac{1}{Y_i^A}\right| 2\,Re\left[\frac{\partial Y_i}{\partial x_j} \cdot (Y_i - Y_i^A)^*\right]$$

$$\frac{\partial^2 E}{\partial x_j \partial x_k} = \sum_i \frac{1}{|Y_i^A|} \cdot 2\,Re\left[\frac{\partial^2 Y_i}{\partial x_j \partial x_k} \cdot (Y_i - Y_i^A)^* + \frac{\partial Y_i}{\partial x_j} \cdot \frac{\partial Y_i^*}{\partial x_k}\right] \qquad (17)$$

For the bipole the eight parameters are chosen as

$$x_1 = R_1,\ x_2 = L_1,\ x_3 = C_1^{-1},\ x_4 = n_1$$

$$x_5 = R_2,\ x_6 = L_2,\ x_7 = C_2^{-1},\ x_8 = n_2$$

$$\frac{\partial E}{\partial x_j} = \sum_i 2\left[Re\left(Tr\ \frac{\partial Y_i}{\partial x_j}\ (Y_i - Y_i^A)^+\right)\right]$$

$$\frac{\partial^2 E}{\partial x_j \partial x_k} = \sum_i 2\,\mathrm{Re}\left[\mathrm{Tr}\left(\frac{\partial^2 Y_i}{\partial x_j \partial x_k}(Y_i - Y_i^A)^+ + \frac{\partial Y_i}{\partial x_j}\cdot\frac{\partial Y_i^+}{\partial x_k}\right)\right] \quad (18)$$

These bipole equations are, of course, to be understood in a matrix sense.

If in either case the matrix of second derivatives of E is denoted by G, and the column vector of first derivatives by g, then in the Newton method, at each iteration,

$$Gp = -g \quad (19)$$

defines the vector p containing the increments to the variables x_i. Close to the solution this procedure has quadratic convergence, approximately doubling the number of significant digits in the answer at each iteration, but far from the solution it may not converge at all. This difficulty is overcome by first ensuring that under all conditions the increment p corresponds to a descent direction, that is a direction along which the value of E can be reduced. The condition for this is

$$\tilde{p}\, g < o \quad (\sim \text{ denotes transpose}) \quad (20)$$

or $$\tilde{p}\, G\, p > o \quad (21)$$

and a sufficient condition for (21) to hold is that G should be positive definite. The solution of linear equations such as (19) is invariably performed by some form of triangular factorisation (Gaussian elimination etc) and in this case the modified Cholesky factorisation of Gill and Murray [9] has been adopted. With this method the matrix is checked during the factorisation and modified if necessary to ensure that the result is always positive definite; close to the solution G is necessarily positive definite and no modifications will be introduced. Computing a descent direction is not sufficient in itself to ensure convergence. It is possible that the calculated increment will overshoot the minimum along its direction of search, and produce an increase in E; this can be overcome by scaling down the increment p if necessary.

Using these methods it is possible to provide an iterative method of solution which will not normally take more than four iterations to find the solution, and which will not fail under any reasonable conditions. The process terminates when values obtained on successive iterations agree to within stringent limits. The only problems which can still arise are convergence to a subsidiary minimum, or failure to make sufficient progress due to poor starting values. Practical experience shows that the first of these problems rarely, if ever, occurs and careful processing of the input data, combined with a system of default values, has virtually eliminated the second.

It might be supposed from the preceding equations that the computational effort involved in this procedure was quite enormous, but there is a high degree of redundancy in the calculation of the derivatives, and if this is exploited cleverly, then the data processing can normally be completed in a few seconds on an HP200 series machine.

The analysis routine contains many subtleties which an engineer using it need not appreciate. However, the routine will produce answers quite happily even when the data supplied is not very well chosen. For this reason it is desirable to provide some simple indication of the suitability or otherwise of the data set. Unsuitable in this sense means data for which the error function E displays low sensitivity to one or more of the equivalent circuit parameters; in this situation such circuit parameters may be poorly estimated. A typical case is where there are no data points close to the resonant frequency; this gives low sensitivity to R_1. To detect this situation the parameters

$$\delta x_i = \left[\frac{E}{\partial^2 E / \partial x_i^2}\right]^{\frac{1}{2}}$$

are computed using the values of E and its second derivatives at the minimum; δx_i represents the change in x_i necessary to produce a change in E equal to the residual error. Strictly speaking δx_i is an estimate of sensitivity rather than error, but a reliable measurement should have $\delta x_i << x_i$ for all parameters.

Experimental Results

To demonstrate the operation of the system a range of measurements were performed on a selection of resonators and bipoles using an HP 3577A network analyser.

The results for eight quartz resonators are shown in Table 1; the frequences cover 5 MHz to 176 MHz, and all

CRYSTAL TYPE	METHOD	FREQUENCY Hz	STATIC CAP. pF	ESR OHMS	MOTIONAL INDUCT. mH
5MHz 5OT AT-QUARTZ	REFL.	5000116.05 0.348	7.071 0.1133	127.741 0.3206	7735.113 57.9856
	TRANS.	5000116.78 0.133	2.328 0.0758	129.009 0.0984	7764.835 32.9337
10MHz 10T SC-QUARTZ	REFL.	10003868.51 1.093	5.720 0.0773	0.496 0.0042	113.713 0.1062
	TRANS.	10003847.49 0.583	2.503 0.0985	0.700 0.0055	113.042 0.0983
21.7MHz 10T AT-QUARTZ	REFL.	21733863.39 0.255	4.029 0.0091	12.126 0.0039	4.978 0.0009
	TRANS.	21733908.45 0.191	2.981 0.0034	12.027 0.0062	4.975 0.0005
45MHz 30T AT-QUARTZ	REFL.	45033162.87 0.816	3.425 0.0298	35.789 0.0693	11.748 0.0154
	TRANS.	45033155.93 0.243	2.759 0.0074	35.508 0.0527	11.791 0.0032
58.4MHz 30T AT-QUARTZ	REFL.	58408227.41 0.918	3.288 0.0273	52.681 0.0589	9.445 0.0168
	TRANS.	58408228.18 0.555	2.337 0.0128	53.428 0.0094	9.508 0.0020
90MHz 5OT AT-QUARTZ	REFL.	90001854.08 2.054	3.197 0.0016	52.507 0.0155	8.627 0.0018
	TRANS.	90001878.69 4.948	2.725 0.0018	53.343 0.0186	8.777 0.0011
116MHz 5OT AT-QUARTZ	REFL.	116002575.06 1.075	3.258 0.0053	68.720 0.0045	6.600 0.0042
	TRANS.	116002574.47 1.196	2.815 0.0083	68.304 0.0186	6.198 0.0000
175.5MHz 7OT AT-QUARTZ	REFL.	175556009.27 4.382	2.950 0.0000	74.789 0.0232	4.618 0.0024
	TRANS.	175556049.31 1.926	2.284 0.0004	75.005 0.0000	4.627 0.0004

Table 1 : Single resonator measurements.

of the devices were measured in both transmission and reflection using the same frequency points and power levels. For a single resonator the programme will allow between 3 and 100 frequency points in the fitting procedure. The time taken to error correct the data is of course directly proportional to the member of points used; the analysis time is also approximately proportional to the number of points. For reference approximate timings are:

Data correction

Reflection	0.03	sec/point
Transmission	0.1	sec/point

Analysis

Single resonator	0.1	sec/point
Bipole	0.3	sec/point

All of these timings apply to a standard HP 9816; with the addition of the optional floating-point card the speed could be improved by a factor of 3-4.

For maximum accuracy and reproducibility it is obviously better to use a large number of points, as this will serve to reduce random errors. All of the measurements in Table 1 used the maximum 100 points, but a choice of ~20 points would be more usual. Typical output plots for a measurement on the 21.7 MHz resonator are shown in Figures 9 and 10 for reflection

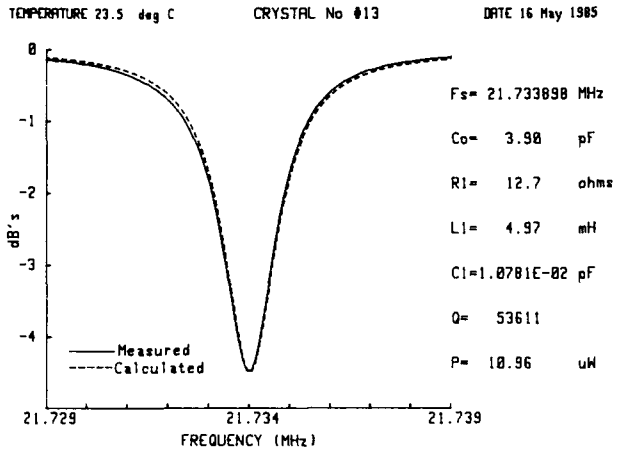


Figure 9 : Resonator reflection measurement.

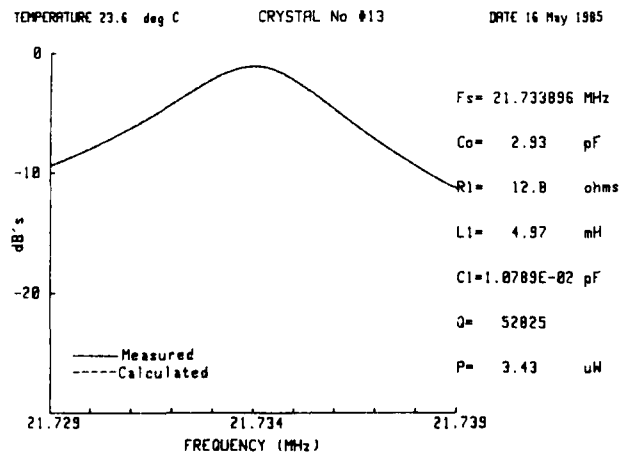


Figure 10: Resonator transmission measurement.

and transmission respectively. In reflection the measured and theoretical values of S_{11} are shown, and it can be seen that the agreement is very close. For transmission S_{12} is used instead, and in Figure 10 the agreement between measurement and theory is so close that the curves are hard to distinguish apart. In the table each box contains two entries representing the mean and the standard deviation of ten repeated measurements. It can be seen that in general the standard deviations are negligibly small, and some of the frequency variation may be due to temperature fluctuations during the measurements. The 5 MHz 50T crystal was the most time consuming to measure, as sweep times of 2-3 mins were required on the network analyser; the motional inductance of this crystal showed a much larger that average variability. The fitting procedure was chosen to provide as much consistency as possible between reflection and transmission measurements, and the agreement is in practice very good. The value of C_0 obtained in reflection is always greater than that obtained in transmission, as it included implicitly one of the electrode to can capacitances. All of the other parameters are very similar in both transmission and reflection, though the differences usually exceed the standard deviations. It must be expected therefore that the measured parameters will depend slightly on the frequency points used and on the method employed, as well as on factors such as drive level.

Table 2 shows measurements taken on two bipoles, one quartz and one lithium tetraborate. As before, mean values and standard deviations from ten repeated

BIPOLE	MODE	FREQUENCY Hz	ESR OHMS	MOTIONAL INDUCT. mH	TRANS. TURNS RATIO
19MHz 10T AT-QUARTZ	1	18974987.8 8.385	53.347 8.8417	23.673 8.8148	-1.012 8.8882
	2	18980483.4 8.373	84.788 8.8227	23.692 8.8185	8.988 8.8882
16.8MHz 10T LITHIUM TETR.	1	16789541.1 5.747	59.696 8.8234	3.883 8.8814	-8.833 8.8882
	2	16874321.4 5.914	88.128 8.8363	4.744 8.8888	1.825 8.8882

Table 2 : Bipole measurements.

measurements are given; in general these are of a similar magnitude to those found for single resonators, except for the frequency fluctuations of the lithium tetraborate device which are presumably caused by its greater temperature sensitivity. Output plots for the two bipoles are given in Figure 11 and 12, and it can be seen that the theoretical fit is excellent; the very much greater coupling of the lithium tetraborate device should also be noted.

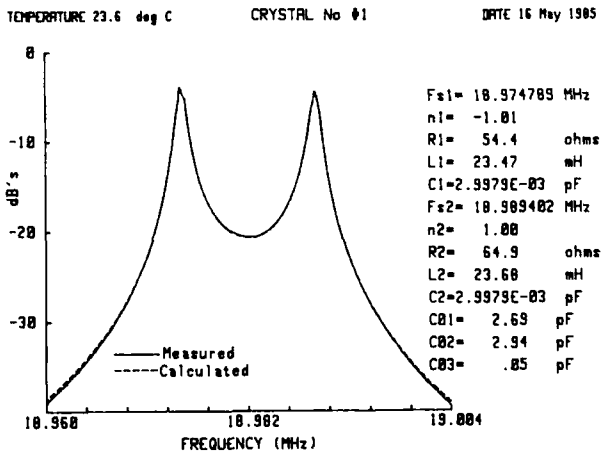


Figure 11 : Quartz bipole measurement.

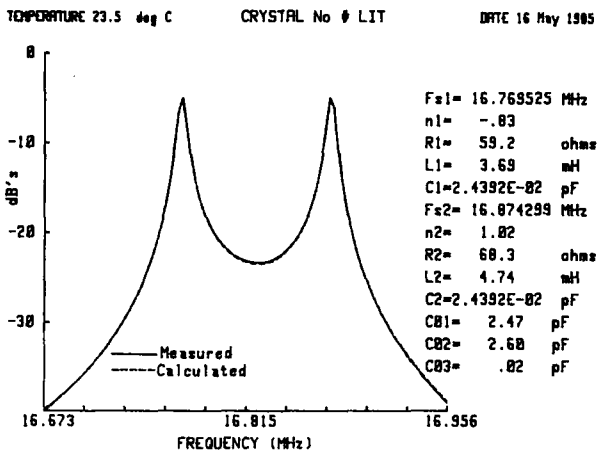


Figure 12 : Lithium tetraborate bipole measurement.

Conclusions

A measurement system for single resonators and bipoles has been developed using as far as possible the best available techniques at every stage of the procedure. It is believed that the system offers the best accuracy and traceability currently available.

As well as providing high precision the method is extremely efficient, and enables crystals to be measured at a reasonably high rate; it might therefore be applicable to certain high grade production tasks, as well as to standards and quality assurance applications.

Acknowledgement

This work has been carried out with the support of the Procurement Executive, Ministry of Defence, sponsored by DCVD.

References

[1] IEC Publication 444, 'Basic method for measurement of resonance frequency and equivalent series resistance of quartz crystal units by zero phase technique in a pi-network', 1973.

[2] R.C. Peach, A.J. Dyer, A.J. Byrne and S.P. Doherty. 'New method for the measurement of quartz crystal parameters'. Proc. 36th AFCS pp 297-301, 1982.

[3] H.S.Pustarfi and W.L. Smith. 'An automatic crystal measurement system'. Proc. 27th AFCS, pp 63-72, 1973.

[4] J. Fitzpatrick, 'Error models for systems measurement'. Microwave Journal, pp 63-66, May 1978.

[5] S. Rehnmark. 'On the calibration process of automatic network analyser systems';. IEEE Trans., MTT-22, 4, pp 457-458, 1974.

[6] L.W. Johnson and R.D. Riess. 'Numerical Analysis'. 2nd Ed. Addison-Wesley, 1982.

[7] P. Lloyd. 'Equations governing the electrical behaviour of an arbitrary piezoelectric resonator having N electrodes'. BSTJ, 46, pp 1881-1900, 1967.

[8] J.P. Aubry, E. Gerard and S. Lechopier. 'S.Y. parameter method for accurate measurement of bulk wave crystal resonators at frequencies up to 2GHz'. Proc. 37th AFCS, pp 306-316, 1983.

[9] P.E. Gill and W. Murray. 'Newton type methods for unconstrained and linearly constrained optimization'. Math. Prog., 7, pp 311-350, 1974.

INVESTIGATION OF QUARTZ CRYSTAL THICKNESS SHEAR AND TWIST MODES USING A NEW NONINTERFEROMETRIC LASER SPECKLE MEASUREMENT METHOD

Stefan Hertl, Ewald Benes, Leopold Wimmer, Michael Schmid

Institut für Allgemeine Physik, Technische Universität Wien
Karlsplatz 13/134, A-1040 Vienna, Austria

Overview

The measurement of the mechanical vibration amplitudes of quartz crystals has been carried out by a variety of different methods. Interferometric methods[1-4] firstly were used to investigate out-of-plane vibrations. Enhanced interferometric setups also allow the measurement of in-plane vibrations[5,6]. Holographic techniques[7] present an instantaneous picture of the whole object, whereas the interferometric methods supply the vibration amplitude distribution point after point.

A very powerful technique to study amplitude distributions of quartz crystals is the use of X-ray diffraction[8]. Using this method, many quartz crystal vibration modes have been investigated so far[9-14]. The measurement setup, however, is complicated and the resolution of the graphic output is rather poor.

Looking for a simple mechanical setup that is insensitive to environmental disturbances, we developed a new method which allows the measurement of in-plane vibration amplitudes of quartz crystals[15]. The detection range is about one to some hundred nanometers. Due to the use of a microcomputer that controls the whole measurement the graphic output can easily be adapted to the user's needs.

Principle of operation and experimental setup

Coherent radiation, such as laser light, when reflected by a surface with a roughness depth greater than the wavelength, has a granular appearance. This phenomenon, caused by the interaction of the scattered waves, is called speckle effect. It is the idea of our method to directly utilize the speckle effect to investigate in-plane movements of quartz crystals oscillating in thickness shear mode.

By positioning a photo diode anywhere into the speckle field the photo current will be proportional to the intensity of the light detected. Provided that the sensitive area of the photo diode is smaller than the average speckle size, the photo current depends on the position of the diode relative to a bright spot of the speckle field. All intermediate states between the maximum and the minimum photo current will be obtained when only a part of the sensitive area is illuminated by a bright speckle.

A lateral translation of the surface causes an analogous movement of the speckle pattern and thus causes a modulation of the photo current due to the varying light intensity at the sensitive area of the photo diode. Hence, the vibration amplitude of the surface point under examination can be obtained by investigating the photo current at the oscillation frequency of the surface. For small amplitudes, there is a proportionality between the signal component of the photo current at the oscillation frequency and the mechanical vibration amplitude of the crystal surface. No special precautions have to be taken to protect the measurement setup from environmental noise. The relative position between the object under investigation and the photo diode is not very critical either.

Unfortunately, the proportionality between the mechanical vibration amplitude and the signal component of the photo current at the oscillation frequency is not constant during a scan across the entire surface of the quartz crystal. Therefore, the primary amplitude data obtained by the measurement system have to be smoothed to extract the fluctuations imposed by the locally varying sensitivity of the measurement method. Two algorithms have been tested successfully; one is the application of cubic splines, and the other method is Fourier transformation and appropriate filtering during retransformation.

The quartz crystal under examination is driven by a frequency synthesizer. The drive level of the synthesizer is in the range between 0 and 13 dBm. The current of the photo diode is input to a spectrum analyzer. The spectrum analyzer is locked at the oscillation frequency of the vibrating quartz crystal, i.e. the synthesizer frequency. The spectrum analyzer at best operates with a bandwidth of 30 Hz. Thus, only the fraction around the oscillation frequency is filtered out of the photo diode signal. Two orthogonal stepper motor drives enable the scan across the entire surface of the quartz crystal. To illuminate the optically rough surface of the vibrating quartz crystal a He-Ne-laser of 2 mW radiation power is sufficient. The whole measurement is supervized by a microcomputer which controls spectrum analyzer and frequency synthesizer via an IEEE-488 interface bus and the stepper motors using a parallel interface coupled to external stepper motor logic.

The mathematical smoothing of the raw data as well as the graphic output is done by the same computer. A block diagram of the experimental setup is shown in Fig.1. Two different presentations of the vibration amplitude distributions were chosen. One shows a three-dimensional picture of the vibration mode with the magnitude of the displacement along the vertical axis (Z-axis). The other presentation uses lines of equal amplitude and can be better compared to the pictures obtained by X-ray topography.

We have measured various vibration modes of plano-convex circular AT-cut quartz crystals of three different vendors. Figures 2-25 show some measured

CH2186-0/85/0000-0535$1.00©1985IEEE

vibration modes. Some measurements show significant coupling of two or more anharmonic modes.

Theory of the vibration amplitude distributions of plano-convex AT-cut quartz plates

Since about 1955 the distributions of the mechanical vibration amplitude and the resonance frequencies of quartz crystals have been calculated theoretically. Mindlin and Deresiewicz[16] developed a general solution for the frequency spectrum and the vibration amplitude distribution of a rectangular plane quartz plate. For this calculation, the anisotropy of the crystal and the piezoelectric coupling have been neglected. For a given ratio of diameter to thickness of the plate, also circular quartz crystals have been treated in this way[17].

Plane quartz crystals show a rather complicated distribution of the vibration amplitudes. If at least one side of the quartz plate is convex, the vibration modes are much more distinct, but still there are modes that cannot be predicted by theory[18]. The following theoretical treatment virtually reviews the analysis of Tiersten and Smythe[19] for a circular, plano-convex AT-cut quartz crystal.

The derivation of the frequency spectrum and the vibration amplitude distributions of thickness shear and twist modes of a plano-convex crystal may be based on the 1978 Standard on Piezoelectricity[20]. Sign changes of certain constants that have been introduced by this standard, however, do not influence the following analysis; that is why all data taken from papers published before 1978 are still valid. The crystallographic directions are denoted by X,Y,Z, whereas the axes of the rotated quartz plate are specified by x,y,z. Only the axes of the rotated quartz crystal are of importance for our treatment, so the letters X,Y,Z are instead used to denote the directions of the measurement setup (see Figs. 2-5). The components of the displacement vector will be denoted by u_x, u_y and u_z. For the elastic and piezoelectric tensors (c_{pq} and e_{kp}, respectively) the compressed matrix notation[20] will be used with the indices p,q ranging from 1 to 6 (k=1,2,3). Some vital simplifications of the general three-dimensional differential equations for the linear piezoelectric continuum have been introduced by Tiersten and Smythe[19]:

1. Piezoelectric coupling is very small in quartz; therefore the influence of the electric variables on the resonance frequency and the vibration amplitude distribution can be neglected for the x- and z-directions.

2. The focusing of vibration energy is almost entirely determined by the contouring of the crystal. (For plane quartz crystals the energy of the oscillation can be concentrated to the central regions of the plate merely by the use of thick electrodes or by depositing a foreign material on the surface of the plate.) Therefore, shape and diameter of the electrodes can be neglected for the treatment of contoured (plano-convex or biconvex) crystals. That is why no boundary conditions at the edge of the crystal and the electrode have to be considered.

3. Some relatively unimportant quantities are neglected, for example any displacement in the z-direction.

Electrode mass as well as the influence of shorted electrodes and piezoelectric stiffening will be considered in the analysis. Under these assumptions, Tiersten and Smythe solved the following differential equations for the displacements u_x, u_y, u_z and the electric potential φ:

$$c_{11}\frac{\partial^2 u_x}{\partial x^2} + (c_{12}+c_{66})\frac{\partial^2 u_y}{\partial x \partial y} + c_{66}\frac{\partial^2 u_x}{\partial y^2} + c_{55}\frac{\partial^2 u_x}{\partial z^2} + e_{y6}\frac{\partial^2 \varphi}{\partial y^2} = \rho_Q \frac{\partial^2 u_x}{\partial t^2} \quad , \tag{1}$$

$$(c_{66}+c_{12})\frac{\partial^2 u_x}{\partial x \partial y} + c_{66}\frac{\partial^2 u_y}{\partial x^2} + c_{22}\frac{\partial^2 u_y}{\partial y^2} = \rho_Q \frac{\partial^2 u_y}{\partial t^2} \quad , \tag{2}$$

$$e_{y6}\frac{\partial^2 u_x}{\partial y^2} - \varepsilon_{yy}\frac{\partial^2 \varphi}{\partial y^2} = 0 \quad . \tag{3}$$

The mass density of quartz is denoted by ρ_Q and is assumed to be constant. The (maximum) thickness of the quartz plate is l_Q. The boundary conditions at $y=\pm l_Q/2$ are due to the mass of the electrodes:

$$T_{yx} = \pm \rho_E l_E \frac{\partial^2 u_x}{\partial t^2} \quad , \tag{4}$$

$$T_{yy} = \pm \rho_E l_E \frac{\partial^2 u_y}{\partial t^2} \quad , \tag{5}$$

$$\varphi = \pm \frac{V}{2} e^{i\omega t} \tag{6}$$

with T_{ij} denoting the stress tensor and ρ_E and l_E denoting mass density and thickness of the electrodes, respectively. Both electrodes are assumed to be of equal thickness. The voltage which is applied to the electrodes has the amplitude V and angular frequency ω.

The locally varying thickness $l(x,z)$ of a plano-convex quartz crystal with a radius of curvature R constitutes another boundary condition. Wilson[21] established the relation

$$l(x,z) = l_Q - \frac{x^2+z^2}{2R} \tag{7}$$

which is valid to first order in (x^2+z^2). Taking all these boundary conditions into account we get the following differential equation for thickness shear and twist modes of a plano-convex quartz crystal after developing $l(x,z)^{-2}$ to first order in (x^2+z^2)[19]:

$$M_n \frac{\partial^2 u_x}{\partial x^2} + c_{55}\frac{\partial^2 u_x}{\partial z^2} - (n\pi/l_Q)^2 \hat{c}_{66}\left(1+\frac{x^2+z^2}{Rl_Q}\right)u_x - \rho_Q \frac{\partial^2 u_x}{\partial t^2} = \rho_Q \omega^2 \frac{e_{y6}}{c_{66}} V \frac{y}{l} e^{i\omega t} \quad . \tag{8}$$

The effective elastic constants M_n and $\hat{c}_{66}$ are[19]

$$M_n = c_{11} + (c_{12}+c_{66})r + 4\,\frac{(r\bar{c}_{66}-c_{66})(c_{22}r+c_{12})}{c_{22}n\pi\varkappa\tan(n\pi\varkappa/2)} \quad , \tag{9}$$

$$\hat{c}_{66} = \bar{c}_{66}\left\{1 - \frac{8k_{y6}^2}{(n\pi)^2} - 4\,\frac{\rho_E l_E}{\rho_Q l_Q}\right\} \tag{10}$$

where n is the overtone number, and $\bar{c}_{66}$ is the piezoelectrically stiffened elastic constant, and k_{y6}^2 is the piezoelectric coupling factor. Furthermore, the quantities r and $\varkappa$ are

$$r = \frac{c_{12}+c_{66}}{\bar{c}_{66}-c_{22}} \tag{11}$$

$$\varkappa = \sqrt{\frac{\bar{c}_{66}}{c_{22}}} \tag{12}$$

Near a resonance frequency, the displacement u_x along the thickness y of the plate is sinusoidal which results in

$$u_x(x,y,z,t) = u_o u_x(x,z)\sin(n\pi y/l_Q)e^{i\omega t} \tag{13}$$

where u_o is an unimportant scaling factor. From now on, we will consider only shorted electrodes, i.e. V=0. For the dimensionless vibration amplitude distribution $u_x(x,z)$ we take[21]

$$u_x(x,z) = X(x).Z(z) \tag{14}$$

and get

$$\frac{M_n}{\rho_Q}X''Z + \frac{c_{55}}{\rho_Q}XZ'' + \left\{\omega^2 - (n\pi/l_Q)^2\frac{\hat{c}_{66}}{\rho_Q}\left(1+\frac{x^2+z}{Rl_Q}\right)\right\}XZ = 0 \quad . \tag{15}$$

With

$$\Omega_1^2 + \Omega_5^2 = \omega^2 - (n\pi/l_Q)^2\,\frac{\hat{c}_{66}}{\rho_Q} \tag{16}$$

we get

$$\left\{\frac{M_n}{\rho_Q}\frac{d^2X}{dx^2} + \left[\Omega_1^2 - (n\pi/l_Q)^2\frac{\hat{c}_{66}}{\rho_Q}\frac{x^2}{Rl_Q}\right]X\right\}Z + \left\{\frac{c_{55}}{\rho_Q}\frac{d^2Z}{dz^2} + \left[\Omega_5^2 - (n\pi/l_Q)^2\frac{\hat{c}_{66}}{\rho_Q}\frac{z^2}{Rl_Q}\right]Z\right\}X = 0 \quad . \tag{17}$$

Each brace of eq.(17) constitutes Hermite's differential equation. Thus, for the vibration amplitude distribution $u_x(x,z)$ we get

$$u_x(x,z) = e^{-\alpha_n\frac{x^2}{2}}H_{p-1}(\sqrt{\alpha_n}x)\,e^{-\beta_n\frac{z^2}{2}}H_{q-1}(\sqrt{\beta_n}z) \quad , \tag{18}$$

where H_m is a Hermite polynomial of the order m, m=0,1,2,3,...; the overtone number is designated by n; and p,q=1,2,3,... identify the regarded anharmonic mode; and

$$\alpha_n^2 = \frac{\hat{c}_{66}}{M_n}\frac{(n\pi)^2}{Rl_Q^3} \quad ; \quad \beta_n^2 = \frac{\hat{c}_{66}}{c_{55}}\frac{(n\pi)^2}{Rl_Q^3} \quad . \tag{19}$$

Finally, considering eq.(16), we get an expression for the eigenfrequencies ω_{npq} in the form

$$\omega_{npq}^2 = (n\pi/l_Q)^2\frac{\hat{c}_{66}}{\rho_Q}\left\{1 + \frac{1}{n\pi}\sqrt{\frac{l_Q}{R}}\left[\sqrt{\frac{M_n}{\hat{c}_{66}}}(2p-1)+\sqrt{\frac{c_{55}}{\hat{c}_{66}}}(2q-1)\right]\right\} \tag{20}$$

Computation of vibration amplitude distributions

Using eq.(18) we have calculated the vibration amplitude distributions for an AT-cut quartz crystal resonator. The elastic constants for the AT-cut were taken from Mindlin[22], the piezoelectrically stiffened elastic constant $\bar{c}_{66}$ and the piezoelectric coupling factor k_{y6}^2 from Ballato[23]. The mass density of quartz is $\rho_Q = 2649\ kgm^{-3}$.

To compare the calculations with our measurements the parameters l_Q and R of an Inficon AT-cut quartz resonator have been used. Since no reliable data could be received from Inficon/Leybold our own measurements were taken. We found a maximum thickness $l_Q = 0.278$ mm and a radius of curvature R = 270 mm. However, the maximum plate thickness could not be determined to the required number of digits, so we used the (not very well known) electrode thickness l_E to set the calculated resonance frequency of mode (1,1,1) equal to the measured frequency of mode (1,1,1). Doing this, we arrived at an electrode mass loading ratio $\rho_E l_E/\rho_Q l_Q$ of 0.00642 which very well corresponds to the electrode thickness assumed for the Inficon crystal. In eq.(18) no damping is considered, so the amplitude distribution is defined without a scaling factor.

