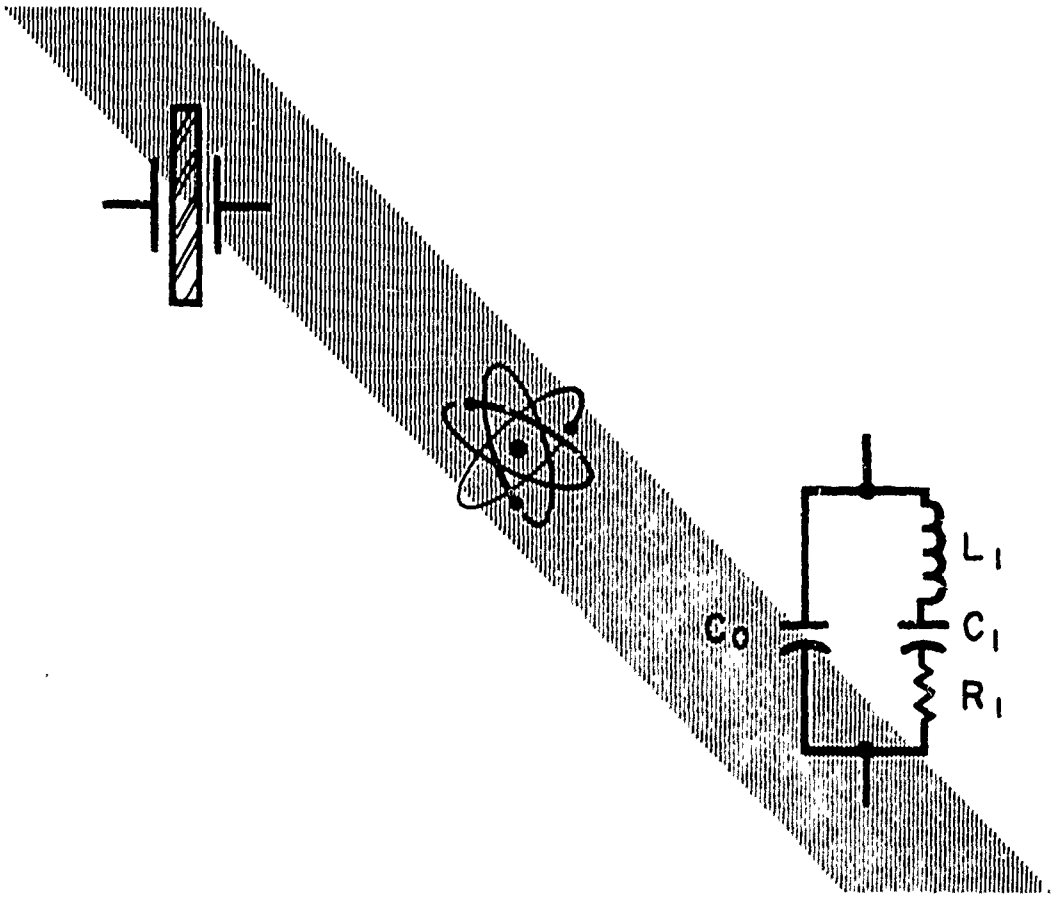
AD 746209

AD 65092

PROCEEDINGS
OF THE
23rd ANNUAL SYMPOSIUM ON FREQUENCY CONTROL



6-7-8 MAY '69

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. Department of Commerce
Springfield VA 22151

U.S. ARMY ELECTRONICS COMMAND
FORT MONMOUTH, NEW JERSEY

DISTRIBUTION STATEMENT

Approved for public release;
distribution unlimited.

Best
Available
Copy

PROCEEDINGS

OF THE

TWENTY-THIRD ANNUAL FREQUENCY CONTROL

SYMPOSIUM

Sponsored by

SOLID STATE AND FREQUENCY CONTROL DIVISION
ELECTRONIC COMPONENTS LABORATORY
U. S. ARMY ELECTRONICS COMMAND
FORT MONMOUTH, NEW JERSEY

Details of illustrations in
this document may be better
studied on microfiche.

ELECTRONIC INDUSTRIES ASSOCIATION
2001 Eye Street, N. W.
Washington, D.C. 20006

6, 7, 8 May 1969

The Shelburne Hotel
Atlantic City, New Jersey

TWENTY-THIRD ANNUAL FREQUENCY CONTROL SYMPOSIUM

Sponsored By

U. S. Army Electronics Command
Solid State and Frequency Control Division
Electronic Components Laboratory
Fort Monmouth, New Jersey

6, 7, 8 May 1969

The Shelburne Hotel
Atlantic City, New Jersey

EXECUTIVE COMMITTEE

General Chairman..... Mr. V. J. Kublin
Executive Assistant..... Mr. M. F. Timm
Technical Program Chairman..... Dr. G. K. Guttwein
Local Arrangements..... Mr. J. M. Stanley

TECHNICAL PROGRAM COMMITTEE

Dr. G. K. Guttwein, Chairman
U. S. Army Electronics Command

Mr. M. Bernstein
U. S. Army Electronics Command

Mr. A. R. Chi
NASA, Goddard Space Flight Center

Mr. C. W. Friend
Wright-Patterson Air Force Base

Mr. J. M. Giannotto
U. S. Army Electronics Command

Dr. E. Hafner
U. S. Army Electronics Command

Dr. D. Halford
National Bureau of Standards

Mr. D. L. Hammond
Hewlett-Packard Company

Dr. H. Hellwig
U. S. Army Electronics Command

Mr. O. P. Layden
U. S. Army Electronics Command

Dr. W. J. Spencer
Bell Telephone Laboratories, Inc.

Mr. R. R. Stone
U. S. Naval Research Laboratory

Dr. G. M. R. Winkler
U. S. Naval Observatory

HONORARY CRYSTAL INDUSTRY COMMITTEE

Mr. J. J. Colbert
Electronic Crystals Corporation

Mr. D. L. Hammond
Hewlett-Packard Company

Mr. J. D. Holmbeck
Northern Engineering Laboratories

Mr. J. H. Sherman, Jr.
General Electric Company

Mr. R. A. Sykes (Ret.)
Bell Telephone Laboratories, Inc.

CHAIRMEN FOR TECHNICAL SESSIONS

Trends in Frequency Control Applications

Mr. O. E. McIntire, Federal Aviation Administration

Fundamental Crystal Studies and Crystal Filters

Dr. W. J. Spencer, Bell Telephone Laboratories, Inc.

Crystal Measuring Problems

Mr. D. L. Hammond, Hewlett-Packard Company

Crystal Design Engineering and Processing

Mr. J. D. Holmbeck, Northern Engineering Laboratories

Oscillators, Synthesizers, and Circuits

Mr. R. W. Frank, General Radio Company

Timekeeping and Distribution

Dr. G. M. R. Winkler, U. S. Naval Observatory

Atomic and Molecular Frequency Control

Prof. R. F. C. Vessot, Columbia University

Optical Frequency Control

Dr. H. Boyne, National Bureau of Standards

INDEX

PAGE

Trends in Frequency Control Applications

Frequency Control Requirements for the Mallard Communication System
J. DelVecchio and J. Dressner..... 1

Application of Precise Time-Frequency Technology in Multi-Function Systems
T. C. Viars..... 8

Frequency Control for Tactical Net SSB Equipment
O. P. Layden..... 14

International Coordination of Radio Time Signal Emissions
H. Smith..... 18

Fundamental Crystal Studies and Crystal Filters

Measurement of the Piezoelectric Coefficient of Quartz Using the Fabry-Perot Dilatometer
V. E. Bottom..... 21

Analysis of Contoured Piezoelectric Resonators Vibrating in Thickness-Twist Modes
M. Onoe and K. Okada..... 26

Resonance Frequencies of Monolithic Quartz Structures
A. Glowinski..... 39

Electric Field Effects in Monolithic Crystal Filters
H. F. Tiersten..... 56

The Development of High Performance Filters Using Acoustically Coupled Resonators on AT-Cut Quartz Crystals
J. F. Werner, A. J. Dyer and J. Birch..... 65

High Frequency Crystal Monolithic (HCM) Filters
H. Yoda, Y. Nakazawa and N. Kobori..... 76

Crystal Measuring Problems

Vector Voltmeter Crystal Measurement System
M. E. Frerking..... 93

On Precision Measurements of Frequency and Resistance of Quartz Crystal Units
C. Franx..... 102

Another Look at Specifying a Crystal
D. W. Nelson..... 111

Temperature Run, MIL-C-3098, Amendment I
R. Pompeo and F. Wolf..... 122

*Abstract only - complete paper not received in time for publication

PAGE

Crystal Design Engineering and Processing

Anomalous Vibrations in AT-Cut Plates
I. Koga..... 128

Comparison of Aging Performance of 5-MHz Resonators Plated with Various Electrode Metals
R. B. Belser and W. H. Hicklin..... 132

A Novel Algorithm for the Design of the Electrodes of Single-Mode AT-Cut Resonators
J. H. Sherman, Jr..... 143

A Flexure-Mode Quartz for an Electronic Wrist Watch
M. P. Forrer..... 157

Laser Machining Thin Film Electrode Arrays on Quartz Crystal Substrates
J. L. Hokanson..... 163

Computerized Process Control for Synthetic Quartz Growth
D. W. Rudd, R. E. Dubois and N. C. Lias..... 171

Oscillators, Synthesizers, and Circuits

Effects of Gamma Irradiation on Frequency Stability of 5th Overtone Crystal Oscillators
C. A. Berg and J. R. Erickson..... 178

A Report on Segmented Compensation and Special TCXO's
D. E. Newell and H. Hinnah..... 187

Temperature-Compensated Csytal-Controlled Oscillators Operating from 800 kHz to 1500 kHz
H. A. Batdorf..... 192

Frequency Standards for Communications
E. N. LeFevre..... 198

Digiphase Frequency Synthesizer
G. C. Gillette..... 201

A Miniature Precision Digital Frequency Synthesizer
R. J. Hughes and R. J. Sacha..... 211

A Carrier Suppression Technique for Measuring S/N and Carrier/Sideband Ratios Greater than 120 dB
C. H. Horn..... 223

Timekeeping and Distribution

Use of the Loran-C System for Time and Frequency Dissemination
LCDR P. E. Pakos..... 236

An Application of Statistical Smoothing Techniques on VLF Signals for Comparison of Time Between USNO and NBS
A. Guetrot, D. W. Allan, L.S. Higgle and J. Lavanceau..... 248

PAGE

Timekeeping and Distribution (Continued)

A Coordinate Frequency and Time System
G. E. Hudson, D. W. Allan, J. A. Barnes, J. Lavenceau, R. G. Hall and G. M. R. Winkler..... 249

Atomic and Molecular Frequency Control

Light Modulation at the Rb^{87} Hyperfine Frequency
H. Tank and W. Happer..... 263

An Optically Pumped Rb^{85} Maser Frequency Standard
W. A. Stern and R. Novick..... 271

A Report on the Performance Characteristics of a New Rubidium Vapr Frequency Standard
D. H. Throne..... 274

Improved State Selection for Hydrogen Masers
R. F. Lacey and R. F. C. Vessot..... 279

Recent Results Concerning the Hydrogen Maser Wall Shift Problem
E. E. Uzgiris and P. W. Zitzewitzt..... 284

A New Method for Measurement of the Population Difference of Hyperfine-Levels of Stored Atoms
C. Audoin, M. Desaintfuscien, P. Piejus and J. P. Schermann..... 288

Atomic Hydrogen Standards for NASA Tracking Stations
H. E. Peters, T. E. McGunigal and E. H. Johnson..... 297

Optical Frequency Control

Frequency Mixing and Multiplication in the Far Infrared and Infrared*
A. Javan..... 305

Pressure Shift and Broadening of Mathane Line at 3.39 Micron Studied by Laser-Saturated Molecular Absorption*
R. L. Barger and J. L. Hall..... 306

Stability Investigations of HCN Laser
V. J. Corcoran, R. E. Cupp and J. J. Gallagher..... 307

Laser Frequency Stabilization Using a Primary Frequency Reference*
S. Ezekiel..... 312

Precision Time Measurements of Optical Pulsars*
P. Boynton, R. B. Partridge and D. T. Wilkinson..... 313

FREQUENCY CONTROL REQUIREMENTS FOR THE MALLARD COMMUNICATION SYSTEM

Jerry M. Dressner
Communications/Automatic Data Processing Laboratory
US Army Electronics Command
Fort Monmouth, New Jersey

and

John P. Del Vecchio
The MALLARD Project
US Army Materiel Command
Fort Monmouth, New Jersey

Summary

The results of work so far performed in the area of synchronization techniques have been promising. It was found that of the four techniques investigated, i.e., master-slave, bit stuffing, frequency averaging and independent atomic clocks, frequency averaging and independent atomic clocks might have application in a future field army communication network.

The simplest technique to achieve synchronization from a conceptual point of view, is the one which utilizes atomic clock frequency standards at each node. The military versions of atomic clocks which are available today are rugged. Their small size and light weight, coupled with their inherently high stability make them suitable for tactical operations. Although the cost of achieving synchronization using atomic clocks is higher than other known methods, the simplicity of system operation which results, and the minimal interconnection problems with other systems may more than compensate for it.

The main advantage of the frequency averaging technique is that it appears relatively inexpensive, has smaller size and weight than the atomic standard, and has a potential for greater reliability. The technique can also readily be implemented with conventional circuitry. Mathematical studies and computer analysis performed thus far indicate that stable operation can quickly be arrived at and maintained with only minor adjustments even with very low accuracy clocks. However, this technique has not yet been conclusively demonstrated for large networks consisting of 150 nodes, under various connectivities and for a variety of disturbing effects. This, of course, is absolutely necessary to the success of its use in a tactical environment.

Introduction

The MALLARD Communications System, presently in its "Concept Formulation" phase, is planned to provide a Joint Service-International deployment of armed forces with the capability of a fully tactical, automatically switched, digital, secure communications network. Each switching center has incoming and outgoing trunk links to other switching centers and provides the necessary switching between links. Channel rates are expected to be 19.2 to 38.4 Kbits and link rates are expected between 1-5 Meg bits. A variety of transmission media are utilized, among which are Line of Sight, Radio Relay, Troposcatter and Satellite Communications. The fully digital nature of the system, and the capability for the accurate interchange of high speed digital information among the switching centers calls for highly accurate system synchronization. This paper discusses the synchronization problem that exists under these conditions, the possible options, and the most promising methods for establishing bit integrity for the entire system.

Background

Project MALLARD is charged with the development of a Post-75 communication system for the military tactical forces. It is an international cooperative project among United States, United Kingdom, Canada, and Australia which had its beginnings several years ago with informal quadripartite meetings staffed by scientists from each of the

countries' military establishments. These meetings developed the basic technical and operational requirements for Post-75 tactical communications and published its findings in a voluminous set of documents, which also outlined and defined an on-going program of research and development. This initial in-house planning was called MALLARD I. MALLARD II continued the program with the initiation of three competitive, industry study efforts designed to demonstrate, independently, the feasibility of designing a complete military tactical communication system and to prove the feasibility of achieving a full field deployment by the 1975-77 time frame. At present, mid 1969, these studies are being concluded and their results are being evaluated and correlated by a special international team of experts. The purpose of this process is to select the best recommendations from each of the studies, and from these, to develop a full system implementation. When evaluation has been completed in mid May 1969 and the system concepts decided, a functional model will be developed and tested. This will be followed by a full engineering development phase and finally production and fielding.

Objective

The basic objective of MALLARD is to develop a secure, tactical, trunk communication system, including the methods and means of subscriber access. The system must be capable of full security, allow the transmission of data, automatically switch subscribers, and possess both quick reaction time and increased transmission reliability over that presently achievable.

SAMPLE NETWORK CONFIGURATION

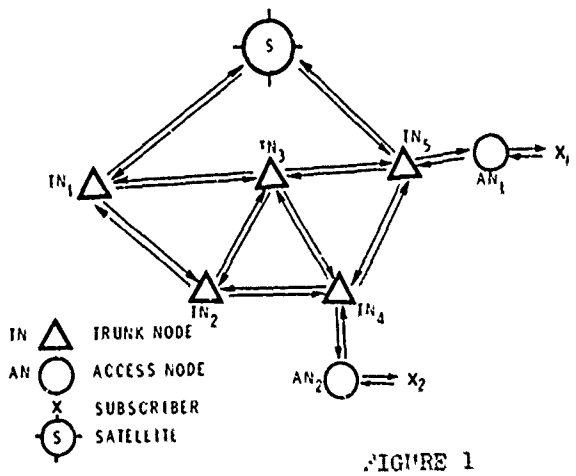


FIGURE 1

The system as depicted in figure 1 is envisioned as composed of an interconnected network of nodes at which the basic operations of routing, switching and multiplexing are performed. The nodes are interconnected by multichannel radio circuits. Subscribers are interconnected to one another through the facilities of the nodes. Groups of co-located subscribers may have their own local nodes, and certain specialized groups of subscribers, such as the Random Access Discrete Address System (RADA) for instance, may also be served through the nodal network. Several types of internodal radio facilities are envisioned; line of sight radio relay, tropospheric scatter and satellite relay are the most important.

In a full tactical field force of 2 corps and 8 divisions, requirements are envisioned to be of the order of 150 nodes and 50,000 separate subscribers. Regardless of system size, it must satisfy a wide range of subscriber types from relatively fixed geographically type such as a headquarters to highly mobile type such as division commander travelling about in a jeep.

The requirement for full security and, both voice and data communication naturally leads to the design selection of a common carrier digital system.

In the node itself, the most important unit is the switch and switch control. Time division switching of pulses which come from other nodes and from the individual subscribers on a synchronous basis are timed by a local clock located in the node's control unit. Hence, the requirement that all pulse streams coming from other nodes must somehow be put into proper phase and channel alignment so that the switching procedure is carried out without inadvertent channel interchange, losses or inversions of adjacent bits. This requirement is one of the bases for the necessity to examine inclose detail the methods which are available to keep the entire internetted system in synchronization.

System Synchronization

Master-slave

Considering only nodes S_1 and S_2 with their interconnections as shown in figure 2, assume 12 subscribers are attached to each node and the nodes are interconnected by a multichannel radio trunk group of six channels.

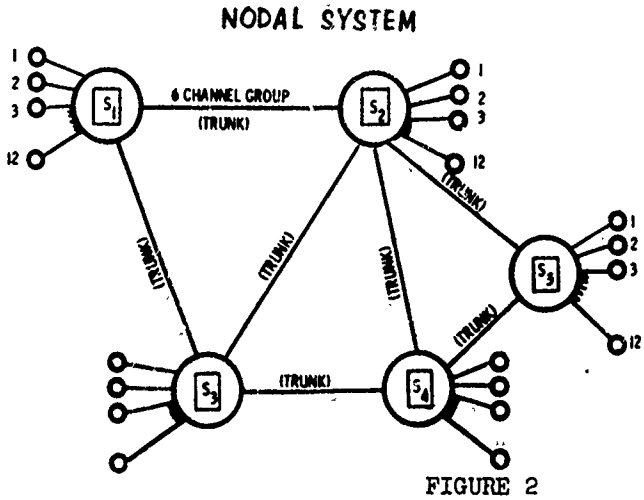


FIGURE 2

Pulse transmission in both directions occurs at some regular rate between the nodes which is held fixed and nonchangeable even if certain of the channels are not in actual use by any of the subscribers. The stability of the rate at which each node's transmission is generated is determined of course by the clock at each node, since the clock at each node controls the time position of each outgoing pulse. Similarly, each node receives, and clocks in the distant transmission at its own local rate. Suppose the clock rates differ slightly. Then the receiving node will eventually "slip" a pulse, by either having to throw one away if the pulses are coming in too fast, or waiting through an empty period if the pulses are coming in too slow. In either case, the continuity of service has been destroyed. What has come to be called "bit integrity" has been lost. For this simple case, it is possible to slave one of the nodes to the other and the problem is solved. Which one slaves to which, in this example, does not matter as long as it results in one of the clocks being forced into synchronization with the other. It has been assumed, in fact, all along that the subscribers devices, usually controlled by a crystal oscillator, are being slaved to their parent nodes. Now consider the addition of a third, fourth and fifth node. Their addition complicates the system and it is no longer a simple matter of slaving the incoming bit stream, since each node receives several such bit streams from several different nodes each with its own distinct clock. Slaving to only one arbitrary node is not enough; the other incoming bit streams are still out of synchronization. The slave principle can in theory still be maintained but it is not very practical. It is done as follows. Designate one node as the Master for the whole system. Designate certain of the paths through the system as control paths for slaving and inform each of the slaved nodes which path to slave on. If this is done correctly, then all stations or nodes will clock to the node designated the Master. This method has a number of drawbacks. A tactical land force must be highly mobile. Elements of this force are constantly picking up and moving. Both nodes and subscribers move around frequently, and nodes may be destroyed or links disrupted. Consequently, the system connectivity is always changing. As a result, the system rules and regulations which determine the master and slave relationship must constantly be updated. This seriously complicates the overall system control and adds to its complexity which for other very good reasons is already complex enough. Therefore, there is reason to ask if there are other perhaps more desirable synchronization alternatives. There are three other methods distinctly enough different from each other to be considered separately.

Frequency Averaging

A logical generalization of the slave principle is called Frequency Averaging. If each node took some average of all the incoming rates and slaved its own clock to that average, it can be shown mathematically that if a steady state exists, it would be one in which the entire system would settle down to a common clock rate and thus achieve, theoretically at least, total and complete system synchronization forever. This concept has been subjected lately to thorough examination and treatment in the communications literature for commercial application. However, its treatment for highly tactical mobile systems such as MALLARD has only recently begun and as discussed later, leave unanswered many important questions.

Independent Atomic Clocks

Another distinct method of system synchronization uses uncontrolled clocks which are highly accurate with respect to one another. Atomic Standards are available with accuracies of the order of plus or minus 5 parts in 10^{12}. Such clocks, if used at each node, would insure that not more than 1 pulse in a 10^{11} would get lost. At an expected internodal transmission rate of 1 megabit per second this amounts to

about one day's worth of "bit integrity".

When complete synchronization of the system is required, it is not meant forever, since both nodes and subscribers move around. This required mobility of the system pervades and dominates the entire set of military requirements and sets this system distinctly apart from previous developments of digital systems for the military. It appears that system synchronization over a period of perhaps a day would be sufficient to meet the basic tactical requirement. This being so, the independent operation of each station with its own uncontrolled but highly stable atomic clock appears as a satisfactory candidate with many useful features to recommend. Again, subscribers are slaved to their parent nodes, and perhaps even the so called minor nodes which connect to only one major node by one path. This exception will obviously cause no difficulty and is in fact economically desirable since, by slaving, they need not be supplied with an expensive atomic standard.

Bit Stuffing

The last possibility discussed here uses the principle of bit stuffing or message redundancy. In this method, the individual nodal clocks remain uncontrolled as in the independent clock method, however, they are not required to be highly stable atomic standards. If left to operate with say 1 in 10^8 per week stability, then bit slippage would occur about 1000 times more often than with atomic standards or roughly once every 100 seconds, and this would not satisfy the basic user requirements. The solution runs as follows. Transmit from each node at a slightly higher rate than is actually necessary to transmit the given subscriber information and fill in the gaps thus created in the transmitted stream with so called dummy bits. These dummy bits are added periodically to the transmitted stream in such a way that upon reception at the distant node, their presence and locations are easily identifiable. What the receiving node then sees is an incoming bit stream at a rate higher than its own clock and of course must periodically throw bits away, but it makes sure it throws one of the useless dummy bits away rather than an important data bit.

If the design of the system is correct; the transmission rates are chosen correctly and the dummy bits encoded into the stream properly, then it is possible to preserve information bit integrity over as long a period as required.

Techniques Constraining System Effectiveness

The Master-Slave principle suffers from complexity of the system control and since its limited application to MALLARD has been quite clearly defined, it will not be further discussed. The method of pulse stuffing, although technically feasible, does not appear as a major candidate because when compared against the two remaining candidates it suffers some major drawbacks. It requires in the first place that a higher bit rate be transmitted (in order to include the dummy bits) than would ordinarily be needed, thus it is wasteful of bandwidth. Second, it requires that each node possess a rather increased degree of complexity because it must detect the dummy bits, compare against the local node rate, and make decisions as to how to modify the dummy structure. Then each node must reformat its outgoing data streams with new patterns for transmission to the next node. Further complicating the picture is the necessity for encoding to counter the effects of system errors which, if not corrected, would increase the probability of mistaking data bits for dummy bits and vice versa. For these reasons, bit stuffing, as a general method of achieving system synchronization does not appear satisfactory.

We have then the two remaining methods - frequency averaging and independent clocks.

Remaining Techniques

Frequency Averaging Concept

The frequency averaging approach to the synchronization problem is based upon a comparison of the average phase of all incoming data and the local clock phase. Phase error, either positive or negative, is used to increase or decrease the clock frequency. These corrections are such as to minimize phase error, so that it is always less than one bit. This system will maintain all local clocks at one frequency, which turns out to be the average of the uncorrected clock frequencies. The system would make use of high quality voltage controlled crystal oscillators. Since crystal frequency drift is usually bounded by maximum and minimum values which are within the electrical pull-in range and well within the manual control range, oscillator lifetime is not limited by frequency drift.

FREQUENCY AVERAGING CONCEPT

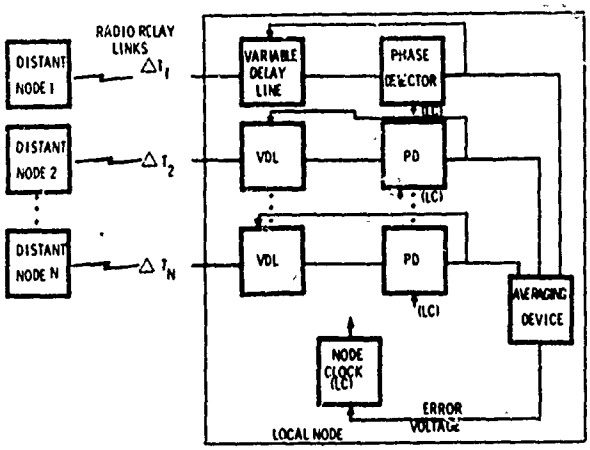


FIGURE 3

As shown in figure 3, the system may be described as follows:

Each variable delay line provides a phase adjustment by means of which changes in link propagation time may be compensated for. The delays may be positive or negative referred to a midpoint value. The phase detector on each incoming line measures the time difference between the output data transitions of each variable delay line and the local clock transitions. The error voltage developed from each phase detector is applied to the variable delay line corresponding to the link; it is also applied, through an averaging device to the node clock. The variable delay line and the clock frequency are thereby changed in such a direction as to reduce the error voltage from the phase detector. This will insure that the phase error is always less than one bit and that meaningful data is not lost. There are several variations of this method. The local clock may be controlled continuously as described above, or alternately it may be switched discontinuously between a high and low value, as a function of the delay line storages. Both of these are methods of changing the frequency of the local clock. The information upon which the phaselock acts can be of two types, single end or both end control. Single end Control signifies that each node controls its clocks solely on the basis of the status of its own delay lines or buffer stores. A more complicated method is available where information concerning the status of remote nodal delay lines is also used to readjust the local clocks. This is called both end control. The intermix of continuous and switched control with either single or both end control can result in four possible types of frequency averaging.

Continuous control of frequency has been shown to be feasible with either single or both-end control. Both end control has been shown to have the advantage of reducing overall storage requirements, but it is necessarily a more complicated implementation. An alternative to variable delay lines is the buffer store which may be used in another equivalent implementation.