Figures 26-30 show some vibration modes that have been computed with the parameters mentioned above. As only the absolute values of the vibration amplitude can be measured, the computed amplitude distributions are shown without considering the phase change between two local maxima of the vibration amplitude. This is of no relevance when lines of equal amplitude are used.

Discussion

A comparison of computed and measured vibration amplitude distributions shows that in general the agreement is excellent. The following items, however, have to be considered:

1. The vibration amplitude distributions were calculated for an Inficon AT-cut quartz crystal. The other types of crystals measured have slightly different parameters that considerably change the resonance frequencies and also have an impact upon the shape of the vibration amplitude.

2. At areas within an anharmonic mode where the vibration amplitude is close to zero or zero due to a phase change the measured amplitude distribution differs from the calculated one. This is a consequence of the fact that we can only measure absolute values of amplitudes which do not reach exactly zero after mathematical smoothing.

3. When two anharmonic modes couple, as it seems to be the case in some of the measurements presented, a similarity between measurement and calculation understandably cannot be found.

Especially from the plots using lines of equal amplitude one can see that the experimental method presented can be utilized also for quantitative analysis. The resolution is better than for the widespread method of X-ray diffraction topography.

Acknowledgements

Many crystal probes have been supplied without charge by Inficon/Leybold-Heraeus Inc.

This work has been sponsored by the "Österreichischer Fonds zur Förderung der wissenschaftlichen Forschung" (Austrian Science Foundation) under Project no. P4610.

References

1 H.Osterberg, J.Opt.Soc.Am. **22**,19(1932)

2 W.D.Dye, Proc.Roy.Soc. **A138**,1(1932)

3 S.Edelman, E.Jones, E.R.Smith, J.Acoust.Soc.Am. **27**,728(1955)

4 V.A.Schmidt, S.Edelman, E.R.Smith, E.T.Pierce, J.Acoust.Soc.Am. **34**, 455(1962)

5 R.Dändliker, J.-F.Willemin, Opt.Lett. **6**,165(1981)

6 R.Jones, C.Wykes, Holographic and Speckle Interferometry (Cambridge University Press, Cambridge, 1983)

7 Y.Tsuzuki, Y.Hirose, K.Iijima, Proc. 25th AFCS, 113(1971)

8 G.W.Fox, P.H.Carr, Phys.Rev. **37**,1622(1931)

9 A.L.Bennett, R.A.Young, N.K.Hearn, Appl.Phys.Lett. **2**,154(1963)

10 W.J.Spencer, Appl.Phys.Lett. **2**,133(1963)

11 W.J.Spencer in Physical Acoustics, vol.5, ed.W.P.Mason (Academic Press, New York, 1968)

12 W.J.Spencer, R.M.Hunt, J. Acoust. Soc. Am. **39**, 929(1966)

13 R.A.Young,C.E.Wagner, Br.J.Appl.Phys. **17**,723(1966)

14 R.D.Mindlin, W.J.Spencer, J. Acoust.Soc. Am. **42**, 1268(1967)

15 L.Wimmer, S.Hertl, J.Hemetsberger, E.Benes, Rev. Sci. Instrum. **55**, 605(1984)

16 R.D.Mindlin, H.Deresiewicz, J.Appl.Phys. **26**, 1435(1955)

17 R.D.Mindlin, H.Deresiewicz, J.Appl.Phys. **25**, 1329(1954)

18 E.Hafner, IEEE Trans.Sonics Ultrason. **SU-21**, 220(1974)

19 H.F.Tiersten, R.C.Smythe, J. Acoust.Soc. Am. **65**, 1455(1979)

20 IEEE Standard on Piezoelectricity 176-1978 (Institute of Electrical and Electronics Engineers, New York, 1978)

21 C.J.Wilson, J. Phys. D 7, 2449(1974)

22 R.D.Mindlin, Quart.Appl.Math. **19**,51(1961)

23 A.Ballato, IEEE Trans.Sonics Ultrason. **SU-25**, 185(1978)

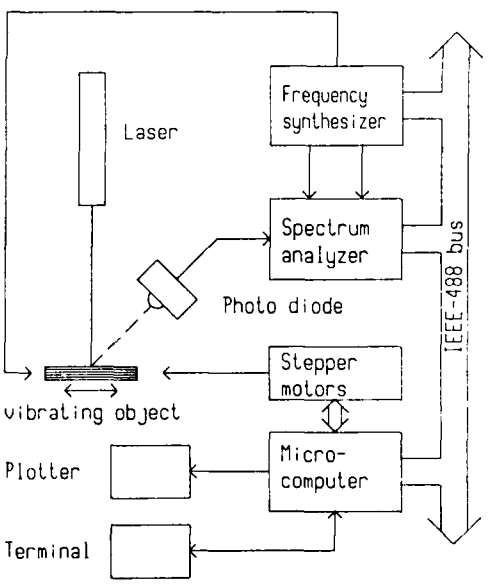


Fig. 1. Block diagram of experimental setup.

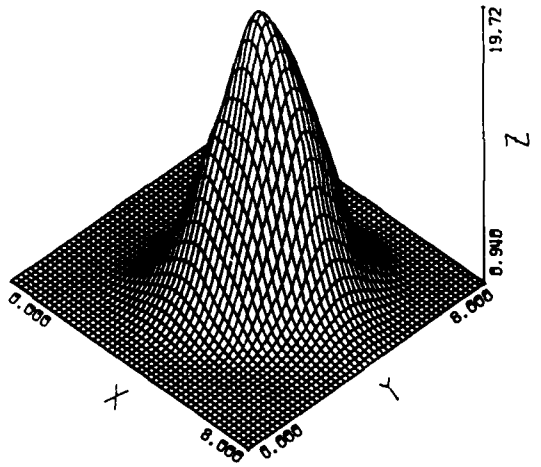


Fig. 2. Sample A; mode (1,1,1), 4.961741 MHz. (x-axis of crystal coincides with vertical diagonal).

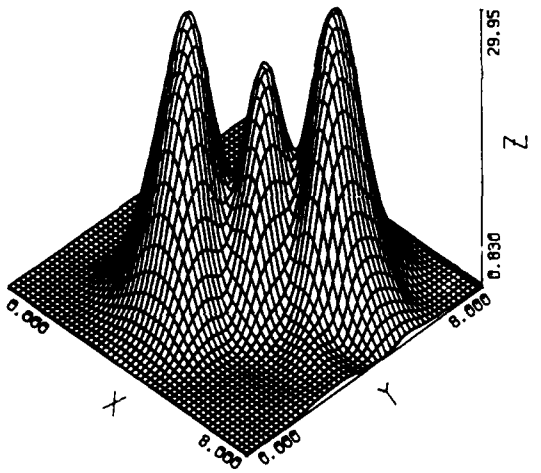


Fig. 3. Sample A; mode (1,1,3), 5.187546 MHz. (x-axis of crystal coincides with vertical diagonal).

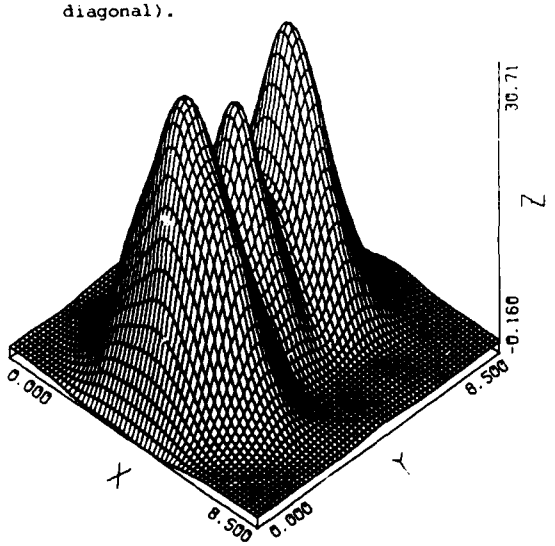


Fig. 4. Sample B; mode (1,1,3), 6.152575 MHz.

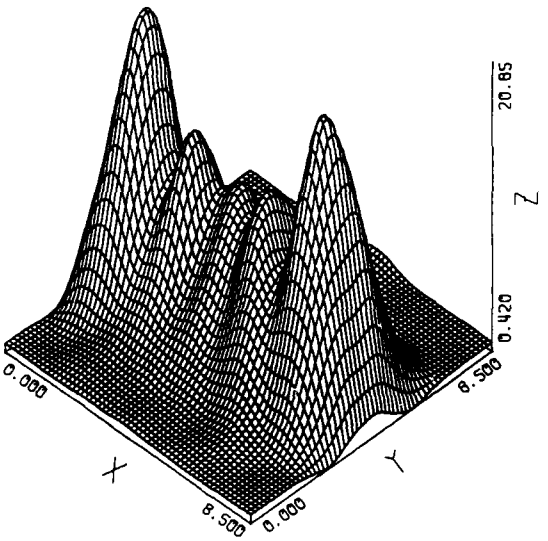


Fig. 5. Sample B; mode (1,5,1), 6.415074 MHz.

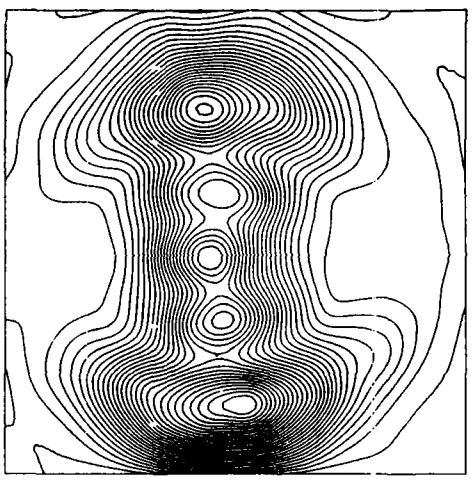


Fig. 6. Sample B; mode (1,1,5), 6.326601 MHz.

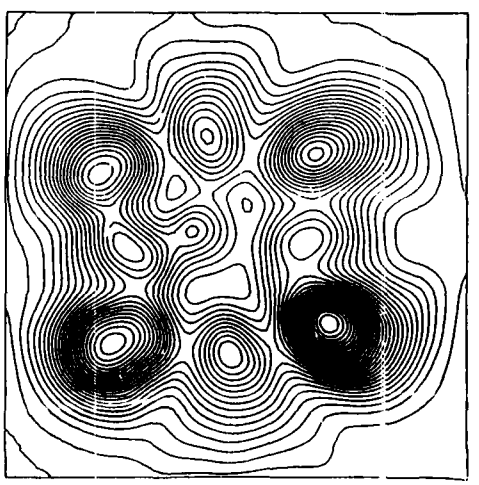


Fig. 7. Sample B; mode (1,3,3), 6.343964 MHz.

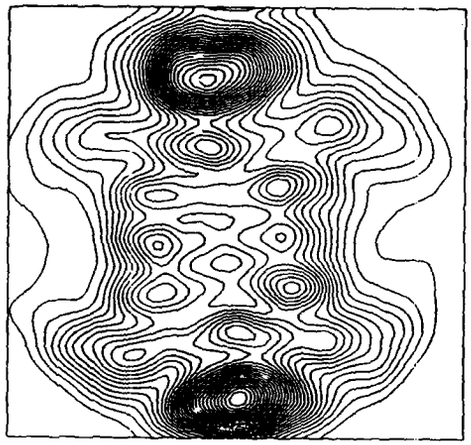


Fig. 8. Sample B; 6.509193 MHz.

Fig. 9. Sample B; 6.648891 MHz.

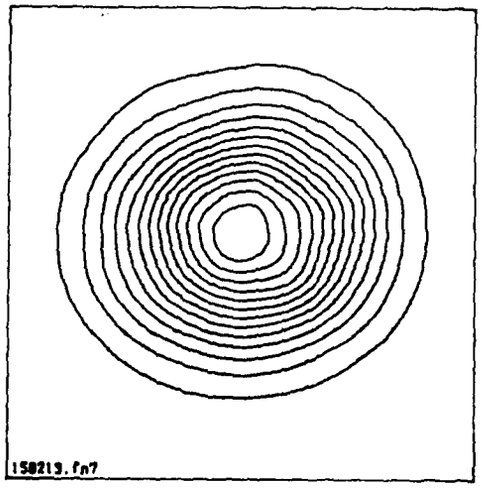


Fig. 10. Sample C; mode (1,1,1), 6.027370 MHz.

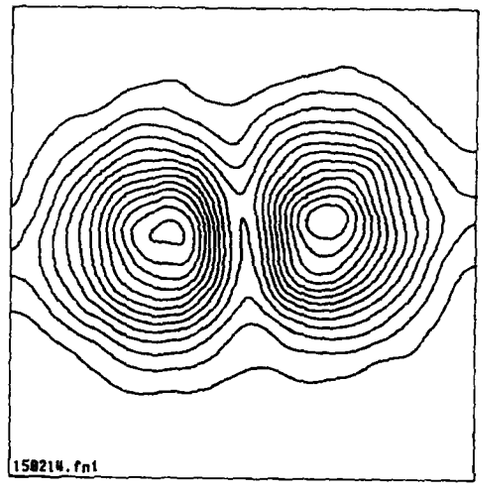


Fig. 11. Sample C; mode (1,2,1), 6.115643 MHz.

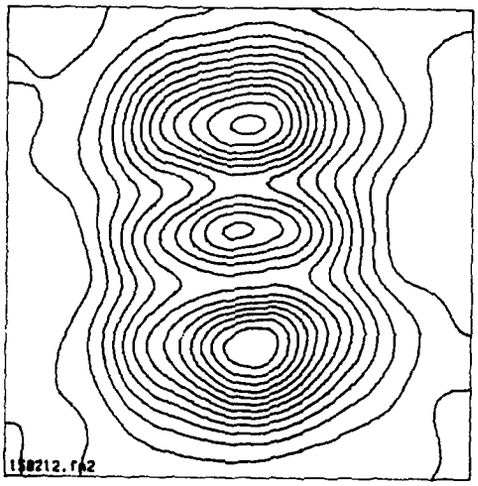


Fig. 12. Sample C; mode (1,1,3), 6.181563 MHz.

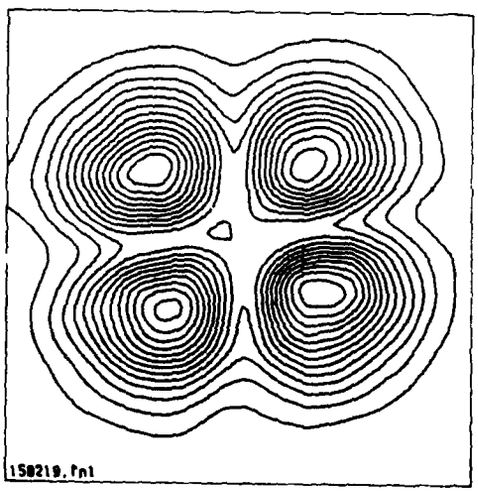


Fig. 13. Sample C; mode (1,2,2), 6.188764 MHz.

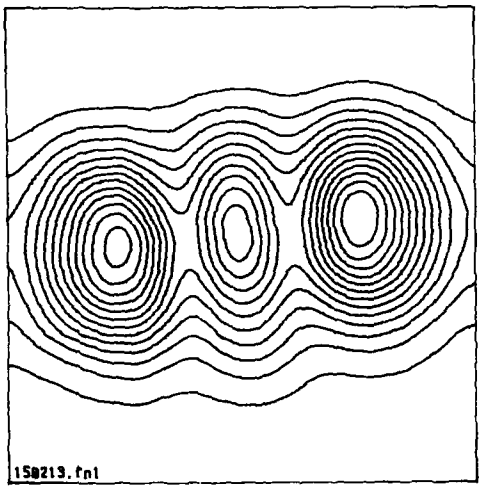


Fig. 14. Sample C; mode (1,3,1), 6.207681 MHz.

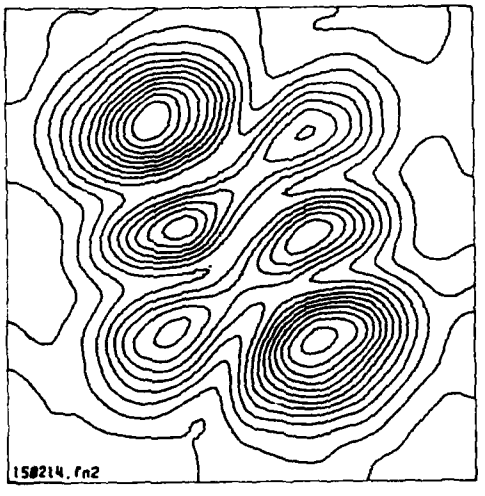


Fig. 15. Sample C; modes (1,2,3), ... 6.264107 MHz.

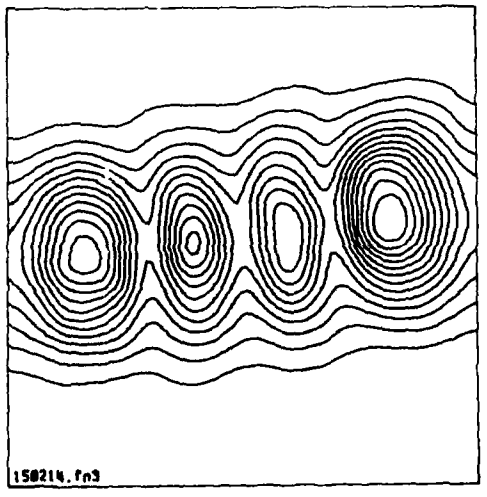


Fig. 16. Sample C; mode (1,4,1), 6.306682 MHz.

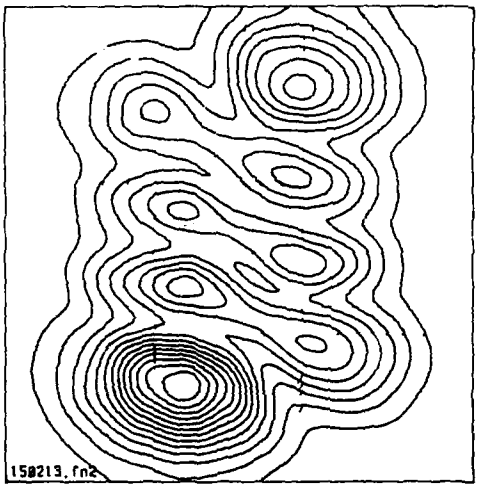


Fig. 17. Sample C; modes (1,2,4), ... 6.339180 MHz.

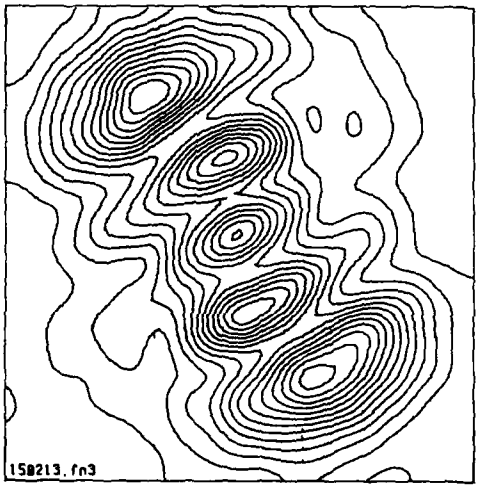


Fig. 18. Sample C; modes (1,1,5), ... 6.344733 MHz.

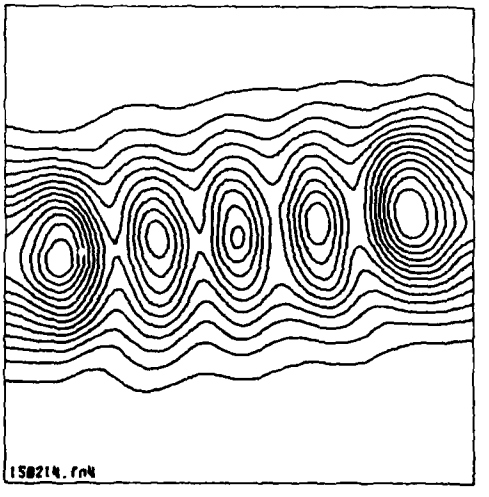


Fig. 19. Sample C; mode (1,5,1), 6.405400 MHz.

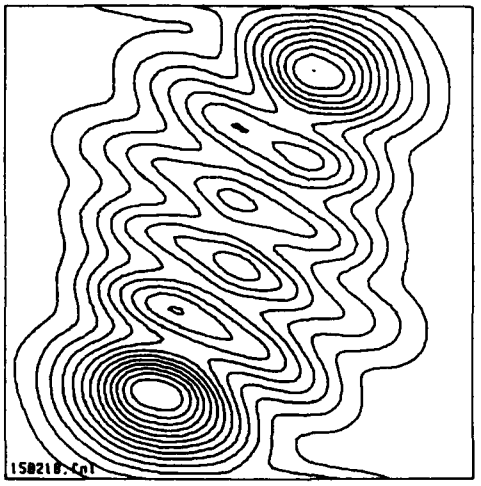


Fig. 20. Sample C; modes (1,1,6), ... 6.415463 MHz.

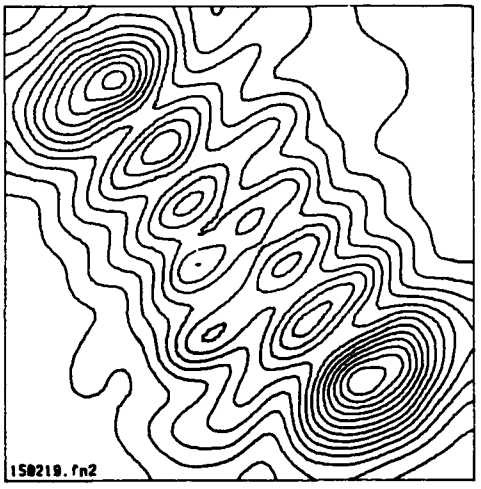


Fig. 21. Sample C; modes (1,1,7), ... 6.496820 MHz.

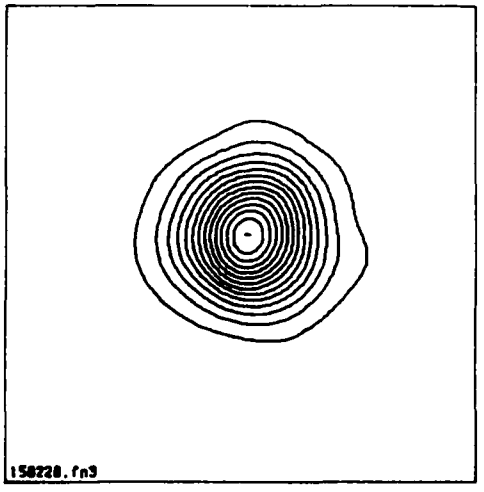


Fig. 22. Sample C; mode (3,1,1), 17.959241 MHz.

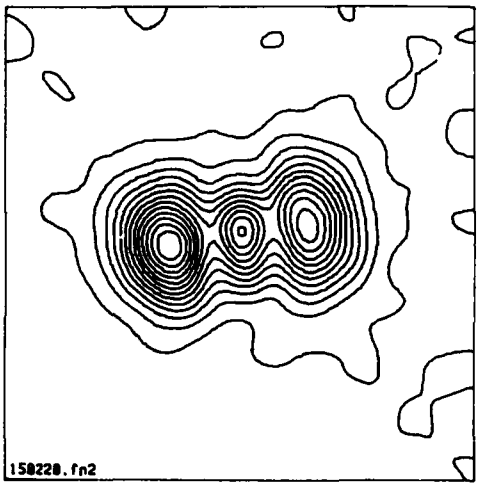


Fig. 23. Sample C; mode (3,3,1), 18.106876 MHz.

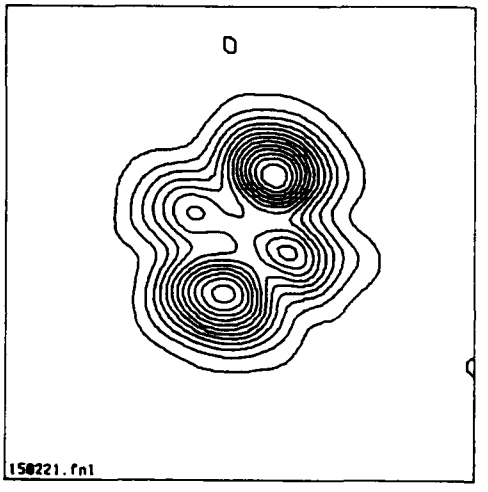


Fig. 24. Sample C; modes (3,2,2), ... 18.112776 MHz.

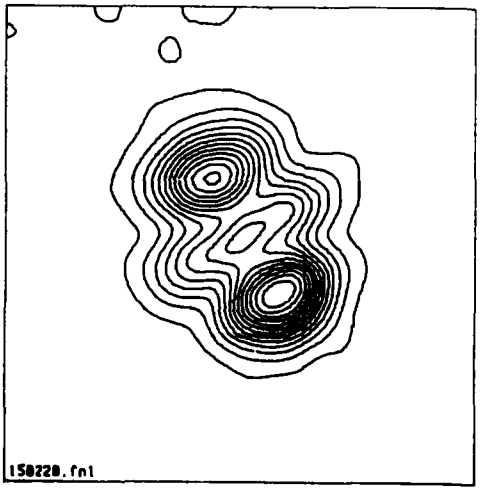


Fig. 25. Sample C; modes (3,1,3), ... 18.114697 MHz.

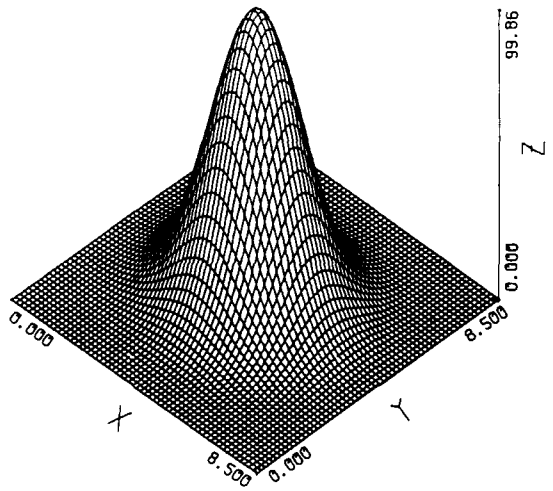


Fig. 26. Calculated mode (1,1,1), 5.984275 MHz.

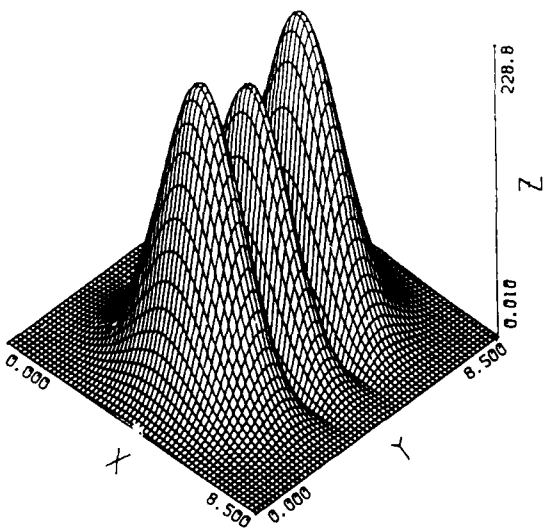


Fig. 27. Calculated mode (1,1,3), 6.165498 MHz.

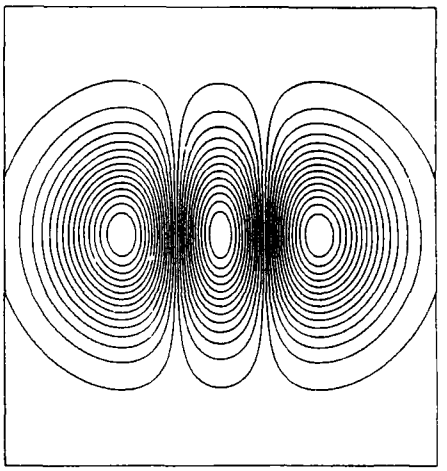


Fig. 28. Calculated mode (1,3,1), 6.212469 MHz.

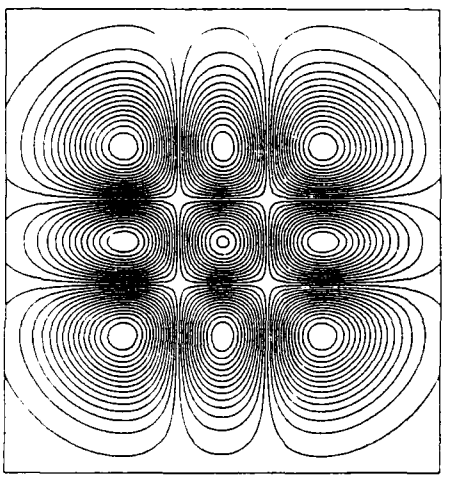


Fig. 29. Calculated mode (1,3,3), 6.387220 MHz.

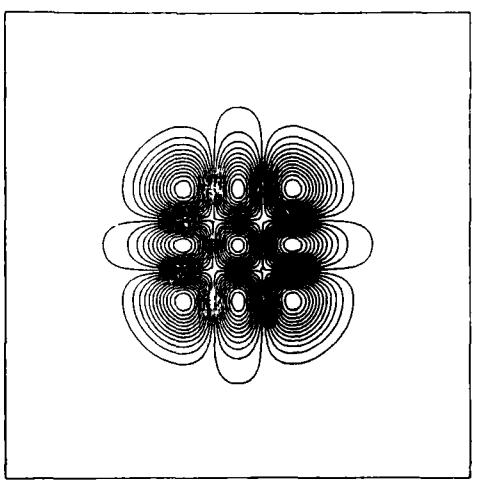


Fig. 30. Calculated mode (3,3,3), 18.161848 MHz.

A STUDY OF FLEXURAL, ANHARMONIC AND THICKNESS-SHEAR MODES OF VIBRATIONS IN QUARTZ RESONATORS USING SCANNING ELECTRON MICROSCOPE

Harish Bahadur* and R. Parshad
National Physical Laboratory
Hillside Road, New Delhi
110 012, India.

SUMMARY

Crystal resonator plates or disks are capable of vibrating in different modes. Primarily, these modes can cause extensional, flexural or shear (thickness or face) motions. In the actual practice of using bonded resonators, due to finite dimensions and reflection from boundaries, there results a complex coupling of these different modes which gives rise to a host of frequencies that a crystal resonator is capable of vibrating at. A variety of experimental techniques have been used for mode identification and also to suppress the undesired modes of vibrations.

Among the new experimental techniques, scanning electron microscopy has been used for last about a decade to study vibrations in quartz crystals. In the technique, an electron beam (accelerated to about 2 kV) is used to monitor the electric potential distribution onto the surface of the resonator plate or disk. The SEM micrographs thus exhibit a particular mode of vibration of the crystal in operation.