The amount of time that a continuous control system will remain in synchronization appears to be unlimited, barring exceptional malfunctions such as excessive frequency drift in a single clock.

Discontinuously switched frequency control proved more difficult to analyze but has been shown to be equally suitable. However, it cannot be implemented as easily as the continuous system and has not been considered a practical contender.

Initial Setability of the system is an important consideration.

For the single end control method, the midpoint of each oscillator may be either manually set to a known standard frequency if one is available, or else since all system oscillators are at the same unknown value of frequency within say one part in 10^7 or 10^8 per wk it may not be necessary to set each oscillator at all. Even so, some procedure for bringing a new node into operation in an already existing network is desirable since it may occur that the existing system frequency and the free running frequency of the new node may differ greatly. Such procedures are under investigation and all depend in some fashion on a method for introducing a variable but fixed bias on the center frequency of each new oscillator.

Limited work had been performed on the analysis of controlled communications systems, both single end and both end under continuous and discontinuous control, for The MALLARD Project. But as stated earlier, the analysis of a communications system containing 150 nodes each highly mobile has not been

completed. Efforts in this area will be continued if the frequency averaging technique is the method of frequency control selected for the MALLARD system.

Independent Atomic Clock

In the independent clock approach, a high stability device must be used at each node, for a system which uses independent clocks relies solely upon the high accuracy and stability of each node's clock in order to achieve bit integrity.

In order to increase the effectiveness of this system's performance, small buffers may be used at each node in conjunction with the high accuracy clocks to act as additional elastance in the system and also to average out any short-term radio media delay variations.

Clocks which may be suitable for this application include the categories of atomic standards, rubidium and ultra stable crystal oscillators. In order for these timing devices to be tactically useful, they must meet the following requirements:

(a) High accuracy and long-term stability.

(b) Short warm up time.

(c) High reliability.

(d) Reasonable weight, size and power consumption.

(e) Low price of the clock and simple implementation.

An atomic frequency standard is indicated for use in this system since it possesses the qualities of very high accuracies, very short warm up time, very smooth fine tuning and high environmental insensitivity. For long-term stability and for the best temperature stability a cesium beam standard has been considered. Rubidium gas cells too have been considered for their high short and long-term stability, and insensitivity to coriolis forces.

Atomic standards can be tuned in steps of 1×10^{11} compared with quartz crystal oscillators where tuning sometimes has an effect on the aging characteristics. The warm up time required for the crystal, especially a high precision crystal unit, is in terms of hours; for atomic standards it is in terms of minutes, The tactical advantage is obvious. There is a disadvantage in Rubidium in that gas cell standards unlike the cesium are actually secondary standards. This requires that it must initially be calibrated. Another Rubidium effect which has recently been observed is a small but non-random drift of the order of a part in 10^{11} per month. The cause of this non-random instability has not yet been fully determined. Its effect on the system performance would be to require periodic rests of perhaps once every several months and some additional buffer storage on each incoming transmission link.

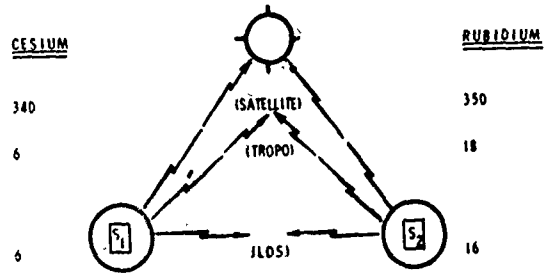
Both types, Cesium and Rubidium, can easily be incorporated into a communication system, however, some arrangement in case of failure of the node clock must be at hand. Should such a clock fail, a procedure for resetting must be at hand as rapidly as possible to maintain continuity of the system.

If each node is provided only a single standard it might slave temporarily to an adjacent node's clock. Some disruption of service might ensue in such a case. If two clocks per node are used with some sort of periodic checking capability between them then it is possible to preserve continuity of service at an increased cost. For instance, the output of one of standards may be fed to a phase shifter, and then to a phase detector. The other output would be fed directly to the phase detector. Both outputs are compared and should a phase difference exist, one of the outputs is periodically adjusted so as to be identical with the other. The two output lines are then fed to the alarm and disconnect circuits before going out on the output lines. Should one of the clocks fail, the alarm unit detects this and simply disconnects the faulty line without introducing a transient into the output lines. The faulty unit is then repaired and installed back in the system.

Performance of the atomic standards can be improved by incorporating small buffers at each receiving terminal. These will be read into at the transmitted (received) rate and read out of at the local clock rate.

Clearly, when the local buffers at each node either fills up completely or depletes completely a reset is necessary during which bit integrity cannot be maintained. Estimates of the amount of buffering which is necessary in order to maintain a given period of bit integrity have been developed in a number of studies. Three basic factors must be accounted for, the relative drift of the clocks, the variations in delay through the propagation medium and the desired

length of time between resets. For one day bit integrity, using cesium clocks, and assuming no a priori knowledge as to direction of drift it is necessary to account for a possible bit offset of 1.7 bits in either direction during a day when transmitting at one megabit/sec. To this must be added the maximum expected media delay changes. For the means of transmission cited earlier namely radio relay, troposcatter and satellite, the only one which introduces a significant delay variation is satellite.



BUFFER REQUIREMENTS FOR 1 DAY RESET FIGURE 4

Consequently, as depicted in figure 4, 340 bits of buffer storage is required for that case against only six for the other two. For Rubidium clocks these are all about 10 bits higher reflecting the lesser stability and small drift of Rubidium. For higher transmission rates of course, either proportionately more storage is required or a smaller period of bit integrity must be accepted.

Conclusions

Figure 5 lists some of the characteristics of Atomic Clocks and Frequency Averaging along with what is considered to be their advantages.

	ATOMIC CLOCKS	FREQUENCY AVERAGING
SIZE	1 CUBIC FOOT	.1 CUBIC FOOT
WEIGHT	30 - 70 POUNDS	4 TO 5 POUNDS
RELIABILITY	20,000 MTBF	40,000 MTBF
HIGH STABILITY	1 PART IN 10^{12}	
ADVANTAGES	SIMPLICITY OF OPERATION	SMALLER SIZE AND WEIGHT
	MINIMAL INTERCONNECTION NECESSARY	INCREASED RELIABILITY
		LOWER COST

FIGURE 5

Other factors for consideration are the costs of each technique which is easily computed on a per node basis. The frequency averaging system must provide a summing device for each node while the atomic clock system does not. Projected estimates for atomic clocks several years hence is $4,000 each, for quantities of 25 or more. If a backup clock is used the cost if $8,000 per node.

The present cost of a crystal oscillator is $500 (without quantity discount), including a varicap control, internal trimmer, stability equal to 1×10^{8} per wk after 24 hour warm up period and a volume of 2 x 2 x 5 inches. The cost of the summing device may be $100 if an operational amplifier is used. Thus frequency averaging costs include 2 oscillators, a summing device and the required control circuitry. The total projected cost is about $1,000. This figure relates only to the nodal cost and does not include the crystal oscillators used by individual subscribers. Further work is contemplated to determine the Combat/Cost Effectiveness of each of these two areas.

APPLICATION OF PRECISE TIME-FREQUENCY TECHNOLOGY IN MULTIFUNCTION SYSTEMS

Theodore C. Viars
Avionics Laboratory, USAECOM
Ft. Monmouth, New Jersey

This paper discusses concepts for a multifunction avionics system based on the use of accurate synchronized clocks at each terminal of the system. An experimental program to identify technological barriers and aid in their resolution will be described. This experimental activity involves development and utilization of a facility known as the Time Ordered Techniques Experimental System (TOTES).

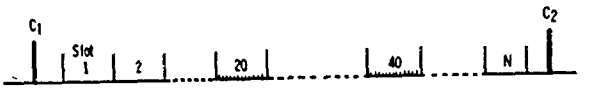
Basic Concept

The multifunction system is basically a time multiplexed digital data link which uses a common wide band frequency channel. It is similar to conventional time division data links in that each user is assigned his own time slot during which he performs active (mostly radiating) functions which provide information to the passive members of the system during that time slot. Each user takes his turn at being the active member. The multifunction system differs from other time division data links in the following respects:

1. A systematic effort is made to fully utilize the radio frequency channel to perform other services, in addition to communication, which are required to provide more effective Army avionic systems. These additional services generally relate to air traffic regulation in both the en-route and terminal situations, and involve functions which can be derived from accurate range measurements between key elements of the dynamic air traffic environment.

2. The time synchronization of the RF link is maintained more precisely than would be required for data transmission alone, in order to provide the accurate ranging information referred to above.

Figure 1 shows a representative time slot structure. Each user is alloted a 5 millisecond slot, which might be budgeted as shown. C_1, C_2..... mark the beginning of new cycles. With the parameters assumed in Figure 1, a system which has . . users would have a cycle time of 2 seconds. For functions requiring a faster data rate, special time slots can be programmed at regular intervals as shown at slots 20 & 40. Examples of such special functions are: proximity warning for air-to-air collision avoidance, station-keeping, and landing approach aids.



TIME BUDGET:

Range: Sync	1.5 Ms
Data (100 @ 100 KB)	1.0
Integral Relay	1.0
Propagation (3 X 100 mi)	1.5
TOTAL	5.0 Ms

Fig 1 - Typical Time Slot Structure

Clock Errors

The maximum timing errors that the system can tolerate vary with the particular function to be performed, and with such factors as data rates and accuracies to be achieved for a given function. Figure 2 shows the degree of timing precision necessary for various functions, where the time scale represents the maximum synchronization error that can be tolerated for the accuracies shown. For example, for functions such as station-keeping which may require range information to the order of $\pm$ 10 to 100 feet accuracy, the system must be synchronized to the order of 10 to 100 nanoseconds. (One-way propagation time is assumed to be about 1 foot per nanosecond). These accuracies must be maintained for the duration of a mission which might last for several hours. They are currently achievable, both by atomic frequency standards, and by less accurate crystal standards in systems where it is practical to resynchronize the remote station clocks to a common system time reference over a wide-band RF channel.

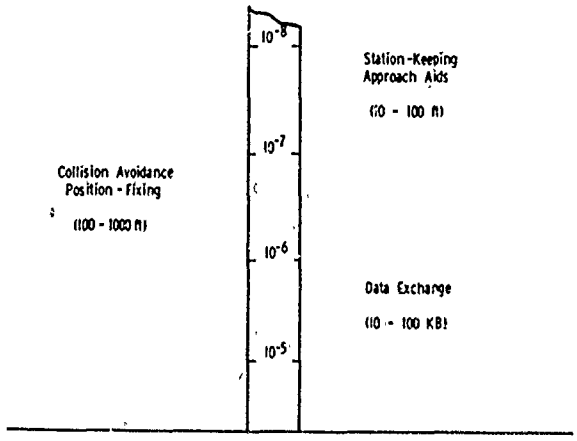


Fig 2 - Allowable Timing Errors (SEC)

Candidate System Approaches

The multifunction system postulated here consists of a time synchronized UHF or SHF radio network, under control of a master station. Since a large portion of the information handled will be related to air traffic regulation, it is logical to integrate the ground-air system via a semi-automatic air traffic control terminal which is optimized to function as an integral part of the system. Some of the services that the system would provide, and possible technical approaches for their implementation, are discussed below.

1. Identification & Collision Avoidance

Identification of each user is implicit by the prearranged order of the time slots. Since each station knows when an active station is scheduled to transmit a ranging pulse, range information is derived by all participants from the single range transmission. By integration of successive range pulses, or by extracting doppler information, range rate is available at each station. Altitude information can be transmitted during the data interval of a time slot. With the information now available: range, range-rate, and altitude, air-to-air collision avoidance can be automated.

2. Aircraft Location

Although accurate range information ($\pm$ 75 to 100 ft.) can be obtained by one way ranging, further information is required to establish aircraft position. Following are some candidate techniques for accomplishing this:

a. Multiple Ranging

This approach would exploit the accurate ranging capability of the system by having secondary beacon sites, located at known points on a baseline. The propagation time of aircraft ranging pulses received at the beacons could be transmitted to the master station over a data link of restricted bandwidth. There the aircraft position would be computed (by triangulation) and utilized in traffic regulation and other operational functions. If required, the aircraft's position could be conveyed back to the aircraft over a data link.

Use of beacons on elevated terrain points in the vicinity of a flight operations center also has the advantage of extending the range of the system to low flying aircraft beyond the terrain mask which would otherwise be screened from the operations center.

b. Directional Receiving Antenna

The azimuth of received signals can be determined by an electronically - scanned antenna at the ground terminal. Errors of a few degrees are presently achievable with this technique. Although this method lacks the accuracy of other techniques, position information is available from a single station.

c. Rotating Beacon at the Flight Operations Center (FOC)

A bearing measurement system can be provided by transmitting an omnidirectional time reference and a rotating directional continuous signal such as the techniques employed in VOR and TACAN. By using reasonably high directivity in the horizontal plane and a vertical fan beam, it should be possible to achieve accuracies on the order of 1 or 2 degrees with relatively simple equipment.

d. Range Δt

This technique would employ a master and two or three beacon stations set up on a short baseline with a spacing of about 600 meters between terminals. By transmitting pulses with known time delays from these stations, hyperbolic position lines similar to LORAN can be established. For aircraft at ranges greater than about ten times the baseline, the hyperbolic lines are nearly straight lines as shown in Figure 3. This means that a given difference in arrival time (Δt) of a set of pulses received at the aircraft can be converted directly into bearing angle without elaborate computation. The aircraft now has bearing information to an accuracy comparable to VOR to supplement the accurate range information available from the time ordered system. Note that this configuration is pilot oriented. It is not difficult to visualize a system in which primary

emphasis at the ground terminals is shifted to receiving pulses transmitted from the aircraft, and measuring Δt (and hence bearing angle) at the master station.

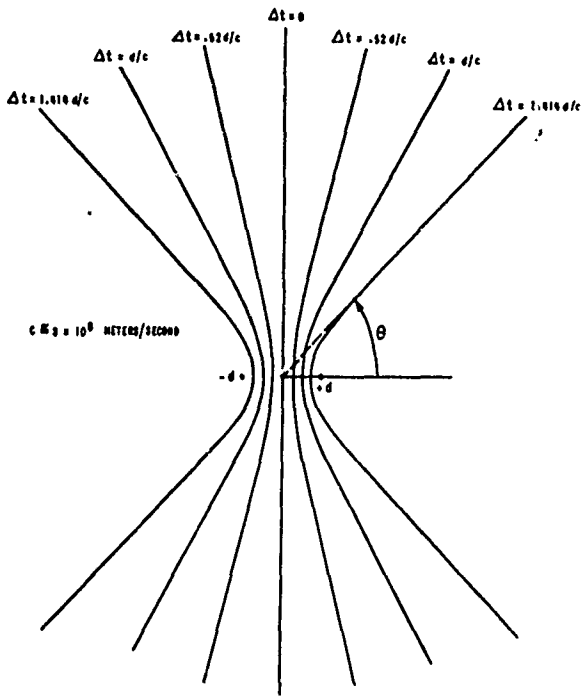


Fig 3 - Typical Time Difference Hyperbolic Lines of Position

Figure 4 shows the relationship of bearing error to time measurement accuracy for the range Δt technique.

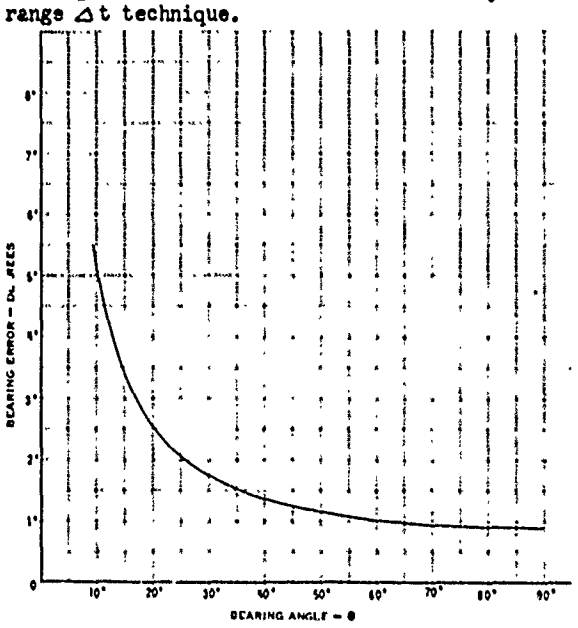


Fig 4 - Bearing Error vs. Bearing Angle (Time Error 30 Ns; 600 Meter Baseline)

3. Localizer

A variation of the range Δt technique outlined above could be employed to provide a more accurate approach aid. Since the hyperbolic lines of position become a straight line on the perpendicular bisector of a line connecting two transmitting stations, it is possible to define this straight line (which could be an extension of the runway centerline) through the region of fixed base stations at an airfield by a constant time difference.

The Time Ordered Techniques Experimental System (TOTES)

The TOTES is part of an exploratory development program to examine the feasibility of applying time-frequency and related advanced techniques in multifunction avionics systems which might become operational in the post-1975 time period. The TOTES provides the physical means for accomplishing the following:

Environmental and techniques investigation.
System development, feasibility demonstration and system evaluation.

The TOTES hardware is being designed for maximum flexibility to enhance its value as a research tool. The primary system consists of three radio frequency (RF) modules with associated precision time references, system timing and processing logic, and data recording equipment. One of the RF modules located at the center, along with the processing and recording equipment, comprises the master station. A remote base station will be located on a tower, and will serve primarily as the second point of a baseline position-location system and as a relay station. Another RF module is located in a shop van, thus permitting a variety of environmental situations to be examined.

TOTES Terminal

A TOTES terminal shown in Figure 5 is comprised of several hardware modules organized along functional lines. Two general types of terminals, master station and slave station, are required. Both terminals have the same type and number of modules; in some instances these modules have minor hardware (but significant functional) differences. These modules are:

1. Radio Frequency Module:

Special purpose "L" band transceiver employing biphase, psuedo-random pulse modulation. Radio transmissions are coherent with self-contained or external precision oscillators. An alternate mode of operation de-activates the psuedo-noise matched filter mode and uses straight envelope detection. This mode is particularly useful in multipath clutter investigations.

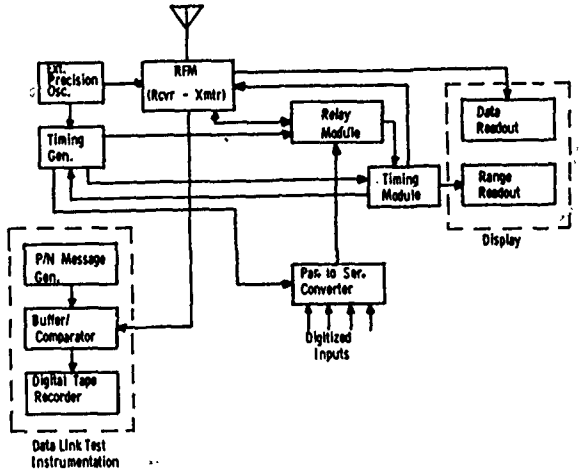


Fig 5 - TOTES Master Station

2. Timing Generator:

Generates a number of programmable timing reference pulses during each time slot; and establishes time slot ordering and identities.

3. Time Ordered System Module:

Provides a capability to accomplish inter-station timing synchronization, measure one-way range propagation delays of received signals, and to control various RFM functions.

4. Integral Relay Module:

Performs data buffering and address functions to establish relays through intermediate slave stations. Relay assignments are transmitted, along with other digital data, in the regular time slot of each affected slave station.

5. Parallel to Serial Converter:

Serializes parallel digital data inputs into a specified format, at a rate established by the Timing Generator.

6. Data/Range Display:

Displays digital data and one-way range along with station identity upon request.

7. Data Link Test Instrumentation:

The data link test instrumentation consists of transmitting and receiving data equipments located at the TOTES terminals between which the digital data exchange capability is to be measured. The transmitting equipment is a programmable length digital sequence and sync preamble generator, which is shnchronized via the first bit of the received sequence, or by the sync preamble. A buffer/comparator compares each bit of the locally generated code with the received code, detects bit errors, and processes, stores, and formats the data for subsequent recording on an incremental tape recorder. Data on the tape can be transferred directly to the ECOM B-5500 digital computer for processing.

Preliminary TOTES Tests

Figures 6 through 9 show the waveforms obtained in preliminary tests of the TOTES RF and timing modules, using conventional pulse code modulation. (Horizontal sweep speed was ½ microsecond per division). Figure 6 shows a 1 microsecond pulse received over an 8.2 mile ground-to-ground path. This photograph represents the integration of about 300 separate pulses. The pulse-to-pulse variation is indicated by the broad fuzzy line at the top of the waveform. This may be due to variations in multipath effects on the individual pulses.



Fig. 6 - 1 Microsecond Pulse Width (Composite of Many Pulses)

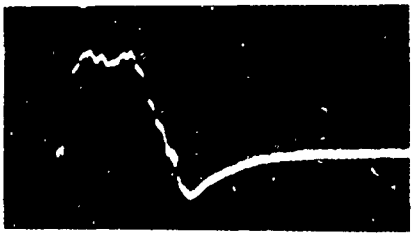


Fig 7 - Single Trace

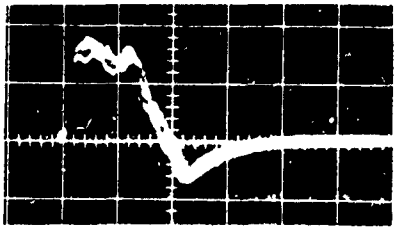


Fig 8 - Composite of 4 Traces

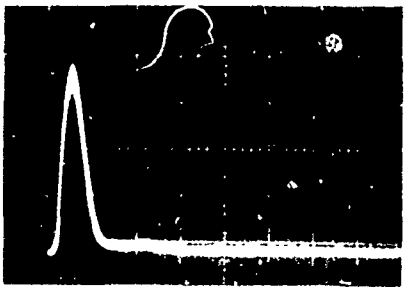


Fig 9 - 250 Nanosecond Pulse

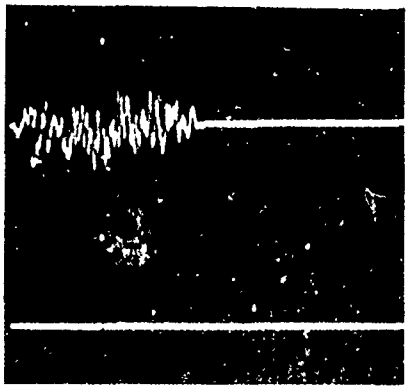


Fig 10 - Output of P/R Decoder (Top Trace) Schmidt Trigger (Bottom Trace)

Figure 7 shows the received waveform for a single pulse; Figure 8 is the composite for 4 pulses. These latter two figures are for a path length of only a few hundred feet, where there was not a direct line of sight path between transmitter and receiver. In Figure 9 the pulse width has been reduced to 250 nanoseconds; transmission path length is about 1400 feet.

Figure 10 shows the output of the matched filter for the pseudo-random mode of operation (top trace). The sweep speed was 5 microsecond per division. The bottom trace shows output of a very fast rise time Schmidt trigger (integral to the TOTES receivers) which was actuated by the decoded sequence. Using the matched filter mode, we expect to achieve a range resolution of 50 to 100 feet.

Conclusions

Synchronized time-frequency technology offers attractive possibilities for developing a multi-function system for support of Army aviation operations in the post-1975 time period.

Advantages of such a system include:

1. Large decrease in weight and space of on-board electronic equipment by sharing such items as frequency synthesizers, receiver-transmitter units, digital processors, antennas and power supplies.

2. Implementation of additional functions which are urgently needed, such as air to air collision avoidance.

3. Increased information handling capacity for Army aviation, required to cope with the demands for rapid response to changing tactical situations in a congested aerial environment.

4. Decreased logistics and maintenance requirements for Army avionics equipment, by reducing the number of different equipment boxes, implementing automatic fault-isolation procedures facilitated by use of digital techniques, and use of replacable functional modules.

Technological barriers which affect feasibility of such system include multipath clutter, and line-of-sight limitations of the microwave frequencies at which the wide-band time ordered system must operate. Problems of more subjective nature involve such factors as system redundancy and system evolution. The Avionics Laboratory has initiated exploratory development activity which will provide answers to some of these questions.

Acknowledgements

Basic ideas for the positioning schemes discussed, and preliminary error analyses relative to their performance, are due to work performed by Cornell Aeronautical Laboratory under Avionics Laboratory contract. The TOTES concepts and implementation to date are the result of joint effort by the Advanced Studies Team, Ground Systems Technical Area of the Avionics Laboratory, with technical support by the Electronic Components Laboratory of USAECOM and Cornell Aeronautical Laboratory.

References

1. Equipment Parameters Required for Army Tactical Airspace Regulation System; by C.W. Wightman, Cornell Aeronautical Laboratory; May 67.

2. Radio Frequency Modules for TOTES Radio Frequency Modules; by E.H. Eckert, Cornell Aeronautical Laboratory; July 1968.

FREQUENCY CONTROL FOR TACTICAL NET SSB EQUIPMENT

Owen P. Layden
Electronic Components Laboratory
U. S. Army Electronics Command

Summary

In WW II, the Korean conflict and later Army tactical net radio communication systems were based on the use of AM and FM radio equipment. These equipments had certain requirements for the frequency control systems which were met by the .005% type crystal used in crystal saving circuitry to generate many signals. In the past number of years the Army has been developing suppressed carrier SSB radio equipment for the HF and lower VHF ranges for tactical net use. This report concerns tactical net, or point to point communication whereas the Mallard system, presented earlier, is tactical trunk communications. In the actual tactical situation there is an interface between these two separate systems. The voice intelligibility of these SSB equipments depends on the frequency difference between the transmitter and receiver and since a carrier is not transmitted this has put stringent frequency stability requirements on the reference oscillators in the transmitter and receiver. In the future these types of radio equipments will have the capability of secure voice communication which puts even more stringent frequency control requirements on the frequency control system.

Discussion

Figure 1 summarizes present and future frequency control requirements for SSB equipment. For present systems the frequency difference allowed between equipments for normal voice communication is 40 Hz which is 20 Hz frequency stability for each equipment since there is no carrier transmitted between transmitter and receiver. This overall stability must be met over the temperature range and recalibration time requirements as shown; -40°C to +75°C and 26 weeks. For HF equipment using worst case frequency, 30 MHz, the overall tolerance is 6.7×10^{-7}. For VHF at 76 MHz the tolerance is 2.6×10^{-7}. These figures have been met or closely approached by the use of oven controlled oscillators and temperature compensated oscillators utilizing HC-27 crystal units. Frequency synthesizers, controlled by the reference oscillator, with channel steps of 1 kHz and 100 Hz have been used. The next generation of equipments will have the capability of secure voice communication which upgrades the frequency control requirements on the reference oscillator as shown in the future column.

In order to have adequate secure voice communication with SSB equipments it is necessary that the transmitter and receiver frequency difference be no more than 5 Hz which allows ± 2.5 Hz per equipment. This total frequency stability of ± 2.5 Hz can be divided broadly into two requirements (1) as a function of temperature and (2) as a function of recalibration time. The temperature range is -40°C to +75°C as before but the recalibration time period is 120 weeks or 2.3 years. Time between equipment calibration is based on the practical aspects of field use. It is not feasible for these types of equipment to be recalibrated often. The usual time period for frequency recalibration has been 26 weeks but the next equipment generation requires that no maintenance, therefore no frequency recalibration, be performed for 120 weeks. The frequency tolerance for SSB secure voice using the 2.5 Hz figure is 8.3×10^{-8} for HF and 3.3×10^{-8} for VHF.

These future requirements are put in graph form on Figure 2. This figure shows the relationship between aging/week and frequency-temperature stability based on a $\Delta f = 2.5$ Hz per equipment and a time of 120 weeks. The total required stability is the addition of the frequency-temperature stability and aging. To meet the HF stability requirement, 8.3×10^{-8}, the combination of aging and temperature stability must fall on or below the HF line. For instance, with aging of 1×10^{-10}/week the oscillator temperature stability would have to be 7×10^{-8} or better. This is shown as point A. Likewise for worse aging, 5×10^{-10}/wk, the temperature stability would have to be improved to 2×10^{-8}, shown as point B. The trade-off between aging and frequency-temperature stability should be noted. By sliding down the requirement line as aging improves worse temperature stability can be tolerated and vice versa. From the user point of view, it makes no difference what the aging or the temperature stability is as long as the combination of these two result in meeting the 8.3×10^{-8} tolerance. For VHF equipment, up to 76 MHz, the VHF line is the limiting condition. Points A and B will be referred to later on Fig. 5 as to the aging needed, that is point A corresponds to aging of 1×10^{-10}/wk and point B to 5×10^{-10}/wk. Also note that the temperature stability for point B will meet VHF requirements if the aging is improved to 1×10^{-10}/wk. This figure shows the need for aging figures of 1 to 5×10^{-10}/wk coupled with oscillator frequency-temperature stability of parts in 10^8.

Figure 3 shows these same requirement lines drawn to a different scale and they appear in the lower left hand corner. Present state of the art is shown as point C which represents aging of 50×10^{-10}/wk and frequency temperature stability of 50×10^{-8} for a TCXO. Experimental units have exhibited aging of 10×10^{-10}/wk and temperature stability of 20×10^{-8} as shown by point D. Even these experimental units are not close to meeting the requirements.

To show the types and relative size of frequency control components discussed in this report Fig. 4 gives some examples. The crystals shown are the glass enclosed HC-27, being used extensively at this time and the newer development, the coldweld crystal. The TCXO, on the left, and oven-oscillator on the right, make use of crystals of the HC-27 type. The synthesizer shown, in the middle, has as an integral part, a TCXO used as the reference oscillator. All output signals have the frequency stability of this reference oscillator.

Low Power Crystal Oscillator

In the area of low power crystal oscillators, Fig. 5 gives general information on frequency-temperature stability as a function of power consumption for three types of oscillators and shows their relation to the needs in this area. Point A and B requirements, taken from Fig. 2, are shown on the right in this figure. The cross hatched areas are the requirements based on a maximum acceptable average power consumption of 250 mW, at the lowest ambient temperature. The point A cross hatched area is the HF requirement if aging is 1×10^{-10}/wk. That is, with this aging the requirement is met if the frequency-temperature stability is 7×10^{-8} or better and power consumption is 250 mW or less. The point B cross hatched area is based on HF requirement with aging of 5×10^{-10}/wk or the VHF requirement with aging of 1×10^{-10}/wk. In both cases requirements are met with a frequency-temperature stability of 2×10^{-8} or better and, again, power consumption of 250 mW or less. In all cases the requirement is met if the frequency-temperature stability vs. power consumption falls within the cross hatched areas.

The TCXO point has resulted from the work done by CTS Knights under contract to ECOM and shows stability of 20×10^{-8} and power consumption of about 50 mW. The TCXO-oven combination work was carried on under previous contracts with USAECOM. The approach was to put a TCXO into an oven that operates at 0°C so that the TCXO never sees a temperature below 0°C. A stability of 10×10^{-8} with power consumption of 175 mW resulted. For the oven-oscillator the point is shown that it meets frequency-temperature stability requirements of 2×10^{-8} but its power consumption is off scale. This figure reinforces the point that the present experimental state of of the art does not approach what is needed. The TCXO has the low power capability but not the frequency stability. At the other end the oven-oscillator exhibits the needed frequency stability but the power consumption is too high. The further development of the TCXO-oven combination has a chance of meeting point A requirements. Other factors not shown on the graph are cost, it must be low, size, it must be small, and warmup time. As to warmup time it is required that the oscillator be within tolerance within one minute from the lowest ambient temperature, -40°C. Higher power, as high as ten watts, is allowed for this short period of time. The TCXO can meet this, the oven cannot. This requirement is quite important for tactical net equipment and will be difficult to meet. It is necessary to apply considerable heat to the crystal quickly which sets up temperature gradients. This causes frequency changes and out of tolerance operation. Present information on oven units is warmup time of a number of minutes is needed.

To obtain an oscillator with a frequency-temperature stability of 2×10^{-8}, point B requirement, which is needed for HF (with 5×10^{-10}/wk aging) and VHF (with 1×10^{-10}/wk aging), use of an oven appears to be the best approach. The problem is to reduce the power consumption down to less than 250 mW. Present design normally has the crystal and circuitry as separate packages which are housed in the temperature controlled chamber. This results in large volume and high power. An approach that appears to have high promise is to house the crystal, temperature sensor, critical circuitry and heater in one evacuated enclosure. With a good vacuum, 10^{-7} to 10^{-8}, maximum heater power in the order of 50 to 100 mW should result. Uncontrolled heat producing circuitry, voltage regulators, isolation amplifiers would be housed in a separate package. To get fast on-frequency warmup it will be necessary to have good thermal contact between the heater and crystal perhaps by depositing heater material on the quartz or by having the heater and crystal very close to each other. The effect on crystal aging of having components in the same evacuated holder as the crystal or depositing heater material on the quartz must be determined. There may be feasible tradeoffs between oven power and crystal aging.

Crystal

The crystal used in this oven-oscillator or TCXO must be of the low aging type or else the good frequency-temperature stability is wasted. Figure 6 gives information on the extent of the low aging crystal and/or oscillator problem. The straight lines show the requirement on aging vs. time for a rate of 5×10^{-10}/wk and 1×10^{-10}/wk for a 120 wk period. An actual aging curve for a temperature compensated crystal oscillator using an HC-27 crystal is shown. This oscillator fails to meet the requirement by about a factor of two. This oscillator is not the best of the group but is reasonably typical of the oscillators under aging.

The crystal aging curve shown is an actual crystal and is also reasonably typical of a number of coldweld crystals. This coldweld sealed crystal was sealed in a vacuum of about 10^{-5}, after a high temperature bakeout at 300°C for two hours. It misses the requirement by about a factor of two and its shape is typical of the group of crystals tested, that is initial aging upward which soon becomes negative and stays negative. The crystal shown is not the best of the group but is what can typically be expected from the group. The crystals and oscillators were held at 60°C throughout the tests.

Frequency Synthesizers

All of the preceding discussion has been concerned with the reference oscillator, circuit and crystal, which is needed to stabilize the output frequency. This reference oscillator is used to drive a frequency synthesizer which generates the many signals needed for the radio equipment. The reference oscillator could be any frequency, however, 5 MHz has become quite common for these applications. This signal is used to generate all output signals needed for the HF, 2-30 MHz, equipment, however, the synthesizer output need not and usually does not cover this range but a range depending on the radio set design.

Figure 7 gives a general representation of the present art in HF and VHF synthesizers for SSB equipment and a view of the future needs in this area. Presently in these small synthesizers spurious signals of 40-60 dB and worse exist in the output signal. An upgrading to the 80 dB level is needed and to the 100 dB level is desirable. Phase jitter, normally not a requirement, is in the range 5-15° now. With future secure voice radio sets, since digital transmission is the means of security, phase jitter of no more than 2 to 5° measured in a 13 millisecond time period is a necessity. Presently, noise figures, measured in a 3 kHz B.W., 100 kHz from the carrier, are in the 90 dB range. Noise must be decreased to the 125 dB level in order to help solve the problem of a transmitter densensitizing nearby receivers. It is necessary to decrease the size and power consumption of these units and it is highly desirable that the figures be as small as possible consistent with performance. Synthesizer design is now based on a particular radio set development and on the particular engineers on the job. Future needs call for modular type design so that modules used for narrow channel spacing SSB equipment can be used to the maximum extent for VHF FM and UHF AM radio systems.

Frequency synthesizer techniques for Army tactical net radio equipment in the phast few years have been of the indirect type. These synthesizers have used spectrum generators along with multi-crystal oscillators in a phase locked loop arrangement to generate the needed signals. Indirect type not using crystal oscillators in this manner have also been developed. Equipment now in development or just developed with small or large channel steps, are utilizing indirect synthesizer techniques with digital counter divide by N circuitry in a phase lock loop (s). Generally speaking these latter synthesizers can be made small but not low power. The high power results to a large extent by the digital counter circuitry so one approach to decreased power is the development of low power divider circuits. Another possibility is the combination of techniques us spectrum generation to obtain the big channel steps and divide by N circuitry for the smaller steps.

Conclusion

The present state of the art in frequency control for tactical net SSB equipment has been covered along with the future needs in this area. The need for crystal aging of 1 to 5 x 10^{-10}/wk and low power crystal oscillator frequency-temperature stability of parts in 10^8 have been established. In addition the needs in the frequency synthesizer area have been shown. Possible approaches to meeting the future requirements have been outlined.

REFERENCE OSCILLATOR REQUIREMENTS

	Present SSB Voice	Future SSB Secure Voice
Δf (Total)	± 40 Hz	± 5 Hz
Δf/Equipment	± 20 Hz	± 2.5 Hz
$\frac{\Delta f}{f}$ (HF, 30 MHz)	6.7 x 10^{-7}	8.3 x 10^{-8}
$\frac{\Delta f}{f}$ (VHF, 76 MHz)	2.6 x 10^{-7}	3.3 x 10^{-8}
Temperature Range	-40°C → +75°C	-40° → + 75°C
Time	26 weeks	120 weeks

FIGURE 1

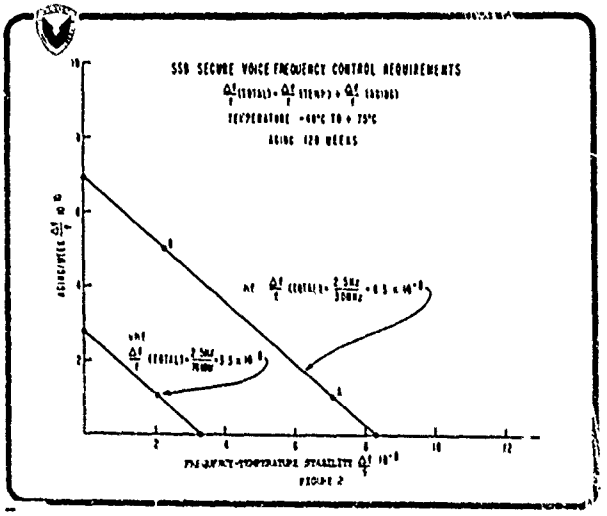
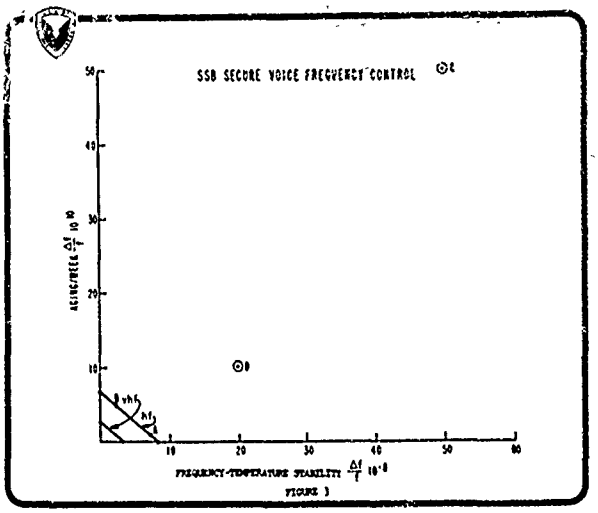
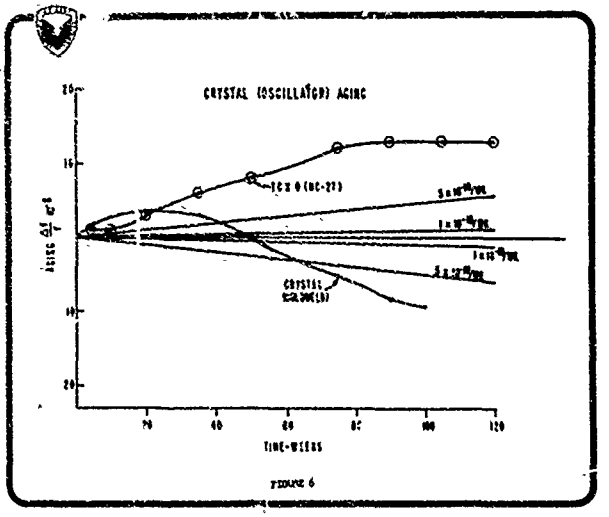


FIGURE 2





Frequency Control Components

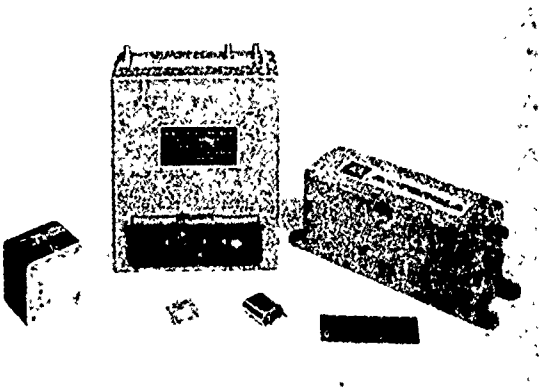


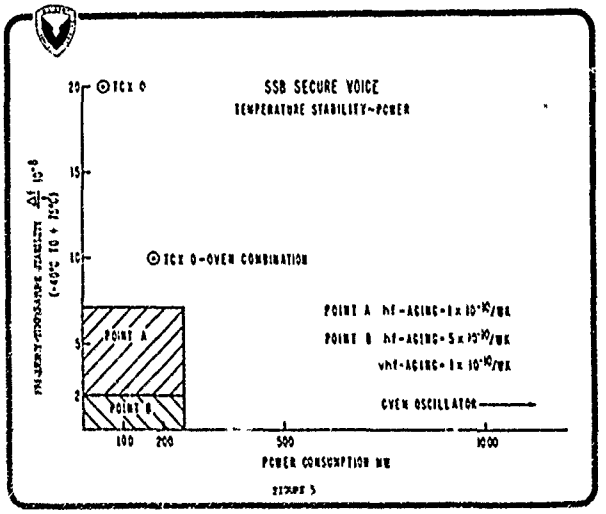
Figure 2

FREQUENCY SYNTHESIZER REQUIREMENTS

	Present	Future
Frequency Range	Use In HF & VHF Radio	Same
Channel Steps	100 Hz	Same
Spurious Signals	40 - 60 dB	80 - 100 dB
Phase Jitter	5 - 15°	2 - 5°
Noise	90 dB *	125 dB*
Power	1 - 4 watts	Low
Size	30 - 40 cu. in.	Small
Design	Individual	Modular

*Measured in a 3 kHz bandwidth, 100 kHz from carrier.

FIGURE 7



INTERNATIONAL CO-ORDINATION OF RADIO TIME SIGNAL EMISSIONS

Humphry Smith
Royal Greenwich Observatory
Herstmonceux Castle,
Hailsham, Sussex.

Universal Time (UT) is the term used in scientific contexts in place of the name Greenwich Mean Time (GMT) used formerly in all fields and still widely employed, especially in navigation and general maritime usage, and in communications (for example, in the Radio Regulations of the International Telecommunication Union). UT is related to the angular position of the rotating Earth with respect to the Sun. In practice, the rotation of the Earth is measured, not by direct observations of the Sun, but by those of star transits, which are more precise. From the star observations sidereal time is determined, and this is converted into mean solar time. Mean solar time observed on the Greenwich meridian (or corrected to Greenwich by the application of a longitude correction) is known as UT.

In the process of time determination from astronomical observations, corrections are applied for a number of well-known causes of variation:- precession, nutation, parallax, proper motions of stars and aberration. Since 1948 the published results of the Greenwich Time Service have also been corrected for the effects of polar variation (a wobble of the Earth on its axis). Since 1953 corrections for polar variation and for a seasonal fluctuation in the rate of rotation of the Earth (due to world-wide meteorological changes) have been applied to all UK time signals. This Greenwich smoothed system of UT was called Provisional Uniform Time (PUT), and in 1955 it was used as the basis for the initial calibration of the caesium atomic beam standard developed at the National Physical Laboratory.

From January 1956, by international agreement, three types of UT were recognised: UT0 based directly on the stellar observations; UT1 incorporating corrections for the effects of polar variation; UT2 incorporating an additional correction for the estimated effect of the seasonal fluctuation. The corrections to be applied are adopted and made available by the Bureau International de l'Heure. All time-keeping observatories were requested to adhere to UT2 for the control of radio time signals.

While the adoption of UT2 removed two important variations of a periodic nature, there remain a suspected progressive retardation (of the order of one [illegible] second per century) and irregular changes which, by their very nature, are inherently unpredictable. It is the cumulative effect of these irregular changes which inhibit the use in many current applications of a time scale based directly on the rotation of the Earth.

The superior stability of quartz oscillators led, as long ago as 1944, to a new method of control of UK time signals. The daily rate of the signals was kept as constant as possible (in terms of a mean of the quartz clocks then available) and precise step adjustments were made, when necessary, on a Wednesday, to maintain agreement with UT. The successful realisation of an atomic standard in 1955 made it possible to define a uniform time scale with greater precision, and the use of modern astronomical instruments, such as the PZT and Danjon astrolabe, reduced the observational scatter in the determination of UT.

The frequency of the caesium beam atomic standard was first measured in terms of the second of PUT (the antecedent of UT2), but by 1958 the frequency had been measured in terms of the second of Ephemeris Time (ET). Ephemeris Time is a uniform astronomical time which may be regarded as being based on the year (the period of the annual motion of the Earth around the Sun) instead of on the diurnal period of rotation of the Earth on its axis. In practice it is determined from observations of the motion of the Moon. It is free from any of the variations associated with the rotation of the Earth and is, therefore, directly comparable with atomic time if we ignore the possibility of very small relativistic effects. For all practical purposes, ET and AT may be regarded as equivalent in rate, and differing by a fixed interval in epoch.

A clear need therefore arose for a service of radio time signals which would provide a constant rate (based on AT) and time markers which would remain within an acceptable limit of UT. The system, now known as UTC, was first proposed by Essen for the control of the standard frequency emission MSF, Rugby, England. In 1959 it was announced that the service would be based on the nominal frequency of caesium in terms of the second of ET, but that the actual frequency employed would be offset by a prearranged amount to bring the time signals into satisfactory agreement with UT2. The amount of the offset remained unchanged for a calendar year, but was revised annually, as necessary, in the light of current astronomical observations. If, during the course of a year, the departure of the signals from UT2 exceeded a stated limit, a step adjustment was made to the radiated time signals.

In 1960 the MSF and GBR signals were co-ordinated on the UTC system, and in 1961 the time services of USA, Canada and UK were co-ordinated. Changes in the offset and the dates of step adjustments were a matter of

informal agreement between the USNO and the RGO. Other administrations decided to keep in step with the co-ordinated stations, and shortly afterwards, UTC was formally recommended for universal adoption (IAU 1961 and CCIR 1963) with the Bureau International de l'Heure acting as the co-ordinating authority. Minor modifications have been made on the advice of IAU, URSI, CCIR, but the system has remained basically the same. In general, it has succeeded in meeting all essential needs. A careful re-appraisal of the system has, however, been called for in the light of the adoption, in 1967, by the Thirteenth General Conference of Weights and Measures, of an atomic definition of the second as the unit of time in the International System of Units (SI). As an immediate consequence of this decision, there is a strong demand that the time pulses should be at intervals of exactly one (S.I.) second; this implies the abolition of the offset in the equipment generating the pulses. In general, the same oscillator controls both the pulses and the carrier frequencies. In any case, there is an equally strong desire to abandon the offset in the carrier frequencies, so that frequency is measured directly in terms of the SI second without any corrections. At present UTC is maintained within 0.1 second of UT2; step adjustments are exactly 0.1 second, and are made on the first day of a month (the date having been announced at least a month in advance); the offset is in multiples of 50×10^{-10}. A proposal has been made by CCIR Study Group VII that the limits should be extended to approximately 0.5 second (in practice, under the CCIR scheme, the limit would be of the order of 0.7 second); there should be no offset; step adjustments should be made on the first day of a month (to be announced) and should be of exactly 1 second.

In order to discuss the implications of the adoption of these proposals, it is necessary to consider what divergencies may be expected between the signals (with a rate corresponding to AT, which, as previously mentioned, corresponds to ET) and UT2. In the year 1900 the length of a mean solar day (that is, a day of 86,400 UT seconds) was very close to 86,400 SI seconds. The day is now some 0.0025 seconds longer. From an examination of the records of the past changes of UT with respect to ET, D. H. Sadler has deduced the following figures as the reasonable limits to the variations which might be expected in the near future (say up to the year 2000).

Length of the mean solar day 86,400 from -0.0015 to +0.0075 SI secs.

Predictability of UT2	3 months	±0.1 seconds
	6 months	±0.4 seconds
	1 year	±0.9 seconds
	2 years	±2.0 seconds

The probable error of prediction is, of course, smaller, but in this instance, it is the possible maximum errors which are of importance. You may have heard of the man who was drowned fording a river of <u>average</u> depth of four feet: the prudent navigator does not hazard his craft on probabilities.

It is important to notice that the present limit of 0.1 second in the departure between UT2 and UTC corresponds to the human limit of time discrimination. There is, therefore, no hardship to the navigator, surveyor or satellite observer using eye-and-ear methods of comparison.

Any increase in the permitted departure will increase the number of users who will need to apply corrections. The relevant corrections may be (i) promulgated beforehand (subject to the prediction limits quoted); (ii) broadcast in coded form in the radio time signal emissions or (iii) communicated subsequently. Since for some applications it is not practicable to make the corrections in arrear, the CCIR scheme proposed the use of coded corrections. It may be noted in passing that if a completely satisfactory method of giving coded corrections could be devised, the limits of departure from UT2 might be further widened, but this would necessitate the application of corrections by many users who are now satisfied with the signals as radiated. Two objections to the use of coded corrections have been pointed out. It is the field user and the navigator who will be required to apply corrections, and they will not have available the elaborate equipment of the laboratory-borne scientist, or engineer. Under poor radio propagation conditions, the coded message may not be 100% readable, and there is a risk of error which could prove serious. The second objection is that many users do not listen to the major radio time signal emissions, but make use of "talking-clocks" or time signals radiated in broadcast radio programmes. It is difficult to see how coded corrections could be applied to all time signal emissions.

On the question of step adjustments, there is again some division of opinion. For co-ordination with world-wide networks of radio navigational systems there is an advantage in making the steps exactly one second: this involves no physical readjustment of the system, but merely a re-numbering of the pulses. The date of a step adjustment would be announced, say, a month in advance. On the other hand, some users prefer fixed dates for adjustments. This minimises the importance of the prior notification (which may fail to arrive) but, unless the limits are to be further extended, necessitates some flexibility in the amount of the adjustment: n x 0.2 second has been suggested. In discussions at the Bureau International des Poids et Mesures the possibility of two co-existent systems was regarded as acceptable. The CCIR Study Group VII meeting at Boulder was strongly in favour of one system.

Any modification of UTC which retains the

present feature of step adjustments is regarded, by many people concerned, as a step towards the eventual adoption of an atomic time scale for all purposes, (with, perhaps, very rare adjustments to avoid the accumulation of a large departure from UT2). Such a drastic change would have wide repercussions and, indeed, might well require legislative action in the definition of legal and civil time. The feasibility of using AT in astronomical survey and navigation is being investigated. Although AT is not exactly identical with ET the independent variable of celestial mechanics, it is possible to calculate geocentric ephemerides of the Sun, Moon, planets and stars to high precision in terms of A.T. A (movable) meridian - the Atomic Meridian - is conceived as rotating round the Earth's axis, but independent of the Earth itself, at a uniform rate of one revolution in 86,400 seconds; all the operations and tabulations directed towards determining position on the Earth's surface can then be carried out in terms of A.T. relative to the Atomic Meridian. Observations enable the position of the Greenwich Meridian (the zero of longitude) to be obtained relative to the Atomic Meridian, and thus longitudes measured from the Atomic Meridian can be converted to true longitudes. Since the Earth is now rotating more slowly than in 1900 the Greenwich Meridian lags increasingly behind the Atomic Meridian; Greenwich (west) longitudes will thus always be less than Atomic (west) longitudes, and Greenwich Hour Angles will also always be less than Atomic Hour Angles (i.e. hour angles measured from the Atomic Meridian).

The national astronomical, navigational and surveying ephemerides can tabulate either right ascension (or S.H.A.) or Atomic Hour Angle, with declination, in terms of A.T.

One result of the enquiries that have been made concerning the acceptability of the CCIR scheme has been a careful reconsideration of the needs of the many users of radio time signals. The users of UT are engaged in a wide field of activities and there is no international organisation which is fully representative of their varied requirements. It has become increasingly apparent that the formulation of a simple universally satisfactory system is no easy task. It is also evident that the widest possible publicity must be given to the repercussions of any proposed scheme before any hasty action is undertaken. Is it necessary to insist upon one universal compromise system, or is it possible to contemplate the existence of two different systems each designed to meet a particular need? If coded corrections are to form an essential part of the scheme to be adopted, what form of coding would prove most reliable in practice? Many interested parties emphasise the urgency of reaching a decision. On the other hand, it would clearly be undesirable to adopt a system which would be subject to further lengthy discussion and successive modifications. There are very many users who, in answer to the question "what is the time", will demand an exact and unequivocal answer.

MEASUREMENT OF THE PIEZOELECTRIC COEFFICIENT OF QUARTZ USING THE FABRY PEROT DILATOMETER

Virgil E. Bottom
Department of Physics
McMurry College
Abilene Texas
79605

Summary

The piezoelectric strain coefficient d_{11} of alpha quartz has been studied by many investigators since the discovery of the effect in 1880. A marked lack of agreement exists among the values reported which range from 6.3 to 7.1 x 10^{-8} cm/statvolt [2.1 to 2.3 x 10^{-10} coul/newton]. Although the unweighted average of all the values reported is about 6.53 authorities have generally accepted the value 6.9 on the basis that the experimental errors tend to make the measured values too low. Curiously, the values tend to fall near the upper and lower ends of the range with relatively few values near the average. No relationship is apparent between the method of measurement and the result obtained.

The present measurements have been made on a single bar of quartz using a Fabry-Perot Dilatometer to measure the strain. The results of 37 independent measurements made over a period of several months yield the value d_{11} = 6.55 x 10^{-8} cm/statvolt in excellent agreement with the average of all previous measurements. The values show the usual tendency to fall near the upper and lower ends of the range with few near the average value. The average deviation is 0.18.

Further work is planned using specimens cut from a number of different crystals including synthetic quartz. Further refinements of the experiemntal technique are also planned.

Key Words (for information retrieval).
Quartz, Piezoelectricity, Dilatometer, Piezoelectric coefficient.