In the present studies, we have used different crystal geometries and electrode shapes for mode identification in quartz resonators. Surface patterns have been presented which represent that the higher overtones of low frequency flexural vibrations can couple with the fundamental thickness-shear vibrations in rectangular plates. Flexural vibrations have been observed propagating both along and perpendicular to X-axis of the crystal resonators. For the first time, evidence of circular flexural waves is found using only high frequency (30 MHz, third overtone) thin unbeveled circular disk resonators. Beveling of these resonators makes the flexural waves linear propagating only along the X-axis. This suggests that in the case of unbeveled thin disks the two sets of linear flexural vibrations couple with a phase difference of $n\,\pi/2$ where n is an odd integer.

Vibration patterns of some natural quartz resonators irradiated by Y-rays have been presented and compared with the patterns obtainable upon thermal bleaching. At the present stage, the small change in the nature of electric potential distribution can only be attributed to a small change in resonance frequency of the crystal in two different stages; irradiated and bleached.

At high magnification in the SEM, where only a very small area of the surface is examined, the micrographs represent an uniform electric potential and the smearing in the patterns can be used either to measure the displacement or track the direction of vibrations. Using the smearing technique evidence of shear vibrations at the bevels is presented.

* Harish Bahadur is a Homi Bhabha Fellow and is presently with the Department of Physics, Oklahoma State University, Stillwater, OK, 74078.

INTRODUCTION

Quartz crystal vibrations play a key role in modern Frequency Standards instrumentation. These crystals form a very important link for deriving very low frequencies of the order of 1 Hz from the atomic frequencies of the order of GHz in Atomic Frequency Standards. It therefore becomes important to study the modes of vibrations in quartz crystals.

Crystal resonators (plates or disks) are capable of vibrating in different modes. Primarily, these modes can cause extensional, flexural or shear (thickness or face) motion. In the actual practice of using bonded resonators, due to finite dimensions and reflection from boundaries, there results a complex elastic coupling of these different modes which gives rise to a host of frequencies that a crystal resonator is capable of vibrating at. It is the task of the designer to make a crystal resonator vibrate in as pure a mode as possible by an appropriate choice of plate geometry (by contouring the body and beveling the edges) and also by restricting the electroded area of the crystal (energy trapping). The beveling consists in tapering the edges of the quartz crystals, rectangular or circular, in one or more directions. In the beveled crystal the major surfaces may be plane, plano-convex or double convex. Mathematical analysis and experimental evidences have already suggested that use of beveling makes the vibration spectrum of the quartz crystal much simpler, there being less unwanted vibrational modes than in the unbeveled crystal. In particular, in the beveled crystal the flexural waves are produced at the bevel or near the edge of the bevel. On the other hand the desirable thickness-shear vibrations are concentrated at the center. The spatial isolation between the thickness-shear and the flexural modes avoids or decreases the undesirable coupling between the two classes of vibrations. The unwanted modes cause undesired energy dissipation which tends to reduce the quality factor (Q) of the crystal for obtaining good frequency stability. Thus, it is always desirable to be able to determine experimentally the different modes of vibrations being generated in the completed crystal device. Knowing this, alterations can be done in the design to get the optimum performance.

A variety of experimental techniques such as use of lycopodium powder, electrical and mechanical probes, X-ray diffraction topography, optical detection, laser holography and scanning electron microscopy etc. have been used for mode identification. A survey of all these techniques along with the present status of theory of vibrations in quartz resonators has been recently reviewed by the authors (1).

Among the new sophisticated experimental technique scanning electron microscopy has been used for last abo a decade to study vibrations in quartz resonators. In the technique, an electron beam usually accelerated to

CH2186-0/85/0000-0544$1.00©1985IEEE

about 2 kV is used to monitor the electric potential distribution onto the resonator surface caused due to its piezoelectric effect. The SEM micrographs thus exhibit a particular mode of vibration of the crystal in operation. At high magnification (X 2000 - 5000), where only a very small area of the surface is examined, the micrographs represent an uniform electric potential and the smearing in the patterns can be used either to measure the displacement or track the direction of vibrations.

Scanning electron microscopy of vibrations of quartz crystals has been reviewed in 1980 (2). In this paper we shall describe further results of our studies in the subject using different electrode shapes and crystal geometries for obtaining additional information on vibrational characteristics of quartz crystals. The results presented here demonstrate the complex elastic coupling that can take place between the various modes particularly the fundamental thickness-shear and the higher order overtones of flexural vibrations. It is expected that the technique and results would be useful for designing the resonators to vibrate in purer modes by using appropriate crystal geometry and electrode design.

Until now, the crystal shape most investigated by using SEM has been the centrally plated AT-cut quartz disks (Fig. 1) executing thickness-shear vibrations. It has been shown (2-6) that low magnification SEM patterns (magnification≃ 20) of periodic or aperiodic bright and dark areas represent varying potential distribution, the dark area being at positive potential and the bright areas being at relatively negative potential. Further, regarding the low magnification patterns, it has been shown (2-6) that in the case of energy trapping (vibrational energy being confined to the centrally plated area), the SEM pattern produced is spatially aperiodic, while in the absence of energy trapping, due to leakage of waves from the centrally plated area and their reflection from boundaries, some kind of spatial periodicity results in the nature of the SEM patterns.

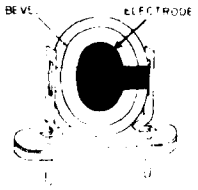


Fig. 1 Mounting arrangement of a circular beveled quartz resonator.

It has been established (2-6) that at high magnification (≃2K), the electron images of different physical features on the resonator surface run into each other due to tangential vibrations and therefore give rise to a smearing effect. For an out-of-plane motion there is no smearing observed (1-3). The smearing effect was indeed made a basis of a technique for accurate measurement of direction and amplitude of tangential vibrations (6,7). Also, it was shown (2,3) that on both the regions of the crystal surface under oscillations showing dark and bright regions of contrast, tangential vibrations exist, thus showing that these regions of contrast are in no way associated with the nodes and antinodes of vibrations.

This paper describes some of the significant new experimental results obtained regarding the nature of SEM patterns related to quartz crystal vibrations for their geometries and electrode shapes used. At the outset it was confirmed that the SEM patterns by and large were reproducible when the electrodes were dismantled and reassembled. Thus the change of patterns, if any, obtained by variation of electrode shape and positioning will have fundamental causes.

EXPERIMENTAL RESULTS

Both rectangular and circular disk (beveled and unbeveled) resonators were used for the investigations. Rectangular AT-cut quartz plates of dimensions 38 x 28 x 0.4 mm^3 capable of oscillating in the fundamental thickness-shear mode in the frequency region of 1.87 MHz were excited with the electrodes of shape shown in Fig. 2. Henceforth, these electrode shapes will be designated as I and II. In these electrode shapes, a sizeable area of the central part of the crystal is left unelectroded, particularly so for the electrode of shape II. The reason for the choice of the electrode shapes with central gap shown in Fig. 2 was the fact that these electrodes were found to suppress the unwanted responses in the vicinity of fundamental thickness-shear mode (11). Hence, the SEM patterns being looked for would give a clearer information regarding the mode shape for the crystals excited in fundamental, overtones or other oscillating mode. In the figure the inner corners of the electrodes adjacent to the gap are marked A thru D. The electrode material was in the form of steel plates of thickness ≃2.12 mm put on the bare quartz and held together by elastic insulating bands. Thus, the electrodes could be easily dismantled and held together again. The X-axis of the quartz crystals was parallel to the length of the plates.

The circular disks used were of plane, unbeveled, beveled and plano-convex geometries having centrally continuously plated circular electrodes. The disks used for investigations were capable of vibrating in the frequency regions centered at 1.87- and 10-MHz. The modes of vibrations studied were fundamental, third harmonic and anharmonics. The quartz crystal resonators were made the frequency determining elements in a self-oscillating circuit. Methods of overtone excitation and generation of anharmonic modes are described elsewhere (8-11).

Studies On Rectangular AT-Cut Resonators:

In the following, first, the nature of SEM patterns for use of electrode shape I is described. Mode patterns obtained were in general complex and aperiodic, the complexity being more than what was obtained for use of the centrally plated circular disks. As an example, Fig. 3 depicts an aperiodic vibration pattern for the third harmonic overtone obtained for a rectangular crystal at the corner 'C'. Some aperiodic patterns using this shape of electrode have been reported in earlier publications (2, 3). In case the pattern obtained was aperiodic, the shape of the pattern, in general, was different at different corners of the electrode (2,3).

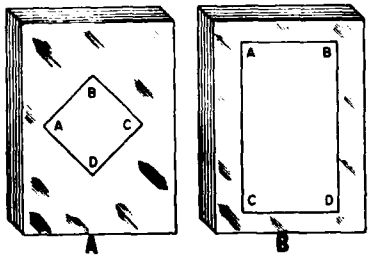


Fig. 2 Electrode arrangement for exciting vibrations in rectangular quartz plate: (A) type I and (B) type II.

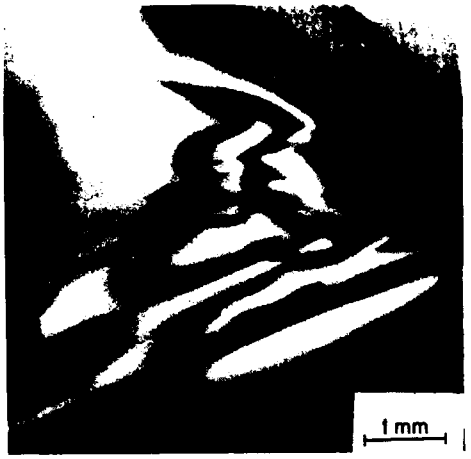


Fig. 3 SEM micrograph of a rectangular quartz plate, with electrodes of type I (Fig.2), showing a complex electric potential distribution on the resonator surface in corner 'C' when the crystal was excited in the third harmonic overtone.

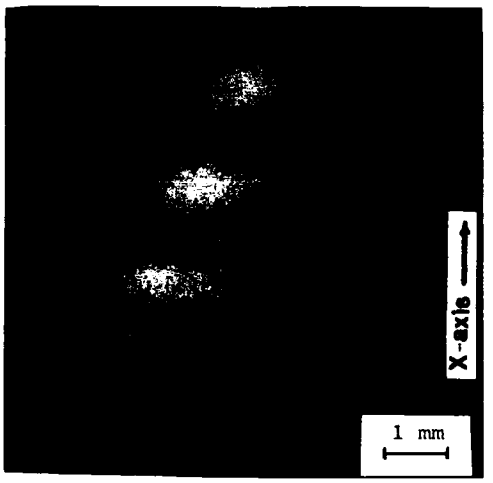


Fig. 4 SEM micrograph of fundamental mode oscillations of a rectangular quartz plate with electrodes of type I (Fig.2).

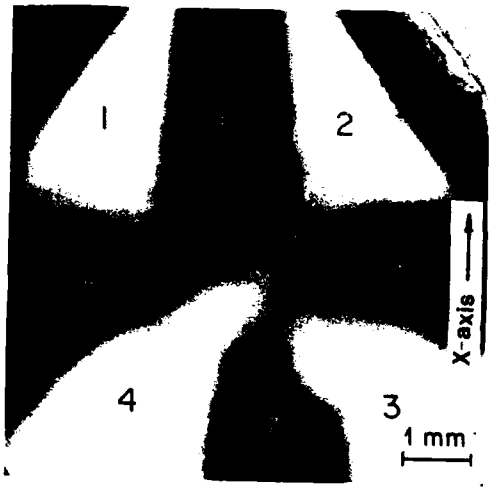


Fig. 5 Third overtone mode oscillations of the same quartz plate for which fundamental oscillations are shown in Fig.4.

There were many variations in the character of SEM patterns with electrode shape I. The patterns might be aperiodic or, in many cases complex, or periodic. For the fundamental and the third harmonic overtone, the nature of mode patterns for the vibrational modes occurred independent of each other. For example, Figs. 4 and 5 depict the patterns for the fundamental and third harmonic overtone. It may be noted that although the pattern for the fundamental (Fig.4) is periodic in structure but that for the overtone (Fig.5) is quite complex. Also, some patterns had a semiperiodic character, such as dots or curls and twists on a background of periodic structure. When the patterns for both the fundamental and the overtone had periodic or semiperiodic structure, corresponding features in the two had a 3:1 dimensional ratio. Figures 6 and 7 illustrate the intensity-modulated patterns for the fundamental and third harmonic overtone of a rectangular quartz plate other than used in Figures 4 and 5. Figure 7 has an added significance since for all the types of electrodes including the circular disks studied earlier by us (4-6), the overtones commonly tend to produce patterns having curls and twists (e.g. see Fig.3 in Ref.4) or equivalent dots (Ref.6, Fig.5). Figure 8 shows the Y-modulated pattern corresponding to Figure 6, wherein the areas appearing bright in the intensity-modulated micrograph stand out more clearly.

The mode patterns occurring on both flat faces of the crystal were compared by turning over the crystal faces (keeping the direction of the X-axis unchanged) to expose each face of it to the electron beam. Interestingly, the new patterns were found to be the mirror images of the older patterns. A rotation of the crystal by 180° in its own plane (direction of X-axis reversed), exhibited the inverted images of the original pattern (12). This suggests the contribution of the X-axis in determining the mode structure.

In some cases, a regular periodic pattern indicative of standing waves for both the fundamental and overtone modes could be obtained. In this case, there was a definite approximate relationship of 3:1 in the band spacings in the patterns for the two vibrational modes. Figures 9 and 10 illustrate this case. In reverse, the definite ratio of band spaing itself confirms that the patterns obtained (Figures 9 and 10) are standing waves since, of-course, the frequency ratio for these modes

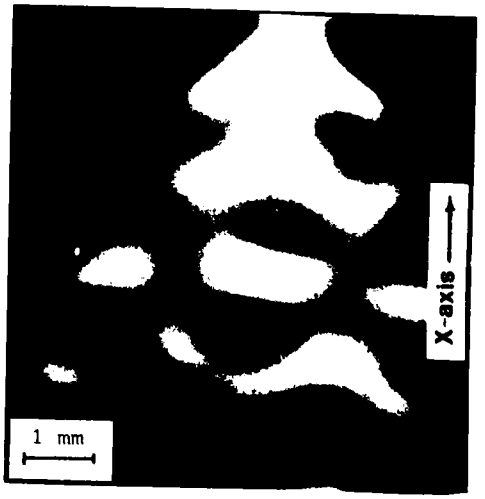


Fig. 6 SEM micrograph of the fundamental mode oscillations of a rectangular quartz plate (other than used in Fig.4) with electrodes of Type I.

are in relation of 1:3. As will be shown later, the micrographs of Figures 9 and 10 depict the generation of an appropriate overtone of flexural vibrations the frequency of which nearly equals to the fundamental thickness-shear vibrations.

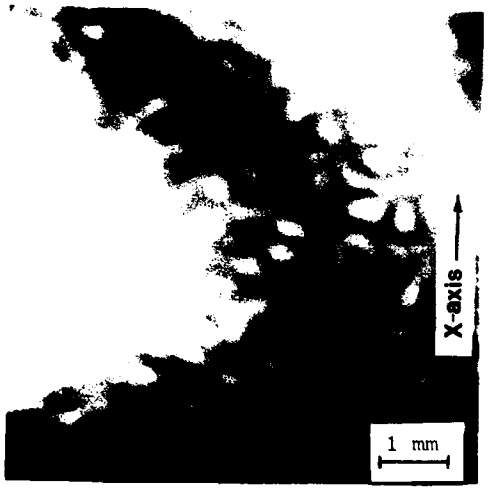


Fig. 7 Third harmonic overtone mode oscillations of the crystal of Fig.6.

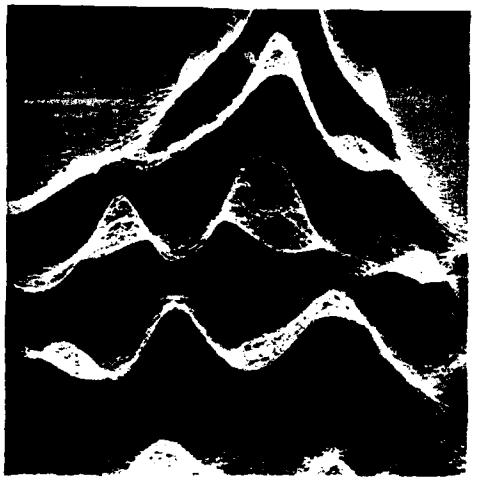


Fig. 8 Y-modulated SEM micrograph of Fig.6.

In the following we shall describe the results using electrodes of shape II. For this more open open type of electrode, the SEM patterns for the fundamental were always periodic, althogh faint, indicating the relative weakness of vibrations, due to the large unelectroded portion (about 70% of the total surface area), compared to those of the electrodes of shape I, and were continuous over the whole unelectroded surface of the crystal (Figure 11). Due to general decreased contrast of the pattern the harmonic overtone patterns were not discernable. This fact should be due to much smaller amplitude

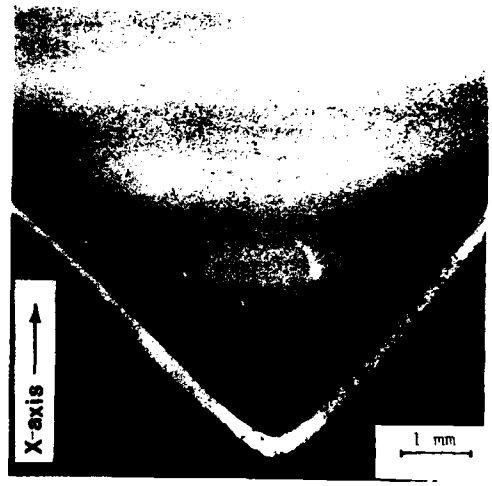


Fig. 9 SEM micrograph of fundamental mode oscillations of a rectangular quartz plate depicting flexural vibrations.

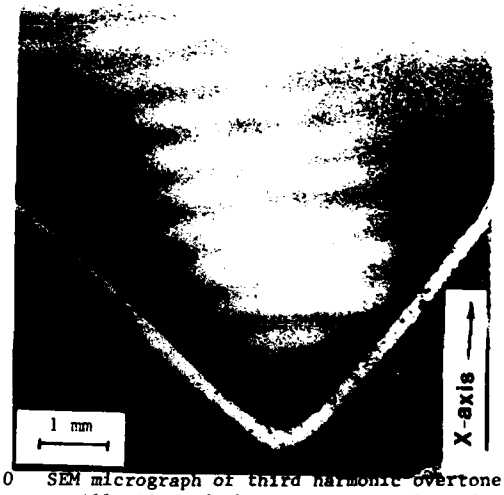


Fig.10 SEM micrograph of third harmonic overtone oscillations of the same quartz plate for which fundamental mode oscillations are shown in Fig.9.

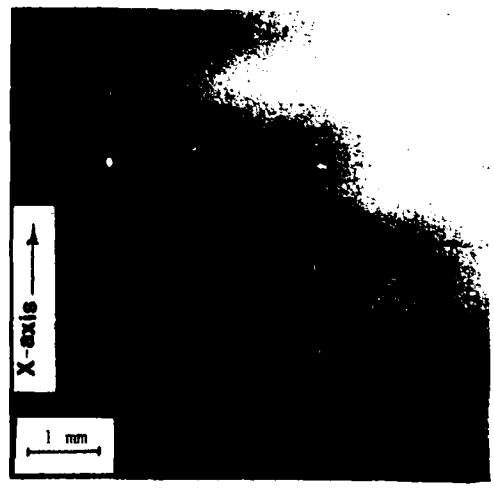


Fig. 11 Fundamental mode oscillations of a rectangular quartz plate with electrodes of type II (Fig.2).

of vibrations of the harmonics than the fundamental and use of the large open area electrode system, this resulting in lack of appreciable activating voltage available for the quartz crystal as a whole. When the electroded area was increased by having a smaller gap in the unelectroded region of the crystal plate (Figure 12), the crystal excitation got increased with the following interesting results. It was found that, exciting the crystal in anharmonic modes, as the frequency was varied, the flexural waves were produced either at right angles to the X-axis or along the X-axis. Figure 13 depicts the SEM pattern of vibrational mode of frequency 1.912778 MHz of a rectangular crystal (of dimensions 38 x 28 x 0.04 mm^3) of 1.87 MHz fundamental frequency in the thickness-shear mode. It may be noted that the vibration are directed parallel to the X-axis. For the same crystal, Figure 14 represents the SEM pattern of mode of frequency 1.202567 MHz representing flexural waves perpendicular to the X-axis.

It was noticed that the contrast for the waves propagating along X-axis was greater than that for those propagating at right angles to the X-axis. This shows that with the overall consideration of boundary conditions and crystal dimensions, the flexural waves tend to propagate more along the X-axis than along any other direction.

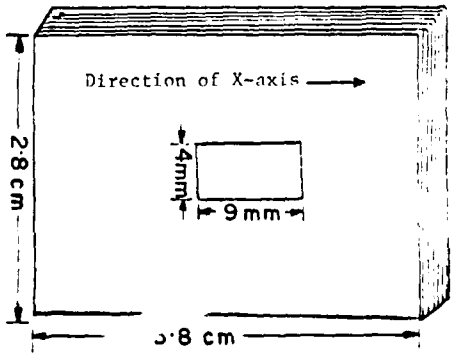


Fig.12 Electrode arragnement for exciting a rectangular quartz plate with less open and more electroded area for increased excitation of the resonator plate.

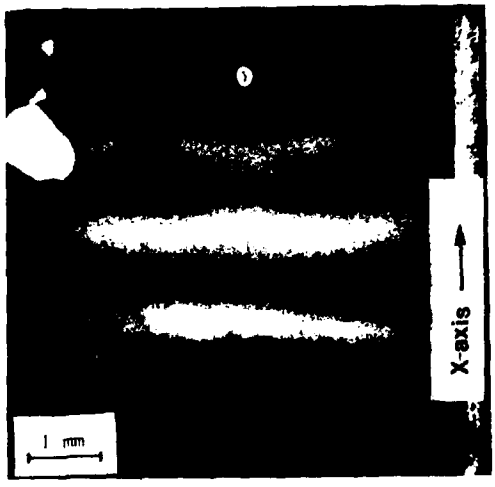


Fig. 13 Oscillating mode of frequency 1.912778 MHz of a 1.87 MHz rectangular quartz crystal depicting genaration of flexural vibrations along X-axis.

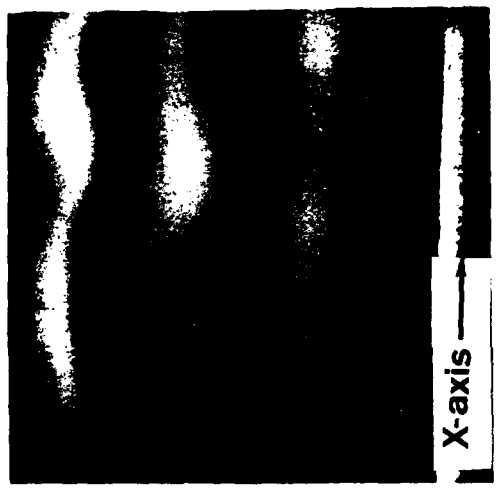


Fig. 14 Oscillating mode of frequency 1.202567 MHz of the same quartz crystal (Fig.13) depicting generation of flexural vibrations perpendicular to X-axis.

Studies On Circular Disk AT-Cut Resonators:

In the above context of generation of flexural waves both along and perpendicular to X-axis, it was interesting to see, using the centrally plated unbeveled circular crystals, that the flexural waves produced were broadly circular. Figures 15 and 16 depict the fundamental mode vibrations of two different unbeveled quartz resonators of about the same fundamental frequency, the exact frequencies for Figures 15 and 16 being 10.001242 and 10.000916 MHz respectively.

It was observed that even in the case of the crystal vibrating in the harmonic overtone of the fundamental thickness-shear vibrations (Figure 17), there were still indications of generation of circular flexural waves. It can be seen that in Fig. 17 modulated on the circular pattern is a dot-like characteristic which has been found to exist in all harmonic overtone vibrations (2-6). Also, in the figure (Figure 17) the disturbing effect of the circular clip of the electrode can be clearly seen by an irregularity of contrast around the clip. All the above effects tend to make the pattern complex. Apart from the above, it was also seen that the anharmonic modes of the harmonic overtones of thickness-shear produce the dot-pattern sometimes degenerating into curls and twists. Figure 18 depicts one such pattern for an anharmonic overtone mode of 5.346226 MHz of a plano-convex crystal vibrating in the fundamental thickness-shear at about 1.87 MHz. Detailed investigations on various anharmonic responses of resonators with different geometries on their mode shapes are contemplated.

It would be noticed that for use of circular electrodes the flexural waves tend to get produced only when the frequency of the oscillating crystal is high enough (~10 MHz). For low frequencies (~5 MHz or lower) only the aperiodic patterns which are unrelated with the flexural waves were produced (2-6).

Use Of Beveled Crystals:

As mentioned in the Introduction, it is well known that beveling of resonators suppresses the generation of flexural waves. In this context, it was of interest to look for generation of flexural waves if any, in the beveled crystals using SEM. It was found that in the beveled crystals using centrally plated circular electrodes, the flexural waves at the fundamental were prod-

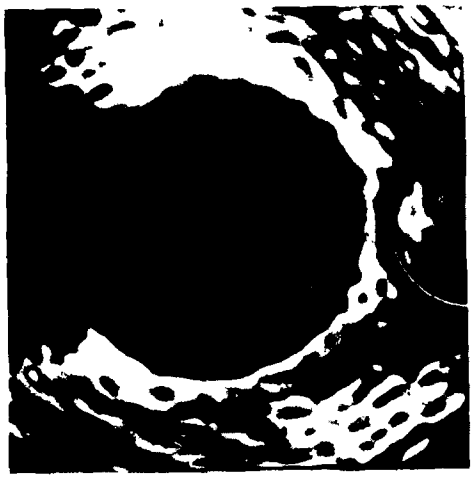


Fig. 15 Generation of circular flexural waves in a centrally plated unbeveled quartz resonator of 10-MHz fundamental thickness-shear frequency.

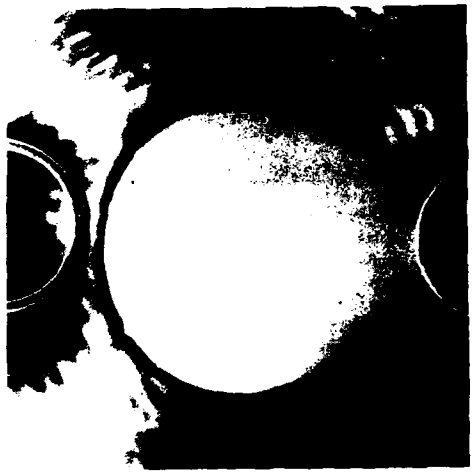


Fig. 16 Generation of circular flexural waves in another centrally plated unbeveled quartz resonator of 10 MHz fundamental thickness-shear frequency.

Fig. 17 Third harmonic overtone oscillation pattern of crystal for which the fundamental oscillations are shown in Fig. 16.

Fig. 18 Anharmonic mode pattern of frequency 5.346226 MHz of a 1.87 MHz plano-convex circular disk.

Fig. 19 SEM micrograph of an anharmonic mode of frequency 9.996068 MHz of a beveled quartz resonator of fundamental frequency 10.001242 MHz.

Fig. 20 SEM micrograph of an anharmonic mode of frequency 9.996459 MHz of the same quartz resonator (Fig.19).

uced along the X-axis. For anharmonic modes also, the flexural waves were produced along the X-axis. Figures 19 and 20 depict the SEM micrographs for the two modes of frequencies 9.996068 MHz and 9.996459 MHz of beveled crystal of fundamental thickness-shear resonance frequency equal to 10.001242 MHz.

Studies On Gamma-irradiated Quartz:

It has found earlier by Hearn and Schwuttke (13) and Hearn (14) using X-ray diffraction topography that in presence of line (planar) defects in the crystal, which are formed during growth process, the vibration patterns tend to align towards the defects. In the present study, it was attempted to see whether SEM could be used effectively to observe such effects. It is known (13-15) that the X-ray diffraction topographs map the defects density distribution. Also, X-ray topographs of dormant quartz plate are known to correspond with their γ-irradiation color patterns (16-18). The areas appearing dark in the topograph also appear dark in the γ-irradiation color-pattern. It happens that in the X-ray topograph, the dark areas indicate the increased diffracted intensity from a defect or region of mechanical strain and therefore appear dark on a photographic film. In the irradiation-color pattern, the unpaired electrons released due to the radiation passing thru the specimen are trapped at the defect sites and form color-centers. Therefore, the X-ray topographs and irradiation-color pattern seem to be identical.

In order to observe the effect of line defects on the nature of SEM patterns, some crystal plates were irradiated uniformly in a Gamma-cell. Many resonator plates were found to show areas of inhomogeneous coloration. Figure 21 shows the irradiation-color pattern of one of the crystal plates exhibiting the presence of bands which may be taken to be due to line or planar defects. This crystal was examined for its mode structure in the SEM. However, in the studies conducted by us, there did not seem to be any relationship in the color-patterns of the crystal and the SEM patterns of the oscillations of the same crystal. As an illustration the micrographs shown in Figures 4 and 5 were obtained for the crystal of which the irradiation-color pattern is shown in Figure 21. The significance of this result would be taken up under 'Discussion of Results'.

Changes in the SEM patterns for crystals irradiated and subsequently bleached were looked for. Figures 22 and 23 respectively illustrate the SEM micrographs of a crystal irradiated and bleached. From the figures, it may be seen that tho the broad character of electric potential distribution (aperiodicity) is more or less the same but the position of bright and dark regions on the surface of the crystal is different. As is well known (19-22), the resonance frequency changes upon irradiation and also the quartz crystals develop a smoky color. Heating of the irradiated crystals at temperature range 250-400^{0}C bleaches out the smoky color and the resonance frequency returns nearly to that of the pre-irradiated state of the crystal (28,29). Due to the above facts, the change of patterns of the irradiated and bleached crystals may be attributed to the change of resonance frequencies of the two states of the crystal.

Studies At High Magnification:

As mentioned in the Introduction and is well known (1-7), the high magnification SEM micrographs can be taken, due to the very small area under investigation, to have uniform potential and therefore represent surface topography and direction of vibrations. In an effort to study the spread of vibrations over the crystal surface, topographical features at high magnifications were studied in the different regions of contrast noted

Fig. 21 Irradiation-color pattern of a rectangular quartz plate (dose=5Mrads using ^{60}Co).

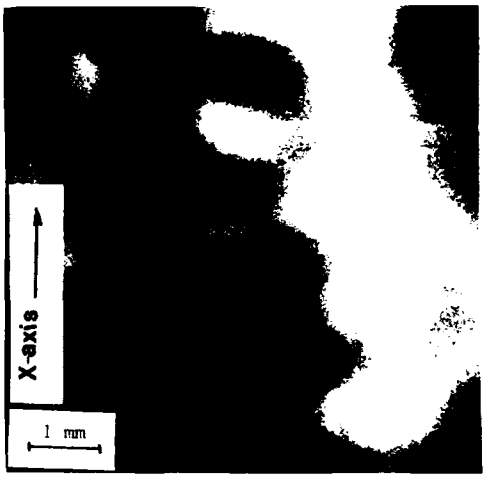


Fig. 22 SEM micrograph of fundamental mode vibration patterns of a gamma-irradiated (5 Mrads) rectangular quartz plate using electrodes of type I.