Introduction

The piezoelectric strain coefficient d is defined as

$$d = \frac{\partial P}{\partial X}_E = \frac{\partial x}{\partial E}_X$$

where P is the polarization, X is the stress, x is the strain, and E is the electric field intensity. In particular the coefficient d_{11} in alpha quartz may be defined either as

$$d_{11} = -\frac{\partial P_x}{\partial X_x} \qquad (1)$$

or

$$d_{11} = \frac{\partial x_x}{\partial E_x} \qquad (2)$$

The piezoelectric coefficient $d_{12} = -d_{11}$ may likewise be defined as either

$$\text{by } d_{12} = -\frac{\partial P_x}{\partial Y_y} = -d_{11} \qquad (3)$$

$$\text{or } d_{12} = \frac{\partial y_y}{\partial E_x} = -d_{11} \qquad (4)$$

where the subscripts denote the directions of the various quantities. Thus we have four relationships by means of which d_{11} can be determined. In this paper these are designated:

Method I	Direct-Longitudinal Eq. 1
Method II	Converse-Longitudinal Eq. 2
Method III	Direct-Transverse Eq. 3
Method IV	Converse-Longitudinal Eq. 4

Units Since the piezoelectric strain coefficient d may be expressed either as

$$\frac{\text{polarization}}{\text{stress}} \quad \text{or} \quad \frac{\text{strain}}{\text{field intensity}}$$

the value may be given either as coulombs newton or meters volt. In the older literature d is usally expressed as cm/statvolt although other units are sometimes found. The value in M.K.S. units may be converted to cgs esu by multiplying the value in M.K.S. units by 3 x 10^4.

Static Measurements An X-cut plate with electrodes covering the major surfaces is used in Methods I and II. In Method I pressure is applied in the thickness direction and the resulting charge is measured by a ballistic galvanometer, electrometer or some other suitable means. Brief descriptions of several of the experimental are given by Sosman (1). In Method II an electric field is applied in the X-direction and the resulting change of thickness is measured.

An X-cut bar is used in Methods III and IV. The bar is cut with its smallest dimension parallel to the X-axis and its longest dimension parallel to the Y-axis. In Method III a longitudinal stress is applied to the bar in the X-direction and the resulting polarization in the X-direction is measured by the same techniques used in Method I. In Method IV the electric field is applied in the X-direction and the resulting longitudinal strain in the Y-direction is measured.

Most of the measurements of d_{11} have been made using Methods I and III. Three determinations have been made using Method IV and only one

appears to have been made using Method II.

Methods I and II involve the measurement of mechanical stresses and electrical charges whereas Methods II and IV involve the measurement of the mechanical strains and electric fields. The mechanical stresses are very much easier to measure than the small strains produced by the piezoelectric effect thereby accounting for the preference shown for Methods I and III. On the other hand measurement of the small electrical charges developed in Methods I and III is quite difficult whereas the measurement of the applied field is relatively simple. The principle source of error in Methods I and III are those involved in measuring the values of the small electric charges.

The following Table I shows the values of d_{11} [in esu] measured by static methods reported in the literature. The methods used by the various experimenters are also indicated.

TABLE I

Source		d_{11}	Method
P. and J. Curie	(2)	6.32×10^{-8}	I, III
Czermak	(3)	6.3	I
Riecke and Voight	(4)	6.45	I
Pockels	(5)	6.27	III
Nachtikal	(6)	6.54	I
J. Curie	(7)	6.90	III
Veen	(8)	6.32	III
Hayashi	(9)	6.31	I
Röentgen	(10)	6.94	I, III
Tsi-Ze	(11)	6.4	II, IV
Dawson	(12) Ave.	6.13	I
Seidl	(13)	6.90	III
Günther	(14)	6.35	IV
Knol	(15)	6.83	I
Osterberg	(16)	6.22	IV
Clay and Karper	(17)	6.80	I
Langevin	(18)	7.10	III
Lüdy	(19)	6.54	III
Average		6.53×10^{-8}	

An analysis of the above values reveals a curious fact. The distribution of the experimental values does not approach a normal distribution curve. The average of the 18 values is 6.53. Cady (20) and Sosman (21) have adopted the value $d_{11} = 6.9$ on the basis that the experimental errors are most likely to tend toward low values. Nevertheless it is difficult to explain why no less than eight of the eighteen measurements should fall between 6.2 and 6.4, five between 6.8 and 7.0 while only four fall in the range 6.4 to 6.8.

Dynamic or Resonant Methods The piezoelectric coefficients can also be measured by utilizing the theoretical relationship between the parameters of the equivalent circuit of the piezoelectric resonator and the physical properties of the material and the dimensions. No reason is known for expecting to find any difference between the static and dynamic values other than a very small difference between the isothermal and adiabatic values. However measurements made by dynamic methods are generally considered less trustworthy than those made by static methods because of the difficulty of determining the distribution of the strain in the vibrating piezoid. Table II shows the values of d_{11} as measured by dynamic methods.

TABLE II

Andreff, et. al	(22)	6.51×10^{-8}
Fujimoto	(23)	6.1
Fréederickz	(24)	5.55
Nussbaumer	(25)	6.84
Van Dyke	(26)	6.70
Cook and Weissler	(27)	6.88
Mason	(28)	6.76
Bechman	(29)	6.93
Average		6.53×10^{-8}

It is doubtful if the average is significant but it is interesting to note that four of the seven values are above 6.7 and two are below 6.2 whereas only one falls in the large intermediate range.

Experimental

In the work described herein d_{11} has been measured using Method IV. The piezoid is a quartz bar having its shortest dimension in the X-direction and its longest dimension in the Y-direction The electric field is applied in the X-direction and the resulting strain in the Y-direction is measured using a Fabry-Perot Dilatometer. This device is capable of measuring displacements as small as 10 Å° directly in terms of the wavelength of the light used. No calibration of the apparatus is required. The complete theory and design of the instrument may be found in reference (36) and consequently only a short resumé will be given here.

Theory of the Fabry-Perot Dilatometer The instrument consists of a Fabry-Perot interferometer in which the distance between the two mirrors is changed by the elongation of the quartz bar. If we let t be the distance between the two mirrors and p the order of interference of the central fringe then from the theory of the Fabry-Perot interferometer (37) we have

$$p = \frac{2t}{\lambda}$$

where λ is the wavelength of the light used.

For the first circular fringes where n = 1

$$(p - 1) = \frac{2t}{\lambda} \cos \theta_1$$

where θ_1 is the angle subtended by the radius of the first fringe. Similarly for the nth fringe

$$(p - n) = \frac{2t}{\lambda} \cos \theta_n$$

Combining the last three equations we have

$$\cos \theta_n = 1 - \frac{n}{p}$$

For small angles $\cos \theta = 1 - \frac{\theta^2}{2} + \ldots\ldots$

So we may write

$$\theta_n = \sqrt{\frac{2n}{p}} = \sqrt{\frac{n\lambda}{t}}$$

If we let S_n be the diameter of the nth fringe then

$$S_n = 2\theta_n f$$

where f is the focal length of the objective lens of the telescope. We may then write

$$(p - n) = \frac{2t}{\lambda}\left(1 - \frac{S_n^2}{8f^2}\right) \qquad (5)$$

The change in length of the sample is communicated directly to the movable mirror causing the distance between the mirrors to change from t to t' where $t' = t \pm \Delta t$. Decreasing the distance between the mirrors caused the diameters to decreace and vice versa.

For the nth fringe, we have

$$(p - n) = \frac{2t'}{\lambda}\left(1 - \frac{S'^2_n}{8f^2}\right) \qquad (6)$$

where S'^2_n is the new diameter of the nth fringe. Equating the last two equations we have to a very high degree of approximation

$$\Delta t = \frac{t}{8f^2}\left(S'^2_n - S_n^2\right) \qquad (7)$$

Similarly for another circular fringe, say the kth fringe,

$$(p - k) = \frac{2t}{\lambda}\left(1 - \frac{S_k^2}{8f^2}\right) \qquad (8)$$

Subtracting Eq. (6) from Eq. (8) we obtain

$$t/8f^2 = \frac{\lambda}{2}\left[(n - k) \,/\, (S_n^2 - S_k^2)\right] \qquad (9)$$

Eliminating $t/8f^2$ between Eqs. (7) and (9) leaves

$$\Delta t = \frac{\lambda}{2}(n - k)\frac{S'^2_n - S_n^2}{S_n^2 - S_k^2} \qquad (10)$$

From Eq. (10) it follows that Δt may be measured directly in terms of the wavelength of the light used and the diameters of the fringes. Since only ratios are involved it is unimportant what units are used to measure the diameters of the fringes or the magnification.

Since the piezoelectric effect is linear, we may double the displacement by applying the electric field to the crystal in both directions. If we take S_n to be the diameter of the nth fringe with the field in one direction and S_n' to be the diameter of the same fringe with the field reversed and use adjacent rings so that $n = k + 1$ we have

$$\Delta L = \frac{\lambda}{2}\frac{S'^2_n - S_n^2}{S_n^2 - S_k^2} \qquad (11)$$

where ΔL is the change in length of the specimen when the field is reversed.

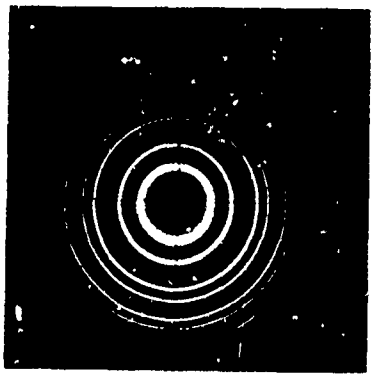
Experimental Procedure

Electrical The electric field is supplied by a regulated power supply of 3000 volts. The voltage is measured with a model ESD Electrostatic Voltmeter made by Sensitive Research Corp. The accuracy has been carefully checked and found to be as good as the scale or about ± 10 volts.

Mechanical The dimensions of the quartz bar are measured to an accuracy of ±0.01 mm using a Gaertner optical microscope. The orientation of the bar is checked by X-ray diffraction using a double crystal X-ray diffractometer.

Optical The diameters of the fringes may be measured in two ways. In the first method the diameters are measured using the micrometer eyepiece of the telescope. This method required several minutes to make a set of measurements and the temperature of the sample or of the instrument may change appreciably during this time. To avoid this problem a photographic technique may be used. A double exposure is made of the fringe pattern with the field first in one direction and then in the other. Using high speed Polaroid film each exposure takes one second. Figure I shows a typical photograph obtained in this way using the light from a low pressure mercury lamp.

Figure I



Results The measurements reported herein were made on a quartz bar having the following dimensions:

Length $L = (4.33 \pm 0.01) \times 10^{-2}$ m

Thickness $e = (4.93 \pm 0.01) \times 10^{-3}$ m

The major surfaces are covered with a conducting film of aluminum applied by evaporation. The length of the bar is parallel to the Y-axis and the thickness to the X-axis within 20 minutes of arc.

All measurements were made at room temperature which varied between 20° and 22° C. Every care was taken to avoid temperature changes during a given measurement.

Fifteen measurements were made using the micrometer eyepiece to measure the diameters of the fringes and twenty-two were made using the photographic technique.

The applied voltage of 3 00 volts was applied in both directions thereby doubling the elongation L. From the definition of the piezoelectric strain coefficient we have

$$d_{11} = \frac{\Delta L / L}{2V/e} = \frac{\Delta L\, e}{2V\, L} = K\,\Delta L \qquad (12)$$

where $K = \frac{e}{2\,V\,L}$ (13)

The value of d_{11} is obtained by substituting the value obtained for ΔL from Eq. (11) into

Eq. (12) using Eq. (13).

Table III shows the results of fifteen measurements made using the micrometer eyepiece to measure the diameters of the fringes.

TABLE III

Aug.	23, 1968	6.43×10^{-8}
	27	6.70
	28	6.74
Sept.	5	6.68
	5	6.62
	6	6.43
	7	6.89
	9	6.80
	10	6.39
	10	6.66
	15	6.42
	18	6.76
	20	6.99
Oct.	7	6.36
	10	6.57
	Average	6.63
	$\sigma = 0.18$	

Table IV shows the results of twenty two measurements made using the photographic technique.

TABLE IV

Dec.	6, 1968	6.39×10^{-8}
	8	6.69
	8	6.73
	10	6.45
	11	6.17
	11	6.28
	12	6.34
	12	6.15
	13	6.37
	16	6.44
	18	6.59
	24	6.37
Jan.	6	6.79
	7	6.69
	8	6.71
	8	6.26
	9	6.75
	10	6.83
	11	6.44
Feb.	8	6.34
	9	6.55
	10	6.74
	Average	6.50
	$\sigma = 0.20$	

Statistically one set of data is not significantly better than the other, the standard deviation being approximately the same for both.

FIGURE II

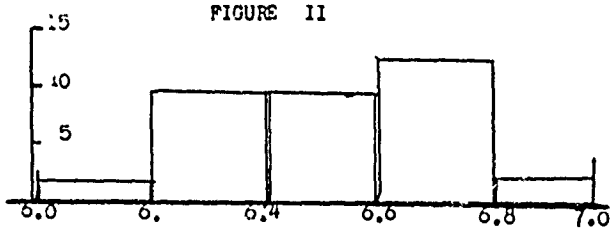


Figure II is a histogram showing the distribution of the values obtained for the two sets of data combined. The distribution can hardly be classed as a normal distribution curve. The average of the thirty-seven measurements is 6.55 with a standard deviation of 0.20 giving a probable error of $\pm$ 0.03. We therefore take the value of d_{11} for this particular sample of quartz to be

$$d_{11} = (6.55 \pm 0.03) \times 10^{-8} \text{ cm/statvolt}$$

or

$$d_{11} = (2.18 \pm 0.01) \times 10^{-12} \text{ m/volt}$$

Discussion

The average of the thirty seven separate measurements on the one sample of quartz is 6.55. It is interesting, and perhaps significant to compare this result with the value 6.53 which is the average of all the measurements reported heretofore by both the static and the dynamic method. These values may be compared with the values which have been accepted by the various authorities as shown in Table V.

TABLE V

Accepted values for d_{11} in quartz.

Sosman	1927	(30)	6.9×10^{-8}
Vigoreux	1939	(31)	6.45
Cady	1946	(32)	6.9
Wooster	1957	(33)	7.3
Bechman	1958	(34)	6.93
Mason	1966	(35)	6.76

It is not at all clear how much of the variation among the results is due to experimental errors and how much is due to variations in the samples of quartz or other unknown factors. Certainly the measurements are difficult and experimental errors are numerous. If the spread of values is mainly attributable to experimental errors [and this is the present opinion of the writer] careful analyses of the experimental procedures are demanded.

In the present work the value of K in Eq. (13) can be determined with an accuracy no worse than 0.5 percent. In any case K is constant for all of the measurements made on a given specimen. The measurements of the diameters of the circular fringes are repeatable from any given photograph with an accuracy no worse than $\pm$ 0.1 percent. For example, in a typical set of data

$$S_2 = 42.13 \pm 0.04 \qquad S_2^2 = 1775 \pm 4$$

$$S_2' = 45.86 \pm 0.04 \qquad S_2'^2 = 2103 \pm 4$$

$$S_3 = 50.62 \pm 0.04 \qquad S_3^2 = 2562 \pm 5$$

from which

$$\frac{S_2'^2 - S_2^2}{S_3^2 - S_2^2} = 0.416 \pm 2 \text{ percent} \qquad (14)$$

Since the error in the determination of d_{11} depends on the accuracy with which Eq. (14) can be evaluated we should expect most of the values to fall within a range of $\pm$ 2 percent of the average value. Examining the data we find that only 12 of the 37 values fall within the range 6.43 to 6.68

whereas 14 are above this range and 11 are below. This result is strongly reminiscent of the results of previous measurements which tend to fall around 6.3 and 6.9 as Sosman (30) has pointed out.

No correction has been made for the fringing of the electric field. Since part of the electric field falls outside the quartz and therefore does not contribute to the strain, the value obtained for d_{11} is smaller than the correct value. The area of the major faces of the sample used in this work is about 5 cm^2 and the thickness about 0.5 cm so that the effect is small and probably less than the errors of measurement. However further refinements should include correction for the effect.

Acknowledgment

The author is indebted to the National Bureau of Standards, Boulder, Colorado for the loan of the Fabry-Perot Dilatometer used in this work.

References

1. Sosman, R. B., "The Properties of Silica", Chemical Catalog Co., Inc., New York, 1927, p. 560

2. P. and J. Curie, Bull. soc. min. de France, Vol. 3, pp. 90-93, 1880

3. Czermak, P. Ber. Wein, Vol. 96, pp. 1217-1244, 1887.

4. Riecke, E. and Voight, W., Nachr. Gott., pp. 247-255, 1891; Weid. Ann., Vol. 45, pp. 523-552. 1892

5. Pockels, F., Abh. Gott., pp. 1-204, 1894

6. Nachtikal, F., Nachr. Gott., pp. 109-118, 1899

7. Curie, Mme. P., "Traite de Radioactivite", Gauthier-Villars and Cie, Paris, 1910

8. Veen, A. L. W. E. Van der, Thesis Delft, 1911

9. Hayashi, F., Dissertation, Gottingen, 1912

10. Roentgen, W. C. and A. Joffe, An. Phk., Vol. 41, pp. 449-498, 1913

11. Tsi-Ze, Ny, C. R., Vol. 184, pp. 1645-1647, 1927

12. Dawson, L. H., Phys. Rev., Vol. 29, pp. 532-541, 1927

13. Seidl, F., Zs. Ph., Vol. 75, pp. 488-503, 1913

14. Gunther N., An. Phk., Vol. 13, pp. 783-801, 1932

15. Knol, K. S., Konink. Akad. Amst., Vol. 35, pp. 99-106, 1932

16. Osterberg, H. and J. W. Cookson, R. S. I., Vol. 6, pp. 347-356, 1935

17. Clay, J. and J. Karper, Physica, Vol. 4, pp. 311-315, 1937

18. Langevin, A., C. R., Vol 209, pp. 627-630, 1939

19. Ludy, W., Helv. Ph. Ac., Vol. 15, pp. 527-552, 1942

20. Cady, W. G., "Piezoelectricity", McGraw-Hill Book Company, Inc., New York, 1946, pp. 219-220

21. Sosman, R. B., "The Properties of Silica", Chemical Catalog Co., Inc., New York, 1927, p. 559

22. Andreeff, A., V. Freedericksz, and I. Kazarnowsky, Zs. Ph., Vol. 54, pp. 477-483

23. Fujimoto, T., Proc. World Eng. Cong., Tokyo, Paper 396, pp. 399-416, 1929

24. Freedericksz, V. and G. Mikhailov, Zs. Ph., Vol. 76, pp. 328-336, 1932

25. Nussbaumer, B., Zs. Ph., Vol. 78, 781-790, 1932

26. Van Dyke, K. S., Proc. I. R. E., Vol. 23, pp. 386-392, 1935

27. Cook, R. K. and P. G. Weissler, Phys. Rev., Vol. 80, 712-16, 1950

28. Mason, W. P., "Piezoelectric Crystals and their Application to Ultrasonics", D. Van Nostrand, New York, 1950, pp.

29. R. Bechman, Phys. Rev., Vol. 110, pp. 1060-62, 1958

30. Sosman [See reference 21]

31. Vigoreus, P. "Quartz Oscillators and their Applications", His Majesty's Stationary Office, London, 1939, p. 12

32. Cady [See reference 20]

33. Wooster, W. A., "Experimental Crystal Physics", Oxford, 1957, p. 79

34. Bechman [See reference 29]

35. Mason, W. P., "Crystal Physics of Interaction Processes", Academic Press, 1966, p. 92

36. Bottom, Virgil E., Rev. Sci. Insts., Vol 35, pp. 374-376, 1964

37. Meissner, K. W., J. Opt. Soc. Am., Vol. 31, p. 405, 1941

ANALYSIS OF CONTOURED PIEZOELECTRIC RESONATORS VIBRATING IN THICKNESS-TWIST MODES

Morio ONOE and Katsuhiro OKADA
Institute of Industrial Science
University of Tokyo
Roppongi, Tokyo, JAPAN

Summary

An analysis of thickness-twist modes in a plate which is contoured along the direction of the wave propagation is presented. The thickness is assumed uniform along the direction perpendicular to the propagation direction. First the plate is partitioned into such small sections along the propagation direction as the variation of thickness within a section is negligible. Second the equivalent transmission line circuit for a section of uniform thickness is obtained. Then characteristics of the whole plate can be represented by the cascade connection of equivalent transmission lines and hence can be obtained by the multiplication of transmission matrices of sections.

This analytical method can handle any continuous shape of contour. The calculation can be mechanically carried out by a computor. The distribution of displacement is obtained at the same time when a resonant frequency is determined, because the displacement at each partition is always available during the process of the calculation.

Plano-convex and double convex cylindrical plates and truncated linearly tapered plates are analyzed by the present method. Such parameters as the anisotropy, the width, the thickness, the lateral length, the radius of curvature, the order of harmonic and inharmonic overtones, and the partition number are taken into account.

It is also pointed out that the present method can be approximately applied to thickness-shear modes.

Experimental observations using AT-cuts of quartz are in good agreement with calculated resonant frequencies and displacements.

Introduction

The major problems in the design of piezoelectric resonators vibrating in thickness modes are how to mount the resonators without impairing the response of the main mode and, at the same time, how to suppress the level of unwanted modes. In the range of frequency more than a few MHz, the energy trapping has been one of the most effective means to solve these problems. For a much lower frequency, however, a parallel plate design based on the energy trapping yields impractically large dimensions. Instead, plate contourings have been extensively used for many years. The design of contouring has been mostly empirical because of mathematical difficulties in an analysis of the effect of contouring. Previous analyses have been limited to such contours as straight bevels[1], Gaussian curves[2] and linear tapers[3].

This paper presents an analysis of thickness-twist modes in a plate which is contoured along the direction of the wave propagation. The thickness is assumed uniform along the direction perpendicular to the propagation direction.

In the analysis, first the plate is partitioned into such small sections along the propagation direction as the variation of thickness within a section is negligible. Second the equivalent transmission line circuit for a section of uniform thickness is obtained. Then characteristics of the whole plate can be represented by the cascade connection of equivalent transmission lines, and hence can be obtained by the multiplication of transmission matrices of sections.

This analytical method has the following features :

(1) Any continuous shape of contour can be handled.
(2) The calculation consists of mostly the multiplication of matrices, and hence can be mechanically carried out by a computor.
(3) The distribution of displacements is obtained at the same time when a resonant frequency is determined. This is because the displacement at each partition is always available during the process of the calculation.

The present method is applied to the analysis of plano-convex and double convex cylindrical plates and truncated linearly tapered plates. Experimental observations using AT-cuts of quartz are in good agreement with calculated resonant frequencies and displacements.

List of symbols

c_{ij} stiffness
H the maximum thickness in y direction

H_m	thickness in y direction of the m-th section
G	transmission matrix
k_{nm}	wave number of the n-th harmonic overtone for the m-th section
L	lateral length
n	order of harmonic overtone
N	number of partition sections
R	radius of curvature
T_i	stress
u	displacement
$\dot{u}$	particle velocity
w_m	width of the m-th section
w	width of the plate
x,y,z	coordinates
X_m	Eq. (9)
Z_m	characteristic impedance of the m-th section, Eq. (10)
c_{11}	voigt's strech modulus
ρ	density
ω	angular frequency
ω_c	cutoff angular frequency, Eq. (5)
Ω	normalized frequency, Eq. (23)

Analysis

Transmission matrix for an elementary section

Using a Cartesian coordinate system (x,y,z), a plate is oriented with its lower surface, which is temporarily assumed to be flat, on the x-z plane, as shown in Fig. 1. The plate is continuously contoured along the z direction and the thickness varies only along this direction. Whereas the length in the x direction can be considered infinitely wide.

The plate is partitioned into N sections by planes perpendicular to the z axis. The number of sections N, is so chosen as to be able to neglect the variation of thickness within each section. The right side of the m-th section is called the m-th partition.

In thickness-twist vibrations propagating along the z direction, only the displacement u in the x direction exists. Such vibrations can be piezoelectrically excited by electrodes on major faces in rotated Y-cut plates of quartz and ceramic plates when the diagonal and the poling axes are parallel to the x axis, respectively.

In the following, an analysis of purely elastic cases is done, because a main interest of the present paper is in quartz, of which electromechanical coupling is very small. Assuming a monoclinic symmetry in the stress-strain relation, the equation of motion for each section can be written in the following form.

$$\rho\ddot{u} = c_{55}\frac{\partial^2 u}{\partial z^2} + 2c_{56}\frac{\partial^2 u}{\partial y \partial z} + c_{66}\frac{\partial^2 u}{\partial y^2} \qquad (1)$$

A dot denotes differentiation with respect to time.

The boundary conditions to be satisfied on the major faces of the plate are :

$$T_2 = T_4 = T_6 = 0 \quad \text{on } y=0 \text{ and } H_m \qquad (2)$$

Although an exact solution of Eq. (1) subject to conditions Eq. (2) was obtained by Mindlin[4], a simpler solution is used here by taking an assumption, $c_{56} = 0$. This assumption is exactly true for ceramic plates and AC-cuts of quartz and approximately correct for AT- and BT- cuts of quartz.

The general solution for the m-th section is

$$u = \sum_{n=0}^{\infty} (A_n \cos k_{nm} z + B_n \sin k_{nm} z) \cdot \cos\left(\frac{n\pi y}{H_m}\right) \cdot \exp j\omega t \qquad (3)$$

where

$$\rho\omega^2 = c_{66}\left(\frac{n\pi}{H_m}\right)^2 + c_{55}k_{nm}^2 \qquad (4)$$

It can be seen that, for a given n, there is a cutoff frequency ω_c, below which k_{nm} becomes imaginery.

$$\omega_c = \frac{n\pi}{H_m}\sqrt{\frac{c_{66}}{\rho}} \qquad (5)$$

The arbitrary constants, A_n and B_n, are to be determined from conditions of continuity of displacement and stress at the partition boundaries. Only nonvanishing stress component at the boundaries is T_5.

$$T_5 = c_{55}\frac{\partial u}{\partial z} + c_{56}\frac{\partial u}{\partial y} \simeq c_{55}\frac{\partial u}{\partial z} \qquad (6)$$

Strictly speaking, a combination of an infinite number of harmonic modes as expressed by Eq. (3) is required to satisfy the boundary conditions, when the thicknesses of adjacent sections are different. By increasing the number of sections, however, the difference in thickness can be made diminishingly small and hence the use of only one predominant mode and the neglect of all other modes becomes permissible. We designate the case of $n = 1$ the foundamental and the case of $n = 3$ the third harmonic overtone and so on. Even overtones are not interesting here, because they cannot be effectively excited by electrodes on

major surfaces of the plate.

For a given n and given sets of stress and particle velocity at both sides (four conditions) of the m-th section, Eq. (3) (two arbitrary constants) yields two linear homogeneous equations, which is expressed in the following matrix form.

$$\begin{pmatrix} T_5 \\ \dot{u} \end{pmatrix}_{m-1} = G_m \begin{pmatrix} T_5 \\ \dot{u} \end{pmatrix}_m \tag{7}$$

where

$$G_m = \begin{pmatrix} \cos x_m & -j\, z_m \cdot \sin x_m \\ -j \dfrac{1}{Z_m} \sin x_m & \cos x_m \end{pmatrix} \tag{8}$$

$$x_m = k_{nm} w_m \tag{9}$$

$$z_m = \frac{c_{55} k_{nm}}{\omega} \tag{10}$$

The particle velocity is used instead of the displacement in conformity with the electromechanical (force-voltage and velocity-current) analogy. The matrix G_m is the same as the transmission matrix of an electrical line having the propagation constant k_{nm} and the characteristic impedance Z_m.