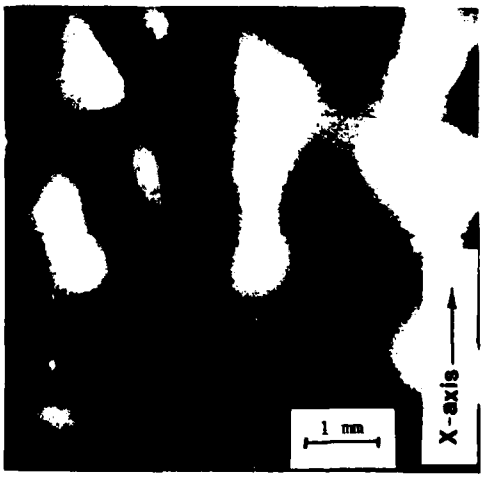


Fig. 23 SEM micrographs of the same crystal resonator (Fig.22) after thermal bleaching to bring the resonance frequency to the value of virgin crystal.

at low magnifications. As an illustration, Figure 24 shows the direction of vibrations in different regions marked 1 thru 8 in Fig.5. It can be seen that the direction of vibrations changes upon proceeding from region of one hue of the pattern observed at low magnification to the other. Also, the amplitudes (displacements) in the bright and dark regions were comparable showing that these regions are not the usual nodes and antinodes characteristic of standing waves. For the patterns which exhibited regular standing waves, the amplitude in the brighter regions was more than that in the darker ones.

Another significant observation made using high magnification was a clear cut evidence, thru smearing, of presence of thickness-shear vibrations at the beveled regions in the circular disk resonators. Figure 25 shows the results of dormant and vibrating states at the bevel.

DISCUSSION OF RESULTS

The following experimental results stand out from the above studies.

For use of electrode of shape I, the mode patterns obtained were complex and aperiodic. In case of these aperiodic patterns, there was no correspondence between the mode structure for the fundamental and the third overtone. Also, in only a few cases, regular patterns indicative of standing waves were obtained. In the case of the latter, it was predominantly the fundamental mode which showed periodicity. And in some cases,more for the overtone than the fundamental, there was a superposition of the periodic and aperiodic characters. Comparision of patterns, using electrodes of shape I and II, showed that the greater the area of the unelectroded region, the greater is the tendency for standing waves to be produced.

It has been shown in Figs. 9 and 10 that the standing waves produced were along the X-axis (the periodic bands being at right angles to this axis). The periodic pattern is supposed to be due to flexural waves. This assumption gets support from the experimental investigations on X-ray topography of vibrating crystal. For example, Figure 26 depicts an X-ray topograph of a quartz crystal executing flexural vibrations, the flexural waves so produced propagating along the X-axis of the crystal. The close similarity in the shape between the X-ray topograph and the SEM patterns may be noted.

Another strong evidence that the standing waves are due to flexural waves may be seen from the following simple computation. In the case of Figure 9, the wavelength of the flexural waves derived from the band spacing and SEM magnification was 0.302 cm. Taking the modulus of elasticity in the X-direction (c_{11}) = 86.05 x 10^6 dynes/cm^2 (25), the velocity V of the waves along the X-direction is $V = (c_{11}/\rho)^{\frac{1}{2}}$ where ρ is the density of quartz material (ρ =2.65), and using the relation $V = \nu\lambda$, the frequency of the flexural waves came out to be about 1.87 MHz, nearly the same as the actual frequency of vibrations of the crystal in the fundamental thickness-shear mode, the frequency of which is primarily determined by the thickness of the crystal and the modulus of elasticity ($c_{66} \simeq$ 29 x 10^{10} dynes/cm^2) in this direction. This shows that the flexural mode generated is an appropriate overtone of the low frequency fundamental flexural waves, the frequency of this overtone being the same as that of the fundamental thickness-shear mode. The coincidence or nearness of the two frequencies enables coupling of the two kinds of waves to form one vibrational system at a coincidental frequency.

It is significant to mention here that the flexural waves are always found to be accompanied by thickness-shear vibrations as detected by the smearing effect at high magnifications at places where flexural waves exist.

It has been shown that by increasing the activating voltage on the crystal by use of less open area (Fig.12), flexural waves can be produced both along and at right angles to the X-axis. This result is in accord with X-ray topographs for rectangular crystals (26). The SEM results further show that there is a greater tendency for the waves to be along X-axis.

In the centrally plated unbeveled crystals, circular waves can be produced. This shows that the boundary conditions for the crystals used are appropriate for producing the flexural waves simultaneously both along the X-axis and perpendicular to it. The boundary condition should favor the generation of proper phase difference ($n\,\Pi/2$ where n is an odd integer) between the two sets of linear flexural waves to be produced.

It has been shown that beveling of crystals produces a clean pattern of flexural waves along the X-axis only. This must be due to the fact that since, as a whole, beveling produces unfavorable conditions for propagation of flexural waves on the crystal surface (27,28) and since, as mentioned already, the waves along the X-axis are more probable than at right angles, the latter are the first to disappear as a result of beveling.

The experimental results indicate that flexural waves are produced more easily for the high frequency crystals ($\simeq$10 MHz). This is understandable since the thinner crystals (being used at higher frequencies) would be mechanically more amenable to flexural deformations.

In contrast to the effective representation of flexural waves in the SEM patterns, it seems from the present and past work that thickness-shear waves do not show up directly in SEM patterns. The reason for this is yet to be cleared up. It seems though not directly confirmed that thickness-shear modes are associated with appearance of aperiodic patterns in the following way.

The aperiodic patterns since they do not show up in X-ray topographs must be due purely to surface phenomena not replicating in bulk. For the explanation of our results, we draw analogy from electromagnetic, where higher order modes, exponentially attenuating with distance and so non-propagating, are produced at the metal discontinuties to satisfy appropriate boundary conditions. In our opinion, the aperiodic SEM patterns are due to higher order acoustic vibrations at electrode metal-crystal interfaces. A consideration in support of the above suggested mechanism for aperiodic waves is that the aperiodic patterns are localized while the periodic patterns, if produced, cover the whole surface area of the resonator.

The present investigations show how almost pure flexural waves can be generated by using an open-type electrode having electric field excitation at the outer areas of the resonator, which, from our results, seem to be the effective area for generation of flexural waves. Earlier, theoretical investigations using plate theory of beveled crystals (29) support this consideration (generation of flexural waves at edges).

The present work also indicates that using electrodes of shape I, there is, particularly for the fundamental, a clear tendency for either the periodic pattern or aperiodic pattern to exist exclusively. This would mean that either the flexural vibrations are generated predominantly or the thickness-shear but not both at the same time.

It will be in order here to compare results on vibr-

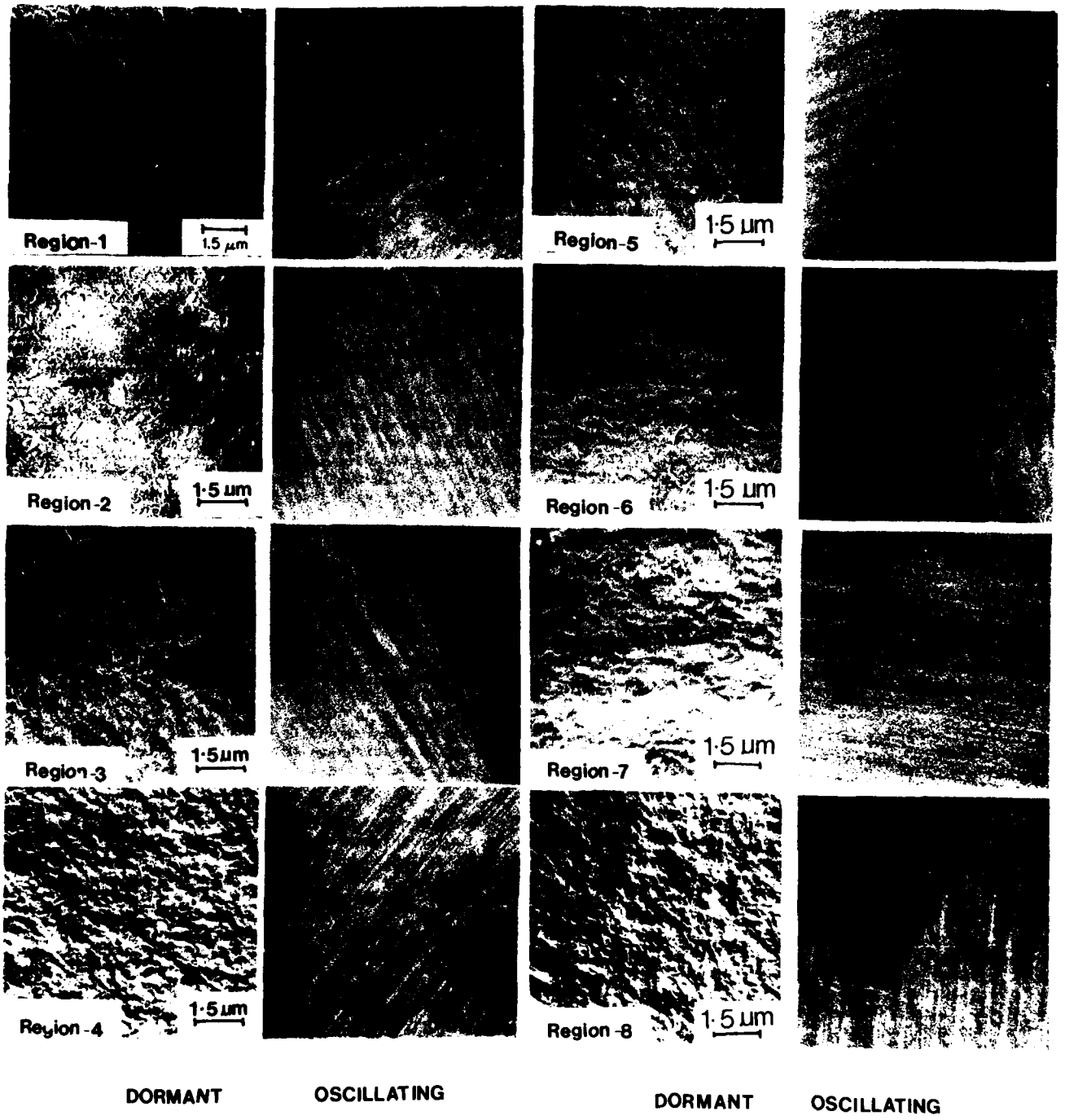


Fig. 24 SEM micrographs recorded at high magnification depicting topographical features of dormant (left) and oscillating (right) quartz crystal for different regions of contrast marked 1 thru 8 in Figure 5. Arrows in the above micrographs indicate the direction of vibrations in different regions (after reference 1).

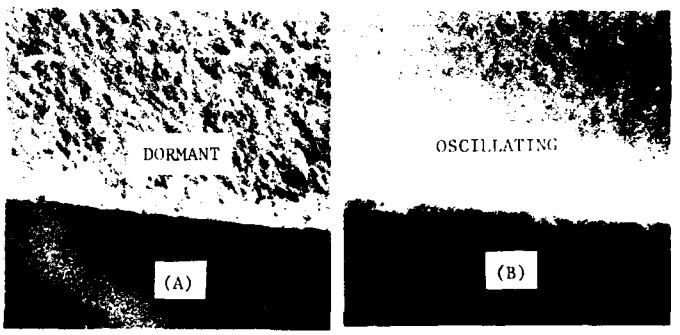


Fig. 25 High magnification SEM micrographs showing topographical features at beveled portion of a circular disk resonator: (A) Dormant and (B) Oscillating in the fundamental mode.

Fig. 26 X-ray diffraction topograph of a rectangular quartz plate executing flexural vibrations (after W.J. Spencer, Physical Acoustics, V, pp. 111-161, Ed. W.P. Mason, Academic Press Inc., New York, 1968).

Fig. 27 X-ray topographs of a vibrating circular disk resonator exhibiting the alignment of vibrations along a line defect in the crystal (after references 13,14).

ations of quartz crystals drawn from X-ray topography and Scanning Electron Microscopy. The X-ray topographs give clear indications of anharmonic thickness-shear modes while flexural waves have not been clearly demarcated especially when centrally plated circular crystals are used. In contrast, the SEM patterns do not show up, for reasons yet to be understood, the patterns for thickness-shear but flexural waves are shown up conspicuously. All this is probably due to the fact that flexural waves are weak compared to the thickness-shear, the X-ray topography does not remain sensitive enough especially due to the background of thickness-shear and the fact that crystals have to relatively strongly excited for recording the X-ray topographs. On the other hand the SEM, since there is no background of thickness-shear patterns, becomes a sensitive tool for detection of flexural vibrations. Thus, it is that the flexural wave patterns have been recorded for the first time using Scanning Electron Microscope.

The SEM study of γ-irradiated quartz has not yielded any fundamental insight at this stage of work except that the change of frequency caused by irradiation is supported by the change in the corresponding SEM patterns. The irradiation-color patterns, in their exact resemblance (15-18) with X-ray topographs of the crystals, represent inhomogeneous inherent mechanical strain in the crystal and as long as the strain pattern does not affect the mode of vibrations of the crystal no relation between the two (color and SEM patterns) is expected. In special cases, however, when the strain is so large that it affects the pattern of vibrations, as, for example, has been observed (13,14) by using X-ray topography (Figure 27), a corresponding effect on the SEM patterns may become noticeable. Future work in this direction would lead to a more exact understanding on the behavior of inherent strain to affect the mode of vibrations in quartz resonators.

It has been shown that the direction of vibrations changes upon proceeding over the area of the crystal, especially in crossing the regions of one hue to the other of the SEM patterns observed at low magnification. The change of direction of vibrations was noted earlier (2,3) for centrally plated circular disk resonators. This fact is not covered by the existing resonator plate theory and must await fundamental explanation. From what has been discussed regarding the role of surface in determining mode shapes, it may well be that the change of direction of vibration is only of local origin, being confined to the surface.

Using smearing effect in high magnification micrographs, there was evidence of generation of thickness-shear vibrations in the beveled part of the resonators. This fact is also not covered by the current theory of resonator plates.

APPLICATION

The p·esent studies can be directly applied in time and frequency standardization for obtaining stable frequencies from quartz crystals. It has been mentioned in the paper that thin crystals would be more prone to flexural deformations and hence generation of unwanted flexural waves. For generating high frequencies in the Frequency Standards, it would be more appropriate to use a thick enough crystals (thickness say greater than 1 mm) and excite them at proper overtone to get the required high frequency rather than to use a crystal thin enough for producing the desired high frequency. This would help avoiding the undesired coupling between the thickness vibrations (used at high frequencies) and the flexural vibrations.

It is of interest to note that Warner (30) has already shown that excitation of quartz crystal resonators in their overtone is desirable for obtaining standard frequencies. Our work adds another fundamental reason for use of thick crystals in frequency standardization. Apart from getting purer thickness vibrations without flexural component, the added advantage of use of thick crystals would be the increase of equivalent inductance of the crystal due to its increased mass. The increased equivalent inductance of the crystal would lead to obtaining increased frequency stability (30).

From what has been discussed, it is expected that the use of a combination of the techniques of scanning electron microscopy and X-ray topography would expose significant information about the vibrational characteristics of quartz crystal resonators.

CONCLUSION

SEM studies of vibrating quartz crystals having different electrode shapes and crystal geometries have been carried out. The results indicate that the SEM patterns, observed at low magnifications, can be spatially aperiodic or periodic or both. The cause of the aperiodic SEM patterns is not known with certainty, tho it seems to be associated with the generation of thickness-shear waves. The regular periodic pattern is due to some higher overtone of the fundamental flexural waves coupling with the thickness-shear vibrations, a coincidental frequency being produced as a result of this coupling. The coupling is confirmed by detection of thickness-shear waves at sites of generation of flexural waves.

Other results obtained are, generation of circular flexural waves for unbeveled crystal disks having centrally plated circular electrodes, existence of thickness-shear waves at bevel sites and existence of flexural waves either along X-axis or at right angles for rectangular crystals having their inner rectangular unelectroded area, and the evidence of change of direction of vibrations at boundaries of different hues in the SEM patterns at low magnifications.

ACKNOWLEDGMENTS

The present studies were conducted as a part of the Homi Bhabha Fellowship program with Harish Bahadur. The authors acknowledge the cooperation of the Indian Institute of Technology, New Delhi for the use of their Scanning Electron Microscope. One of us, Harish Bahadur expresses his appreciation for his association with Prof. Joel J. Martin at Oklahoma State University.

REFERENCES

1. Harish Bahadur and R. Parshad,"Acoustic Vibrational Modes in Quartz Crystals: Their Frequency, Amplitude and Shape Determination", Physical Acoustics, 16, (Eds. W.P. Mason and R.N. Thurston), Academic Press Inc., New York, pp. 37-171 (1982).

2. Harish Bahadur and R. Parshad,"Scanning Electron Microscopy of Vibrating Quartz Crystals - A Survey", IEEE Trans. Sonics and Ultrasonics, SU-27,pp.303-317 (1980).

3. Harish Bahadur and R. Parshad,"Scanning Electron Microscopy of Vibrating Quartz Crystals - A Review", Scanning Electron Microscopy/1980/I, SEM Inc., AMF O'Hare (Chicago), IL, pp.509-522 (1980).

4. Harish Bahadur, Allen Hepworth, V.K. Lall and R.Parshad,"Electron Contrast Effects from Oscillating Quartz Crystals seen by the Scanning Electron Microscope", IEEE Trans. Sonics and Ultrasonics, SU-25, pp. 309-313 (1978).

5. Harish Bahadur, Allen Hepworth, V.K. Lall and R. Parshad,"Some Observations Using SEM for Studying the Ultrasonic Vibrations in Quartz Crystals", Proc. Annual Frequency Control Symposium, USAERADCOM, New Jersey, 32, pp.207-219 (1978).

6. Harish Bahadur, Allen Hepworth, V.K. Lall and R. Parshad,"Studies on SEM Surface Patterns of Oscillating Quartz Crystals and Application to Energy Trapping", Scanning Electron Microscopy/1979/I,SEM Inc., AMF O'Hare, Chicago, IL, pp.333-338 (1979).

7. Harish Bahadur and R. Parshad, "Accurate Measurement of Amplitude of Tangential Vibrations of Quartz Crystals", Ind. J. Pure and Appl. Phys., 18, pp.905-906 (1980).

8. Harish Bahadur and R. Parshad,"Use of Transistor Heavy Biasing for a Novel Method of Generation of Quartz Crystal Overtones and Mixed Frequency Oscillations", Proc. IEEE, 68, pp.1345-1347 (1980).

9. Harish Bahadur and R. Parshad,"A New Method for Determination of the Frequency Spectrum of a Quartz Crystal and Its Application for Generation of a Number of Precise Frequencies and Time Intervals", Proc. European Conf. Precise Electrical Measurements, University of Sussex, London, Sept. 5-9, 1977, pp. 79-82 (1977).

10. Harish Bahadur and R. Parshad,"New Method of Detection of Weak Anharmonic Modes of a Quartz Crystal and Its Application for Generating Multicrystal Resonance Frequencies", J. Appl. Phys. 48, pp.3141-3142 (1977).

11. Harish Bahadur and R. Parshad,"A New Method for Determination of Frequency Spectrum of a Quartz Crystal and Its Application", IEEE Trans. Sonics and Ultrasonics, SU-25, pp.24-29 (1978).

12. Harish Bahadur, V.K. Lall and R. Parshad,"Further Results on Scanning Electron Microscopy of Vibrating Quartz Crystals", Scanning Electron Microscopy/1981/I AMF O'Hare, Chicago, IL. pp.333-342 (1981).

13. E.W. Hearn and G.H. Schwuttke,"Defects and Frequency Mode Patterns in Quartz Plates", Proc. Annual Frequency Control Symposium, U.S. Army Electronics Command, Fort Monmouth, NJ, 24, pp.64-73 (1970).

14. E.W. Hearn,"Topographical and Electrical Analysis of Oscillating Quartz", IBM Tech. Rept., TR 22.1003, IBM Hopewell Junction, New York, NY, May 1970.

15. W.J. Spencer and K. Haruta,"Defects in Synthetic Quartz", J. Appl. Phys. 37, pp.549-553 (1966).

16. U. Bonse,"Elastic Strain and Coloration Pattern in Natural Quartz Crystals", Z. fur Physik, 184, pp. 71-84 (1965).

17. Harish Bahadur "Investigations On Oscillating and Structural Characteristics of Quartz Crystals", Doctoral Dissertation, University of Delhi, 1978.

18. Harish Bahadur and R. Parshad, IEEE Trans. Nucl. Sci., NS-32, pp.1169-1179 (1985).

19. D.B. Fraser,"Impurities and Anelasticity in Crystalline Quartz", Physical Acoustics, (Ed. W.P. Mason), 5, Academic Press Inc., pp.59-110 (1968).

20. B.R. Capone, A. Kahan, R.N. Brown and J.R. Buckmelter,"Quartz Crystal Radiation Effect", IEEE Trans. Nucl. Sc. NS-16, pp. 217-221 (1970).

21. Harish Bahadur and R. Parshad,"Some New Findings On the Effect of Nuclear and X-Irradiation On the Oscillating Characteristics of Quartz Crystals", Ind. J. Phys. 53A, pp. 239-256 (1979).

22. Harish Bahadur and R. Parshad,"Some New Results on the Frequency Characteristics of Quartz Crystals Irradiated by Ionizing and Particle Radiations", Physica Status Solidi (a)67, pp.683-692 (1981).

23. C. Frondel,"Effect of Radiation On the Elasticity of Quartz", Am. Miner., 30, pp.432-460 (1945).

24 C. Frondel,"Final Frequency Adjustment of Quartz Oscillator Plates", Am. Miner., 30,pp.416-431 (1945).

25. W.G. Stoddard,"Design Equations for Plano-convex AT-Filter Crystals", Proc. Annual Frequency Control Symposium, U.S. Army Electronics Command, 17, pp. 272-282 (1963).

26. W.J. Spencer and G.T. Pearman,"X-Ray Diffraction from Vibrating Quartz Plates", Adv. X-Ray Analysis, 13, pp.507-525 (1970)., Plenum Press, New York.

27. R. Bechmann,"Single Response Thickness-shear Mode Resonators Using Circular Beveled Plates", J. Sci. Instr. 29, pp.73-76 (1952).

28. P.C.Y. Lee and S. Chen,"Vibrations of Contoured and Partially Plated Contoured, Rectangular, AT-Cut Quartz Plates", J. Acous. Soc. Am. 46, #5.(Pt. II), pp.1193-1202 (1969).

29. G.T. Pearman,"Thickness-Twist Vibrations in Beveled AT-Cut Quartz Plates". J. Acous. Soc. Am. 45, pp. 928-934 (1969).

30. A.W. Warner,"Design and Performance of Ultra-Precise 2.5 Mc Quartz Crystal Units", Bell System Tech. J., 39, pp.1193-1217 (1960).

ENHANCED COMPOSITE RESONATOR ANALYSIS AND ITS APPLICATION TO THE QUARTZ CRYSTAL MICROBALANCE

Ewald Benes
Institut für Allgemeine Physik, Technische Universität Wien
Karlsplatz 13, A-1040 VIENNA, Austria

Klaus-Christoph Harms
AVL Gesellschaft für Verbrennungskraftmaschinen und Meßtechnik m.b.H.
Herderg. 11, A-8010 GRAZ, Austria

Gernot Thorn
Leybold-Heraeus G.m.b.H.
Wilhelm-Rohn-Str. 25, D-6450 HANAU 1, FRG

Abstract

The sensor head of a quartz crystal thin film thickness monitor can be regarded as a two-material quartz-film composite resonator and it has already been shown in earlier works that the specific acoustic impedance ratio z of the deposited film and the crystal strongly influences the mass sensitivity of heavily preloaded crystals. In contrast to the purely mechanical composite resonator treatment of MILLER and BOLEF, in our work a one-dimensional composite resonator analysis is presented which takes the influence of the electrical load between the resonator electrodes into consideration. The treatment starts from an equivalent electrical circuit based upon a transmission line analogy and is the most general (inner-acoustic losses included) one-dimensional treatment of the composite resonator. It is exact to the extent that the one-dimensional treatment accurately models the physical situation under consideration. Specializing to the loss-free case and to open circuited electrodes (vanishing electrical load admittance) the well-established resonance equation after LU and LEWIS is yielded, which now turns out to be the description of the parallel composite resonator frequencies. In the case of short-circuited electrodes a series resonance equation is obtained which describes the entire harmonic series resonance spectrum of the quartz-film composite.

The piezoelectric excitability alteration of the resonance frequencies of the entire quasiharmonic spectrum is shown to be a direct continuous function of the mass load. Additionally, the series resonance spectra of several plano-convex AT-cut crystals with different load steps and materials (Al, Ti, Cu) has been measured and analyzed. It is shown that the introduced series resonance equation is of significance for the accurate determination of the z-value from the measurement of two frequencies of the resonance spectrum. Furthermore, the mass loads and film thicknesses indicated with the loaded crystal probes by use of a two-frequency (Auto-Z-Match) thin film thickness monitor in the prototyping state are compared directly with the results obtained by a conventional Z-Match technique and by the older so-called frequency- and period-measurement techniques, as well as with the AAS (Atomic Absorption Spectrometry) mass determination as a sufficiently sensitive independent reference method. The results clearly indicate the superiority of the auto-slope-accomodation technique based upon the two-frequency measurement method.

Introduction

Thin film thickness monitoring during vacuum deposition is generally accomplished by the use of oscillating quartz crystal microbalances. In the past, the major steps forward in this field were each correlated with improvements of the mass load versus frequency relation upon which the thickness monitor is based.

Early investigations on quartz crystal resonators by Sauerbrey[1,2] and Lostis[3] indicated that, for small mass changes, the decrease in a thickness shear vibration mode frequency of a crystal upon which a thin film is deposited is linearly proportional to the deposited mass. The theoretical derivation of the mass load versus frequency equation given by Sauerbrey was based on the substitution of the areal mass density Δm_Q of an assumed additional quartz layer by the areal mass density m_F of the deposited foreign material. This substitution was defended by the brilliant heuristic argument that the elastic properties of the foreign film cannot contribute to the resonance frequency as long as the film is entirely located within the antinode region of the resonator displacement amplitude (compare Fig. 1) and hence no shear deformation of the foreign material itself takes place. With this basic assumption, which is obviously valid as long as the thickness of the deposited material is very small compared with that of the crystal, one yields the following relation between the mass load (reduced areal mass density) m_F/m_Q and the reduced frequency decrease $\Delta f/f_Q$

$$\frac{m_F}{m_Q} = \frac{f_Q - f}{f_Q} = \frac{\Delta f}{f_Q}, \qquad (1)$$

where $m_F = \rho_F l_F$ with ρ_F and l_F being the density and the thickness of the deposited foreign layer, $m_Q = \rho_Q l_Q$ with ρ_Q and l_Q being the density and the thickness of the quartz crystal, f_Q and f are the resonance frequencies of the unloaded and the loaded crystal, respectively, and Δf is the frequency decrease caused by the mass load. This formula was the original justification for the use of quartz crystal resonators as mass-sensing devices, since it indicates that the resultant frequency decrease is independent of all properties of the added material except its mass. Equation (1) is supported by early experimental data,[4] but it has been shown to be accurate only up to a mass load of $m_F/m_Q < 2\%$.

Introducing the fundamental substitution

$$l_Q = \frac{\lambda_Q}{2} = \frac{v_Q}{2f_Q} = \frac{N_{AT}}{f_Q}, \qquad (2)$$

CH2186-0/85/0000-0556$1.00©1985IEEE

where λ_Q is the wavelength of the thickness shear wave excited in the crystal, v_Q is the phase velocity of that wave and $N_{AT} = v_Q/2$ the frequency constant of the commonly used AT-cut crystal, the thickness l_F of the deposited film can be determined with

$$l_F = \frac{\rho_Q N_{AT}}{\rho_F f_Q} \frac{m_F}{m_Q} . \qquad (3)$$

Quartz crystal thin film thickness monitors of the first generation used Eq. (3) in connection with (1) and are commonly referred to as devices using the "frequency-measurement technique".

Attempts by Stockbridge[5] to justify Eq. (1) theoretically have viewed the additional mass of the deposited film traditionally as a perturbation of the unloaded quartz crystal resonator. This more rigorous approach predicts the mass load versus frequency relation (1) as the first order term of a series expansion. The method, however, did not find practical use since the higher order terms require experimentally measured constants with no direct physical interpretation. Moreover, the basic assumption is a negligible amount of potential energy stored in the deposited foreign layer during oscillation, which again restricts the analysis to small film thicknesses with negligible thickness shear deformation.

In the years following Sauerbrey's original work, the usable mass load range was dramatically extended by improvements in both the crystal design (increase in the adhesive strength of the electrodes and utilization of plano-convex crystals) as well as the feedback gain control range of the driving circuits. At that time, it was shown by Behrndt[6] that the period τ of the crystal oscillation increases in proportion to mass loading according to

$$\frac{m_F}{m_Q} = \frac{\tau - \tau_Q}{\tau_Q} = - \frac{f - f_Q}{f} \qquad (4)$$

where τ and τ_Q are the oscillation period of the loaded and the unloaded crystal, respectively. Applying this relation, the usable mass load range with tolerable thickness display error was stated to be $m_F/m_Q < 10\%$. Second generation quartz crystal thickness monitors[7] based on Eq. (4) in connection with (3) are commonly referred to as "period-measurement technique" devices.

At this point, it should be emphasized that the terms "frequency- and period-measurement technique" are not very appropriate. The technique would be better described by the equation used to deduce the interesting areal mass density, respectively the deposited film thickness, than by the primary measurand acquired (frequency or period of oscillation). In the past, there actually has been a thin film thickness monitor manufacturer that used a frequency meter hardware, but, nevertheless, used the frequency term of Eq. (4). Hence, this monitor must be regarded as a "period-measurement technique" device with respect to the extended usable crystal load range.

Presently, the last step in the refinement of the proper mass-load versus frequency equation is based upon a one-dimensional continuous wave acoustic analysis of the quartz/film-composite resonator by Miller and Bolef.[8] This theoretical treatment was adopted by Lu and Lewis[9] who reduced the formidable expression of Miller and Bolef through lengthy but simple algebra to the more handsome form

$$z_Q \tan\frac{f}{f_Q}\pi + z_F \tan\frac{f}{f_F}\pi = 0 \qquad (5)$$

where f is the composite resonance frequency and f_Q and f_F are the mechanical resonance frequencies of the crystal and the film, respectively, z_Q and z_F are the specific acoustic impedances of the crystal and the film, respectively, with regard to the piezoelectrically excited shear wave. The explicit mass load versus frequency relation

$$\frac{m_F}{m_Q} = - \frac{z_F f_Q}{z_Q \pi f} \arctan\left\{\frac{z_Q}{z_F} \tan\frac{\pi f}{f_Q}\right\} \qquad (6)$$

is yielded from Eq. (5) by the following elementary substitutions:

$$f_Q = \frac{1}{2l_Q}\sqrt{\frac{c_Q}{\rho_Q}} , \qquad f_F = \frac{1}{2l_F}\sqrt{\frac{c_F}{\rho_F}} \qquad (7)$$

$$z_Q = \rho_Q v_Q = \sqrt{\rho_Q c_Q} , \qquad z_F = \rho_F v_F = \sqrt{\rho_F c_F} \qquad (8)$$

where v_Q and c_Q are the shear wave velocity and the shear stiffness constant effective in the crystal with respect to the propagation and displacement direction of the piezoelectrically excited acoustic wave, and v_F and c_F are the respective parameters in the film . The validity of Eq. (6) is experimentally justified up to mass loads of $m_F/m_Q < 70\%$. It is easy to show that Eqs. (4) and (1) can be derived from Eq. (6) by successive approximations.[9] If the acoustic impedance ratio is unity, Eq. (6)[10] is identical to Eq. (4) without any approximating step. Third generation crystal thickness monitors based on Eq. (6) are commonly referred to as Z-Match[11] technique devices.

The only drawbacks of these superior third generation thickness monitors are the need for the knowledge of the acoustic impedance ratio $z = z_F/z_Q$;[12] and the necessity of manual keying-in that z-value for every deposited material. Since the z-value actually effective in thin films deposited with different rates is not always exactly known, the high accuracy potential of thickness monitors offering the Z-Match feature often cannot be utilized to full advantage.

In this work an advanced composite resonator model is described which is based upon a transmission line analogy. Using this mathematical description, it is shown how the z-value can be determined accurately from the measurement of two resonance frequencies of the quasiharmonic spectrum. Finally, the accuracy of a thin film thickness monitor which uses these calculated effective z-values of the crystal-film composite (Auto-Z-Match[11] technique) is demonstrated for some deposition material examples with different specific acoustic impedance values.

Theoretical model

Equivalent electrical circuit

Following the transmission line concept after Krimholtz, Leedom and Matthaei,[13] a piezoelectric composite thickness shear resonator can be represented by an equivalent circuit according to Fig. 2. In this figure F_1,[14] v_1,[15] F_2, v_2 and F_3, v_3 are the forces and particle velocities at the acoustic terminals, those are the crystal back and front side as well as the film front side, respectively; V and I are the voltage and the current at the electrical port, respectively. The element values are given by the following equations:

$$Z_Q = A\sqrt{\rho_Q c_Q}, \quad Z_F = A\sqrt{\rho_F c_F}, \tag{9}$$

$$C_o = \varepsilon A/l_Q, \tag{10}$$

$$\Gamma = \frac{\varepsilon Z_Q \omega}{2e \sin(\gamma_Q l_Q/2)}, \tag{11}$$

$$Z_x = j(e/\omega\varepsilon)^2 \frac{\sin\gamma_Q l_Q}{Z_Q}, \tag{12}$$

$$\gamma_Q = j\omega\sqrt{\rho_Q/c_Q}, \quad \gamma_F = j\omega\sqrt{\rho_F/c_F}, \tag{13}$$

where Z_Q and Z_F are the acoustic impedances of the crystal and the film, respectively, regarding the cross section given by the electrode area A, c_Q and c_F are the elastic shear stiffness constants in the crystal (with respect to the propagation and particle displacement direction of the piezoelectrically excited wave) and the film, ρ_Q and ρ_F are the densities, γ_Q and γ_F the regarded wave propagation constants in the respective media, ε and e are the permittivity and the piezoelectric constant in the crystal plate in thickness direction, l_Q is the thickness of the crystal plate, ω the corresponding angular frequency, Γ the transformation ratio of the transformer between the acoustic and the electric domain, and C_o is the capacity built by the electrode areas and the quartz crystal as dielectricum.

The nomenclature chosen is very close to that of Schüssler.[16] The generalization to dissipative media can be made by interpreting c, z, and γ as complex parameters according to Holland.[17]

The equivalent electrical circuit presented in Fig. 2 - which is applied here to an AT-cut quartz crystal resonator - offers two advantages compared with the lumped-element equivalent circuit originally developed by Mason.[18] First, a clear distinction can be drawn between the lumped-element electrical behaviour and the wave acoustic (overtone) behaviour of the crystal. Second, it is easy to demonstrate the effect of various acoustic loads (deposited films) and electrical loads (especially open and short circuited electrodes).

From the basic electric transmission line theory,[19] the first electro-acoustic analogy,[20] and the multiple port network theory,[21] the impedance matrix representation of the crystal and the correspondence between the electrical and acoustical parameters at the terminals can be derived from the three-port box in Fig. 2 (compare also the formally identical representation of a ceramic thickness shear resonator plate described by Schüssler[16]):

$$\begin{pmatrix} F_1 \\ F_2 \\ V \end{pmatrix} = \begin{pmatrix} Z_Q\coth\gamma_Q l_Q & -Z_Q/\sinh\gamma_Q l_Q & NZ_C \\ Z_Q/\sinh\gamma_Q l_Q & -Z_Q\coth\gamma_Q l_Q & NZ_C \\ NZ_C & -NZ_C & Z_C \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ I \end{pmatrix} \tag{14}$$

$$N = eA/l_Q, \tag{15}$$

$$Z_C = -j/\omega C_o. \tag{16}$$

The transmission matrix representation of the passive (non-piezoelectric) film and its acoustic terminal parameters can be determined as

$$\begin{pmatrix} F_2 \\ v_2 \end{pmatrix} = \begin{pmatrix} \cosh\gamma_F l_F & Z_F\sinh\gamma_F l_F \\ \frac{1}{Z_F}\sinh\gamma_F l_F & \cosh\gamma_F l_F \end{pmatrix} \begin{pmatrix} F_3 \\ v_3 \end{pmatrix} \tag{17}$$

Resonance equations

Substituting F_2 and v_2 in Eq. (14) by F_3 and v_3 using Eq. (17), as well as introducing the boundary condition for vacuum,

$$F_1 = F_3 = 0, \tag{18}$$

we yield after rearranging:

$$(v_1/I)Z_Q\coth\gamma_Q l_Q - (v_3/I)Z_Q\cosh\gamma_F l_F/\sinh\gamma_Q l_Q = -NZ_C \tag{19}$$

$$(v_1/I)Z_Q/\sinh\gamma_Q l_Q - (v_3/I)(Z_Q\coth\gamma_Q l_Q\cosh\gamma_F l_F + Z_F\sinh\gamma_F l_F) = -NZ_C \tag{20}$$

$$(v_1/I)NZ_C - (v_3/I)NZ_C\cosh\gamma_F l_F - V/I = -Z_C \tag{21}$$

where v_1/I and v_3/I are the particle velocities at the acoustic terminals referred to the current into the electrical port. The voltage/current ratio V/I defines the electrical impedance Z_e' appearing at the electrical port. In order to yield the resonance frequencies of the composite resonator, we assume no externally applied voltage source, we consider, however, an arbitrary external electrical termination impedance Z_e. Thus, as a consequence of the validity of Kirchhoff's rules, the admitted (eigen-)frequencies force the resonator to match its impedance to the external electrical load:

$$V/I = Z_e' = -Z_e. \tag{22}$$

Solving the linear equation system (19) to (21) with regard to V/I by elimination of v_1/I and v_3/I, we yield the most general resonance equation of a lossy piezoelectric thickness-shear composite resonator with arbitrary external electrical load Z_e:

$$\{N^2 Z_C/Z_Q\}\{[1 - (\cosh\gamma_Q l_Q)^{-1}]^2/[\tanh\gamma_Q l_Q + z\tanh\gamma_F l_F] + \tanh\gamma_Q l_Q\} - 1 = Z_e/Z_C \tag{23}$$

where the reduced specific acoustic impedance

$$z = z_F/z_Q = Az_F/Az_Q = Z_F/Z_Q \tag{24}$$

has been introduced. Equation (23) is exact to the extent that the thicknesses and the mass loads of both electrodes are vanishing and that the one-dimensional treatment used accurately models the physical situation under consideration. Equation (23) also describes the special case of viscous (liquid) overlayers upon the crystal if the imaginary part of the c-values are interpreted according to the viscous damping mechanism. Thus, Eq. (23) fully contains the viscoelastic overlayer model introduced by Kanazawa.[22]

Using the definition equations for N and Z_C and

$$Z_Q = A\sqrt{\rho_Q c_Q} \tag{25}$$

$$k^2 = e^2/c_Q\varepsilon \tag{26}$$

where c_Q again is the effective (piezoelectrically stiffened) elastic constant and k is the electromechanical material coupling factor, we can express the first factor in Eq. (23) by:

$$N^2 Z_C/Z_Q = -jk^2/Q \tag{27}$$

with the abbreviation

$$Q = \pi f_n/f_Q \tag{28}$$

where f_n is the quasiharmonic resonance frequency of the quartz/film composite regarding the mode number n and f_Q is the fundamental mechanical resonance frequency of the crystal.

Specializing to the case of open-circuited electrodes $Z_e \to \infty$, we can satisfy the resonance condition by setting zero the denominator in Eq. (23). By means of this we yield immediately the resonance equation after Lu and Lewis[9]

$$z_Q \tanh\gamma_Q l_Q + z_F \tanh\gamma_F l_F = 0 \ , \tag{29}$$

however, extended to dissipative media if the propagation constants γ_Q, γ_F and the specific acoustic impedances z_Q, z_F of the crystal and the film, respectively, are interpreted as complex parameters according to Holland.[17] Thus, it has been shown that the use of the introduced equivalent circuit and matrix representation not only provides a much shorter derivation of the equation after Lu and Lewis from an entirely different starting point, but also this treatment takes into account the possible regarding of dissipative media and the influence of arbitrary electrical loads on the resonance frequencies.

Specializing to the case of short-circuited electrodes $Z_e = 0$, we yield the series resonance equation of the quartz/film composite resonator:

$$Kz_Q \tanh\gamma_Q l_Q + z_F \tanh\gamma_F l_F = 0 \tag{30}$$

$$K = \frac{1 - (2k^2/jQ)\tanh(\gamma_Q l_Q/2)}{1 - (k^2/jQ)\tanh\gamma_Q l_Q} \tag{31}$$

where K may be interpreted as factor which takes into account the influence of the short-circuited electrodes on the effective specific acoustic impedance

$$z_Q^s = Kz_Q \ . \tag{32}$$

Because of the high quality factors of crystal resonators, the change to the loss-free treatment is fully justified and means a not not· able decrease of the model's accuracy. Hence we y interpret the specific acoustic impedances z_Q and z_F of the crystal and the film, respectively, as pure real and the respective wave propagation constants γ_Q and γ_F as pure imaginary parameters

$$\gamma_Q = j\pi f_n/f_Q l_Q \ , \qquad \gamma_F = j\pi f_n/f_F l_F \tag{33}$$

Additionally, changing from the hyperbolic to the trigonometric functions, Eqs. (30) and (31) become

$$K(Q) z_Q \tan Q + z_F \tan F = 0 \tag{34}$$

$$K(Q) = \frac{1 - (2k^2/Q)\tan(Q/2)}{1 - (k^2/Q)\tan Q} = 1 + \frac{(k^2/Q)\tan Q[\tan(Q/2)]^2}{1 - (k^2/Q)\tan Q} \tag{35}$$

where we used the abbreviations

$$Q = \pi f_n^s/f_Q \ , \qquad F = \pi f_n^s/f_F \ . \tag{36}$$

The superscript s indicates the regarded series resonance case (short-circuited electrodes). As a consequence of the very small electromechanical coupling factor k of a quartz crystal, the numerical value of K(Q) is usually very close to unity, except for the poles at frequencies corresponding to l_F/l_Q ratios of the film and quartz thicknesses for which an integer multiple of a quarter wave-length in the crystal coincides with the crystal thickness. (Even multiples of a half wave-length in the crystal, ever, do not coincide with poles of K(Q)). $K(Q).z_Q$ may be interpreted as (frequency dependent) effective specific acoustic impedance of the crystal which takes into account the reaction of the short-circuited electrodes on the effective elastic constant as a result of the piezoelectric effect.

If we consider a vanishing piezoelectric coupling factor k=0, which characterizes a non-piezoelectric material,[23] we obtain exactly the well established resonance equation after Lu and Lewis[9]

$$z_Q \tan Q + z_F \tan F = 0 \ . \tag{37}$$

The same equation is obtained by specializing to the loss-free case Eq. (29), which was derived directly from the condition of open-circuited electrodes. In order to explicitly indicate the mechanical or parallel resonance frequency type determined by Eq. (37), we use the superscript p:

$$Q = \pi f_n^p/f_Q \ , \qquad F = \pi f_n^p/f_F \ . \tag{38}$$

Equation (34) is more relevant for quartz crystal thin film thickness monitors than Eq. (37), since, all presently used oscillator driving circuits in such devices excite the crystal at a low impedance resonance which is very close to the crystal's series resonance. However, it can be shown by a numerical analysis that the relative difference in the evaluated mass load or film thickness by use of the series resonance equation (34) instead of the commonly used parallel resonance equation (37) is only about 0.3 %. This is within the usual uncertainty of the calibration measurements with reference methods, therefore it cannot be recognized, as long as only one (in common crystal thickness monitors the fundamental) resonance frequency is utilized.

In order to completely check the correspondence of specialized versions of the series resonance equation (34) with established relations, we investigate the case of vanishing film thickness $l_F \to 0$ which means $f_F \to \infty$. In this case, Eq. (34) degenerates to

$$\tan(Q/2) = Q/2k^2 \ . \tag{39}$$

This is exactly the relation, according to Onoe, Tiersten and Meitzler,[24] between the fundamental me-

chanical resonance frequency f_Q of an infinite crystal plate and the corresponding series resonance spectrum $f_Q^{s(n)}$ (for vanishing film thickness the regarded series resonance frequency f_n^s of the composite resonator becomes the series resonance frequency of the unloaded crystal) if the reduced mass load $2m_E/m_Q$ of the electrodes is neglected. Since the deviations of $f_Q^{s(n)}/n$ from f_Q are very small, Eq. (39) is more often used in the approximation form

$$f_Q^{s(n)} = nf_Q(1 - 4k^2/n^2\pi^2) \tag{40}$$

which can be yielded from Eq. (39) setting

$$f_Q^{s(n)}/f_Q = n + (f_Q^{s(n)} - nf_Q)/f_Q \tag{41}$$

and using

$$\tan(n\pi/2 + \Delta) = -\cot\Delta \approx -1/\Delta \tag{42}$$

for $\Delta \ll n\pi/2$ and n restricted to odd numbers. (There is naturally no series resonance frequency defined for an even symmetry crystal vibration mode).

In order to demonstrate the influence of the mass load and the z-value on the entire resonance spectrum (fundamental frequencies and overtones), the parallel resonance Equation (37) is written in its explicite form

$$m = \frac{\rho_F l_F}{\rho_Q l_Q} = -z\frac{f_Q}{\pi f_n}\operatorname{Arctan}\left(\frac{1}{z}\tan\frac{\pi f_n}{f_Q}\right) , \tag{43}$$

where the subscript n of the composite frequency f_n indicates the specific number of the quasiharmonic frequency considered. For the numerical evaluation of Eq. (43), one must keep in mind that the reverse goniometric function is indefinite periodic

$$\operatorname{Arctan} x = \arctan x + k\pi , \tag{44}$$

where $-\pi/2 < \arctan x < \pi/2$ is the principal value of the reverse tan-function and $k = 0, \pm1, \pm2, \ldots$ an integer number. Hence Eq. (43) can be written

$$m = z\frac{f_Q}{f_n}\left\{n - n_Q - \frac{1}{\pi}\arctan\left(\frac{1}{z}\tan\pi\frac{f_n}{f_Q}\right)\right\} , \tag{45}$$

with n_Q being the integer number nearest to f_n/f_Q

$$-0.5 < (n_Q - f_n/f_Q) < 0.5 . \tag{46}$$

In Fig. 3, the plot of Eq. (45) is shown for n=1,2,3 and three z values corresponding to an assumed quartz crystal film (z=1) and to a W and a Mg film representing materials of extremely high (z=6.27) and extremely low (z=0.63) specific acoustic impedance, respectively. The plot of the series resonance version of Eq. (45) shows nearly the same picture, the deviation remains within the graphic representation accuracy.

One can see from Fig. 3 that the resonance frequency spectrum of a composite resonator for a given reduced areal mass density m generally ($z \neq 1$) does not consist of equally distanced frequencies, in contrast to that of a homogeneous resonator (z=1). Therefore, in this paper, resonance frequencies of a composite resonator are termed as "quasiharmonic" to distinguish them both from the harmonic frequencies of a homogeneous resonator, as well as from the anharmonic resonance frequencies,[25,26] which are not solutions of the one-dimensional model.

Piezoelectric excitability

The piezoelectric excitation of a resonance vibration of an unloaded crystal is only possible for odd overtones. The piezoelectric effect produces for a standing wave with an even symmetry no voltage between the electrodes; vice versa, there is no reciprocal interaction of an applied voltage coming into force to drive such an even mechanical deformation standing wave mode. In the case of a composite resonator, the fractional deformation symmetry share of the crystal changes continuously with increasing mass load (compare Fig. 1) and a corresponding continuous change in piezoelectric excitability has to be expected.

In order to show the excitation condition for the composite resonator more quantitatively, the voltage amplitude $\hat{V}$ produced by a given displacement amplitude $\hat{x}$ of the crystal's backside surface is derived in the following. Starting from Eq. (19), (20) and (21) and assuming the case of parallel resonance (open electrodes), I = 0, we yield, after rearranging, the three equations:

$$\frac{v_1}{V}\coth\gamma_Q l_Q - \frac{v_3}{V}\frac{\cosh\gamma_F l_F}{\sinh\gamma_Q l_Q} = 0 , \tag{47}$$

$$\frac{v_1}{V}\frac{1}{\sinh\gamma_Q l_Q} - \frac{v_3}{V}\{\coth\gamma_Q l_Q \cosh\gamma_F l_F + z\sinh\gamma_F l_F\} = 0 , \tag{48}$$

$$\frac{v_1}{V}NZ_C - \frac{v_3}{V}NZ_C\cosh\gamma_F l_F = 1 . \tag{49}$$

Eliminating v_3/V by suitable combination of Eqs. (47) and (49), one yields for the ratio of the sound particle velocity and the corresponding voltage

$$\frac{v_1}{V} = \frac{1}{NZ_C(1-\cosh\gamma_Q l_Q)} . \tag{50}$$

Changing to the loss-free case and using Eq. (7), we can substitute

$$\gamma_Q = j\pi\frac{f_n^p}{f_Q^p l_Q} . \tag{51}$$

Doing this, we can go over from the hyperbolic to the trigonometric function. Additionally substituting the particle velocity v_1 by the corresponding displacement x_1

$$v_1 = j\omega x_1 \tag{52}$$

we finally get

$$\frac{x_1}{V} = \frac{\varepsilon}{e}\frac{1}{1-\cos(\pi f_n^p/f_Q^p)} = \frac{\hat{x}_1}{\hat{V}} \tag{53}$$

for the ratio of the mechanical displacement amplitude $\hat{x}_1$ on the crystal backside surface and the amplitude of the driving voltage $\hat{V}$.

In order to define a driving oscillator circuit independent quantity to measure the ease of excitability of a certain composite resonance the mass load dependent $\hat{V}/\hat{x}_1$ ratio of Eq. (53) is referred to the $\hat{V}/\hat{x}_1$ ratio of the unloaded crystal excited at the

fundamental resonance frequency (f_n^p/f_Q^p=n=1). This relative value is suggested to be termed as electromechanical resonator efficiency

$$\eta_p = \frac{1 - \cos(\pi f_n^p/f_Q^p)}{2} . \qquad (54)$$

The sub- and superscript p indicates the assumed case of piezoelectric excitation of the parallel resonance frequencies.

Figure 4 shows the continuous change of this electromechanical resonator efficiency η_p from unity to zero corresponding to the special cases of purely odd and even f_n/f_Q ratios. It can also be seen from this figure that after exceeding a certain material dependent mass load, it is easier to excite the second quasiharmonic than the third one. This effect explains quantitatively the observation already described by Cady that crystals with one thicker electrode are also excitable at even overtone frequencies.[27,28] It also explains the experience occasionally made by users of quartz crystal deposition monitors in the case of utilizing heavily loaded crystals that the sensor oscillator output frequency sometimes tends to jump to a frequency roughly twice the fundamental one. Such jump tendencies, of course, can be easily avoided with properly designed sensor oscillator circuits by implementing a band pass transfer function tuned to the fundamental frequency range in the feedback circuit.

In Fig. 4, the frequency regions occupied by composite resonators with mass loads of e.g. 30% consisting of deposition materials with extremely different acoustic impedance ratios z, like W and Mg, are indicated. Since the excitation limit of a piezoelectric resonator driven by an oscillator circuit depends directly on the feedback loop gain, it is necessary to make a proper assumption about the automatic gain control (AGC) range of the oscillator circuit in order to determine the excitation limit. In Figs. 3 and 4, the non-excitable frequency regions are shown shaded for an assumed oscillator circuit gain control range of 1:10 corresponding to a minimum required electromechanical resonator efficiency $\eta_{pmin} = 0.1$. These indicated theoretically excitable quasiharmonic frequency ranges may be reduced in reality for deposition materials offering high material inherent acoustic losses, since Figs. 3 and 4 are based upon the assumption of a lossless composite resonator.

Attention should be paid to the fact that, in the parallel resonance case, there is a direct, damping-independent relation between the voltage V at the electrodes and the crystal displacement x, while in the series resonance case there is a direct, damping-independent relation between the current I through the electrodes and the crystal displacement x. Hence,

$$\eta_s = \hat{I}/\hat{x} \qquad (55)$$

is the appropriate definition for the series resonance electromechanical resonator efficiency. η_s also changes continuously with increasing mass load, but an additional decrease with increasing quasiharmonic number n appears.

Determination of the specific acoustic impedance ratio

Corresponding to the different possible n-values Eqs. (34) and (37) can be regarded as sets of linearly independent relations. Thus, in principle, the evaluation of both unknown composite resonator parameters $m=m_F/m_Q$ and $z=z_F/z_Q$ is possible if two quasiharmonic frequencies are measured.

By insertion of two measured quasiharmonic composite frequencies f_{n1} and f_{n2}, corresponding to the quasiharmonic numbers n_1 and n_2, in Eq. (34) or (37), one yields two equations for the assumed to be unknown reduced specific acoustic impedance z and reduced areal density m. The z-value implicitly determined by these equations can be used for the conventional mass load or film thickness calculation from the fundamental resonance frequency. The only hardware adaption needed for this z-determination is a quartz crystal driving oscillator circuit with the capability of switching from the excitation of the fundamental to that of a quasiharmonic overtone frequency. To keep the frequency range, which has to be accepted by the sensor oscillator frequency counter, within the same order, the overtone frequency is preferably divided by the quasiharmonic number n before it is connected to the counter input. In Fig. 5, the described principal arrangement is shown for the example n_1=1 (fundamental resonance frequency f_1) and n_2=3 (third quasiharmonic resonance frequency f_3). The switching from the fundamental to the third quasiharmonic is only necessary once before each deposition cycle since this z-value can be used during the entire deposition cycle.

At this point, it should be mentioned that a reasonable z-determination accuracy of this two frequency measurement method must not be expected as long as the crystal is not sufficiently loaded by the foreign material, since small loads cause no sufficient deviation from the trivial harmonic spectrum. In any case, this z-determination from two frequencies has a high exacting demand to the mathematical model.[29] Thus, besides the introduction of the series resonance equation, further improvements of the model are necessary and described in the following sections.

Acoustic impact of electrodes. The acoustic influence of the electrodes can be taken into consideration by a four layer model (Fig. 6) and the corresponding extension of the equivalent circuit introduced in Fig. 2. This enhanced equivalent circuit is shown in Fig. 7. In principle, a rigorous derivation of a four layer resonance equation can be performed starting from a matrix representation corresponding to the equivalent circuit of Fig. 7. This derivation is, however, very lengthy and tedious and the result is basically the same as can be achieved by utilizing the following heuristic arguments.

The backside electrode may simply be considered by its areal mass density m_E, because this electrode remains located at the displacement antinode (which coincides with the shear stress node) position even for high deposited film thicknesses l_F. The same is not true for the frontside electrode, since this electrode moves with increasing l_F toward increasingly shear stressed zones (compare Fig. 6). Thus, one must expect the possibility of a noticeable contribution of the elastic or acoustic properties of this electrode to the resonance frequency. Hence, the four-layer model can be reduced to a two-layer model if the backside electrode (leftside in Fig. 7) combined with the crystal is regarded as piezoelectrically "active" resonator part Q* and the frontside electrode combined with the deposited film is regarded as "passive" resonator part F*. This passive composite layer may be hypothetically treated as homogeneous film which can be represented by its effective specific acoustic impedance z_{F*} and its fundamental resonance frequency f_{F*} which both will be very close to the respective deposited film para-

meters z_F and f_F, provided the thickness of the deposited film is significantly higher than that of the frontside electrode.

The piezoelectrically active part can be represented by z_Q, which is not influenced by the remaining antinode electrode, and by the fundamental resonance frequency

$$f_{Q*} = f_{QE}/(1-m_E/m_Q) \tag{56}$$

where f_{QE} is the fundamental frequency of the crystal with the mass load influence of both electrodes on frequency lowering taken into account (compare Eq. (1)), and m_E/m_Q is the reduced areal mass density of one (the frontside) electrode.

Since the unknown parameter f_{F*} is eliminated by the two-frequency z-determination technique, the acoustic impact of the electrodes can be considered by simply introducing f_{Q*} instead of f_Q in Eq. (34) and expecting an effective z*-value of the two-layer model which deviates slightly from the true physical z-value.

Deviations from the one-dimensional model. The influence of various crystal shapes and electrode configurations as well as other subtle effects on the resonance frequency spectrum has already been investigated in literature.[25,26,28,30-33] We may adopt these theoretical treatments, which consider the influence of the real crystal geometry, for an improved description of the active part in the composite resonator model. This active resonator part is represented in Eq. (34) by the fundamental mechanical resonance frequency f_Q of the infinite crystal plate. In the case of the trivial mechanical resonance spectrum of an infinite crystal plate, the parameter f_Q can be mode number n independently expressed by the parallel resonance frequencies $f_Q^{p(n)}$:

$$f_Q \equiv f_Q^{p(n)}/n \ . \tag{57}$$

The form of Eq. (39) or (40) shows that already the series resonance frequencies of an infinite crystal plate are not integral numbers of the fundamental. Considering the influence of the plano-convex crystal geometry according to Tiersten and Smythe,[26] the spectra of both the parallel and the series resonance frequencies are deviating significantly from the trivial equidistant spectrum of an infinite purely mechanical resonator. Using Eqs. (A10) to (A13) from the Appendix, the calculated reduced series and parallel frequency spectra of an infinite plate and a plano-convex crystal as well as the electrode mass load influence are shown in Fig. 9 for comparison. The numerical values for the crystal parameters used are given in the Appendix. For the radius of curvature R=0.27m, for the maximum thickness of the plano-convex crystal as well as for the thickness of the compared plane plate $l_Q=2.771\times10^{-4}$m, and for the mass load caused by one electrode $m_E/m_Q=6\times10^{-3}$ were chosen since these values are nominal for quartz crystals presently used in thin film thickness monitors. From Eqs. (A10) and (A12) in the Appendix can be learned, that the plano-convex geometry and the mass load $2m_E/m_Q$ caused by both electrodes influences the series as well as the parallel resonance spectrum by the same factors $(1+PC_n)$ and $(1-2m_E/m_Q)$, respectively. Hence, the correspondence between the series and the parallel spectra both for the plane and the plano-convex crystal are given in excellent approximation by

$$f_Q^{s(n)} = f_Q^{p(n)}(1-4k^2/n^2\pi^2) \tag{58}$$

From the described differences between the infinite plate and the plano-convex crystal, we may conclude the effective parameter f_{Q*n} of the piezoelectrically active resonator part in the series resonance equation (34):

$$f_{Q*n} = f_{Q*}^{p(n)}/n = f_{QE}^{s(n)}/n(1 - 4k^2/n^2\pi^2 - m_E/m_Q) =$$

$$= (1 - m_E/m_Q)f_Q^{p(n)}/n \ . \tag{59}$$

Thus, we have expressed the effective mechanical resonance frequency of the active part of the two-layer model by the series frequency spectrum of the real crystal $f_{QE}^{s(n)}$ with included mass load effect of the electrodes. This is important, since these $f_{QE}^{s(n)}$ frequencies are the only parameters of the unloaded crystals that are directly and with reasonable accuracy measurable. These are the frequencies which are immediately provided by the two-frequency oscillator circuit according to Fig. 5 if the unloaded crystal is driven.

Expressing m explicitly by Eq. (34) for two different quasiharmonic mode numbers n_1 and n_2, we get two linearly independent equations. Dividing the first by the second one, we eliminate m and yield

$$\frac{f_{n2}^s}{f_{n1}^s}\,\frac{(n_1-n_{Q1})\pi - \arctan\{z^{-1}K(Q_1)\tan Q_1\}}{(n_2-n_{Q2})\pi - \arctan\{z^{-1}K(Q_2)\tan Q_2\}} - 1 = 0 \tag{60}$$

where K(Q) is defined by Eq. (35),

$$Q_1 = \frac{\pi f_{n1}^s}{f_{Q*n1}} \ , \quad Q_2 = \frac{\pi f_{n2}^s}{f_{Q*n2}} \ , \tag{61}$$

and the parameters f_{Q*n1} and f_{Q*n2} are interpreted according to Eq. (59). Using this enhanced equation for the z-determination, we must keep in mind that z is now the effective reduced specific acoustic impedance z_{F*}/z_Q, where z_{F*} is the acoustic parameter representing the deposited film combined with the frontside electrode, which fictively are regarded as one homogeneous preload. Therefore we may expect a small variation of this calculated effective z-value of the whole preload.

Measurement results

z-determination from two resonance frequencies

The crystals investigated were ordinary plano-convex AT-cut specimen with Au-electrodes and a fundamental resonance frequency of slightly below 6 MHz, as they are widely used in present thin film thickness monitors. Aluminum, titanium and Copper were chosen as deposition materials representing examples with low, medium and rather high z-values, respectively. With these deposition materials a number of crystals has been loaded with different film thicknesses. The depositions were made with a Leybold-Heraeus electron evaporation gun ESV 14 with rates of 15 nm/s, 10 nm/s and 5 nm/s for the aluminum, titanium and copper, respectively.