Overall characteristics

Overall characteristics of the whole plate is expressed by the cascade connections of sections or by the multiplication of the transmission matrices.

$$\begin{pmatrix} T_5 \\ \dot{u} \end{pmatrix}_0 = G \begin{pmatrix} T_5 \\ \dot{u} \end{pmatrix}_N \tag{11}$$

where

$$G = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} = G_1 G_2 \cdots G_N \tag{12}$$

Both sides of the plate are traction free and hence T_5 is equal to zero. The resonant condition, which corresponds to an infinite amplitude of particle velocity, is :

$$G_{12} = 0 \tag{13}$$

If the shape of the plate is geometrically symmetrical with respect to the center mirror plane, it is necessary to carry out the multiplication of transmission matrices only over one half of the whole plate. Let the resultant matrix for the left half be

$$G_L = \begin{pmatrix} G_{11}' & G_{12}' \\ G_{21}' & G_{22}' \end{pmatrix} \tag{14}$$

then the matrix for the right half takes the following form :

$$G_R = \begin{pmatrix} G_{22}' & G_{12}' \\ -G_{21}' & G_{11}' \end{pmatrix} \tag{15}$$

The product of two matrices yields the resultant matrix for the whole plate, of which "12" component is $G_{11}'\ G_{12}'$. Hence the resonant condition is

$$G_{11}' = 0 \tag{16}$$

or

$$G_{12}' = 0 \tag{17}$$

An examination of the distribution of displacement,

as described later, shows that the former yields antisymmetric modes, whereas the latter yields symmetric modes with respect to the center mirror plane.

In the beginning of this chapter, the lower surface of the plate was assumed to be flat. This assumption is unnecessary and the lower surface can be continuously contoured in any manner along the z axis. This is because the resonant conditions depend on the distribution of thickness, but not or the coordinates of surfaces.

Displacement

The displacement at the r-th partition is obtained from the following equation

$$\begin{pmatrix} 0 \\ \dot{u} \end{pmatrix}_0 = G_1 \cdot G_2 \cdots G_r \cdot \begin{pmatrix} T_5 \\ \dot{u} \end{pmatrix}_r \tag{18}$$

$$= \begin{pmatrix} G_{11,r} & G_{12,r} \\ G_{21,r} & G_{22,r} \end{pmatrix} \cdot \begin{pmatrix} T_5 \\ \dot{u} \end{pmatrix}_r$$

Hence

$$(\dot{u})_r = G_{11,r} \cdot (\dot{u})_0 \tag{19}$$

An integration with respect to time yields

$$(u)_r = G_{11,r} \cdot (u)_0 \tag{20}$$

Since the displacement at a side of the plate may be very small, it is often convenient to express the displacement in terms of either the displacement or the stress at the central s-th partition.
Namely,

$$(u)_r = \frac{G_{11,r}}{G_{11,s}} \cdot (u)_s \tag{21}$$

or

$$(u)_r = -\frac{G_{11,r}}{G_{12,s}} \cdot (T_5)_s \tag{22}$$

Numerical calculations

Numerical calculations are carried out for AT-cuts of quartz and isotropic plates with plano-convex and double convex cylinarecal contours as well as linerly tapered contours. The value of C_{66} / C_{55} is 0.438 for AT-cuts of quartz and 1 for isotropic plates, respectively.

Effect of the partition number

Fig. 2 shows an effect of the number of partitioned section on the calculated resonant frequencies of a double convex cylindrical AT-cut of quartz. The width to the curvature ratio is 0.303 which is also equivalent to 0.606 for a plano convex cylindrical plate. S and A denote symmetric and antisymmetric modes, respectively. The parameter is the width to thickness ratio. The ordinate is the normalized frequency defined as follows.

$$\Omega = \frac{\omega H}{\pi \cdot \sqrt{\frac{C_{66}}{\rho}}} \tag{23}$$

It can be seen that the higher the order of inharmonic overtone, the slower the convergence. Another calculation shows the convergence becomes slower for larger value of the ratio W/R.

The pertition number of 40 seems good enough to obtain an accuracy better than a few tenth per cent for the lowest mode and somewhat worse for higher inharmonic moder. In the following most calculations uses the pertition number of 50.

Plano convex cylindrical plates

Effect of the W/H ratio Fig. 3-9 shows normalized resonant frequencies of plano convex cylindrical AT-cuts of quartz as functions of the width to thickness ratio. Parameters are the order of harmonic overtone and the width to the radius of curvature ratio. In the figure, S and A stand for the symmetric and the antisymmetric modes, respectively.

Fig. 3 is for a plate with W/R=0.01 or a nearly flat plate. Resonant frequencies look very similar to that of a perfectly flat plate, in which the lowest (S1) mode coincides with the abscissa.

Fig. 4-6 are for the foundamental, the third overtone and the fifth overtone modes, respectively, in a plate with W/R=0.151. Whereas Fig. 7-9 are the same in a plate with W/R=0.322.

These figures show that first the separation between inharmonic modes becomes narrower when the order of harmonic overtones increases.

Second, when the W/H ratio decreases resonant frequencies generally increases. For small values of W/H, however, each inharmonic mode exhibits a

stationary portion and the lowest (S1) mode changes to decrease and goes through $\Omega=1$ at W/H=0.

Jumonji pointed out that this anomality is caused by couplings with face shear modes in the xy place and derived the following equation for obtaining its uncoupled resonant frequencies.(5)

$$\Omega = \frac{nH}{H-R\cdot\left[\frac{R}{W}\cdot\theta+\frac{1}{4}\cdot\sin 2\theta-1\right]}$$

$$\text{where} \quad \sin\theta = \frac{W}{R} \qquad (24)$$

The denominator corresponds to an effective thickness of a flat plate, of which crossectional area in the yz plane is equal to the area of the contoured plate having the same width.

Fig. 10 shows resonant frequencies of the face shear modes for W/R=0.151 and 0.322. It can be seen that the tangents at $\Omega=1$, W/H=0 are equal to those in Fig. 4-9, respectively.

Piezoelectric behaviors of the plate are significantly affected by those couplings. For example, the A1 mode for a high value of W/H is hardly excited by an electrode symmetrical with respect to the center mirror plane. However, when the W/H is lower than the value, which yields a coupling with the face shear mode, the mode is now strongly excited by a symmetrical electrode. This is because this portion of the A1 mode is really an extension of the S1 mode. Similar or inverse phenomena exist in other modes.

Effect of the W/R ratio Fig. 11-14 shows normalized resonant frequencies of the foundamental and the third overtone modes of AT-cuts of quartz as functions of the width to the radius of curvature ratio. The parameter is the width to thickness ratio.

Again the higher the order of harmonic overtone, the narrower the separation between inharmonic modes. The effect of the W/R ratio on resonant frequencies is small for the ratio up to about 0.1 and thereafter rapidly increases with the ratio.

Displacement The displacement of the S1 and A1 modes is shown in Fig. 15-16, respectively. The ordinate is normalized by the maximum displacement. Only the left half portion is shown because of the symmetry. The parameter is the width to thickness ratio. It can be seen that the displacement is confined to the centeral portion when the W/H ratio increases. For W/H=25, the displacement at the end of the plate is in the order of 10^{-17}. This trapping of vibration energy in the vicinity of the center makes the mounting of the plate and the suppression of unwanted modes easy.

Case of istropic plates. In order to see the effect of unisotropy, similar calculations are carried out for isotropic plates. Fig. 17 and 18 shows resonant frequencies of the foundamental and the third overtone modes, respectively, as functions of the width to thickness ratio. Fig. 19 and 20 shows the same as functions of the width to the radius of curvature ratio. It can be seen that the separations between inharmonic overtone modes in a isotropic plate are much narrower than those in an AT-cut of quartz.

Double convex cylindrical plates

Fig. 21 shows normalized resonant frequencies of a double convex cylindrical AT-cuts of quartz as functions of the width to thickness ratio. Both major surfaces are contoured with the same radius of curvature. The W/R ratio is 0.303.

A double convex plate can be deformed into a plano convex plate without changing the distribution of thickness. The radius of curvature of the contoured surface of the deformed plate is no longer a constant. However, the radius at the center is the one to be really counted because the vibration energy is confined in the vicinity of the center. Hence the curves in Fig. 21 are close to curves for a plano-convex plate with the W/R ratio of 0.303x2. Calculations show that differences are less than 0.5 percent.

Linearly tapered plates

Truncated linearly tapered plates are studied. Fig. 22 shows resonant frequencies as functions of the width to thickness ratio. The parameter is the degree of taper. Curves for the W/H ratio larger than 10 were also calculated by Loutzenheiser and Denkmann using Mindlin's plate equation.(3) Both curves are in good agreement with each other.

In contrast to the case of convex plates, only the lowest mode exhibits a sign of a coupling with a face shear mode in the xy plane at a low W/H ratio. This is because an uncoupled resonant frequency of the face shear mode in a linearly tapered plate is nearly constant and hence higher inharmonic modes does not couple with the face shear mode. Fig. 23 shows a detailed behavior of the lowest mode for various degree of taper.

Thickness-shear modes

Although the subject of the present paper is the thickness-twist vibrations in a contoured plate, a brief remark on the thickness-shear modes is included here. Details, however, will be presented in a subsequent paper.

In order to accommodate the case of thickness-shear vibrations in AT-cuts of quartz, the x and z axes in Fig. 1 are interchanged. Then the diagonal axis, the desplacement and the direction of propagation are all along the x axis.

The transmission matrix (8) can be applied approximately to the case of the thickness shear modes, provided that C_{55} in Eq. (4) and (10) is replaced by Voight's stretch modulas of the plate in the x direction, γ_{11}.[6] Fig. 24 and 25 shows normalized resonant frequencies of the thickness-shear modes in a plano convex cylindrical plate as functions of the width to thickness ratio.

Experiments

Experimental procedure

Both plano-convex and double convex cylindical AT-cuts of quartz are provided. One major surface of a plate is fully covered by a thin metalic electrode. Whereas the electrode on the opposite surface is divided into two electrodes at the center, so that symmetric modes can be excited by both electrodes in parallel and antisymmetric modes can be excited by either one of electrodes, which is called hereafter the one half electrode.

A combination of a frequency synthesizer, an amplifier-detector and an X-Y recorder is used in the measurements of resonant frequencies. First all the responses are recorded over the interested frequency range. Then frequencies of strong resonances are precisely measured.

Effect of the lateral length

In the analysis, the length of a plate perpendicular to the direction of propagation is assumed to be infinite. In order to see the effect of finite lateral length, several samples with the W/R ratio of 0.322 which differ only in lateral length are provided.

Fig. 26 shows a comparison of resonant frequencies as functions of the width to thickness ratio. The lowest mode seems to be insensitive to the lateral length. Inharmonic overtones for the L/W ratio of 1 are separated from those for L/W of 1.5 and 2, which are very close together. Hence, in the following, a L/W ratio greater than 2 is used.

Comparisons between experimental and calculated results.

Resonant frequencies of the foundamental and the third overtone modes in plano-convex cylindrical AT-cuts of quartz with the W/R ratio of 0.322 are measured and plotted in Fig. 7 and 8, respectively.

Black dots denote the modes strongly excited by an electrode symmetrical with respect to the center mirror plane. Whereas white or hollow dots denote the modes which are hardly excited by the symmetrical electrode, but strongly excited by the one half electrode.

Sometimes a resonance may split into two equaly strong responses or accompany weak responses. The former is shown by two equal dots and the latter by smaller dots.

The agreement between observed and calculated results seems good. As expected, the lower the order of inharmonic overtone, the better the agreement.

The distribution of piezoelectrically induced charge, which is proportional to the displacement, on the surface of a plate with the W/R ratio of 0.32 and the W/H ratio of 12.5 is measured by a probe[7] and plotted as dotted lines in Fig. 15 and 16. Fig. 15 is for the S1-mode and Fig. 16 is for the A1-mode. Since the impedance, to which the probe faces, varies with the thickness of the plate, the output voltage of the probe does not yield a faithful distribution of the displacement. Hence no attempt of quantitative comparisons with calculated values is made. Even so, both figures clearly show the trapping of the vibration energy in the vicinity of the center due to cylindrical contouring.

Observed resonant frequencies of the foundamental mode in double convex AT-cuts of quartz with the W/R ratio of 0.303 are plotted in Fig. 21. The agreement is good for the lowest mode and is getting worse as the order of inharmonic modes increases. This may be due to the larger W/R ratio, which is effectively 0.303 x 2, than the ratios used in Fig. 7 and 8.

Observed resonant frequencies of the foundamental and the third overtone modes in plano-convex AT-cuts of quartz vibrating in thickness-shear modes are plotted in Fig. 24 and 25, respectively. The agreement in the case of thickness-shear modes is not worse than the agreement in the case of thickness-twist modes.

Conclusion

The analysis presented in this paper makes possible the calculation of resonant frequencies and displacements of thickness-twist modes in a plate contoured in any continuous shape. Such parameters as the anisotropy, the width, the thickness, the lateral length, the radius of curvature, the order of harmonic and inharmonic overtones, and the partition number are taken into account. Calculated curves for plano-convex and double convex cylindrical plates are compared with experimental observations using AT-cuts of quartz. The accuracy obtained is good enough for the practical use in the design of resonators.

The displacement at both sides of a plate practically diminishes by the application of cylindrical contouring. Hence the vibration energy is trapped in the vicinity of the center, which makes the mounting of the plate and the suppression of unwanted modes easy.

Next truncated linearly tapered plates are studied. Calculated resonant frequencies are in good agreement with similar calculations by Lontzenheizer and Dankmann using Mindlin's plate equations.

The present method is also useful for the analysis of thickness-shear modes in a contoured plate.

Experimental observations using AT-cuts of quartz show a good agreement with calculated values.

Acknowledgement

This study was done while one of the authors (K.O.) was on leave of absence from Nippon Electric Co. Ltd. The authors wish to thank Dr. M. Kogo and late Mr. H. Kura of the Nippon Electric Co. for their continuous encouragement and to Mr. H. Jumonji for stimulating discussions.

Literature

(1) R.D.Mindlin, J. App. Phys. vol. 25, 1, pp. 12-20, 1954

(2) R.P.Jerrard, Quart. App. Math. vol. 18, 1, pp. 173-181, 1960

(3) C.B.Loutzenheizer and W.J.Denkmann, J. Acoust. Soc. Am. vol. 41, 4, part 2, pp. 962-968, 1967.

(4) R.D.Mindlin, Int. J. Solids Struct. vol. 1, pp. 141-145, 1965

(5) H.Jumonji, Private communication.

(6) R.D.Mindlin and P.C.Y.Lee, Int. J. Solids. Struct. vol. 2, pp. 125-139, 1966.

(7) H.Fukuyo, Bull. Tokyo Inst. Tech., A-1, 1955.

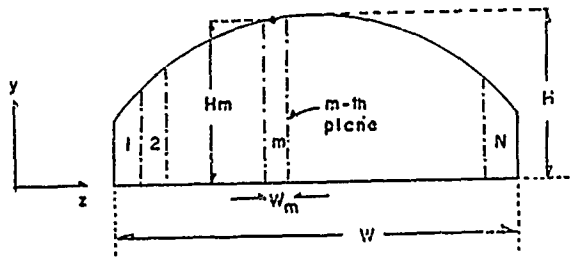


FIGURE 1 -A contoured plate with coordinate axes.

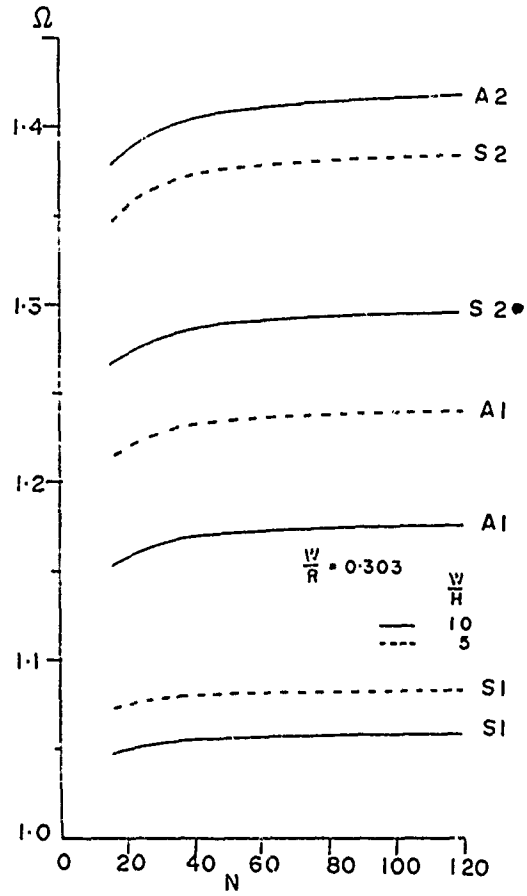


FIGURE 2 -Effect of the partition number on resonant frequencies.

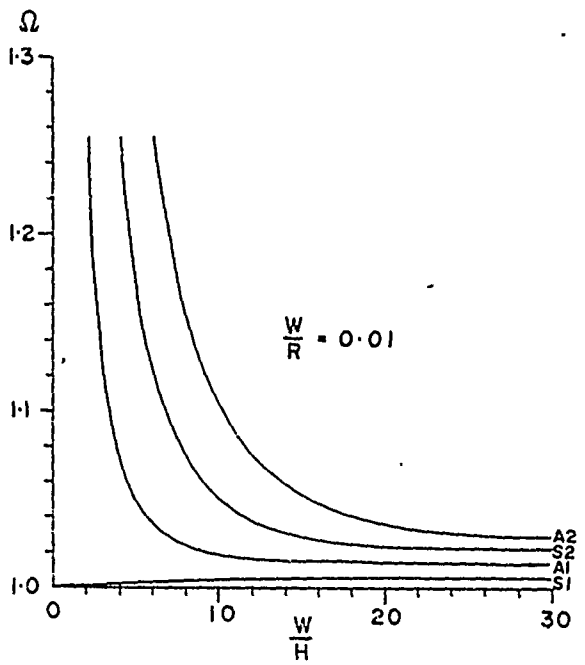


FIGURE 3 - Resonant frequencies of the fundamental modes versus the width to thickness ratio W/R = 0.01

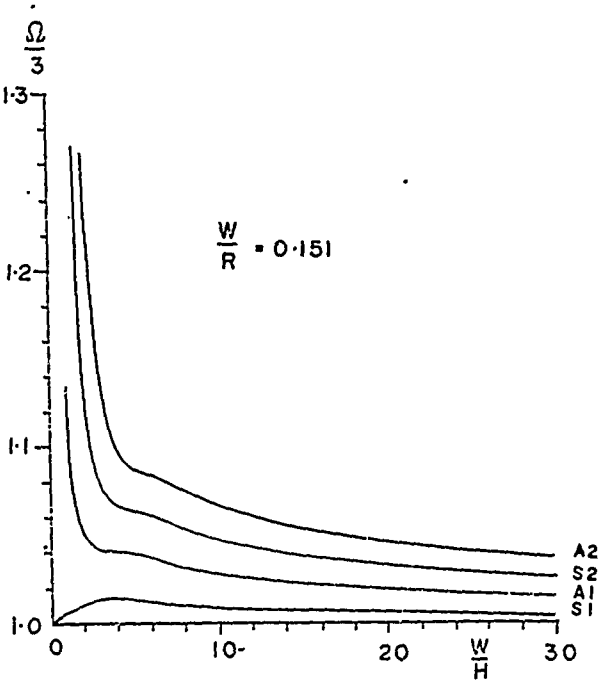


FIGURE 5 - Resonant Frequencies of the third overtone modes versus the width to thickness ratio W/R = 0.151

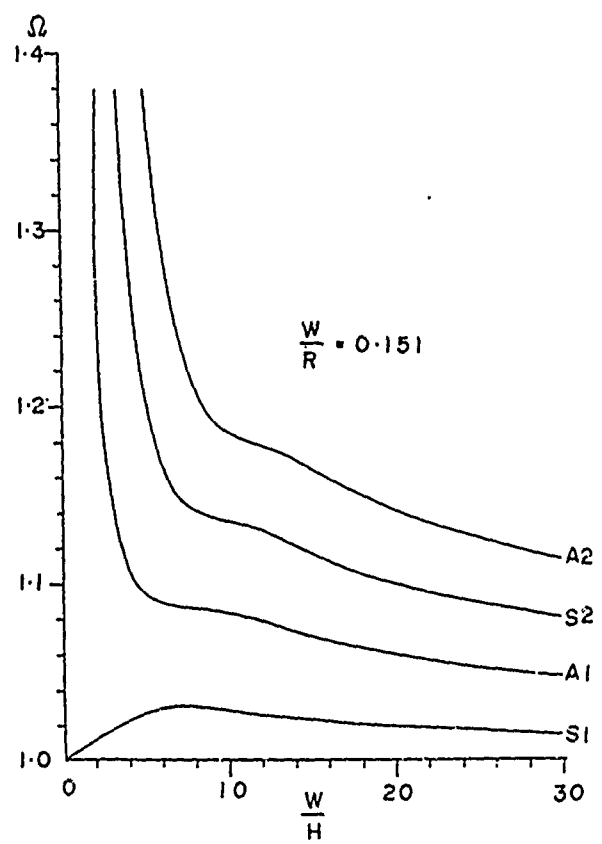


FIGURE 4 - Resonant frequencies of the fundamental modes versus the width to thickness ratio W/R = 0.151

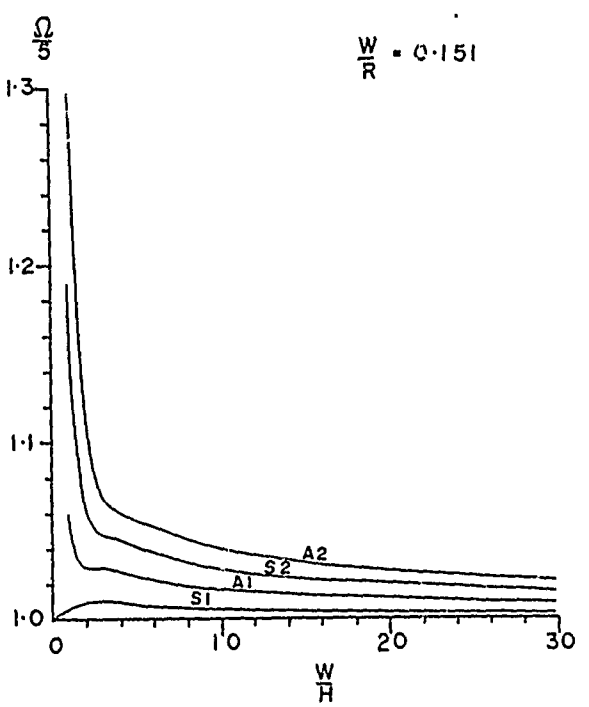


FIGURE 6 - Resonant frequencies of the fifth overtone modes versus the width W/R = 0.151

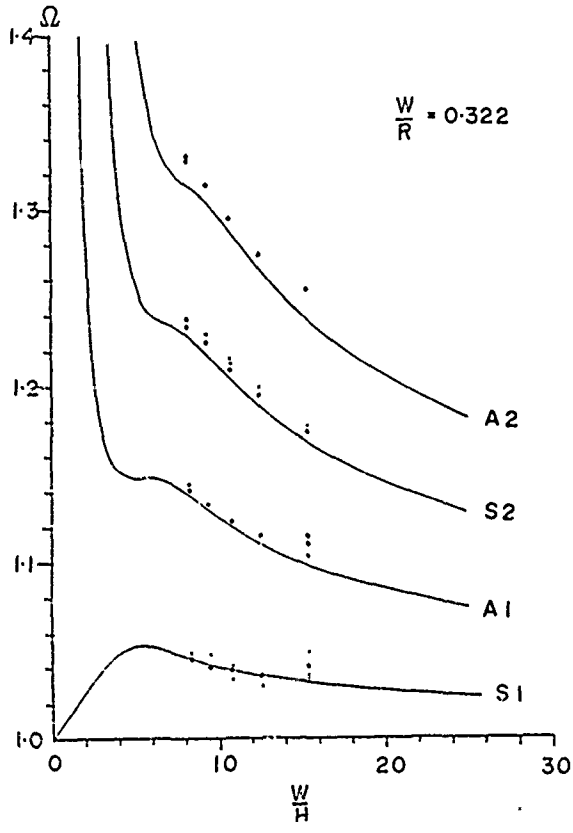


FIGURE 7 - Resonant frequencies of the fundamental modes versus the width to thickness ratio W/R = 0.322

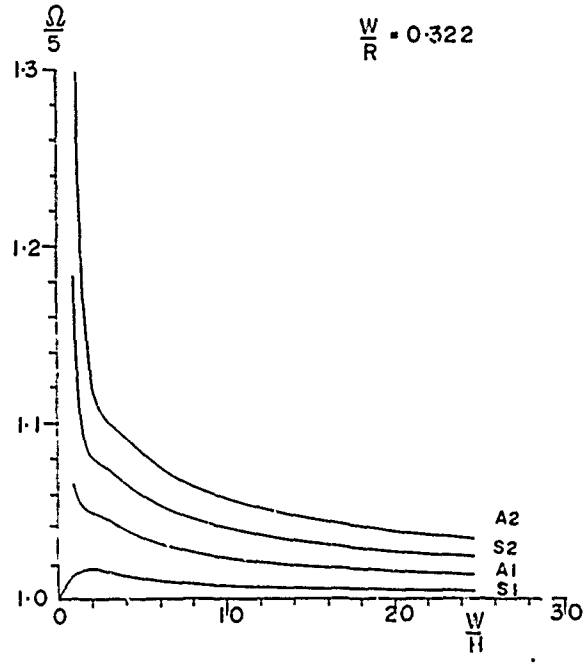


FIGURE 9 - Resonant frequencies of the fifth overtone modes versus the width to thickness ratio W/R = 0, 322

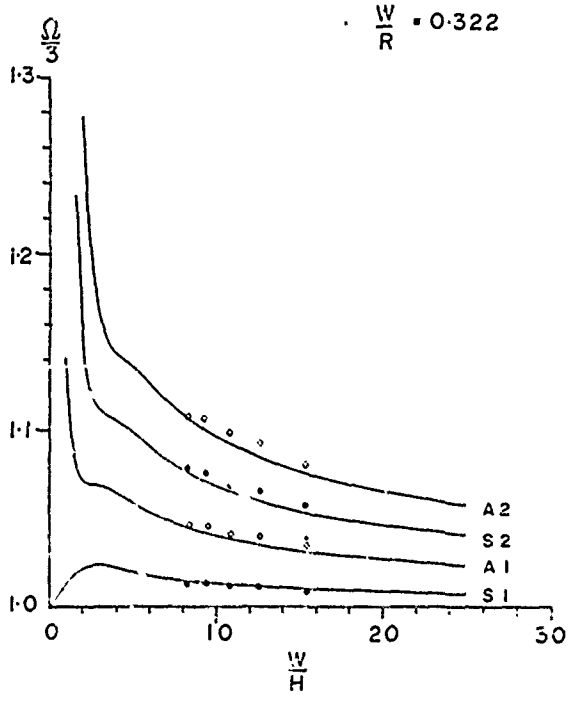


FIGURE 8 - Resonant frequencies of the third overtone modes versus the width to thickness ratio W/R = 0.322