The fundamental and the third quasiharmonic

series resonance frequencies of the loaded crystal probes have been driven with an oscillator concept according to Fig. 5 and measured with a conventional frequency meter. Using these frequencies and the nominal electrode mass load value $m_E/m_Q = 6x10^{-3}$ as input parameters of Eq. (60), the effective z-values of the deposited films are calculated and plotted in Fig. 10 as a function of the film thickness. The film thickness values indicated at the x-axis were taken from the display of an Inficon IC 6000 crystal thin film thickness monitor, that is, in fact, by use of Eqs. (6) and (3).

Accuracies of the different quartz crystal microbalance techniques

Since the user of a thin film thickness monitor is not interested in z-values at all, the thickness display values of the different crystal film thickness measurement techniques have been compared directly. In order to show the relative thickness display deviations of the older techniques from the Z-Match technique, which is the presently most advanced monitor technique available on the market, the relative deviations of the regarded thickness displays 1_{FX} from the thickness display 1_F of a Z-Match technique concept are plotted in Figs. 11 to 13 for the different deposition materials Al, Ti, Cu and as a function of the totally deposited film thickness upon the crystals.

In addition, the relative deviations of the different quartz crystal based film thickness monitoring techniques are also compared with an independent reference method. As sufficiently sensitive reference method the Atomic Absorption Spectrometry (AAS) has been chosen. The measurements were performed by use of a Varian AA 575 Spectrometer. In comparison with the conventional microbalance, the AAS is of higher and material selective sensitivity; hence, it is insensitive to contaminations occuring during the handling of the loaded crystal outside the vacuum chamber. Since the AAS method is sensitive to the total mass solved, the determined mass values have to be divided by the deposition area upon the crystal to get the areal mass density m, which is the primary measurand of the crystal head. These deposition area measurements have been performed by a precision mirror scale with nonius. As a consequence of the not very well defined edges of the deposition area obtained with standard crystal head apertures, the diameter of the circle area has an uncertainty of 1.5 %. This means for the area accuracy 3 %. In connection with the AAS accuracy of 1 % this results in an worst case (absolute) accuracy of 4 %. In the Figs. 11 to 14 these AAS measurement points with its worst case tolerance are also drawn in.

It should be emphasized at this point that the reduced deviations of the film thicknesses shown in Figs. 11 to 14 are identical with the reduced deviations of the areal mass densities

$$(1_{FX} - 1_F)/1_F = (m_{FX} - m_F)/m_F \qquad (62)$$

where 1_{FX} and 1_F are the film thicknesses, m_{FX} and m_F are the areal mass densities yielded by the regarded determination method and by the conventional Z-Match technique, respectively. Hence, the reduced value presentation chosen in Figs. 11 to 14 is not effected by the badly defined thin film volume mass densities.

In practice, the accumulated film thickness upon the crystal is a result of numerous deposition processes (charges) and, thus, we are interested in the thickness display accuracy obtained for an additional layer of an approximate thickness of 1µm upon crystals with different preloads. For the preload and the additional layer the different materials Cu and Ag have been chosen, thus, the areal mass densities of the preload and the additional layer could be determined separately by the AAS-method. The results are shown in Fig. 14.

Summary and conclusion

The theoretical background of the quasiharmonic resonance frequency spectrum of a composite piezoelectric resonator and a new technique for the determination of the specific acoustic impedance of a deposited passive layer upon a quartz crystal using two frequencies of the quasiharmonic spectrum have been deduced. The question about the particular piezoelectrical excitability of the mechanically possible composite resonator frequencies has been analyzed theoretically. For this purpose, an equivalent circuit representation of the quartz/film-composite was used, which was applied to the AT-cut thickness shear quartz crystal resonator loaded with a foreign film.

An exact one-dimensional treatment of the piezoelectric quartz/film composite resonator has been presented. The desired series resonance equation has been adapted to the real physical situation by substituting the trivial mechanical resonance spectrum of the model's piezoelectrically active resonator part by the accurately measurable se.ies resonance spectrum of the commonly used plano-convex crystal.

The z-values calculated from the measured fundamental and third quasiharmonic resonance frequencies show significant deviations from the respective bulk values of the regarded material. The z-values of less loaded samples always seem to be too low. Although a small variation of the effective z-value with increasing film thickness could be expected from grain variations and from the fact that we determine the effective z-value of the electrode+film composite, these effects surely cannot explain the strong deviations shown in Fig. 10. Besides, it has been shown in an earlier work of one of the authors[29] that the z-values obtainable by the two-frequency method by use of Eq. (60) are very little dependent on the film thickness, if the correct (measured) electrode mass load m_E/m_Q is used in the calculation. In the present work the electrode thickness and mass load has not been determined, but a nominal value was taken. This value, obviously, is wrong for the investigated samples and mainly causes the unlikely z-values calculated from less loaded crystals. At this point, one should keep in mind that the acoustic or elastic properties of a very thin film upon the crystal have vanishing influence upon the resonance frequency of the sample, vice versa, it must not be expected that the z-values can be calculated from the resonance frequencies of such samples with reasonable accuracies. In other words, the calculation of the z-value from less loaded samples is extremely sensitive to the model used and to the accuracies of the input parameters. Nevertheless, these "effective" z-values obtained by use of an actually uncontrolled nominal electrode mass load value have been taken to demonstrate the fascinating final thickness display accuracy of the Auto-Z-Match technique under practical conditions (Fig. 11 to 14).

From the low scatter of the AAS measurement points in Fig. 11 to 13 can be learned that the reproducibility of the AAS m-determination is obvi-

ously much better than the worst case conservative absolute error estimation. Only in Fig. 14, the scatter of the AAS-measurement points is higher, since the small film thicknesses of the additional Ag-layers cannot be determined with the same accuracy as the total load in Figs. 11 to 13.

Finally, it can be stated that, by use of the presented composite resonator treatment, the Auto-Z-Match technique offers, within a maximum deviation of less than 1 %, the same accuracy as the Z-Match technique. Naturally, in all cases where the effective z-value of the deposited film is not (e.g. for alloys) or not exactly known, the Auto-Z-Match offers an inherently higher accuracy. Within the measurement accuracy, the areal mass densities and the film thicknesses obtained by use of the crystal (Auto-)Z-Match technique always coincide with the perfectly independent AAS-values. In contrast to that, by use of deposition materials with rather high acoustic impedance values, the error of the frequency- or period- measurement technique increases with increasing preload up to 50 % or 20 %, respectively. The improvement in terms of thickness display accuracy yielded by use of the introduced advanced composite resonator formula instead of the equation after Miller and Bolef could be shown to be up to 4 % for the Auto-Z-Match procedure, while it is only 0.3 % for the Z-Match technique.

Acknowledgements

Some of the measurement equipments in this work were financially supported by the Austrian Science Research Funds (Fonds zur Förderung der wissenschaftlichen Forschung) and by the Hochschuljubiläumsfonds der Gemeinde Wien. The authors thank Dr. J. Wernisch, Dr. W. Hofer, and M. Gröschl from the University of Technology in Vienna, Austria, for the AAS mass determinations, for valuable discussions regarding the proper equivalent circuit representation of piezoelectric composite resonators, and for the computations, respectively. Last but not least, the authors acknowledge the contribution of H. J. Siegmund and H. Maidhof from the Leybold-Heraeus company, BRD, which provided the loaded crystal probes. Leybold-Heraeus is also acknowledged for the financial support of the final steps of this work.

Appendix: Resonance frequency spectrum of a plano-convex crystal

The comprehensive thickness shear resonance frequency spectrum, that are the harmonic as well as the anharmonic resonance frequencies, of a plano-convex rotated Y-cut quartz crystal was derived by Wilson[25] and an enhanced version was presented by Tiersten and Smythe.[26] Following the resulting equations of the latter authors and using our nomenclature and a binominal expansion, we can write

$$f_{QE}^{s(npq)} = \frac{n}{2l_Q}\sqrt{\frac{\bar{c}_{66}^*}{\rho_Q}}\left\{1 + \frac{1}{2n\pi}\sqrt{\frac{l_Q}{R}}\left[\sqrt{\frac{M_n}{\bar{c}_{66}^*}}(2p-1) + \sqrt{\frac{c_{55}}{\bar{c}_{66}^*}}(2q-1)\right]\right\} \quad (A1)$$

$$\bar{c}_{66}^* = \bar{c}_{66}\left(1 - \frac{8k^2}{n^2\pi^2} - \frac{4m_E}{m_Q}\right) \quad (A2)$$

$$k^2 = k_{y6}^2 = \frac{e_{y6}^2}{\bar{c}_{66}\varepsilon_{yy}} \quad (A3)$$

$$m_E = \rho_E l_E, \quad m_Q = \rho_Q l_Q \quad (A4)$$

$$M_n = c_{11} + (c_{12}+c_{66})r + 4\frac{(\bar{c}_{66}r - c_{66})(c_{22}r + c_{12})\cot(x n\pi/2)}{c_{22}x n\pi} \quad (A5)$$

$$r = (c_{12}+c_{66})/(\bar{c}_{66}-c_{22}) \quad (A6)$$

$$x = \sqrt{\bar{c}_{66}/c_{22}} \quad (A7)$$

where $f_{QE}^{s(npq)}$ is the series resonance frequency (with the harmonic mode number n and the anharmonic mode numbers p, q and with the mass load effect of the electrodes taken into account) of the rotated Y-cut crystal, l_Q is the maximum thickness of the crystal (see Fig. 8), l_E is the thickness of one electrode (equal thicknesses of back and front side electrode are assumed), c_{11}, c_{12}, c_{22}, c_{55}, c_{66}, are the elastic stiffness constants in the rotated coordinate system, $\bar{c}_{66} = c_Q$ is the piezoelectrically stiffened elastic shear constant effective in the case of open-circuited electrodes,[34] $\bar{c}_{66}^*$ is the piezoelectric shear stiffness constant effective in the case of short-circuited electrodes and taking into account the frequency lowering caused by the electrode mass load $2m_E/m_Q$; ρ_Q and ρ_E are the densities, m_Q and m_E the areal mass densities of the crystal (assumed here as infinite plate with thickness l_Q) and one electrode, respectively; l_Q is the maximum thickness of the plano-convex crystal, R is the radius of curvature of the convex sphere; k is the electromechanical coupling factor, e_{y6} the piezoelectric constant, and ε_{yy} the permittivity effective in the rotated system. The assumptions made are a small electromechanical coupling factor k and small electrode mass loads $2m_E/m_Q$. These are very well fulfilled for quartz crystals and the usual electrode thicknesses. Equation (A1) was excellently justified by experimental data of Tiersten and Smythe[26] as well as by unpublished measurements of the authors. Since we are only interested in the harmonic frequencies, we set p=q=1, introduce

$$f_Q = \frac{1}{2l_Q}\sqrt{\frac{c_Q}{\rho_Q}} \quad (A8)$$

and use the approximation

$$\sqrt{1+x} \approx 1 + x/2, \quad x<<1 \quad (A9)$$

to yield

$$f_{QE}^{s(n)} = nf_Q\left(1 - \frac{4k^2}{n^2\pi^2} - \frac{2m_E}{m_Q} + PC_n\right) \approx nf_Q\left(1 - \frac{4k^2}{n^2\pi^2}\right)\left(1 - \frac{2m_E}{m_Q}\right)(1 + PC_n) \quad (A10)$$

$$PC_n = \frac{1}{2n\pi}\sqrt{\frac{l_Q}{c_Q R}}\left(\sqrt{M_n} + \sqrt{c_{55}}\right) \quad (A11)$$

where the fact has been used that the right terms of the sums are small compared to unity. For vanishing k we yield the mechanical resonance frequencies of a plano-convex crystal

$$f_{QE}^{p(n)} = nf_Q\left(1 - \frac{2m_E}{m_Q} + PC_n\right) \approx nf_Q\left(1 - \frac{2m_E}{m_Q}\right)(1 + PC_n) \quad (A12)$$

For vanishing electrode mass load $2m_E/m_Q$ we yield the relation between the trivial parallel resonance spectrum cf the infinite plate nf_Q and the parallel reso-

nance spectrum of the plano-convex crystal

$$f_Q^{p(n)} = nf_Q(1 + PC_n) . \qquad (A13)$$

The numerical values of the parameters effective for the AT-cut quartz crystal are according to Mindlin:[35]

$c_{11} = 86.74\times10^9$ Pa $c_{55} = 68.81\times10^9$ Pa

$c_{12} = 8.26\times10^9$ Pa $c_{66} = 29.01\times10^9$ Pa

$c_{22} = 129.77\times10^9$ Pa

and Ballato:[36]

$c_Q = \bar{c}_{66} = 29.24\times10^9$ Pa

$\rho_Q = 2.649\times10^3$ kg/m^3

$k^2 = k_{y6}^2 = 7.744\times10^{-3}$.

References and footnotes

1 G. Z. Sauerbrey, "Wägung dünner Schichten mit Schwingquarzen", Phys. Verhandl. **8** (7), p 193 (1957).

2 G. Z. Sauerbrey, "Verwendung von Schwingquarzen zur Wägung dünner Schichten und zur Mikrowägung", Z. Phys. **155**, pp. 206-222 (1959).

3 P. Lostis, "Étude, réalisation et contrôle de lames minces introduisant une différence de marche déterminée entre deux vibrations rectangulaires", Revue d'Optique **38** (1), pp. 1-28 (1959).

4 H. K. Pulker, "Untersuchung der kontinuierlichen Dickenmessung dünner Aufdampfschichten mit einer Schwingquarzmeßeinrichtung", Z. Angew. Phys. **20** (6), pp. 537-540 (1966).

5 C. D. Stockbridge, Vacuum Microbalance Techniques, edited by K. Behrndt (Plenum, New York, 1966), Vol. 5, p. 193.

6 K. H. Behrndt, "Long-Term Operation of Crystal Oscillators in Thin-Film Deposition", J. Vac. Sci. Technol. **8** (5), pp. 622-626 (1971).

7 D. Hammer, E. Benes, and H. K. Pulker, "A digital quartz deposition monitor using a microprocessor", Thin Solid Films **32**, pp. 47-50 (1976).

8 J. G. Miller and D. I. Bolef, "Acoustic Wave Analysis of the Operation of Quartz-Crystal Film-Thickness Monitors", J. Appl. Phys. **39**, pp. 5815-5816 and pp. 4589-4539 (1968).

9 C. Lu and O. Lewis, "Investigation of film-thickness determination by oscillating quartz resonators with large mass load", J. Appl. Phys. **43** (11), pp. 4385-4390 (1972).

10 The principal value of the reverse tan-function is termed "arctan", thus, $\arctan\{\tan(\pi f/f_Q)\} = \pi f/f_Q - \pi$.

11 Z-Match is a registered trade mark of Inficon/Leybold-Heraeus Inc., East Syracuse, New York.

12 As compared to the definition given by Lu and Lewis (Ref. 9) in this paper the specific acoustic impedance ratio z is reciprocally defined. This is in consistency with the reduced parameters used here, each of which has the respective quartz parameter as denominator.

13 R. Krimholtz, D. A. Leedom, and G. L. Mathaei, "New equivalent circuits for elementary piezoelectric transducers", Electron. Lett. **6** (13), pp. 398-399 (1970).

14 Here and throughout this paper sinusoidal time dependence is assumed. Time-varying quantities are expressed as phasors with suppressed exp(jωt) dependence of time t with ω being the angular frequency and j the imaginary unit. For example, the complex phasor force F at the regarded acoustic port corresponds to an instantaneous force given by Re{Fexp(jωt)}, where Re means "the real part of". The absolute value of the amplitude is indicated by $\hat{F}$.

15 In the previous part of this paper, v was used for the (constant) phase velocity of the acoustic wave, while here v is used for the time dependent paricle velocity, which, in the regarded case of shear waves, is the time derivative of the mechanical particle displacement in a direction transversal to the sound propagation.

16 H. Schüssler, "Darstellung elektromechanischer keramischer Wandler als Dickenscherschwinger mit piezoelektrischer und piezomagnetischer Anregung", Archiv für Elektronik und Übertragungstechnik, Electronics and Communication **A.E.Ü.-22**, p. 399 (1968).

17 R. Holland, "Representation of Dielectric, Elastic, and Piezoelectric Losses by Complex Coefficients", IEEE Trans. Sonics Ultrason. **SU-14**, p. 18 (1967).

18 W. P. Mason, Electromechanical Transducers and Wave Filters, (D. van Nostrand Co., New York, 1948).

19 H. G. Unger, Theorie der Leitungen, (F. Vieweg & Sohn, Braunschweig, 1967).

20 E. Skudrzyk, The Foundations of Acoustics, (Springer, New York, 1971).

21 H. Marko, Theorie linearer Zweipole, Vierpole und Mehrtore, (Hirzel Verlag, Stuttgart, 1971).

22 K. K. Kanazawa, "A General Solution for the Change in Resonant Frequency of a Quartz Oscillator due to Viscoelastic Overlayers", J. Appl. Phys., to be published (1 July 1985).

23 For the determination of $z_Q=\sqrt{\rho_Q c_Q}$ and $f_Q=(1/2l_Q)\sqrt{c_Q/\rho_Q}$, nonetheless, the piezoelectrically "stiffened" elastic constant has to be used by definition. The short-circuiting of the electrodes causes a frequency lowering and, thus, a decrease of the "stiffening" of the effective elastic constant (compare Eq. (8) and Eq. (A2) in the appendix).

24 M. Onoe, H. F. Tiersten, and A. H. Meitzler, J. Acoust. Soc. Am. **35**, p 36 (1963).

25 C. J. Wilson, "Vibration modes of AT-cut convex quartz resonators", J. Phys. D **7**, pp. 2449-2454 (1974).

26 H. F. Tiersten and R. C. Smythe, "An Analysis of Overtone Modes in Contoured Crystal Resonators", Proc. of the 31st Annual Symposium on

Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey, USA, pp. 44-47 (1977).

27 W. G. Cady, Piezoelectricity, (McGraw-Hill, New York, 1946; Dover, New York, 1964), p. 308.

28 A. Ballato, T. J. Lukaszek, and G. J. Iafrate, Ferroelectrics, **43**, pp. 25-41 (1982).

29 E. Benes, "Improved quartz crystal microbalance technique", J. Appl. Phys. **56** (3), pp. 608-626 (1984).

30 F. Boersma and E. C. van Ballegooijen, J. Acoust. Soc. Am., **62** (2), pp. 335-340 (1977).

31 C. van der Steen, F. Boersma, and E. C. van Ballegooijen, J. Appl. Phys. **48** (8), pp. 3201-3205 (1977).

32 E. C. van Ballegooijen, F. Boersma, and C. van der Steen, J. Acoust. Soc. Am. **62** (5), pp. 1189-1195 (1977).

33 A. V. Apostolov and S. H. Slavov, Appl. Phys. A, **29**, p. 33 (1982).

34 Numerical subscripts are abbreviated subscripts according to Voigt's notation.

35 R. D. Mindlin, Quart. Appl. Math., **19**, p. 51 (1961).

36 A. Ballato, "Frequency-Temperature-Load Capacitance Behaviour of Resonators for TCXO-Application", IEEE Trans. Sonics Ultrason. **SU-25** (4), pp. 185-191 (1978).

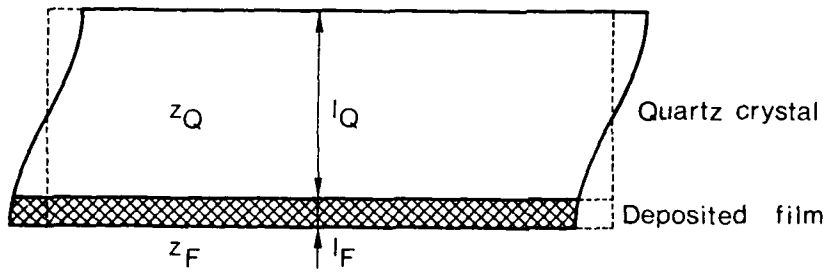


FIG. 1. Schematic representation of the one-dimensional composite thickness shear resonator model.
l_Q, l_F...thicknesses of the crystal and the film, respectively
z_Q, z_F...specific acoustic impedances of the crystal and the film, respectively

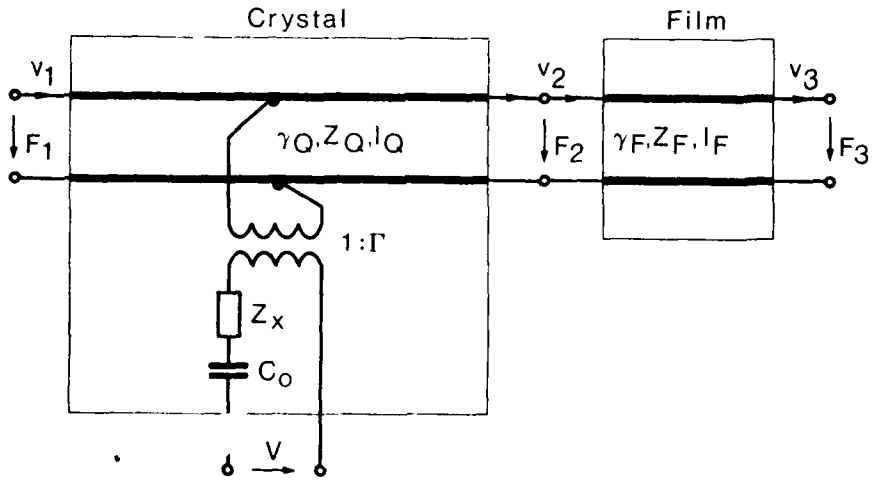


FIG. 2. Equivalent electrical circuit of a one-dimensional piezoelectric composite resonator (electrode mass and thickness neglected) following the transmission line analogy after Krimholtz et al.. Applied to the case of a thickness shear vibrator, the element values are given by Eqs. (9) to (13).

FIG. 3. (On next page). Plot of the reduced areal mass density $m=m_F/m_Q$ versus the reduced composite resonance frequencies (Eq. (45)), where n is the mode number of the regarded quasiharmonic vibration and z is the reduced specific acoustic impedance. The z-parameter values are chosen according to an assumed fictive epitaxial quartz crystal film (z=1, dash-dotted curve), as well as to the deposition materials Mg (z=0,63) and W (z=6,27), representing materials with extremely low and high specific acoustic impedance, respectively. For a given mass load m, the derivation of the reduced frequency spectrum f_n/f_Q of a quartz/tungsten composite resonator is indicated by arrows. The hatched area indicates the region which cannot be excited piezoelectrically by use of a typical oscillator driving circuit.

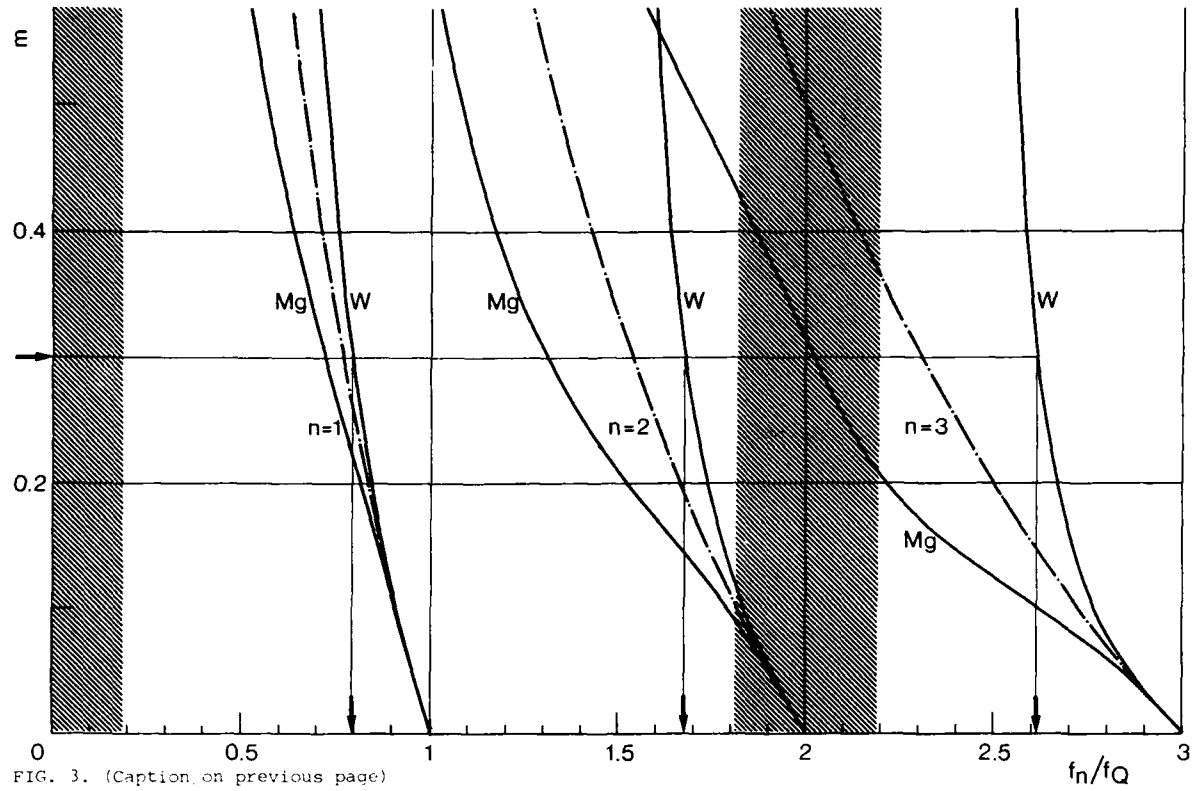


FIG. 3. (Caption on previous page)

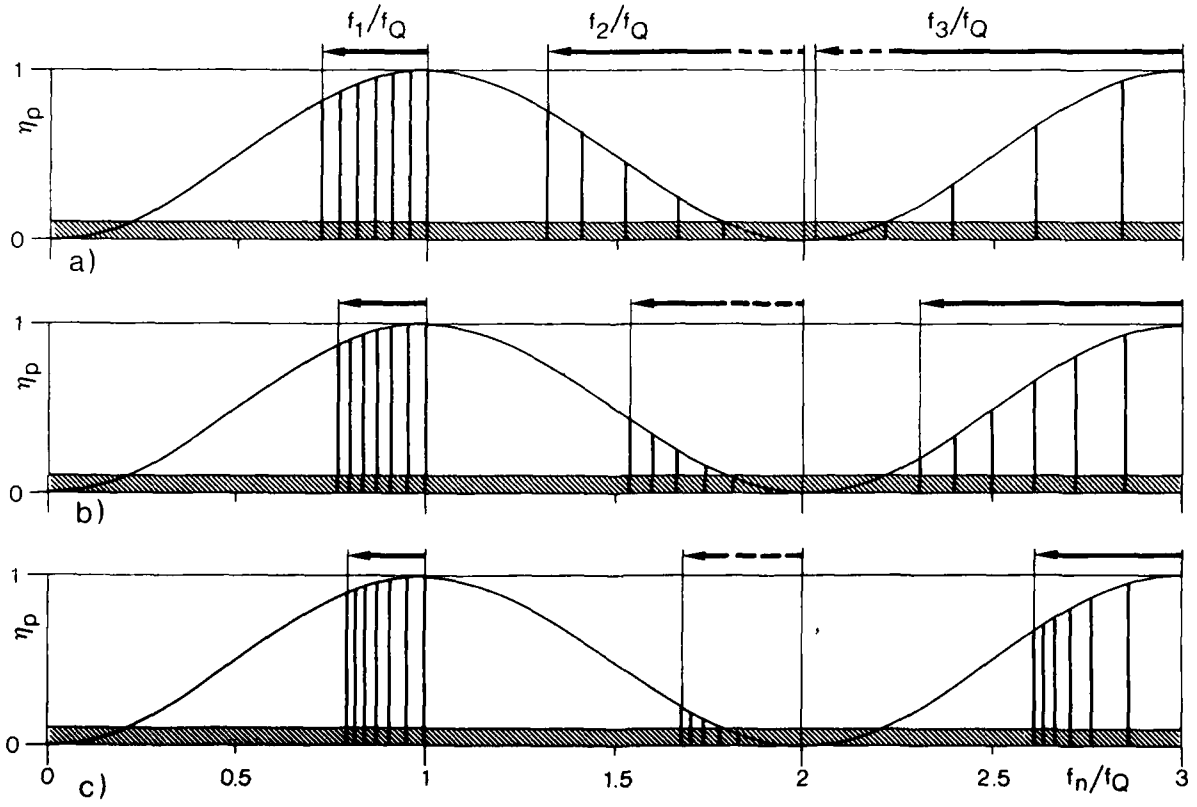


FIG. 4. Calculated electromechanical oscillator efficiency η_p versus reduced frequency f_n/f_Q of a composite resonator loaded with **a)** Magnesium, z=0.63 **b)** AT-cut crystal film, z=1 **c)** Tungsten, z=6.27 . Vertical lines represent composite resonance frequencies for equidistant mass load steps of Δm=5%. The arrow-lines show the material and mode number dependent ranges of the reduced frequencies for an assumed mass load range of 30%. The hatched area indicates the region of insufficient electromechanical resonator efficiency if an oscillator driving circuit with an feedback gain control range of 1:10 is assumed.

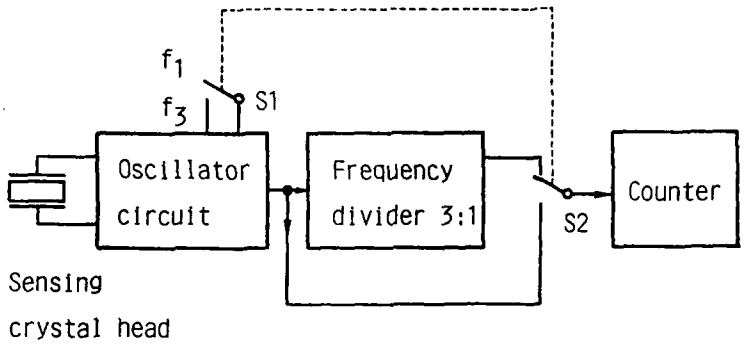


FIG. 5. Principal arrangement of the necessary hardware modification for the two-frequency measurement method. The switching of the oscillator feedback circuit bandwidth from the range of the fundamental frequency f_1 to the range of the third quasi-harmonic f_3 by S1 is actuated simultaneously with the switching of S2 via an output port of the controlling microprocessor.

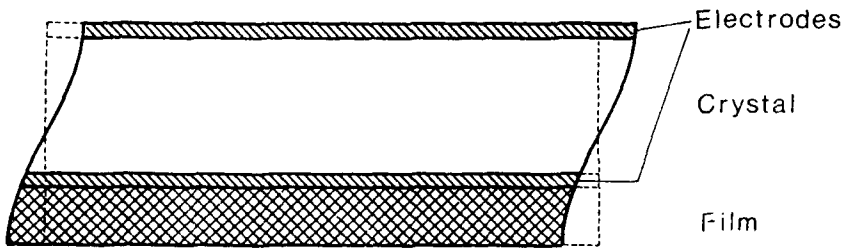


FIG. 6. One-dimensional model of a crystal loaded by a foreign film with included consideration of electrodes. It can be seen that the backside (upper) electrode remains within the displacement antinode (the shear stress node), while the frontside (lower) electrode moves with increasing foreign film thickness toward shear stressed zones.