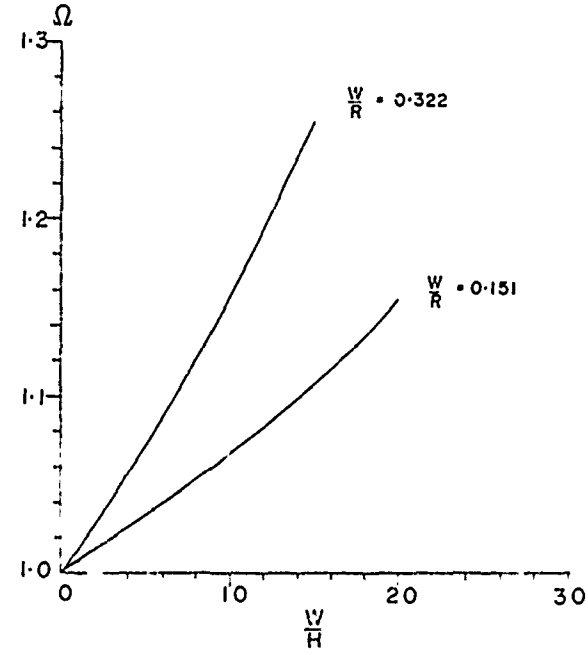


FIGURE 10 - Resonant frequencies of face shear modes in the xy plane

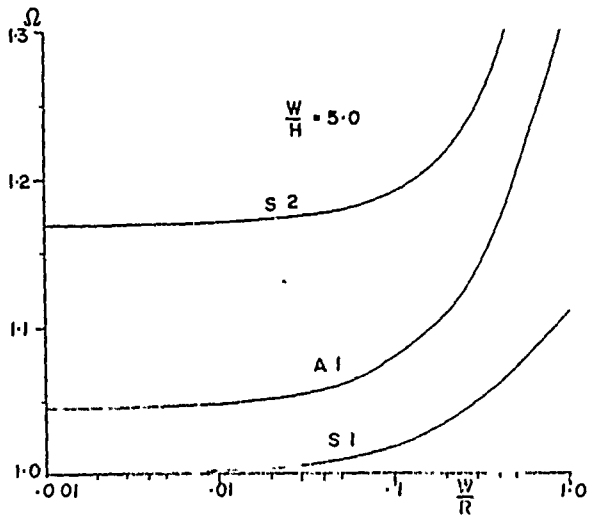


FIGURE 11 - Resonant frequencies of the fundamental modes versus the width of the radius of curvature ratio W/H = 5

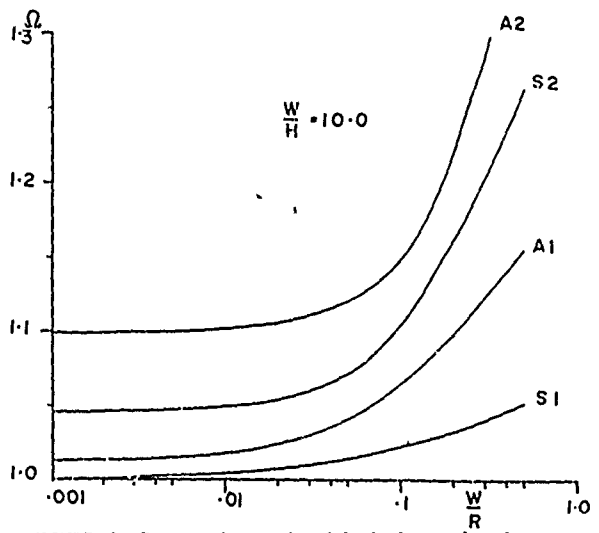


FIGURE 12 - Resonant frequencies of the fundamental modes versus the width to the radius of curvature ratio W/H = 10

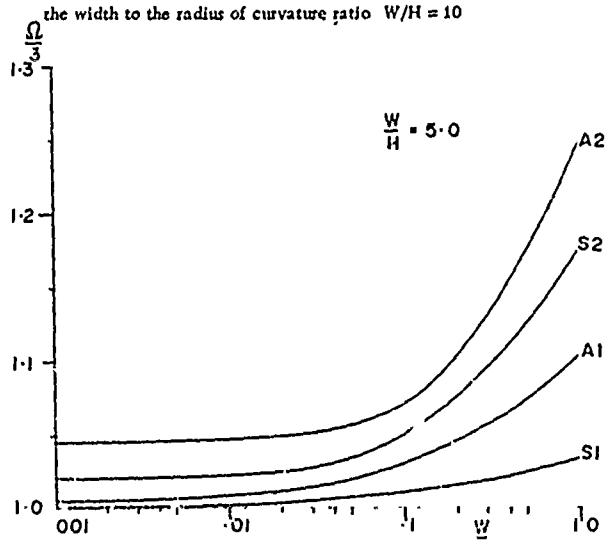


FIGURE 1. - Resonant frequencies of the third overtone modes versus the width to the radius of curvature ratio W/H = 5

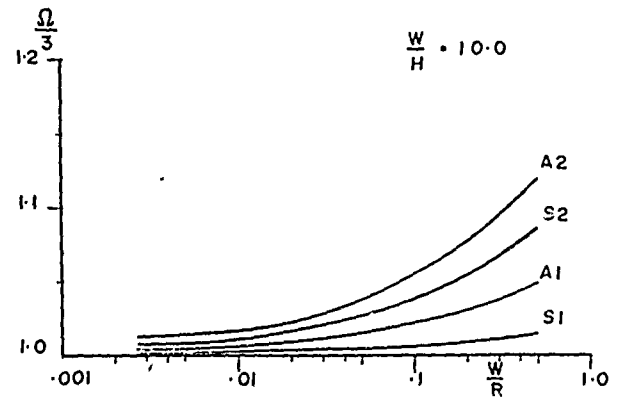


FIGURE 14 -Resonant frequencies of the third overtone modes versus the width to the radius of curvature ratio W/H = 10

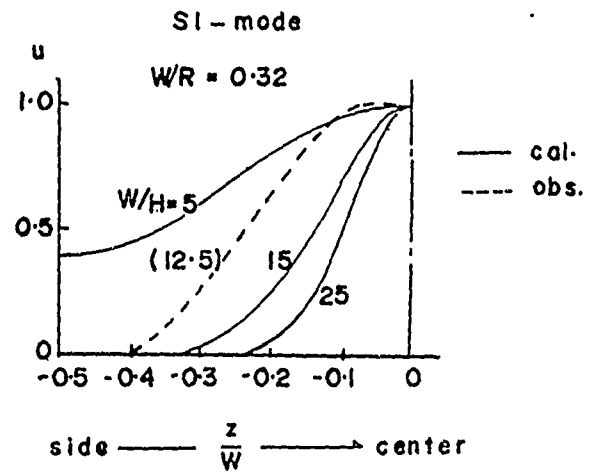


FIGURE 15 - Displacement of the S1 mode. W/R = 0.322. W/H is a parameter.

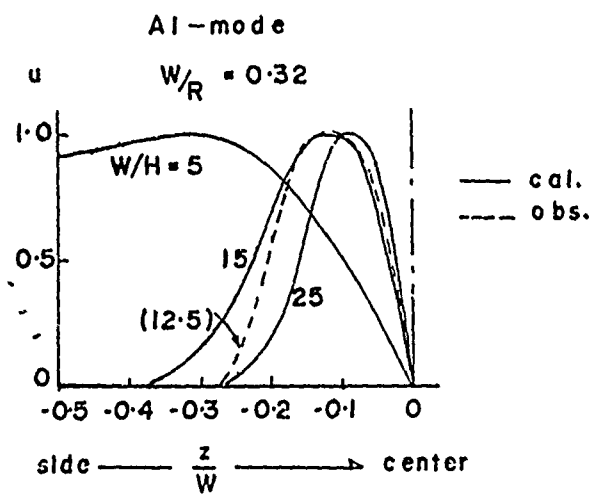


FIGURE 16 - Displacement of the A1 mode W/R 0.322 W/H is a parameter

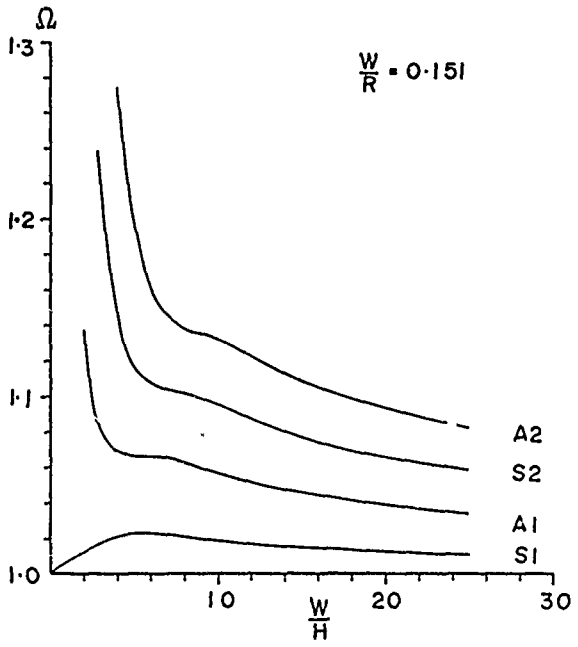


FIGURE 17 - Resonant frequencies of the fundamental modes in an isotropic plate versus the width to thickness ratio W/R = 0.151

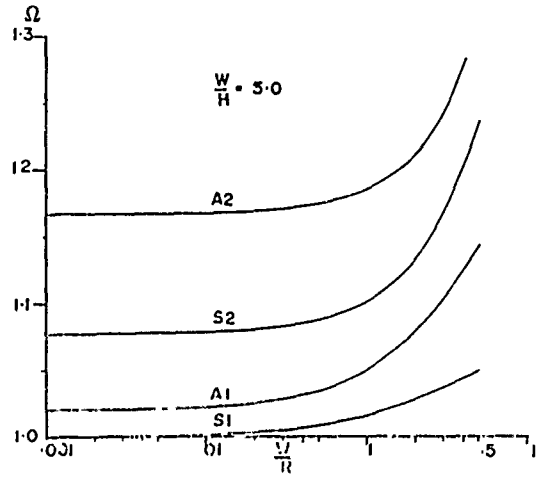


FIGURE 19 - Resonant frequencies of the fundamental modes in an isotropic plate versus the width to the radius of curvature ratio W/R = 0.151

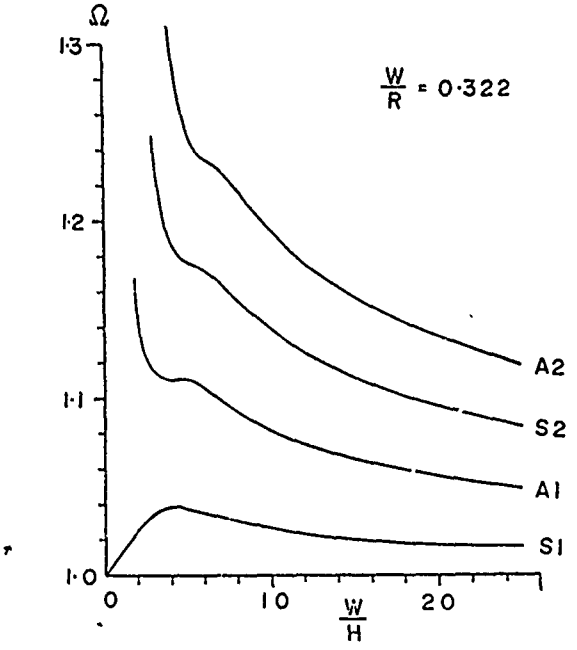


FIGURE 18 - Resonant frequencies of the fundamental modes in an isotropic plate versus the width to thickness ratio W/R = 0.322

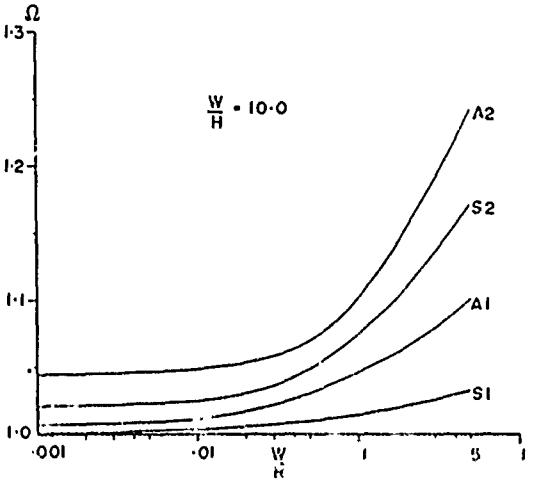


FIGURE 20 - Resonant frequencies of the fundamental modes in an isotropic plate versus the width to the radius of curvature ratio W/R = 0.322

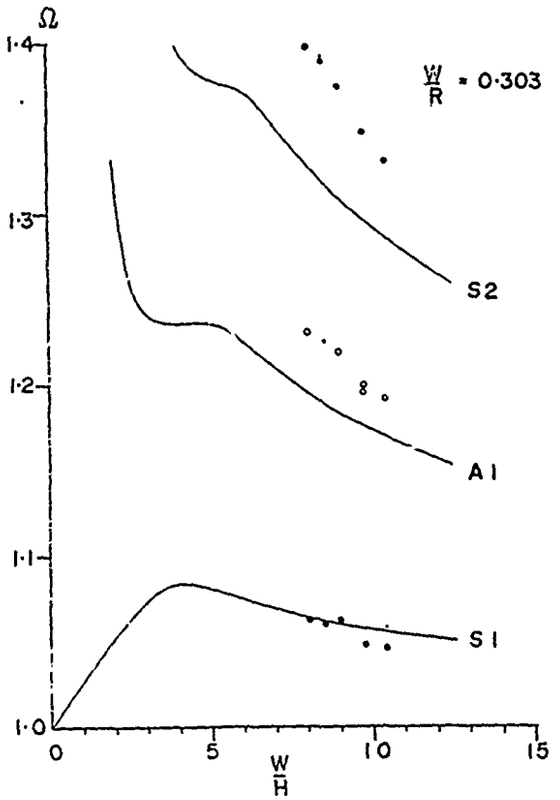


FIGURE 21 - Resonant frequencies of the fundamental modes in a double-convex AT-cut of quartz W/R = 0.303

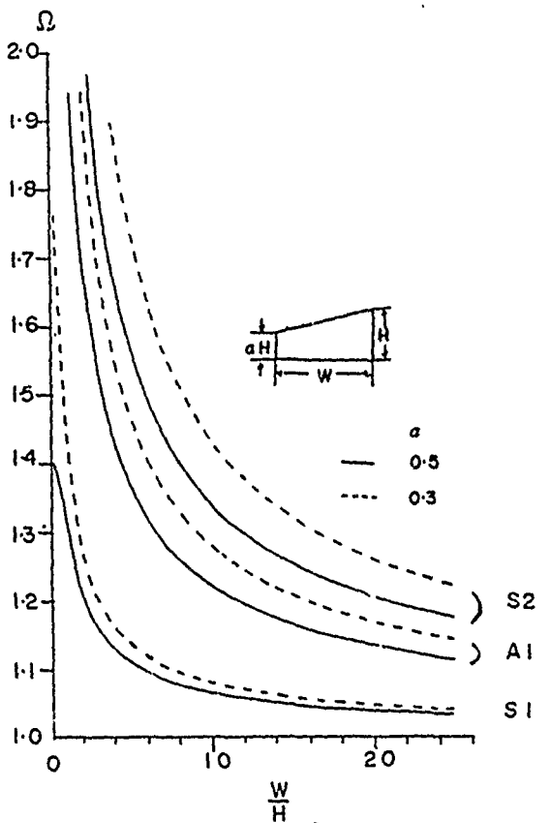


FIGURE 22 - Resonant frequencies of a truncated linearly tapered plate versus the width to thickness ratio.

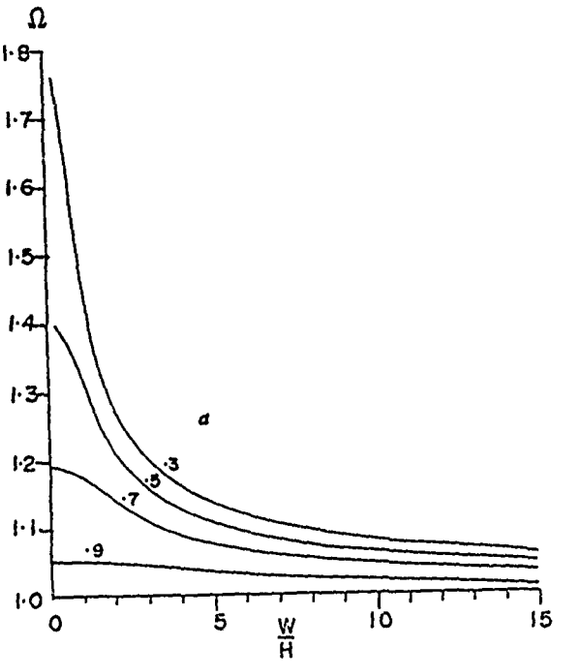


FIGURE 23 - Resonant frequencies of the lowest mode in a linearly tapered plate versus the width to thickness ratio

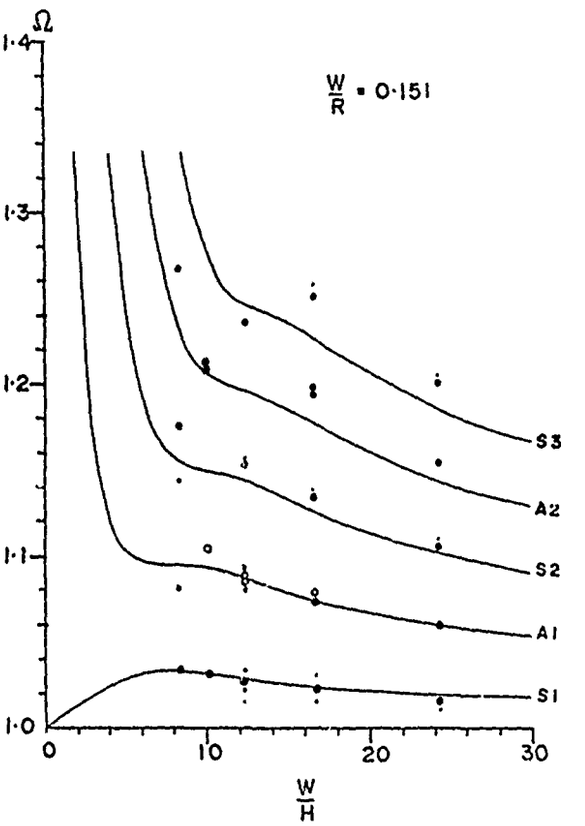


FIGURE 24 - Resonant frequencies of the fundamental thickness-shear modes in a plano-convex cylindrical plate versus the width to thickness ratio

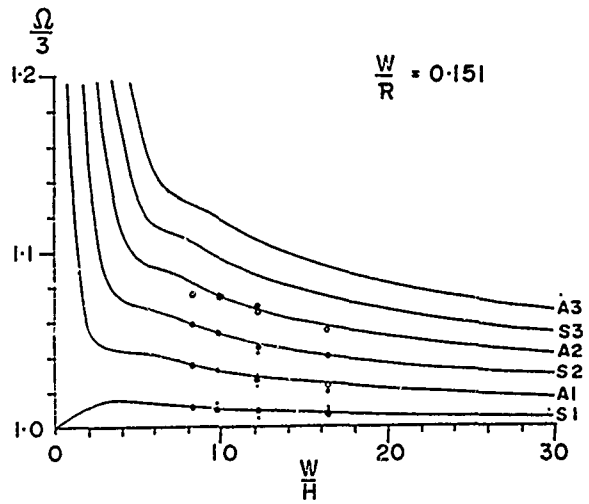


FIGURE 25 - Resonant frequencies of the third overtone thickness-shear modes versus the width to thickness ratio

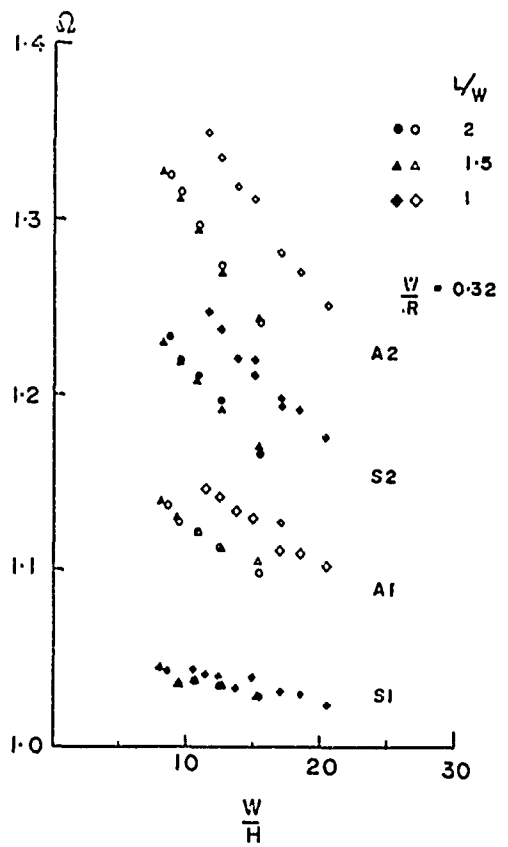


FIGURE 26 - Comparison of resonant frequencies of plates with different lateral length

RESONANCE FREQUENCIES OF MONOLITHIC QUARTZ STRUCTURES

Albert Glowinski

and

René Lançon

Centre National d'Etudes des Télécommunications
Issy-les-Moulineaux, France

Summary

The monolithic structures considered in this paper, are purely mechanical approximations of monolithic quartz filters. Theoretical discussion of resonance frequencies of these structures leads to several results, as prediction of frequency values, number of resonances, and leads to a spectral purity criterion, which is an extension of criterion, established from energy trapping concept, related to crystal single resonator[1].

Introduction

In the theoretical study of monolithic quartz filters, the following problem should be solved : in a thin plate of piezoelectric and elastic crystal, including several resonators (i.e facing electrodes pair) separated by coupling regions, functions of space coordinates and time are searched, for mechanical displacement and electrical potential, solutions of equations and satisfying boundary conditions. Considering then two electrodes pairs as electrical access of the filter, it will be possible to obtain for instance, the transfer matrix of the electrical fourpole.

Unfortunately simple solutions of equations, such as exponential functions, cannot satisfy boundary conditions. Therefore we have considered as a simpler theoretical model, monolithic structures having the same geometrical and mechanical properties as the monolithic filters, but of non-piezoelectric material, in order to solve the resonance frequencies problem. At least two results can be expected : first, knowledge of values and number of resonance frequencies, as a function of structure parameters, result which should be a close enough approximation of the real piezoelectric problem for quartz, since the electro-mechanical coupling coefficient is low; then, knowledge of a method of analysis which, slightly modified, may be adapted to the piezoelectric case.

We consider a two-dimensions anisotropic elastic structure constituated by n regions representing n resonators separated by (n-1) regions (inter-resonator regions), different regions having different mechanical and geometrical properties. From particular solutions of motion equations, depending upon two constants for each region, continuity conditions at the boundaries between adjacent regions yield a system of homogeneous linear equations in the above constants. The system determinant is the resonance frequencies equation of the structure. For structures having convenient symetry, these determinants possess interesting properties : a structure equivalent to a n-resonators filter leads to a 4n-order determinant, and it can be shown that, in the development of the determinant, appears only the product of two 2n-order minors ; in addition, recurrence relationship is established between the minors of the determinants of the n and (n-1)-resonator structures. It is then possible to predict the number of roots of the 4n-determinant from the number of roots of the 4(n-1)-determinant. Finally it is shown that the number of roots depends upon the position in a n-dimensions space, of a point characteristic of the monolithic structure.

Solution of equations and boundary conditions

The n-order structure, corresponding to a n-resonators filter is shown on Fig. 1.

There are (2n+1) distinct regions including n-regions representing n resonators, (n-1) inter-resonators regions and two extreme regions, both infinite. Using numeration from 0 to 2n, regions R_k and R_{2n-k} are completely identical, k varying from 0 to n ; referred to X'X and Y'Y axis, these regions are bounded along X'X by $-D_k$ and $-D_{k+1}$ for R_k, and D_{k+1} and D_k for R_{2n-k} ; their common length is :

$$L_k = D_k - D_{k+1} \qquad (1)$$

Following several authors,[1] we shall assume that the distinction between regions does not come from plate-thickness difference, but from specific mass difference : each couple (R_k, R_{2n-k}) is determined by the thickness h, independent of k, the abscissa $\pm(D_k, D_{k+1})$ and the length L_k, and the specific mass ρ_k.

Since we assume an elastic material having the properties of AT-cut quartz plates, we consider mechanical displacement u(x,y,t) along x-axis and the stress equations :

$$T_1 = c_{11} \frac{\partial u}{\partial x} \qquad (2)$$

$$T_6 = c_{66} \frac{\partial u}{\partial y} \qquad (3)$$

where c_{11} and c_{66} do not depend upon k ; hence exist only thickness-shear and x-longitudinal stress. Newton's law in R_k

$$\frac{\partial T_1}{\partial x} + \frac{\partial T_6}{\partial y} = \rho_k \frac{\partial^2 u}{\partial t^2} \qquad (4)$$

leads, with exponential type solutions :

$$u_k = (A_k \sin\alpha_k x + B_k \cos\alpha_k x) \sin\beta_k y \exp(-i\omega t) \qquad (5)$$

with the steady-state assumption, to :

$$c_{11} \alpha_k^2 + c_{66} \beta_k^2 = \rho_k \omega^2 \qquad (6)$$

Boundary conditions

There are two kinds of boundary conditions : traction-free assumption on the faces of the plate, and continuity of displacement and normal stress across the separation between two adjacent regions.

First condition on $y = \pm h/2$:

$$T_6 = c_{66}\beta_k (A_k \sin\alpha_k x + B_k \cos\alpha_k x) \cos\beta_k \frac{h}{2} \exp(-i\omega t) = 0 \qquad (7)$$

yielding :

$$\beta_k = (2p-1) \frac{\pi}{h} \qquad (8)$$

then the wave number α_k :

$$\alpha_k^2 = \frac{\rho_k}{c_{11}} (\omega^2 - \omega_{pk}^2) \qquad (9)$$

where

$$\omega_{pk} = (2p-1) \frac{\pi}{h} \left(\frac{c_{66}}{\rho_k}\right)^{\frac{1}{2}} \qquad (10)$$

From now on, we shall omit the $\exp(-i\omega t)$ factor and simplify Eq. (10), omitting (2p-1) :

$$\omega_k = \frac{\pi}{h} \left(\frac{c_{66}}{\rho_k}\right)^{\frac{1}{2}} \qquad (11)$$

noticing that the following analysis, valid in the vicinity of the fundamental mode, can be extended for the (2p-1)-mode after reintroduction of this coefficient.

In R_k and R_{2n-k}, α_k is pure imaginary number when ω is lower than ω_k ; for ω greater than ω_k, α_k is real and propagation occurs along X-axis. According to common use, ω_k is the cut-off frequency of R_k, R_{2n-k}.

Continuity conditions on D_k

Continuity of u_k, Eq. (5), induces continuity of T_{6k}, since β is constant in the whole structure, Eq. (8) ; then we have just to express the continuity of u_k and T_{1k} :

$$T_{1k} = c_{11}\alpha_k (A_k \cos\alpha_k x - B_k \sin\alpha_k x) \sin\beta y \qquad (12)$$

Omitting factors independent of k, we have to write continuity of :

$$A_k \sin\alpha_k x + B_k \cos\alpha_k x \qquad (13)$$

$$\alpha_k (A_k \cos\alpha_k x - B_k \sin\alpha_k x) \qquad (14)$$

In R_o and R_{2n} we shall modify the notations : actually in both these infinite regions, no reflections may occur. and we shall use the following functions :

$$u_o = A_o \exp\{-i(\alpha_o x + \omega t)\} \qquad (15)$$

$$u_{2n} = A_{2n} \exp\{i(\alpha_o x - \omega t)\} \qquad (16)$$

with

$$\alpha_o^2 = \frac{\rho_o}{c_{11}} (\omega^2 - \omega_o^2) \qquad (17)$$

If $\omega > \omega_o$

$$\alpha_o = \left\{\frac{\rho_o}{c_{11}} (\omega^2 - \omega_o^2)\right\}^{\frac{1}{2}} \qquad (18)$$

in order to get propagation to the left in R_o and to the right in R_{2n}.