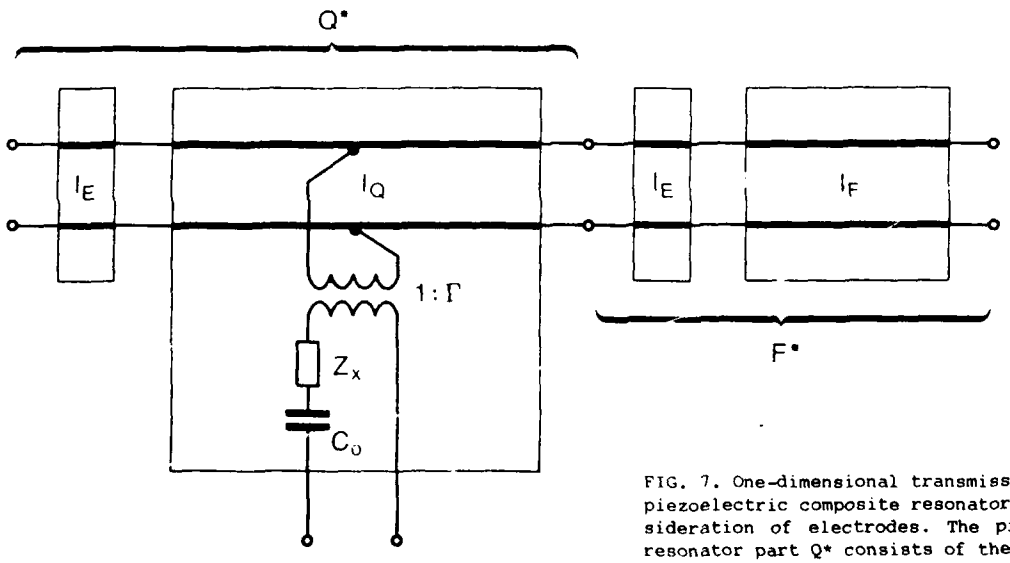


FIG. 7. One-dimensional transmission line model of a piezoelectric composite resonator with included consideration of electrodes. The piezoelectric active resonator part Q* consists of the naked quartz crystal and the backside (left) electrode, the passive part F* is composed of the frontside electrode and the deposited film. l_Q, l_E and l_F are the thicknesses of the crystal, of one electrode (both electrodes are assumed to have equal thicknesses), and of the deposited film, respectively.

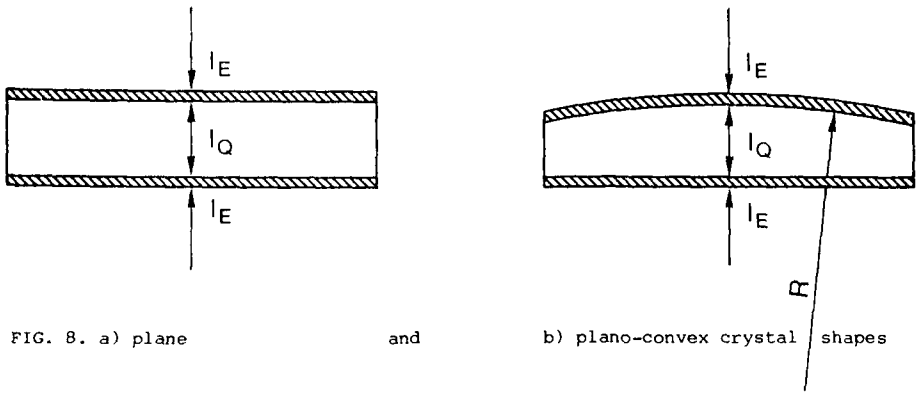
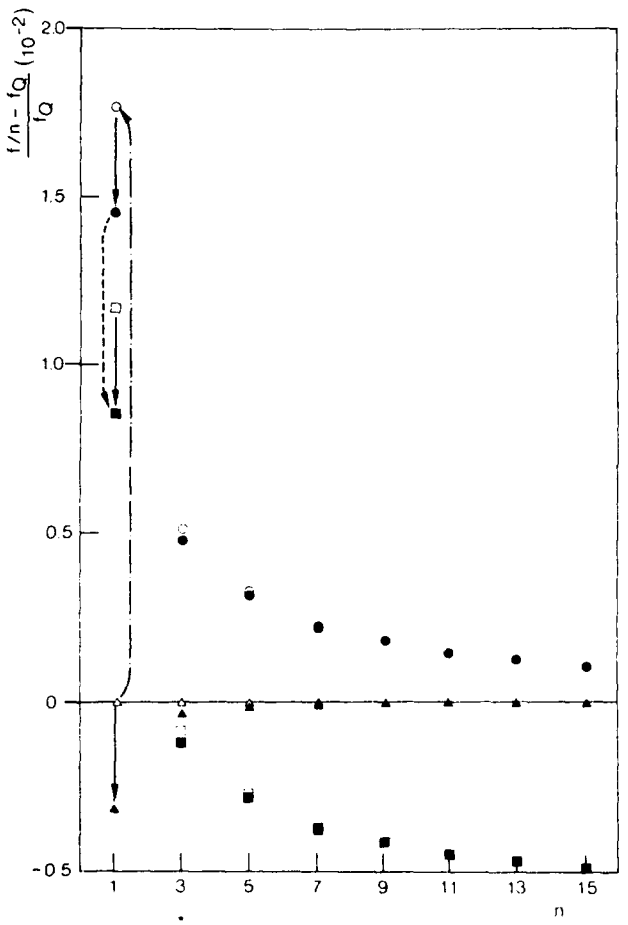


FIG. 8. a) plane and b) plano-convex crystal shapes



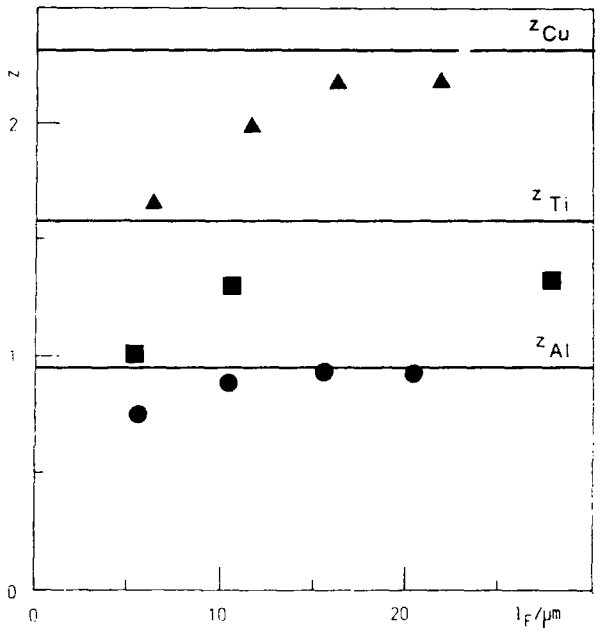


FIG. 10. Thin Film specific acoustic impedance values for the the deposited materials

▲ Cu, ■ Ti, ● Al versus the film thickness.

The horizontal lines indicate the corresponding bulk z-values. The effective thin film z-values were calculated by use of Eq. (60).

FIG. 9. Calculated reduced frequency deviations $(f/n-f_Q)/f_Q$ from the trivial mechanical resonance spectrum of an infinite crystal plate.

△ $f = f_Q^{p(n)}$, ▲ $f = f_Q^{s(n)}$....Trivial mechanical (parallel) and series resonance spectrum (short-circuited electrodes), respectively, of an infinite plano-parallel plate with massless electrodes.

○ $f = f_Q^{p(n)}$, ● $f = f_Q^{s(n)}$....Parallel and series resonance spectrum, respectively, of a plano-convex crystal with massless electrodes.

□ $f = f_{Q^*}^{p(n)}$, ■ $f = f_{Q^*}^{s(n)}$....Parallel and series resonance spectrum, respectively, of a plano-convex crystal with the influence of an assumed electrode mass loading of $2m_E/m_Q=6\times10^{-3}$ taken into account (compare Fig. 8). The used analytical relations between the various spectra are Eq. (58) (short arrows), Eq. (A13) (dash-dotted arrow), and Eq. (59) (dashed arrow).

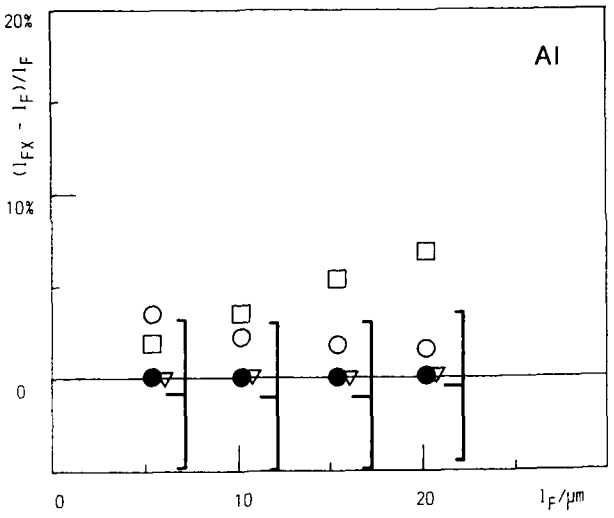


FIG. 11.

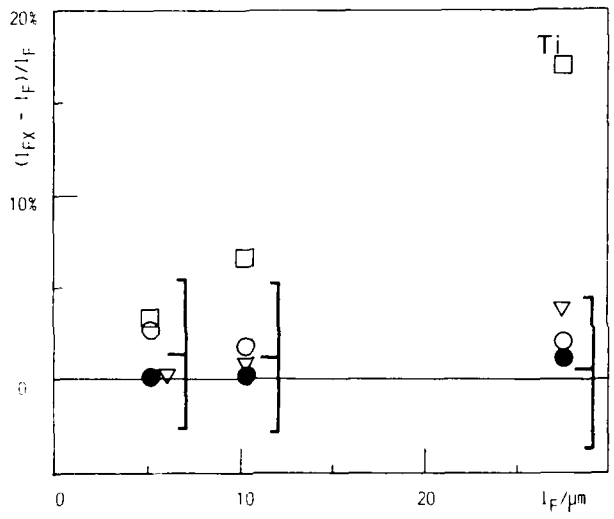


FIG. 12.

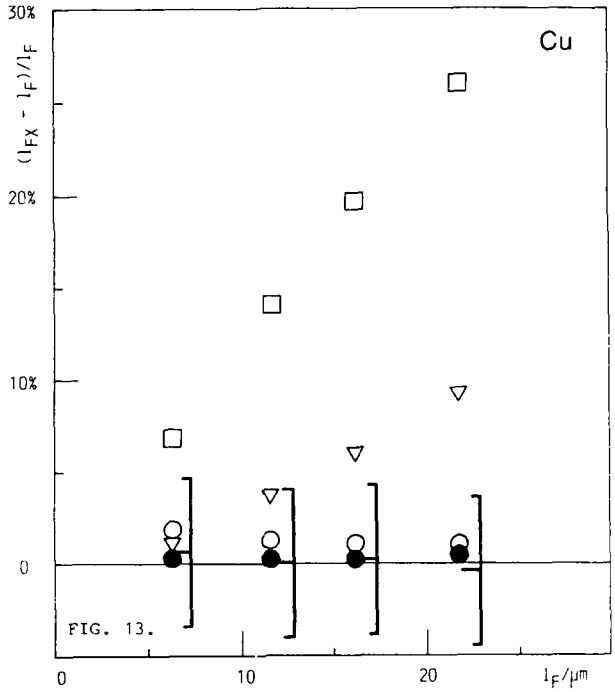


FIG. 13.

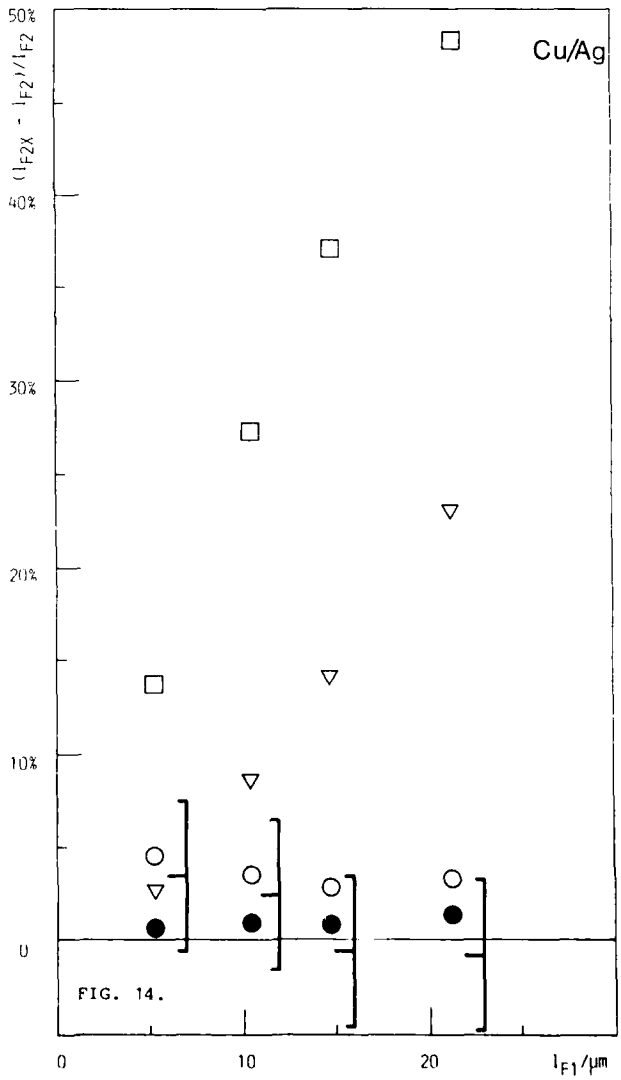


FIG. 14.

FIG. 11. to 14. Comparison of the thickness indications of the different quartz crystal evaluation methods

- □ frequency measurement technique,
- ▽ period measurement technique,
- ○ Auto-Z-Match by use of the Miller and Bolef composite resonator treatment,
- ● Auto-Z-Match by use of the composite resonator treatment presented in this paper,
- ⊣ and of the independent AAS reference method

for the deposition materials Al (FIG. 11), Ti (FIG. 12), Cu (FIG. 13), and for an additional Ag-layer upon a Cu-preload (FIG. 14).
l_{FX} and l_F are the thickness indications yielded with the regarded evaluation method and with the Z-Match technique, respectively. l_{F2X} and l_{F2} are the corresponding thickness indications for an additional thin layer (e.g. Ag) and l_{F1} is the thickness of the preload (e.g. Cu).

TRANSIENT ANALYSIS OF PIEZOELECTRIC TRANSDUCER RESPONSE

ABDULFATTAH H. BANAH

College of Engineering
King Abdulaziz University
P.O. Box-9027, Jeddah-21413
Saudi Arabia

Abstract

This paper introduces a simple, dynamic method for analysing transducer response. The approach uses the Feynman diagrams method and a novel model of the transducer equivalent circuit, which is a reentrant transmission line configuration. The system is initially step excited. This excitation is traced through all possible pathways to the place and time of observations. The analysis are carried out in detail for two specific cases;

1- The ideally bonded transducer in a line.

2- The transducer which is mechanically free on one side.

The results are in good agreement with the theoretical finding of previous workers.

INTRODUCTION

Piezoelectric transducers are usually used as detectors and radiators of acoustic power from very low frequencies to above 1 GHz. The characteristics of such a device depends on the type of mechanical vibration in which it operates.

The equivalent circuits show slight differences according to whether the vibration is mainly thickness dilatation, shear, flexural, or torsional.

In this paper the transducer may be assumed as to vibrate in pure thickness dilatation mode, although other types of transducers may also be analyzed by the same theory without any further justifications.

The commonly used equivalent circuits of a transducer [1 - 3] are rather complicated and difficult to analyze. The tracing of the exciting impulse through successive reflections from both faces of the transducer is the origin of such complications. A novel equivalent circuit has been developed by the author [4]which overcome such difficulties. This new developed technique will be adopted to our problem.

Fig.1 shows an early equivalent circuit of a piezoelectric transducer polarized along the thickness direction. In Fig.2 the lumped parameters on the mechanical side are replaced by a reentrant transmission line which is necessary for the acoustic signals to travel from one face of the transducer to the other face. For thickness mode operation it is required that the lateral dimensions of the transducer must be much greater than the thickness dimension [5,6].

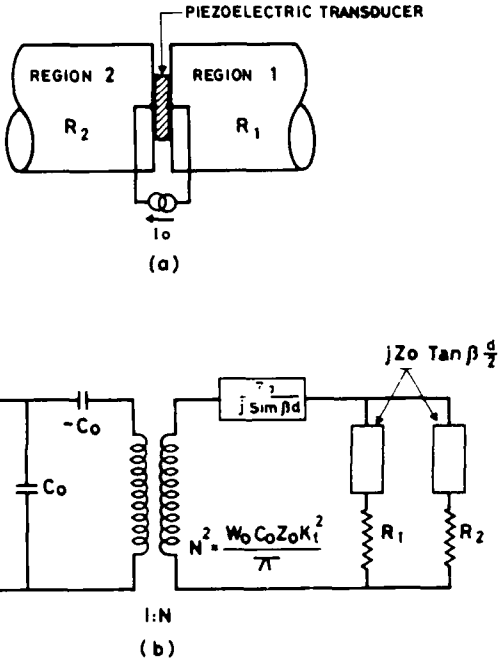


Fig.1 Electromechanical arrangement for a transducer bonded on both faces (a) Physical model (b) Equivalent circuit of a piezoelectric plate for thickness mode operations [1].

In addition the transducer may be operated in fundamental and overtone resonance mode. However, it cannot be excited at frequencies for which its thickness equals an even multiple of mechanical half wavelengths due to phase cancellation of the piezoelectrically induces stress in order to simplify the analysis to avoid serious problems in transducer responses and unpredictable results [1], we may assume that the transducer is ideally bonded to a substrate.

MATHEMATICAL MODEL

The system response may now be derived with the new model of Fig.2. The circuit is initially step excited at C, coupled directly to terminals 11' and 22'. The signal across 11' propagates through the delay line and appears across 22'. The received signal is partially coupled to b and partially reflected back toward 11' either through the delay line (mechanical coupling) or through direct connection (electrical coupling). The

CH2186-0/85/0000-0571$1.00©1985IEEE

Process is repeated and the signal keeps oscillating back and forth between terminals 11' and 22'. A similar sequence of events will follow for the signal that first "appears" across 22'. Hence the total contribution arriving at b will be [7]

$$V_b(t) = \frac{I_o}{N} w_c Z_o u(t)(1 - w_d t)$$

$$+ \sum_{n=1}^{\infty} \frac{I_o}{N} \frac{2w_c^2 Z_o^3}{(Z_o+R_1)(Z_o+R_2)} \left\{ 2(\rho_1 \rho_2)^{\frac{n-1}{2}} \delta_{n,n_o} \right.$$

$$\left. + \frac{R_1^2 + R_2^2 - 2Z_o^2}{(Z_o+R_1)(Z_o+R_2)} (\rho_1 \rho_2)^{\frac{n-2}{2}} \delta_{n,n_e} \right\}$$

$$\times u(t - n\tau)(t - n\tau) \quad .. (1)$$

where

ρ_1 = reflection coefficient at 11'

$$= \frac{R_1 - Z_o}{R_1 + Z_o}$$

ρ_2 = reflection coefficient at 22'

$$= \frac{R_2 - Z_o}{R_2 + Z_o}$$

Z_o = characteristic impedance of the material

$$w_c = \frac{w_o}{\pi} K_t^2 = \frac{K_t^2}{\tau}$$

$$N^2 = \frac{w_o C_o Z_o K_t^2}{\pi} = w_c C_o Z_o$$

K_t = electromechanical coupling coefficient

τ = transit time through the transducer

$$w_d = \left[\frac{R_1 + R_2 + 2Z_o}{(Z_o+R_1)(Z_o+R_2)}\right] w_c Z_o$$

$$\delta_{n,n_i} = \begin{cases} 1 & \text{if } n = n_i \\ 0 & \text{otherwise} \end{cases}$$

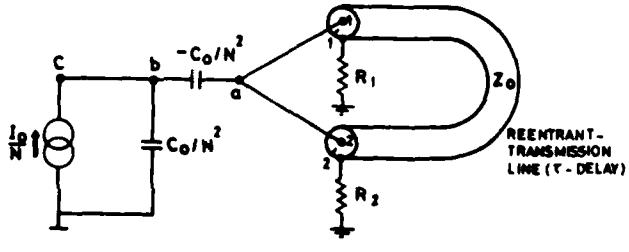


Fig.2 Piezoelectric plate with reentrant transmission line model [7]

Equation (1) describes the transient response referred to the mechanical side of the ideal transformer.

The first term signifies a jump of the voltage from zero to $\left(\frac{I_o}{N} W_c Z_o u(t)\right)$ at time t=0 as expected. Immediately afterwards, the transient stress waves and their reflections modify the voltage response.

The second term is mainly caused by the presence of the negative capacitor in Fig.2 [7] subsequent effect of this capacitor is also included in the odd and even terms of the series. Each term of the series represents a linear sawtooth pulse delayed until time τ. The amplitude of the delayed sawtooth pulses are given by the terms containing the reflection coefficients, and as these are generally less than unity, the amplitude of successive reflections diminish.

RESULTS

Equation (1) will now be evaluated for two simple cases of practical interest. These are:

1) The symmetric element for which the lines on both sides have the same acoustic impedance ($R_1 = R_2$), and
2) The plate which is symmetrically free on one side ($R_2 = 0$).

1) For the symmetric transducer ($R_1 = R_2$) Eq. (1) becomes:

$$V_b(t) = \frac{I_o}{N} w_c Z_o u(t)(1 - w_1 t)$$

$$+ \sum_{n=1}^{\infty} \frac{I_o}{N} \frac{2w_c^2 Z_o^3}{(Z_o+R)^2} 2(\rho)^{n-1} \delta_{n,n_o}$$

$$+ \frac{2(R^2 - Z_o^2)}{(Z_o + R)^2} (\rho)^{n-2} \delta_{n,n_e}$$

$$\times u(t-n\tau)(t-n\tau) \quad ... (2)$$

where $R = R_1 = R_2$,

$$\rho = \rho_1 = \rho_2 = \frac{R - Z_o}{R + Z_o}$$

and $w_1 = w_d \Big|_{R_1=R_2=R} = \frac{2w_c Z_o}{(Z_o+R)}$

Fig.3 shows a plot of Eq. (2) for the following cases:

i. $R = 0$, $\rho = -1$, and $w_1 = 2w_c = 2\frac{K_t^2}{\tau}$

ii. $R = Z_o$, $\rho = 0$, and $w_1 = w_c = \frac{K_t^2}{\tau}$

iii. $R \to \infty$, $\rho = +1$, and $w_1 \to 0$

iv. $R = \frac{1}{3} Z_o$, $\rho = -\frac{1}{2}$, and $w_1 = \frac{3}{2} w_c = \frac{3}{2}\frac{K_t^2}{\tau}$

v. $R = 3\,Z_o$, $\rho = +\frac{1}{2}$, and $w_1 = \frac{1}{2}\,w_c = \frac{1}{2}\,\frac{K_t^2}{\tau}$

2) For a plate which is mechanically free on one side (R_2=0) Eq. (1) reduces to the form:

$$V_b(t) = \frac{I_o}{N}\,w_c\,Z_c\,u(t)\,(1 - w_2\,t) + \sum_{n=1}^{\infty} \frac{I_o}{N}\,\frac{2w_c^2\,Z_o^2}{(Z_o+R_1)}\,2(-\rho_1)^{\frac{n-1}{2}}\,\delta_{n,n_o} + \frac{R_1^2 - 2\,Z_o^2}{Z_o(Z_o+R_1)}\,(-\rho_1)^{\frac{n-2}{2}}\,\delta_{n,n_e}\,u(t-n\tau)\,(t-n\tau) \quad \ldots (3)$$

where $w_2 = \frac{R_1 + 2\,Z_o}{Z_o + R_1}\,w_c$

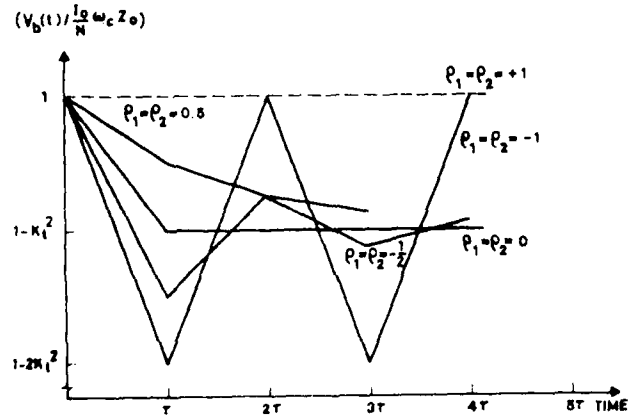


Fig. 3 Transient voltage response of symmetrical loading ($R_1=R_2$) versus time, with reflection coefficient as parameter

The odd and even series in Eq. (3) decrease geometrically as determined by $(-\rho_1)$. The amplitude of the two series are different: in fact the even series will completely disappear when $R_1^2 = 2\,Z_o^2$.

The pertinent features of linear sawtooth pulses are shown in Fig.4 for the following cases:

i. $R_1 = 0$, $\rho_1 = -1$, and $w_2 = 2\,w_c = 2\,\frac{K_t^2}{\tau}$

ii. $R_1 = Z_o$, $\rho_1 = 0$, and $w_2 = \frac{3}{2}\,w_c = \frac{3}{2}\,\frac{K_t^2}{\tau}$

iii. $R_1 \to \infty$ $\rho_1 = +1$, and $w_2 = w_c = \frac{K_t^2}{\tau}$

v. $R_1 = 3Z_o$ $\rho_1 = +\frac{1}{2}$, and $w_2 = \frac{5}{4}\,w_c = \frac{5}{4}\,\frac{K_t^2}{\tau}$

vi. $R_1 = \sqrt{2}Z_o\,\rho_1 = \frac{\sqrt{2}-1}{\sqrt{2}+1} = 0.1716$, and $w_2 = \sqrt{2}\,w_c$

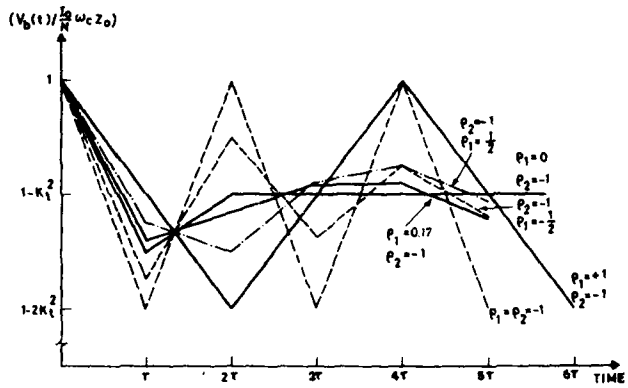


Fig. 4 Voltage response of transducer bonded to one side (R_2=0) versus time, with reflection coefficient as parameter

DISCUSSION AND CONCLUSION

The transient responses of a piezoelectric transducer under various conditions of acoustic loading are shown in Figs.3 and 4 Each response represents the sum of sequence of interactions of the excitation as it propagates between the end faces of the transducer. Such interactions are readily predicted by Feynman diagram method {6}. Its main advantage is that it breaks down the analysis into a series of steps (Paths) that have a ready physical interpretation.

Fig.5 shows all possibl sequences of Feynman diagrams applicable to the circuit in Fig.2. The semi circles in the figure refer to acoustic propagation through the transmission line. The coupling parameters are defined in Table 1. Noting that γ_{c1} and γ_{c2} have the dimension of an impedance; the other coefficients are dimensionsless. Thus in Fig. 5b the excitation enters the line at 1(2), propagates to 2(1), is reflected back to 1(2) and couples to b after a time 2τ. The total response at b contains all possible path propagations that start at 1 and 2

$$V_b^{(2)} = (I_o/N)\,\{\gamma_{c1}\,\rho_2(1+\rho_1)\gamma_{1b} + \gamma_{c2}\rho_1(1+\rho_2)\gamma_{2b}\} \quad (4)$$

similarly,

$$V_b^{(3)} = (I_o/N)\{\gamma_{c1}\rho_1\rho_2(1+\rho_2)\gamma_{2b} + \gamma_{c2}\rho_1\rho_2(1+\rho_1)\gamma_{1b} \quad (5)$$

It is interesting to observe that the odd and even sequences of the series may be grouped separately, identifying the results as geometrical series, and a closed form expression (see equation (1)) may be obtained.

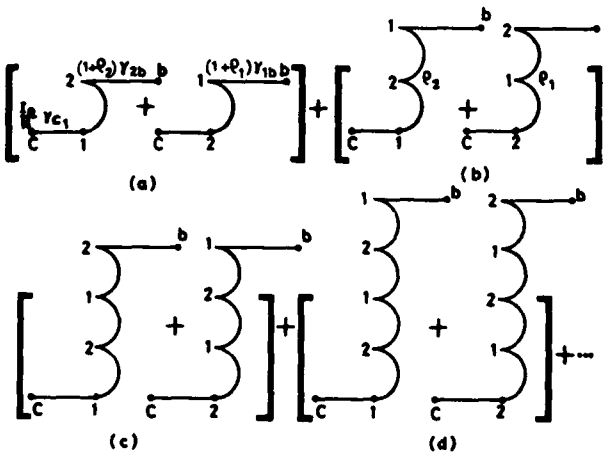


Fig 5 All possible paths of Feynman diagrams that lead to response observed at b, at time t = (n+1)τ .

Table-1: Coupling coefficients and the parameters relevant to circuit in Fig.2.

γ_{c1}	$(N^2/PC_o)\ [Z_o/(R_1 + Z_o)]$
γ_{c2}	$(N^2/PC_o)\ [Z_o/(R_2 + Z_o)]$
γ_{1b}	$(N^2/PC_o\ R_1)$
γ_{2b}	$(N^2/PC_2\ R_2)$
ρ_1	$(R_1-Z_o)/(R_1 + Z_o)$
ρ_2	$(R_2-Z_o)/(R_2 + Z_o)$
P	Laplace's operator

REFERENCES

1 W.P. MASON, Electromechanical Transducers and Wave Filters, Van Nostrand Co., New York, 399 (2948).

2 R. KRIMHOLTZ, D.A. LEDOM, and G.L. MATHAEI, Elect. Lett.6, 398 (1970).

3 M.G. SILK, Ultrasonis (GB) 21, 27 (1983).

4 A.H. BANAH, A. KORPEL, and R.F. VOGEL, IEEE Ultrasonic Symposium, 444 (1981).

5 N.F. FOSTER, G.A. COQUIN, G.A. ROZGONYI and F.A. VANNATTA, IEEE Trans. Sonics and Ultrasonic, 15, 28 (1968).

6 J.H.GOLL, IEEE Trans., Sonics and Ultrasonic, 26, 285 (1979).

7 A.H. BANAH, A. KORPEL, and R.F. VOGEL, J. Acoust. Soc. Am. 73, 677 (1983).

A FILLED THERMAL SYSTEM UTILIZING A GAS DENSITY SENSING QUARTZ CRYSTAL TUNING FORK

Roger W. Ward
Errol P. EerNisse

QUARTZTRONICS, INC.
1020 ATHERTON DRIVE, BLDG C
SALT LAKE CITY, UTAH 84123
801-266-6994

SUMMARY

The resonant frequency of a quartz crystal tuning fork decreases as the density of the (inert) gas surrounding it increases. This gas density effect is shown to be equivalent to adding mass per unit length to the tines of the tuning fork. A simple theory is derived to illustrate the gas density effect on a tuning fork. Deviations from the perfect gas laws (Van der Waal's effects) are accounted for.

The gas density effect on common quartz crystal tuning forks, such as those used in wristwatches, has been applied to sensing the temperature of the gas filled bulb in a classic filled thermal system. Using this technique, a thermometer operating from -50°C to 500°C with a frequency output format, and having a resolution better than 1°C, has been constructed.

The theory of the gas density/filled thermal system is derived and compared to actual experimental results.

BACKGROUND

It has been known for some time that a vibratory element, such as a quartz crystal, when exposed to a gas changes its frequency as the gas pressure changes [1-5]. Crandell [6] disclosed that the frequency of a vibrating string or rod changes with a change in density of the medium in which the string is placed, and stated that the change in frequency resulting from a change in density is not great unless the density of the medium is comparable to that of the rod; ie, a high density liquid. Mechanically driven metal tuning forks have been used to measure the density of various liquids (ie, see Stemme [7]) whose densities range ±25% from the density of water.

The principal mechanism by which a flexural or torsional mode vibratory element exposed to a fluid (liquid or gas) is caused to change its frequency is not an intrinsic change in pressure but rather is a change in the density of the fluid. The effect is equivalent to that of adding mass to the vibratory element (increase in density) or to taking mass from the vibratory element (decrease in density) to respectively reduce or increase the frequency of vibration. When the surface area of the object normal to the direction of movement or vibration of the object is increased, the "pushing" of the fluid by the object becomes more difficult, resulting in a slowing of the vibration, and vice versa. Recognizing this mechanism allows for optimizing the sensitivity of vibratory devices used for measuring density directly, or for indirectly measuring other parameters such as pressure, temperature, etc. In particular, the greater the surface area of the vibratory element perpendicular to the direction of motion, the greater the sensitivity of the frequency of vibration to fluid density changes. Also, by proper selection of the working fluid in which the vibratory element is placed, different measurement objectives are achieved. For example, in a gas density device, use of a more dense gas such as argon gives rise to greater sensitivity, whereas use of a less dense gas, such as helium, allows detection of density (and thus pressure or other parameter) changes over a wider pressure (density) range.

THEORY

The above becomes more apparent from a mathematical analysis of the effect fluid density has on a vibratory element, such as the one illustrated in Figure 1. The pressure due to the frontal area of a rectangular bar having a thickness e and a width w, which is vibrating in the w direction, moving ambient fluid is given by:

CH2186-0/85/0000-0575$1.00©1985IEEE

$$P = -2\pi f\ r\ U_o eC \sin(2\pi ft), \quad (1)$$

where f is the frequency of vibration, r is the density of the ambient fluid, U_o is the peak vibration velocity of the bar in the w direction, and C is a shape factor for edge effects and has a value near unity. The force, F, per unit length acting on the bar due to the pressure is 2Pe, since both sides of the bar are acting on the fluid. Therefore,

$$F = -4\pi f\ r\ U_o\ e^2\ C \sin(2\pi ft). \quad (2)$$

Since the velocity U of the vibrating bar is related to the displacement $Y = Y_o \sin(2\pi ft)$ by:

$$U = 2\pi f\ Y_o \cos(2\pi ft), \quad (3)$$

then the force F per unit length is given by:

$$F = -8\pi^2\ f^2\ r\ Y_o\ e^2\ C \sin(2\pi ft). \quad (4)$$

The conventional equation for equilibrium of a laterally vibrating beam with the force F, from equation 4, added to the inertial forces is:

$$EI\ \frac{d^4y}{dx^4} + (r_q ew + 2re^2C)\ 4\pi^2\ f^2\ Y = 0, \quad (5)$$

where E, I, and r_q are Young's Modulus, moment of inertia, and density of the vibrating beam respectively. The resonant frequency f_o, for r=0 (no ambient fluid), is thus perturbed by the added term $2re^2C$ in equation 5. The added term is of the form of an added mass per unit length, just as $r_q ew$ is the mass per unit length of the bar. The resulting resonant frequency f, from equation 5, is given by:

$$(f/f_o)^2 = \frac{1}{1 + \frac{2reC}{r_q w}}. \quad (6)$$

For small perturbations,

$$f/f_o = 1 - \frac{reC}{r_q w}, \quad (7)$$

or,

$$(f - f_o)/f_o = df/f = \frac{eCr}{wr_q}. \quad (8)$$

For a given vibrating beam, eC/wr_q is a constant, k; hence,

$$df/f = kr. \quad (9)$$

From (8) it is apparent that the frequency of a vibrating element is substantially linearly dependent upon the density of the working fluid and that increased sensitivity of a vibratory element can be achieved by both increasing the thickness-to-width ratio of the vibrating bar and selecting a more dense fluid as the working medium.

Figure 2 shows the experimental results of several commercially available tuning forks having different e/w ratios, frequencies, and overtones, pressurized with Ar and He, compared to the theory of (8), with C=1. The agreement is good; a smaller value of C would improve the agreement.

FILLED THERMAL SYSTEM

Conventional filled thermal systems consist of a gas filled bulb connected via a capillary tube to a pressure gauge, which is typically a Bourdon tube pressure gauge which is calibrated to display bulb temperature. The bulb is exposed to the process temperature, while the Bourdon tube is not. The principle of operation is based upon Charles' and Boyles' gas law:

$$\frac{P_o V_o}{T_o} = \frac{P_1 V_1}{T_1} \quad (10)$$

where P and V are the pressure and volume of the system at temperatures T_o and T_1.

Since $V_o = V_1$, and the volume of the bulb is much greater (typically 100X) than the capillary tube and Bourdon tube, then the system pressure is given by:

$$P_1 = \frac{P_o T_1}{T_o}. \quad (11)$$

Figure 3 shows a gas density/temperature transducer adapted to replace the Bourdon tube of a classic filled thermal system. This structure is comprised of a bulb having a typical volume of 1/2 in^3, and a hollow housing where the tuning fork is mounted. A capillary tube, which has an internal volume substantially less than the volume of the bulb, interconnects the bulb and crystal cavity. The tuning fork is coupled to an oscillator circuit. The system is made of a material possessing high strength at high temperatures, such as stainless steel or Inconel™. Noble gases are the preferred fill gas.

To measure temperature changes, the bulb is exposed to the medium whose temperature is to be measured so that

the gas contained in the housing will increase in temperature and expand, or decrease in temperature and contract, depending upon the temperature of the medium. With such expansion or contraction, the fluid molecules are either driven from the bulb toward the crystal cavity or vice versa. Thus, the gas density in the vicinity of the crystal changes with a change in temperature of the gas in the bulb. Because of the remoteness of the crystal from the bulb and the insulative nature and size of the capillary tube, the temperature of the gas in the crystal cavity does not appreciably change with a change in temperature of the gas in the bulb. With a change in gas density in the crystal cavity, the frequency of vibration of the tuning fork changes to provide a measure of the bulb's temperature change.

The selection of argon or other noble gases below argon on the periodic chart for use as the working gas is made because such gases have critical temperatures below -100°C and so will not condense at any pressure above that temperature. Thus, such gases behave as nearly ideal gases for all pressures and temperatures normally encountered where temperature measurements are desired.

THEORY OF THE FILLED THERMAL/GAS DENSITY EFFECT

Consider the idealized filled thermal system of Figure 3: assume $V_3=0$ (infinitely small capillary volume--this assumption is equivalent to assuming $T_3 = T_2$; hence, the capillary volume is incorporated into V_2. As long as $V_1 >> (V_2+V_3)$ this is a valid real world assumption).

The total number of molecules in the system is N, and the total volume is V. Hence,

$$N = N_1 + N_2 \qquad (12)$$
$$V = V_1 + V_2 .$$

Then,

$n_o = \frac{N}{V}$ = number of molecules per unit volume.

If $T_1 = T_2$, then

$$n_o = \frac{N_1}{V_1} = \frac{N_2}{V_2} = \frac{N}{V} . \qquad (13)$$

The pressure throughout the system must be uniform (at equilibrium). That is,

$$P = P_1 = P_2 \quad \text{for any } T_1, T_2 .$$

Now, assume T_1 increases and T_2 remains constant. Then ΔN_1 molecules leave V_1 and travel to V_2 such that the pressures at V_1 and V_2 remain equal as explained by the gas laws:

$$PV_1 = (n_o V_1 - \Delta N_1) RT_1$$
$$PV_2 = (n_o V_2 + \Delta N_1) RT_2, \qquad (14)$$

where R is the universal gas constant, T in °K.

Solving for ΔN_1:

$$\Delta N_1 = n_o V_2 \frac{T_1/T_2 - 1}{1 + V_2T_1/V_1T_2} . \qquad (15)$$

Since the density, r, of the gas is mN/V, where m is the molecular mass of the gas, and the number of molecules now at V_2 is $N_2 + \Delta N_1$, then the new density r_o' at V_2 is

$$\frac{mN_2 + m\Delta N_1}{V_2} = r_o' = r_o + \frac{r_o\, T_1/T_2 - 1}{1 + V_2T_1/V_1T_2} . \qquad (16)$$

From equation 8,

$$df = \frac{f_o Ce}{r_q w}\, dr. \qquad (17)$$

Hence,

$$\frac{df}{f_o} = \frac{r_o Ce}{r_q w} \frac{T_1/T_2 - 1}{1 + V_2T_1/V_1T_2} . \qquad (18)$$

From (18) it is seen that a filled thermal system using a quartz tuning fork gas density sensor has a maximum sensitivity to bulb temperature changes if:

- e/w (tine thickness to tine width) is large
- r_o (initial gas density when $T_1 = T_2$) is large (high initial pressure or dense gas)
- V_1 (bulb volume) is large
- V_2 (crystal cavity volume) is small.

Notice that if $V_2 = 0$ (or $V_2 << V_1$), there is no volume dependence in (18).

VAN DER WAAL

Van der Waal modified the "perfect gas laws" empirically to explain the

observed deviation of the behavior of "real gases" from the behavior of a perfect gas. He added terms to the gas law equations to account for the attractive forces between molecules, as well as accounting for their finite volume (ie, their general incompressibility).

The effect of Van der Waal's equation is to modify the density calculated for perfect gases to conform to the density observed for real gases. The ratio of "perfect" density to observed density is called the "compressibility", Z, of a gas. Tables of Z for various gases are found in handbooks [8].

Tables of Van der Waal's constants are found in various handbooks. It is difficult to derive a simple expression for df/f_o (equation 18) using a complete Van der Waal's formulation.

For real gases it is usually observed that for pressures from 0-100 atm, the real density change caused by a given pressure change is greater than that predicted for perfect gases, while at pressures greater than 100 atm the changes are less than predicted. To first order, a Van der Waal's correction may be obtained by dividing the right-hand side of equation 10 by Z.

EXPERIMENTAL FTS

A filled thermal system has been built with a commercially available 10kHz tuning fork with e/w = 0.48 and filled to 350 psi with argon ($r_o = 4.25 \times 10^{-2}$ g/cm^3) at 20°C. Figure 4 shows the experimental results. The theory curves in Figure 4 were calculated from (18) with $f_o e C r_o / r_q w = 65.1$ Hz, the experimentally measured slope, and with $V_2/V_1 = 0.1$. A correction for the compressibility factor for Ar has not been included in the theoretical curve; its inclusion would result in better agreement between theory and experiment. T_2 was held constant at room temperature.

The data in Figure 4 shows a thermometer operating over a span range of -50°C to 500°C with a resolution of better than 1°C. This is better than any filled thermal system now available.

Limitations on the span range include 1) gas condensation at the cold extreme, 2) deviation from the gas laws at the hot extreme (approaching plasma temperatures), and 3) limitations imposed by the sensor bulb material (temperature and pressure capabilities).

From (18) it is evident that the cold end temperature, T_2, must be known, or held constant, for accurate T_1 determination, since an increase in T_2 causes molecules to move toward V_1. In actual use, it is not practical to hold T_2 constant, so a temperature sensor must be provided near the gas density sensor. The resolution and accuracy required of T_2 is the same as that desired for T_1. This temperature sensor may be a thermistor, a platinum resistance device, a thermocouple, or a torsional tuning fork temperature sensor [9]. Then, by means of appropriate circuitry, a microprocessor, and proper calibration, both T_1 and T_2 are accurately determined.

With a starting pressure of 350 psi of Ar (a heavy molecule at high initial pressure causes r_o, and therefore the sensitivity, to be large) at room temperature, the pressure span for the system from -50 to + 500°C is 290 to 930 psi. The gas density changes -17% to 267%, causing the 10kHz crystal to change frequency by 100Hz=10,000ppm=1%!

It is important to realize that it is not the pressure change that the crystal's frequency change is due to, but the gas density change caused by the migration of molecules from the hot end of the system to the cold end. This has been experimentally verified by measuring f vs temperature for a quartz tuning fork mounted in a rigid, sealed container back-filled with 350 psi of Ar. Even though the pressure change during the f-T run was 20%, the density remained constant (constant N/V), and the frequency-temperature curve was identical to the same crystal run under atmospheric conditions.

CONCLUSION

The effects of gas density upon the resonant frequency of a quartz crystal tuning fork have been theoretically calculated and experimentally measured. A frequency format filled thermal system therometer, operating from -50 to 500°C, with a resolution better than 1°C, has been constructed. A theoretical model for this system agrees well with the experimental data.

REFERENCES

[1] M.A. Cotter, U.S. Patent No. 4,126,049.

[2] J. W. Stansfeld, U.S. Patent No. 4,232,544.

[3] P. N. Potter, U.S. Patent No. 4,178,804.

[4] Klaus H. Behrndt, Editor, Vacuum Microbalance Techniques, New York: Plenum Press, 1966.

[5] A. Genis and D. E. Newell, "Using The X-Y Flexure Watch Crystal as a Pressure-Force Transducer", Proceedings, Thirty First Annual Frequency Control Symposium, U.S. Army Electronics Command, Ft. Mammouth NJ, pp. 71-77, (1977). National Technical Information Service, Sills Bldg., 5285 Port Royal Road, Springfield, VA 22161, Accession Nr. ADA 088221.

[6] Irving B. Crandell, Theory of Vibrating Systems and Sound, Van Nostrand, 1926, pp. 124-133.

[7] E. Stemme, J. Ekelöf, and L. Nordin, "Measuring Liquid Density with a Tuning Fork Transducer", IEEE Trans. on Instruments and Measurement, Vol. IM-32, No. 3, pp. 434-437, Sept. 1983.

[8] Handbook of Tables for Applied Engineering Science, 2nd Ed., Cleveland: CRC Press, 1976, pp. 57-67.

[9] R.J. Dinger, "The Torsional Tuning Fork as a Temperature Sensor", Proceedings, 36th Annual Frequency Control Symposium, U.S. Army Electronics Command, Ft. Mommouth N.J., pp. 265-269, (1982). EIA, 2001 Eye St., Washington DC 20006.

$$df/f = \frac{eCr}{wr_q}$$

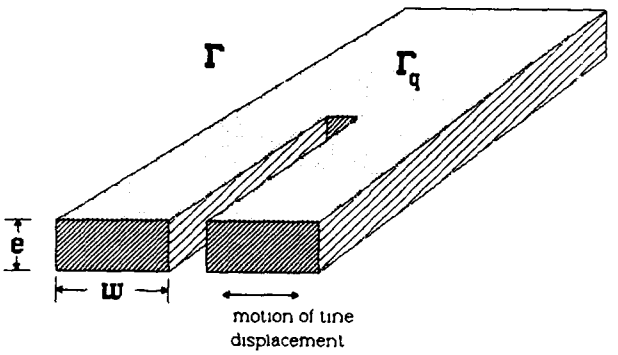


FIGURE 1: Definition of terms for derivation of equation 8.

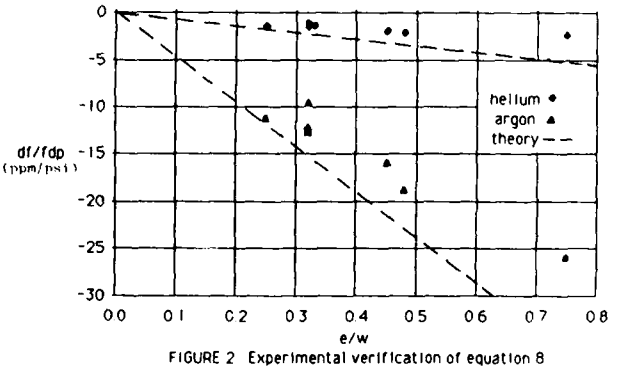


FIGURE 2 Experimental verification of equation 8

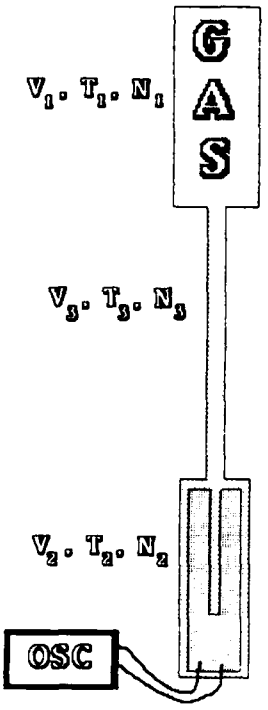


FIGURE 3 A filled thermal system configuration using a quartz crystal tuning fork gas density sensor

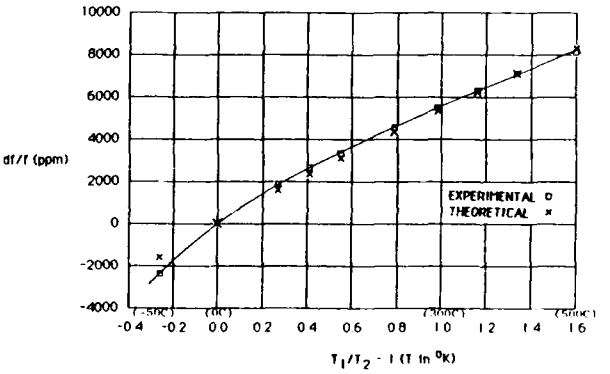


FIGURE 4 Frequency change of a 10kHz tuning fork in an argon backfilled filled thermal system Theoretical results shown with $V_1/V_2=0$).

AUTHOR INDEX

AUTHOR	PAGE
Adachi, T	176
Adams, C.A.	475
Akatsuka, K.	145
Allan, D.W.	107
Armington, A.F.	230
Aubry, J.P.	202, 212, 504
Bahadur, H.	266, 544
Balascio, J.	230
Ballato, A.	462
Banah, A.H.	571
Barnes, J.A.	107
Bauch, A.	8
Benes, E.	535, 556
Benjaminson, A.	140, 171
Bergquist, J.C.	85
Bernier, L.G.	127
Bernot, A.J.	271
Bloch, M.B.	153
Bourquin, R.	405
Brandmayr, R.J.	276
Brauge, J.	504
Brendel, R.	113
Brousseau, R.	64
Brunet, M.	202
Callerame, J.	519
Camparo, J.C.	29
Capelle, B.	234
Caret, G.	212
Chatterjee, A.	59
Chen, C.Y.	259
Chiba, K.	54
Clark, R.L.	166
Clifford, G.C.	367
Cyr, N.	64
Daniel, H.U.	85
Debaisieux, A.	202
De Marchi, A.	3
Derbyshire, A.	18
Détaint, J.	234
Dökmeci, M.C.	431
Doukhan, J.C.	234
Dowsett, J.	301
Driscoll, M.M.	197
Drullinger, R.E.	3, 13, 18
Dulmet, B.	405
Dworsky, L.N.	367, 486
Dwyer, F.G.	301
Ebata, Y.	361
EerNisse, E.P.	575
Eguchi, H.	345
Ezekiel, S.	88
Feldman, M.	18
Filler, R.L.	311
Flamouropoulos, T.	183
Fragneau, M.	504
Frueholz, R.P.	29
Fujiwara, Y.	351
Gagnepain, J.J.	113, 135
Gardiol, F.E.	127
Genestier, G.	405
Gerard, E.	202
Glaze, D.J.	13, 18
Goldfrank, B.	473
Groslambert, J.	113
Gualtieri, J.G.	247
Halliburton, L.E.	259
Harms, K.C.	556
Hashi, T.	54
Hashimoto, M.	43
Hatch, E.R.	462
Hayes, R.R.	80
Heindorff, T.	8
Heinecke, R.A.	301
Hemesath, N.B.	150
Hemmati, H.	85
Hemmer, P.R.	88
Hermann, J.W.	375
Hertl, S.	535
Hilliard, D.	18
Hirose, M.	176
Ho, J.	193
Hou, J.P.	453
Howard, M.D.	481, 491
Howe, D.A.	3, 18
Hunt, J.R.	292, 481
Hwang, H.B.	266
Imae, M.	145
Ishigami, S.	345
Itano, W.M.	85
Jacobs, S.F.	75
Jenselme, L.	333
Jumas, J.C.	234
Kahan, A.	255
Kaitz, G.S.	223
Knoernschild, G.F.	150
Komiyama, B.	159
Kong, A.	356
Korman, J.C.	338, 342
Kosinski, J.A.	400
Krisnaswamy, S.V.	372
Kuramochi, N.	448
Kusters, J.A.	223, 475
Lahiri, S.K.	514
Lau, K.F.	356
Laudise, R.D.	270
Lee, P.C.Y.	415, 453
Lefevre, R.	333
Leuchs, G.	85
Lewis, L.L.	18
Lewis, R.L.	150
Lipson, H.G.	255
Lukaszek, T.	462
Marotel, G.	212
Martin, J.J.	266
Mathur, B.S.	59
Mattison, E.M.	72,75
McAvoy, B.R.	372
Meirs, M.	24
Michaud, A.	64
Miki, C.	145
Miura, M.	386
Miyazaki, Y.	448
Mizan, M.	462
Montress, G.K.	519
Moore, R.A.	372
Morikawa, T.	145
Morley, P.E.	317, 491
Morris, S.E.	527
Nakata, H.	386
Narahara, C.	361
Ohtsu, M.	43
Okazaki, M.	381
Ono, M.	351
Ontai, G.P.	88
Oura, N.	448
Ozawa, H.	43
Parker, T.E.	97, 519
Parshad, R.	544
Pascaru, I.	18, 24
Peach, R.C.	392, 527
Philippot, E.	234
Reddy, A.R.	514
Richardson, C.K.	189
Rosenberg, A.	88
Rovera, G.D.	3
Sachdeva, D.S.	59
Sakai, M.	351
Salvo, H.L., Jr.	372
Sanbao, X.	132
Sato, M.	345
Satoh, H.	361
Saxena, G.M.	59
Scansaroli, M.N.	367
Schmid, M.	535
Schwartzel, J.	234
Sekimoto, H.	386
Servajean, D.	333
Shanley, C.W.	367, 486
Shen, W.	72
Sherman, J.H., Jr.	427
Shirley, J.H.	18, 22
Silvermetz, D.	24
Smythe, R.C.	138,292,317,481,491
Stanciulescu, D.	18
Stern, F.	301
Stevens, S.	325, 436
Sullivan, D.B.	13
Suzuki, H.	361
Suzuki, S.	448
Tapp, S.D.	259
Tellier, C.R.	282
Têtu, M.	64, 119
Thorn, G.	556
Tiersten, H.F.	325, 436
Tilton, R.	462
Tohma, H.	381
Tomikawa, Y.	381
Tremblay, P.	64, 119
Truelove, A.H.	301
Tsaclas, J.	342
Tsuzuki, Y.	176
Uratsuka, M.	145
Uzunoglu, V.	183
Vessot, R.F.C.	72, 75
Vig, J.R.	276
Villeneuve, B.	64
Vulcan, A.	153
Wakatsuki, N.	351
Walls, F.L.	91
Wang, J.S.	356
Wang, H.T.M.	80
Ward, R.W.	575
Warner, A.W.	342, 473
Whalin, J.A.	367
White, M.H.	183
Wimmer, L.	535
Wineland, D.J.	85
Wulin, D.	132
Yamamoto, K.	345
Yamashina, M.	448
Yen, K.H.	356
Yong, Y.K.	415
Yoshimura, K.	145
Zarka, A.	234

SPECIFICATIONS AND STANDARDS GERMANE TO FREQUENCY CONTROL

INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS

Order through: IEEE Service Center
445 Hoes Lane
Piscataway, NJ 08854
(201) 981-0060

176-1978 Piezoelectricity

177-1966 Piezoelectric Vibrators, Definitions and Methods of Measurements for (ANSI C83.17-1970)

180-1962 Ferroelectric Crystal Terms, Definitions of

319-1971 Piezomagnetic Nomenclature

ELECTRONIC INDUSTRIES ASSOCIATION

Order through: Electronic Industries Association
2001 Eye Street, NW
Washington, DC 20006
(202) 457-4900

(a) Holders and Sockets

RS-192-A, Holder Outlines and Pin Connections for Quartz Crystal Units (Standard Dimensions for Holder types).

RS-367, Dimensional and Electrical Characteristics Defining Receiver Type Sockets (Including crystal sockets).

RS-417, Crystal Outlines (Standard dimensions and pin connections for current quartz crystal units-1974).

(b) Production Tests

RS-186-E, (All Sections), Standard Test Methods for Electronic Component Parts

RS-512, Standard Methods for Measurement of Equivalent Electrical Parameters of Quartz Crystal Units, 1 kHz to 1 GHz, 1985.

(c) Application Information

Components Bulletin No. 6, Guide for the Use of Quartz Crystals for Frequency Control

(d) RS-477, Cultured Quartz (Apr. 81)

INTERNATIONAL ELECTROTECHNICAL COMMISSION (IEC) *

Order through: American National Standards Inst.(ANSI)
1430 Broadway
New York, NY 10018

*ANSI can quote prices on specific IEC publications on a day to day basis only. All IEC and ISO standards have been removed from its Standards Catalog. Call ANSI, (212) 354-3300 for prices.

IEC PUBLICATIONS ISSUED:

122:- Quartz crystal units for frequency control and selection

122-1 (1976) Part 1: Standard values and test conditions.

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 connectors.

122-3A (1979) First supplement

122-3B (1980) Second supplement.

122-3C (1981) Third supplement.

283 (1968) Methods for the measurement of frequency and equivalent resistance of unwanted resonances of filter crystal units.

302 (1969) Standard definitions and methods of measurement 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.

368A (1973) First supplement

Amendment No. 1 (1977)

Amendment No. 2 (1982)

368B (1975) Second supplement.

368-1 (1982) Part 1: General information, standard values and test conditions.

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 (1973) Basic method for the measurement of resonance frequency and equivalent series resistance of quartz crystal units by zero phase technique in a π network.

Note: This publication, where revised, will be issued as a second edition of Publication 444-1.

444-2 (1980) Part 2: Phase offset method for measurement of motional capacitance of quartz crystal units.

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 and methods.

679-2 (1981) Part 2: Guide to the use of quartz crystal controlled oscillators.

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.

DEPARTMENT OF DEFENSE

Order through: Naval Publication & Form Center
5801 Tabor Avenue
Philadelphia, PA 19120

MIL-C-3098 Crystal Unit, Quartz, Gen Spec for
MIL-H-10056 Holders (Encl), Crystal, Gen Spec for
MIL-STD-683 Crystal Units, Quartz/Holders, Crystal
MIL-O-55310 Oscillators, Crystal, Gen Spec for
MIL-F-18327 Filters, High Pass, Low Pass, Band Pass Suppression and Dual Functioning, Gen Spec for
MIL-O-39021 Oven, Crystal, Gen Spec for
MIL-O-55240 Oscillators, Audio Frequency
MIL-F-28734 Frequency Standards, Cesium Beam, Gen Spec
MIL-F-288111 Frequency Standard, Cesium Beam Tube
MIL-C-24523 (SHIPS), Chronometer Quartz Crsytal

PROCEEDINGS

ANNUAL FREQUENCY CONTROL SYMPOSIA

NO.	YEAR	DOCUMENT NUMBER	OBTAIN FROM *	COST
10	1956	AD298322	NTIS	$41.50
11	1957	AD298323	"	44.50
12	1958	AD298324	"	46.00
13	1959	AD298325	"	49.00
14	1960	AD246500	"	32.50
15	1961	AD265455	"	28.00
16	1962	PB162343	"	35.50
17	1963	AD423381	"	43.00
18	1964	AD450341	"	43.00
19	1965	AD471229	"	47.50
20	1966	AD800523	"	47.50
21	1967	AD659792	"	41.50
22	1968	AD844911	"	44.50
23	1969	AD746209	"	25.00
24	1970	AD764210	"	28.00
25	1971	AD746211	"	28.00
26	1972	AD771043	"	26.50
27	1973	AD771042	"	34.00
28	1974	ADA011113	"	31.00
29	1975	ADA017466	"	34.00
30	1976	ADA046089	"	40.00
31	1977	ADA088221	"	44.50
32	1978		EIA	20.00
33	1979		"	20.00
34	1980		"	20.00
35	1981		"	20.00
36	1982	ADA130811	NTIS	41.50
37	1983	83CH1957-0	IEEE	59.00
38	1984	84CH2062-8	"	59.00
39	1985	85CH2186-5	"	59.00

*NTIS - National Technical Information Service
Sills Building
5285 Port Royal Road
Springfield, VA 22161

* EIA - Annual Frequency Control Symposium
c/o Electronic Industries Association
2001 Eye Street
Washington, DC 20006

*IEEE - Institute of Electrical & ELectronics Engineers
445 Hoes Lane
Piscataway, NJ 08854

Remittance must be enclosed with all orders. Prices are subject to change without prior notice.

The Proceedings of the 33rd (1979) and 34th (1980) Symposia contain a bibliography of the world-wide literature on precision frequency control and selection compiled by Dr. E. A. Gerber for the years 1968-1978 with part of 1980, respectively.

A subject and author index for the Proceedings of the 10th through the 38th Symposia appears as a supplement to the 38th Proceedings volume.