If $\omega < \omega_o$

$$\alpha_o = i\gamma_o \qquad (19)$$

in order to get an attenuated wave to the left in R_o and to the right in R_{2n}, with

$$\gamma_o = \left\{\frac{\rho_o}{c_{11}} (\omega_o^2 - \omega^2)\right\}^{\frac{1}{2}} \qquad (20)$$

We shall write continuity conditions on $-D_1$, $-D_k$, $-D_n$, D_n, D_k and D_1 :

Conditions on $-D_1$:

$$A_o \exp i\alpha_o D_1 = -A_1 \sin\alpha_1 D_1 + B_1 \cos\alpha_1 D_1 \qquad (21)$$

$$-i\alpha_o A_o \exp i\alpha_o D_1 = \alpha_1 (A_1 \cos\alpha_1 D_1 + B_1 \sin\alpha_1 D_1) \qquad (22)$$

Conditions on $-D_k$:

$$-A_{k-1}\sin\alpha_{k-1}D_k+B_{k-1}\cos\alpha_{k-1}D_k = -A_k\sin\alpha_k D_k+B_k\cos\alpha_k D_k \quad (23)$$

$$\alpha_{k-1}\{A_{k-1}\cos\alpha_{k-1}D_k+B_{k-1}\sin\alpha_{k-1}D_k\} = \alpha_k\{A_k\cos\alpha_k D_k+B_k\sin\alpha_k D_k\} \quad (24)$$

Conditions on $-D_n$:

$$-A_{n-1}\sin\alpha_{n-1}D_n+B_{n-1}\cos\alpha_{n-1}D_n = -A_n\sin\alpha_n D_n+B_n\cos\alpha_n D_n \quad (25)$$

$$\alpha_{n-1}\{A_{n-1}\cos\alpha_{n-1}D_n+B_{n-1}\sin\alpha_{n-1}D_n\} = \alpha_n\{A_n\cos\alpha_n D_n+B_n\sin\alpha_n D_n\} \quad (26)$$

Conditions on D_n :

$$A_n\sin\alpha_n D_n+B_n\cos\alpha_n D_n = A_{n+1}\sin\alpha_{n-1}D_n+B_{n+1}\cos\alpha_{n-1}D_n \quad (27)$$

$$\alpha_n\{A_n\cos\alpha_n D_n-B_n\sin\alpha_n D_n\} = \alpha_{n-1}\{A_{n+1}\cos\alpha_{n-1}D_n-B_{n+1}\sin\alpha_{n-1}D_n\} \quad (28)$$

Conditions on D_k :

$$A_{2n-k}\sin\alpha_k D_k+B_{2n-k}\cos\alpha_k D_k = A_{2n-k+1}\sin\alpha_{k-1}D_k+B_{2n-k+1}\cos\alpha_{k-1}D_k \quad (29)$$

$$\alpha_k\{A_{2n-k}\cos\alpha_k D_k-B_{2n-k}\sin\alpha_k D_k\} = \alpha_{k-1}\{A_{2n-k+1}\cos\alpha_{k-1}D_k-B_{2n-k+1}\sin\alpha_{k-1}D_k\} \quad (30)$$

Conditions on D_1 :

$$A_{2n-1}\sin\alpha_1 D_1+B_{2n-1}\cos\alpha_1 D_1 = A_{2n}\exp i\alpha_o D_1 \quad (31)$$

$$\alpha_1\{A_{2n-1}\cos\alpha_1 D_1-B_{2n-1}\sin\alpha_1 D_1 = i\alpha_o A_{2n}\exp i\alpha_o D_1 \quad (32)$$

On the 2n separations between the (2n+1) regions, we wrote 4n homogeneous linear equations between the A_k and B_k : there are $2(2n-1)$ A_k and B_k for k from 1 to (2n-1), and A_o and A_{2n} : we have an homogeneous system of 4n linear equations between the 4n unknowns A_k and B_k.

The 4n-order determinant Δ_n of this system gives the resonance frequencies equation of the n-order structure.

General theory of Δ_n determinants

Canonical form of Δ_n

The determinant of the system of equations in A_k and B_k, may be written in such a way that the symetry of the problem will appear more clearly. Details of sign inversions to perform on different rows and columns will not be indicated here (see appendix 1).

Let

$$\phi_k = \alpha_k D_k \quad (33)$$

$$\phi'_k = \alpha_k D_{k+1} \quad (34)$$

$$\theta_k = \phi_k - \phi'_k = \alpha_k L_k \quad (35)$$

for k = 1, 2, ..., (n-1), and

$$\theta_n = \phi_n = \alpha_n D_n \quad (36)$$

Δ_n is given in Fig. (2).

Development of Δ_n

We develop Δ_n in respect of 2n-order minors in the following way (cf. Eq. (37)) :

Only two 2n-order minors extracted from the 2n first rows give non zero products. They are built on the array (I), of 2n rows and (2n-1) columns, by addition, in 2nth column, of corresponding terms of column 2n or (2n+1) of Δ_n : these two minors are called M_{2n} and M_{2n+1}

(37) $\Delta_n =$

	C_{2n}	C_{2n+1}	
(I)	0	0	
	$-\sin\phi_n$	$-\cos\phi_n$	0 ...
	$\alpha_n\cos\phi_n$	$-\alpha_n\sin\phi_n$	0 ... 0
0 ... 0	$-\sin\phi_n$	$\cos\phi_n$	
... 0	$\alpha_n\cos\phi_n$	$\alpha_n\sin\phi_n$	(II)
	0	0	

The cofactors minors of M_{2n} and M_{2n+1} are the 2n-order minors built on the array (II) by addition in first column of the respectively corresponding terms of columns (2n+1) or 2n of Δ_n;

we call N_{2n+1} and N_{2n} these cofactors. We get then

$$\Delta_n = M_{2n}N_{2n+1} - M_{2n+1}N_{2n} \quad (38)$$

It is easy to verify that, after an even number of permutation of rows and columns :

$$M_{2n} = N_{2n} \quad (39)$$

$$M_{2n+1} = -N_{2n+1} \quad (40)$$

yielding finally :

$$\Delta_n = 2M_{2n}N_{2n+1} \quad (41)$$

We change the notation of M_{2n} and N_{2n+1} into T_n and S_n which have the following forms :

$$S_n = \begin{vmatrix} & 0 \\ (I) & \vdots \\ & 0 \\ & \cos\phi_n \\ & \alpha_n \sin\phi_n \end{vmatrix} \quad (42)$$

$$T_n = \begin{vmatrix} & 0 \\ (I) & \vdots \\ & 0 \\ & -\sin\phi_n \\ & \alpha_n \cos\phi_n \end{vmatrix} \quad (43)$$

where (I) is the array extracted from the 2n first rows and (2n-1) first columns of Δ_n.

Resonance frequencies equation is then :

$$\Delta_n = 2S_nT_n = 0 \quad (44)$$

giving two series of solutions referring to roots of S_n and T_n : it is shown in appendix-2, that the S_n roots are related to the symetrical modes, and the T_n roots to the antisymetrical modes.

Recurrence relationship

If we develop S_n and T_n, given by Eq. (42) and (43), we find the following formulas :

$$S_n = S_{n-1}\alpha_{n-1}\cos\theta_n + T_{n-1}\alpha_n\sin\theta_n \quad (45)$$

$$T_n = -S_{n-1}\alpha_{n-1}\sin\theta_n + T_{n-1}\alpha_n\cos\theta_n \quad (46)$$

where the S_{n-1} and T_{n-1} are the determinants defined for the (n-1)-structure obtained from the n-structure, by removing the center region R_n, and joining R_{n-1} and R_{n+1} into a single region, the center region R'_{n-1} of the (n-1)-structure. More details appear in appendix 3.

Study of Δ_n roots

We showed that the Δ_n roots divide in two sets, S_n roots and T_n roots. Equations can be written:

$$G_n = F_n \quad (47)$$

from S_n, Eq. (45) ; and

$$H_n = F_n \quad (48)$$

from T_n, Eq. (46), where we call :

$$G_n = -\alpha_n\tan\theta_n \quad (49)$$

$$H_n = \frac{\alpha_n}{\tan\theta_n} \quad (50)$$

$$F_n = \alpha_{n-1}\frac{S_{n-1}}{T_{n-1}} \quad (51)$$

Several properties are demonstrated in appendix 4 : G_n and H_n are real decreasing functions whatever ω is ;

F_n is real when, and only when, α_o is pure imaginary : thus all the roots of Eq. (47) and (48) are lower than ω_o ;

F_n is a real increasing function for any ω lower than ω_o.

To obtain the roots of Eq. (47) and (48), we discuss functions V_n and W_n

$$V_n = F_n - G_n \quad (52)$$

$$W_n = F_n - H_n \quad (53)$$

depending upon the variable X, normalized frequency given by :

$$X = \frac{\omega^2}{\omega_o^2} - 1 \quad (54)$$

the interval to be considered is (-1, 0), corresponding to the frequency interval (0, ω_o) containing, as just seen above, all the roots of V_n and W_n. Inside this interval, V_n and W_n are real increasing functions, and it will be possible to predict the number of their roots, from the

number of their poles and the sign of their values for X = - 1 and X = 0.

To simplify the notation we shall write L_n instead of D_n. Each couple (R_k, R_{2n-k}) may be defined by a parameter Ω_k

$$\Omega_k = K \frac{L_k}{h} \left(\frac{1}{m_k^2} - 1\right)^{\frac{1}{2}} \tag{55}$$

where K depends only on the elastic constants of the material :

$$K = \pi \left(\frac{c_{66}}{c_{11}}\right)^{\frac{1}{2}} \tag{56}$$

$$m_k = \frac{\omega_k}{\omega_o} \tag{57}$$

From Eq. (52) and (53), it is obvious that poles of $V_n(X)$ are poles of $F_n(X)$ and poles of $G_n(X)$. If P(F) is the number of poles of a function F(X), we get :

$$P(V_n) = P(F_n) + P(G_n) \tag{58}$$

Between two successive poles, V_n has one zero, since it is increasing. We shall discuss later on the case where F_n and G_n have a common pole. Between -1 and the first pole of V_n, there is still one zero, since $V_n(-1)$ is negative. V_n has thus $P(V_n)$ zeros between -1 and its last negative pole. Between this pole and 0, presence of a zero of V_n depends on the sign of $V_n(0)$: if $V_n(0) \geq 0$, there will exist one pole in this interval.

Let

$$\lambda_n (\Omega_1, \ldots, \Omega_n) = V_n(0) = F_n(0) - G_n(0) \tag{59}$$

and

$$\mu_n (\Omega_1, \ldots, \Omega_n) = W_n(0) = F_n(0) - H_n(0) \tag{60}$$

and let us define two numbers ϵ_n and η_n, as follows

$$\epsilon_n = 0 \quad \text{if } \lambda_n < 0 \; ; \; \epsilon_n = 1 \quad \text{if } \lambda_n \geq 0 \tag{61}$$

$$\eta_n = 0 \quad \text{if } \mu_n < 0 \; ; \; \eta_n = 1 \quad \text{if } \mu_n \geq 0 \tag{62}$$

If Z(F) is the number of zeros of a function F(X), we come to :

$$Z(S_n) = P(V_n) + \epsilon_n \tag{63}$$

and, with the same discussion for W_n :

$$Z(T_n) = P(W_n) + \eta_n \tag{64}$$

Actually, if F_n and G_n have a common pole, it can be shown easily, that this pole is a root of $S_n = 0$.

We get finally

$$Z(\Delta_n) = P(V_n) + P(W_n) + \epsilon_n + \eta_n \tag{65}$$

It is shown in Appendix 4 that

$$P(G_k) = E\left(\frac{Q_k + 1}{2}\right) \tag{66}$$

$$P(H_k) = E\left(\frac{Q_k}{2}\right) \tag{67}$$

with

$$Q_k = E\left(2 \frac{\Omega_k}{\pi}\right) \tag{68}$$

E(N) meaning the integer part of a positive number N.

If Ω_k is pure imaginary, i.e. when $\omega_k < \omega_o$ or $m_k > 1$, we shall say that Q_k is zero.

It is also shown in Appendix 4 that

$$P(F_k) = Z(W_{k-1}) \tag{69}$$

leading to

$$Z(S_n) = \sum_{1}^{n-1} E\left(\frac{Q_k}{2}\right) + \sum_{1}^{n-1} \eta_k + E\left(\frac{Q_n + 1}{2}\right) + \epsilon_n \tag{70}$$

and

$$Z(T_n) = \sum_{1}^{n-1} E\left(\frac{Q_k}{2}\right) + \sum_{1}^{n-1} \eta_k + E\left(\frac{Q_n}{2}\right) + \eta_n \tag{71}$$

Functions $G_n(X)$ and $H_n(X)$ are never equal, their representative curves do not cross eachother, and the poles of one of them are the zeros of the other. Zeros of $V_n(X)$ and $W_n(X)$ are then interlaced in (-1, 0). First resonance occurs always on a symetrical mode as $G_n(X)$ is lower than $H_n(X)$ between -1 and P_{n1}. These remarks show that

$$Z(S_n) \geq Z(T_n) \geq Z(S_n) - 1 \tag{72}$$

We get on the other hand, from Eq. (70) and (71)

$$Z(\Delta_n) = \sum_{1}^{n-1} 2E\left(\frac{Q_k}{2}\right) + \sum_{1}^{n-1} 2\eta_k + Q_n + \epsilon_n + \eta_n \tag{73}$$

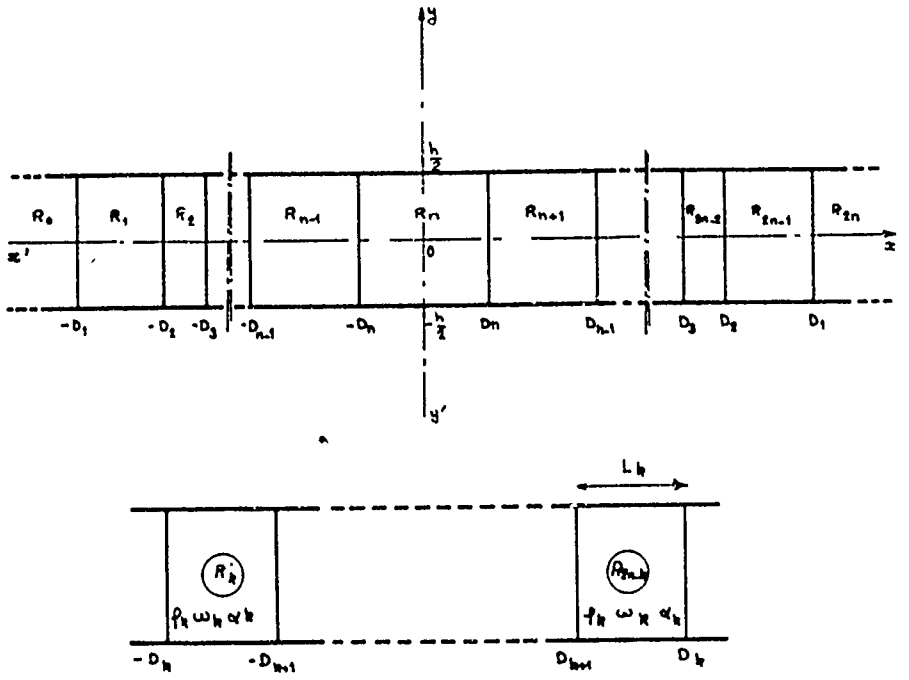


Fig. 1 Monolithic structure.

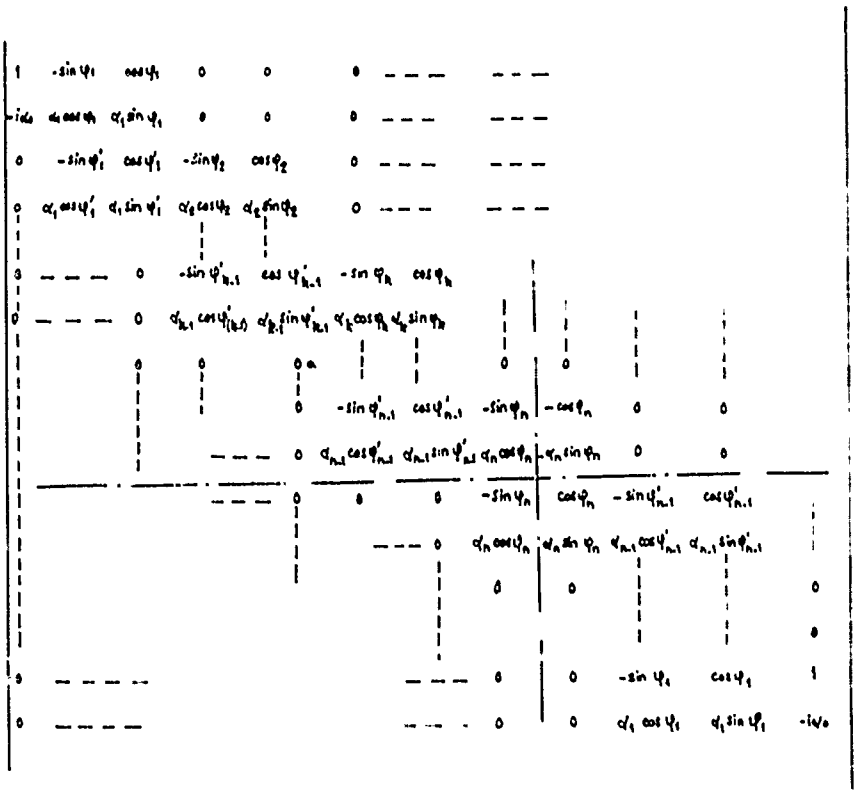


Fig. 2 Determinant Δ_n

If we consider, in a n-dimensions space, the point M_n of coordinates $\Omega_1, \ldots, \Omega_k, \ldots, \Omega_n$, it will be possible to know $Z(\Delta_n)$. The Q_k are immediatly obtained from the Ω_k. The μ_k, ϵ_n and η_n depend of the position of M_n, and of the M_k $(\Omega_1, \ldots, \Omega_k)$, in respect of surfaces having the following equations :

$$\lambda_n (\Omega_1, \ldots, \Omega_n) = 0, \text{ and } 1/\lambda_n = 0 \qquad (74)$$

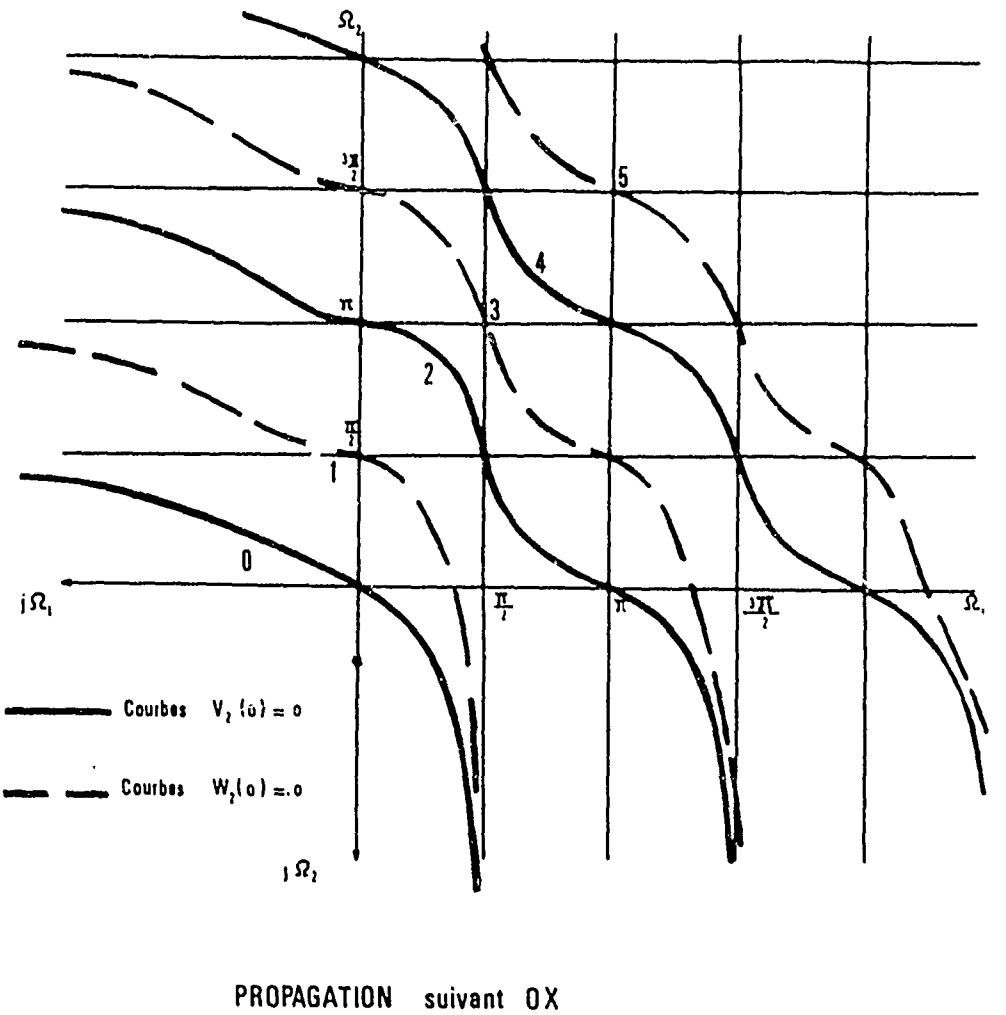
surfaces across which sign of λ_n changes

and μ_k (or $1/\mu_k$) = 0 $\quad k = 1, \ldots, n \qquad (75)$

surfaces where sign of μ_k changes.

The M_k are characteristic points of the k-structures deduced from the n-structure, as indicated before. These points are projections of M_n on k-dimensions sub-spaces.

The surfaces divide the space in bands where $Z(\Delta_n)$ is constant. Since the actual representation is rather troublesome in general case, we give, as illustration, the chart in (Ω_1, Ω_2) plane, for structures of 2nd order (Fig. 3).



PROPAGATION suivant OX

Fig. 3 Chart in (ω_1, ω_2) plane of resonance number of a 2nd-order-structure with $L_1/L_2 = 3.7$.

Application to monolithic filters

Corrections in piezoelectric crystals

Without solving general equations of piezoelectric structures, some corrections may be introduced to take in account influence of piezoelectricity, and of some coupling effects. These corrections will be mainly useful in numerical computation of resonance frequencies.

We know first that, in a piezoelectric crystal, effective elastic constants to be used differ slightly from the elastic constants c^E : [2]

$$c_{eff.} = c\ (1 + \kappa^2) \qquad (76)$$

κ is the electromechanical coupling factor of the material ; corrections depend on the direction of propagation, and of electric field.

Furthermore, for structures along Z-axis of AT-cut plate, the above discussion applies no more, due to coupling between T_5 and T_6, since c_{56} is non-zero. As first approximation, we use, instead of c_{11} for X-axis, $c_{55} - \frac{c_{56}^2}{c_{55}c_{66}}$ for Z-axis.

We know also that a piezoelectric resonator operates on $\lambda/2$, for the parallel resonance, and not for the serie resonance. At this frequency, Eq. (8) is no more valid, but we must use

$$\lambda_p = 2\ \xi_p/h \qquad (77)$$

where ξ_p is the p^{th} root of the equation [3,4]

$$z = \kappa^2 \tan z \qquad (78)$$

For low coupling factors, ξ_p is very close to $(2p-1)\ \pi/2$. For AT-cut of quartz, we find :

$$\xi_1 \simeq \frac{\pi}{2}\ (1 - 32.10^{-4}) \qquad (79)$$

This lowering effect on frequency, due to short-circuit conditions at serie-resonance, has to be taken in account for computation of the m_k. A resonator operating near the serie-resonance, has a cut-off frequency lowered by plate-back, and by the electric charge effect of Eq. (79). In monolithic filters with more than two coupled resonators, it must be noticed that only access resonators are subjected to the cut-off frequency lowering effect of Eq. (79), while this effect does not apply for an intermediate resonator, except if its electrodes are short circuited. Otherwise, intermediate resonators have a cut-off frequency related with operation near parallel resonance.

Computation of resonance frequencies of monolithic filters

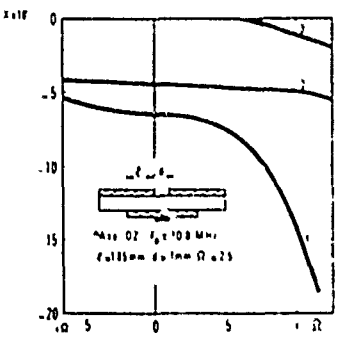
General discussion of Δ_n roots, with above corrections, can be used to compute resonance frequencies. The method is an iteration process giving roots of S_n and T_n from roots of T_{n-1}.

Effectively, poles of V_n are poles of G_n and roots of T_{n-1}. After ordering poles of V_n coming from G_n and T_{n-1}, we have just to apply any interpolation method to find the zeros of S_n between two successive poles of V_n. We do the same way with T_n from H_n and T_{n-1}. Tests of sign should be made on λ_n and μ_n to know if there is one root between the last pole and zero.

Filters tuning

Some authors[5] showed that the n first modes of a n-resonators filter are related to modes having 0 to (n-1) nodal lines, a nodal line in the center region being characteristic of antisymetrical modes.

Theoretical curves of resonance frequencies of a 2 resonators filter show that the first antisymetrical frequency is almost independent of the center region, or m_2 (Fig. 4).



Measurements achieved on 2 resonator plates with special electrodes configuration (Fig. 5) confirm this result. On these structures, m_2, then ξ_2, may be changed by additionnal plating. Variations due to change of m_2 are small for the first antisymetrical mode compared to the first symetrical mode.

Some experiments have been performed to extend this property : frequency of a given mode

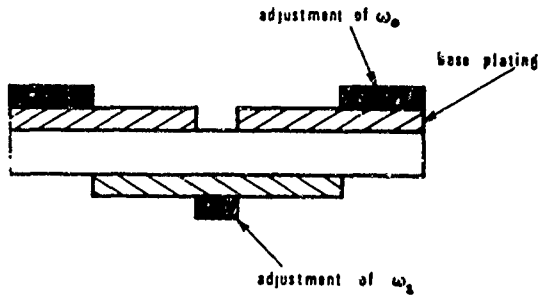


Fig. 5 Two-resonator plate.

is weakly dependent on variations produced in regions where this mode has a nodal line. Experimental curves are shown Fig. 6 and 7.

By specific action on cut-off frequencies of resonators, or of some interresonators region, it is possible to tune, almost independently, the different resonance frequencies.

In relation with the frequency tuning possibilities, we tried to define a monolithic filter by its resonance frequencies in short-circuit, rather than by an equivalent network, since these frequencies are the only easily measurable parameters. This approach in filter synthesis is being done in CNET, and could apply as well to filters of HCM type, developped most in Japan, for tuning of capacitors[6].

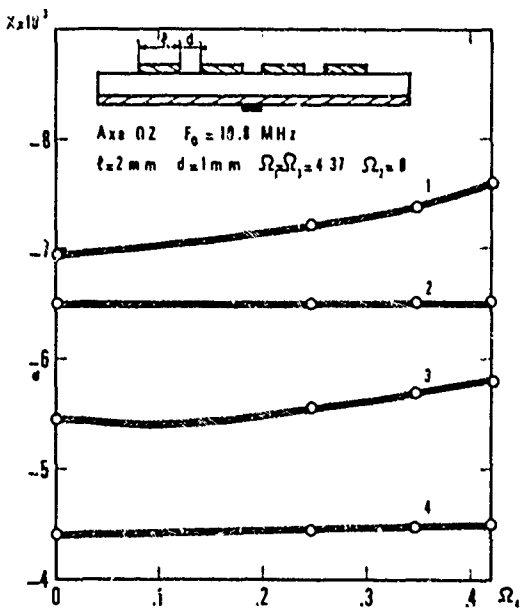


Fig. 6 Variation of the first four resonance frequencies of 4 resonator structure, as function of Ω_4.

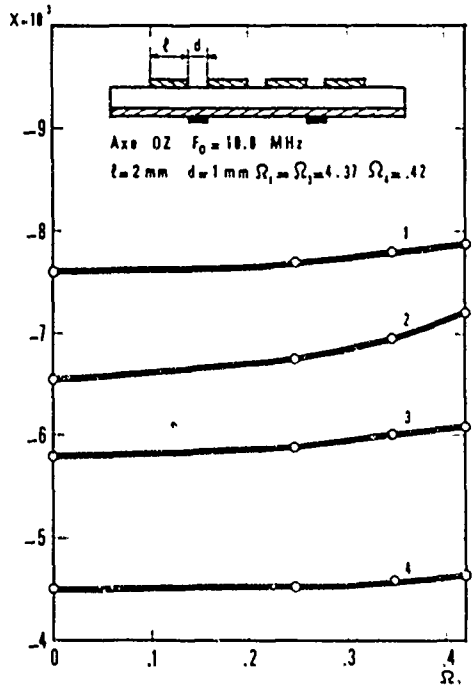


Fig. 7 Variation of the first four resonance frequencies of 4 resonator structure, as function of Ω_2.

This method could be of great interest specially for filters requiring weak ripples in passband, i.e. accurate tuning. Filters using two plates with two coupled resonators have been tuned, to achieve frequencies computed for ripple of .1 and .2 db, and follow the theoretical curves quite well.

Conclusion

Mathematical discussion of determinants relative to structures with n coupled resonators, lead to some general properties of resonance frequencies of these structures.

It had been shown that the design of spurious free filters was not quite equivalent to the design of single response resonators evaporated on the same plate. The problem is actually connected with properties of the whole monolithic structures.

Purity criterion are not of simple use, nor the tuning of resonance frequencies. But progress in knowledge of physical phenomenons could help to improve practical methods of fabrication of monolithic filters.

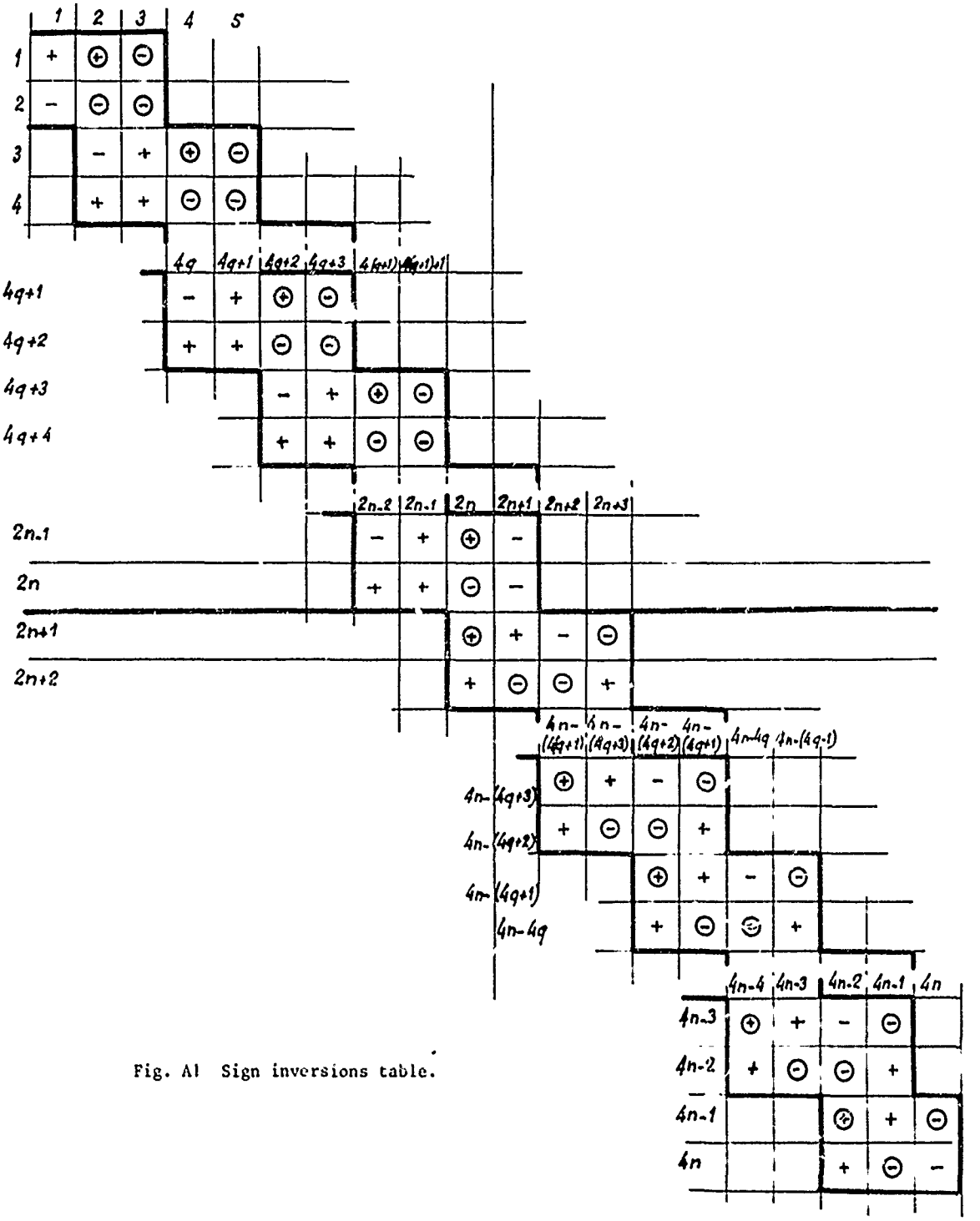


Fig. A1 Sign inversions table.

Appendix I

Canonical form of Δ_n

The determinant given in Fig. 2 is obtained, from the determinant of the linear system in A_k, B_k of boundary conditions, by sign inversions on some rows and columns.

With

$$A'_o = A_o \exp i\alpha_o D_1 \qquad A_{(1)}$$

$$A'_{2n} = A_{2n} \exp i\alpha_o D_1 \qquad A_{(2)}$$

the linear system of equations, from Eq. (21) to (32), is :

$$A'_o + A_1 \sin\phi_1 - B_1 \cos\phi_1 = 0 \qquad A_{(3)}$$

$$-i\alpha_o A'_o - A_1\alpha_1 \cos\phi_1 - B_1\alpha_1 \sin\phi_1 = 0 \qquad A_{(4)}$$

$$-A_{k-1}\sin\phi'_{k-1} + B_{k-1}\cos\phi'_{k-1} + A_k \sin\phi_k - B_k \cos\phi_k = 0 \qquad A_{(5)}$$

$$A_{k-1}\alpha_{k-1}\cos\phi'_{k-1} + B_{k-1}\alpha_{k-1}\sin\phi_{k-1} - A_k\alpha_k\cos\phi_k - B_k\alpha_k\sin\phi_k = 0 \qquad A_{(6)}$$

$$-A_{n-1}\sin\phi'_{n-1} + B_{n-1}\cos\phi'_{n-1} + A_n\sin\phi_n - B_n\cos\phi_n = 0 \qquad A_{(7)}$$

$$A_{n-1}\alpha_{n-1}\cos\phi'_{n-1} + B_{n-1}\alpha_{n-1}\sin\phi'_{n-1} - A_n\alpha_n\cos\phi_n - B_n\alpha_n\sin\phi_n = 0 \qquad A_{(8)}$$

$$A_n \sin\phi_n + B_n\cos\phi_n - A_{n+1}\sin\phi'_{n-1} - B_{n+1}\cos\phi'_{n-1} = 0 \qquad A_{(9)}$$

$$A_n\alpha_n\cos\phi_n - B_n\alpha_n\sin\phi_n - A_{n+1}\alpha_{n-1}\cos\phi'_{n-1} + B_{n+1}\alpha_{n-1}\sin\phi'_{n-1} = 0 \qquad A_{(10)}$$

$$A_{2n-k}\sin\phi_k + B_{2n-k}\cos\phi_k - A_{2n-k+1}\sin\phi'_{k-1} - B_{2n-k+1}\cos\phi'_{k-1} = 0 \qquad A_{(11)}$$

$$A_{2n-k}\alpha_k\cos\phi_k - B_{2n-k}\alpha_k\sin\phi_k - A_{2n-k+1}\alpha_{k-1}\cos\phi'_{k-1} + B_{2n-k+1}\sin\phi'_{k-1} = 0 \qquad A_{(12)}$$

$$A_{2n-1}\sin\phi_1 + B_{2n-1}\cos\phi_1 - A'_{2n} = 0 \qquad A_{(13)}$$

$$A_{2n-1}\alpha_1\cos\phi_1 - B_{2n-1}\alpha_1\sin\phi_1 - i\alpha_o A'_{2n} = 0 \qquad A_{(14)}$$

the two rows (2k-1) and 2k of the determinant, k integer from 2 to n, are

$$0..0 \; -\sin\phi'_{k-1} \quad \cos\phi'_{k-1} \quad \sin\phi_k \quad -\cos\phi_k \; 0..0 \qquad A_{(15)}$$

$$0..0 \; \alpha_{k-1}\cos\phi'_{k-1} \; \alpha_{k-1}\sin\phi'_{k-1} \; -\alpha_k\cos\phi_k \; -\alpha_k\sin\phi_k \; 0..0 \qquad A_{(16)}$$

the conjugated rows having the rank 4n-2k+1 and 4n-2k+2 :

$$0..0 \; \sin\phi_k \quad \cos\phi_k \quad -\sin\phi'_{k-1} \quad -\cos\phi'_{k-1} \; 0..0 \qquad A_{(17)}$$

$$0..0 \; \alpha_k\cos\phi_k \; -\alpha_k\sin\phi_k \; -\alpha_{k-1}\cos\phi'_{k-1} \; \alpha_{k-1}\sin\phi'_{k-1} \; 0..0 \qquad A_{(18)}$$

Inside appendix I, we shall call L_j and C_j, the jth row or column.

In L_{2k-1} and L_{2k}, terms non identically null appear only in C_{2k-2} to C_{2k+1} ; likewise in $L_{4n-2k+1}$ and $L_{4n-2k+2}$ non-zero terms appear only in C_{4n-2k} to $C_{4n-2k+3}$. The pattern of the determinant of equations $A_{(3)}$ to $A_{(14)}$ is shown Fig. $A_{(1)}$, where only signs are indicated, encircled, terms the sign of which must be changed in order to get the determinant Δ_n of Fig. 2.

Sign inversions of C_{2n}, C_{2n+3} and L_{2n+2} lead to proper result for rows L_{2n-1} to L_{2n+2}, with 3 inversions.

We shall discuss L_j, C_j, for $j < 2n$, and $j > 2n$.

Low indices

In Fig. $A_{(1)}$, is shown set of four rows L_{4q+1} to L_{4q+4} ; noticing that L_{4q+3} to L_{4q+6} have the same pattern, these two sets are L_{2k+1} to L_{2k+4}, with k even or odd.

The method :

sign inversion on L_{2k+3}, L_{2k+4}, C_{2k+2}, C_{2k+3} yields the convenient signs on these four rows, k having a given parity. $A_{(19)}$

According to n parity two ways will lead to Δ_n (Fig. 2.)

1. n is even

$$n = 2p \qquad A_{(20)}$$

the two first rows to be changed, starting from the center of Fig. $A_{(1)}$, are L_{2n-3} and L_{2n-2} : k must satisfy

$$2k+3 = 2n-3 \qquad A_{(21)}$$

yielding

$$k = 2(p-1)-1 \qquad A_{(22)}$$

Method $A_{(19)}$ must be applied with odd k

$$k = 2q+1 \qquad A_{(23)}$$

set of four rows L_{4q+3} to L_{4q+6} has convenient signs after application of $A_{(19)}$, q decreasing from (p-2) to 0 : the last sign inversions concern L_3 to L_6. After inversion of L_1, L_2 and C_1, convenient signs are obtained on the array L_1 to L_{2n-2}, after an odd number of sign inversions.

2. n is odd

$$n = 2p-1 \qquad A_{(24)}$$

Eq. $A_{(21)}$ yields in that case

$$k = 2(p-2) \qquad A_{(25)}$$

$A_{(19)}$ must be applied with even k :

$$k = 2q \qquad A_{(26)}$$

This time, set of four rows L_{4q+1} to L_{4q+4} will be transformed into convenient form, by application of $A_{(19)}$, for q decreasing from (p-2) to 0 : last inversions concern L_1 to L_4, leading to convenient signs after an even number of inversions.

High indices

The sets of two rows $L_{4n-(4q+3)}$, $L_{4n-(4q+2)}$, or $L_{4n-(4q+1)}$, L_{4n-4q}, correspond to $L_{4n-(2k+1)}$, L_{4n-2k}, with k odd or even

$$k = 2q+1 \qquad A_{(27)}$$

or

$$k = 2q \qquad A_{(28)}$$

Two methods may be applied to get required signs :

Sign inversions on $C_{4n-(2k+1)}$, C_{4n-2k} and $L_{4n-(2k+1)}$ $A_{(29)}$

Sign inversions on $C_{4n-(2k+2)}$, $C_{4n-(2k-1)}$ and L_{4n-2k} $A_{(30)}$

The first column, to be changed in sign, starting from the center, is C_{2n+3}. Depending on n parity, we still find two cases :

1. n is even

$$n = 2p \qquad A_{(31)}$$

k must satisfy

$$4n-(2k+1) = 2n+3 \qquad A_{(32)}$$

in order to apply $A_{(29)}$ or $A_{(30)}$ from L_{2n+3} ; we get

$$k = 2(p-1) \qquad A_{(33)}$$

and $A_{(29)}$ must be first applied to change sign of C_{2n+3}.

It can be easily verified that the two methods alternate, due to column indices relation-ship : $A_{(29)}$ on $L_{4n-(2k+1)}$ and L_{4n-2k} for k even decreasing from 2(p-1) ; $A_{(30)}$ on $L_{4n-(2k+1)}$ and L_{4n-2k} for k odd decreasing from 2(p-3).

Set of L_{4n-1} and L_{4n} reffers to even k, then to application of $A_{(29)}$ for k = 0 ; sign inversion will not be performed on C_{4n}, in order to get the required signs.

An even number of inversions has been performed on high indices rows.

2. n is odd

$$n = 2p-1 \qquad A_{(34)}$$

in this case Eq. $A_{(31)}$ yields to

$$k = 2p-3 \qquad A_{(35)}$$

the same discussion as for n even, can be applied but alternate application of $A_{(29)}$ and $A_{(30)}$ is performed with, respectively, k odd decreasing from 2p-3, and k even from 2p-4.

The rows L_{4n-1} and L_{4n} reffer to even k, and the last sign inversion is $A_{(30)}$ for k = 0, with inversion on C_{4n} instead, of course, of C_{4n+1}. An even number of inversions has been performed on high indices rows.

To sum up, sign inversions have been performed on the rows and columns of the determinant in order to get the required signs, in an even number if n is even, in an odd number if n is odd.

Since we are only concerned in the roots of the continuity equations determinant, we shall discuss the roots of Δ_n, obtained by the above method, which differs from the primitive determinant, eventually by the sign only.

Appendix 2

Study of modes related to S_n and T_n

For a resonance frequency of the n-structure $\Delta_n = 0$, and we can compute the A_k and B_k, as a function of one of them, A'_o for instance.

We shall not compute actually these coefficients, but we are going to show how A_n and B_n are related to S_n and T_n.

Deleting the last equation (32) of the linear system, the new system can be written, with matrix notation :

$$(\Delta'_n)\ Y = A'_o Y_o \qquad A_{(34)}$$

where $(4n-1)$-order square matrix (Δ'_n) is the array Δ_n where last row and first column have been deleted ; Y and Y_o are column matrices :

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_{4n-2} \\ Y_{4n-1} \end{pmatrix} = \begin{pmatrix} A_1 \\ B_1 \\ \cdot \\ \cdot \\ \cdot \\ B_{2n-1} \\ A'_{2n} \end{pmatrix} \qquad A_{(35)}$$

$$Y_o = \begin{pmatrix} 1 \\ -i\alpha_o \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \qquad A_{(36)}$$

Solution of system $A_{(34)}$ is given by

$$Y_i = \frac{\delta'_i}{\Delta'_n} A'_o \qquad A_{(37)}$$

δ'_i being the determinant obtained by substitution in Δ'_n of the ith column by Y_o.

We shall compute δ'_{2n-1} and δ'_{2n} which are related to A_n and B_n. δ'_{2n-1} can be written :

$$\delta'_{2n-1} = \begin{vmatrix} -\sin\varphi_1 & \cos\varphi_1 & 0 \cdots 0 & 1 & & 0 & 0 \\ \alpha_1\cos\varphi_1 & \alpha_1\sin\varphi_1 & 0 \cdots 0 & -i\alpha_0 & & & \\ -\sin\varphi'_1 & \cos\varphi'_1 & 0 \cdots 0 & 0 & & & \\ & & \cos\varphi'_{n-1} & 0 & -\cos\varphi_n & 0 & \\ & & \alpha_{n-1}\sin\varphi'_{n-1} & 0 & -\alpha_n\sin\varphi_n & 0 \cdots & \\ & & & 0 & & & 0 \\ & & & 0 & & -\sin\varphi_1 \ \cos\varphi_1 & 1 \end{vmatrix} \qquad A_{(38)}$$

C'_{2n-1} C'_{2n}

All 2n-order minors of δ'_{2n-1}, built on its 2n first rows are equal to zero, except the minor built on the 2n first columns. After transfer of the (2n-1)th column in first position we find that this 2n-order minor is S_n, except eventually the sign. A_n is thus proportional to S_n, and it can be shown likewise that B_n is proportional to T_n.

If $A_n = 0$, we get easily

$$u_k(-x) = u_{2n-k}(x) \qquad A_{(39)}$$

whereas, if $B_n = 0$

$$u_k(-x) = -u_{2n-k}(x) \qquad A_{(40)}$$

Roots of S_n are resonance frequencies of modes defined by $A_{(39)}$, symetrical modes, roots of T_n referring to antisymetrical modes defined by $A_{(40)}$.

Appendix 3

Recurrence formulas on S_n and T_n

S_n and T_n are given by Eq. (42) and (43) ; development of S_n, for instance, in respect to its last column, yields

$$S_n = \alpha_n \sin\phi_n \Delta_{1n} - \cos\phi_n \Delta'_{1n}$$

where Δ_{1n} and Δ'_{1n} are (2n-1)-order determinants:

$$\Delta_{1n} = \begin{vmatrix} & \Delta^\circ_{1n} & \\ \cdots & \cdots & \cdots \\ 0 \cdots 0 & -\sin\phi'_{n-1} & \cos\phi'_{n-1} \end{vmatrix} \qquad A_{(42)}$$

$$\Delta'_{1n} = \begin{vmatrix} & \Delta^\circ_{1n} & \\ \cdots & \cdots & \cdots \\ 0 \cdots 0\alpha_{n-1}\cos\phi'_{n-1} & \alpha_{n-1} & \sin\phi'_{n-1} \end{vmatrix} \qquad A_{(43)}$$

and Δ°_{1n} the array extracted from the (2n-2) first rows and (2n-1) first columns of Δ_n, or S_n, or T_n. Development of Δ_{1n} and Δ'_{1n} in respect to the last row yields

$$\Delta_{1n} = \cos\phi'_{n-1}\Delta_{2,n} + \sin\phi'_{n-1}\Delta'_{2,n} \qquad A_{(44)}$$

$$\Delta'_{1n} = \alpha_{n-1}(\sin\phi'_{n-1}\Delta_{2,n} - \cos\phi'_{n-1}\Delta'_{2,n}) \qquad A_{(45)}$$

It can be shown further :

$$\Delta_{2,n}=\alpha_{n-1}\cos\phi_{n-1}\Delta_{3,n}+\sin\phi_{n-1}\Delta'_{3,n} \qquad A_{(46)}$$

$$\Delta'_{2,n}=\alpha_{n-1}\sin\phi_{n-1}\Delta_{3,n}-\cos\phi_{n-1}\Delta'_{3,n} \qquad A_{(47)}$$

leading to

$$\Delta_{1n}=\alpha_{n-1}\cos\theta_{n-1}\Delta_{3,n}+\sin\theta_{n-1}\Delta'_{3,n} \qquad A_{(48)}$$

$$\Delta'_{1n}=-\alpha_{n-1}\Delta''_{1n}=-\alpha_{n-1}(\alpha_{n-1}\sin\theta_{n-1}\Delta_{3,n}-\cos\theta_{n-1}\Delta'_{3,n}) \qquad A_{(49)}$$

We see then, that

$$\Delta_{1n}=\begin{vmatrix} & & & 0 \\ & \Delta & & \vdots \\ & & & 0 \\ 0\cdots 0 & -\sin\phi'_{n-2} & \cos\phi'_{n-2} & -\sin\theta_{n-1} \\ 0\cdots 0\alpha_{n-2}\cos\phi'_{n-2} & \alpha_{n-2}\sin\phi'_{n-2} & \alpha_{n-1}\cos\theta_{n-1} & \end{vmatrix} \qquad A_{(50)}$$

$$\Delta''_{1n}=\begin{vmatrix} & & 0 & \\ & \Delta & \vdots & \\ & & 0 & \\ & & \cos\theta_{n-1} & \\ & & \alpha_{n-1}\sin\theta_{n-1} & \end{vmatrix} \qquad A_{(51)}$$

where Δ is the array built on the $2(n-1)$ first rows and $2(n-1)-1$ first colums of Δ_n.

We can notice then, that in Δ, terms related to R_{n-1} do not appear, and that in Δ_{1n} and Δ''_{1n}, only the dimension L_{n-1} is to be considered to define R_{n-1}; By iteration of the development we could show that in Δ_n or S_n or T_n, only the L_k are to be considered, except D_n in the center region.

The determinants Δ_{1n} and Δ''_{1n} are, then, T'_{n-1} and S'_{n-1}, relative to the (n-1)-resonators structure inferred from the n-resonators structure, by removing R_n, and joining the two half-structures obtained into a single one, such as R_{n-1} and R_{n+1} give a single region R'_{n-1} in the new structure ; R'_{n-1} is defined by the abscissa $\pm D'_{n-1}$

$$D'_{n-1}=L_{n-1} \qquad A_{(52)}$$

As for the L_k of other regions, they remain constant. We shall note S_{n-1} and T_{n-1}, the determinants S'_{n-1} and T'_{n-1} corresponding to the (n-1)-structure inferred as said above, from the n-structure.

We find then

$$S_n=\alpha_n\sin\theta_n T_{n-1}+\alpha_{n-1}\cos\theta_n S_{n-1} \qquad A_{(53)}$$

and likewise

$$T_n=\alpha_n\cos\theta_n T_{n-1}-\alpha_{n-1}\sin\theta_n S_{n-1} \qquad A_{(54)}$$

Appendix 4

Study of functions $F_k(X)$, $G_k(X)$, $H_k(X)$

Functions $G_k(X)$ and $H_k(X)$

Let remind, for k = 1, 2,...,n

$$\theta_k(X)=L_k\alpha_k(X) \qquad A_{(55)}$$

with, for simplitiy

$$L_n=D_n \qquad A_{(56)}$$

and

$$\alpha_k(X)=\frac{K}{hm_k}(X-X_k)^{\frac{1}{2}}$$

with the following notations

$$m_k=\frac{\omega_k}{\omega_o} \qquad A_{(57)}$$

$$X_k=m_k^2-1 \qquad A_{(58)}$$

normalized cut-off frequency in (R_k, R_{2n-k}) ; and

$$K=\pi\left(\frac{c_{66}}{c_{11}}\right)^{\frac{1}{2}} \qquad A_{(59)}$$

When $X < X_k$, α_k and θ_k are pure imaginary, and we shall introduce real quantities γ_k and τ_k:

$$\alpha_k=i\gamma_k=i\frac{K}{hm_k}(X_k-X)^{\frac{1}{2}} \qquad A_{(60)}$$

$$\theta_k=i\tau_k=i\frac{K}{m_k}\frac{L_k}{h}(X_k-X)^{\frac{1}{2}} \qquad A_{(61)}$$

giving for G_k and H_k, the expressions :

$$G_k(X)=\gamma_k\tanh\tau_k \qquad A_{(62)}$$

$$H_k(X)=\frac{\gamma_k}{\tanh\tau_k} \qquad A_{(63)}$$

when $X > X_k$, we have

$$G_k=-\alpha_k\tan\theta_k \qquad A_{(64)}$$

$$H_k = \frac{\alpha_k}{\tan\theta_k} \qquad A_{(65)}$$

and, at last,

$$G_k\ (X_k) = 0 \qquad A_{(66)}$$

$$H_k\ (X_k) = \frac{1}{L_k} \qquad A_{(67)}$$

We notice immediatly that G_k and H_k are always real functions of X ; furthermore, they are decreasing functions : it can be shown that $G_k'(X)$ has the sign of :

$$-\frac{1}{\cos^2\theta_k}\left(1 + \frac{\sin 2\theta_k}{2\theta_k}\right) \qquad A_{(68)}$$

and that $H_k'(X)$ has the sign of :

$$-\frac{1}{\sin^2\theta_k}\left(1 - \frac{\sin 2\theta_k}{2\theta_k}\right) \qquad A_{(69)}$$

both negative expressions, whatever is θ_k, real or pure imaginary. We get also :

$$G_k(-1) = \frac{K}{h}\tanh K\frac{L_k}{h} \qquad A_{(70)}$$

$$H_k(-1) = \frac{K}{h}\frac{1}{\tanh K\frac{L_k}{h}} \qquad A_{(71)}$$

both positive quantities.

We shall notice, at last, that, since $\tan^2\theta_k \neq -1$ for θ_k real or imaginary, $G_k(X)$ and $H_k(X)$ cannot be equal, and their curves do not cross eachother.

Function $F_k(X)$

We have just seen before that G_n and H_n are always real . Roots of Eq. (47) and (48) should belong to the domain of the variable X where F_n is real. But, from Eq. (51) we may write

$$F_n = -H_{n-1}\frac{F_{n-1} - G_{n-1}}{F_{n-1} - H_{n-1}} \qquad A_{(72)}$$

As H_{n-1} and G_{n-1} are real and unequal quantities, it is easily shown that F_n is real if F_{n-1} is real, and only in that case, then, by recurrence, if F_1 is real.

To get $F_1(X)$, we only have to consider the 1-order structure, i.e. single resonator, and its determinant Δ_1 :

$$\Delta_1 = \begin{vmatrix} 1 & -\sin\theta_1 & -\cos\theta_1 & 0 \\ -i\alpha_o & \alpha_1\cos\theta_1 & -\alpha_1\sin\theta_1 & 0 \\ 0 & -\sin\theta_1 & \cos\theta_1 & 1 \\ 0 & \alpha_1\cos\theta_1 & \alpha_1\sin\theta_1 & -i\alpha_o \end{vmatrix} = -2S_1T_1 \qquad A_{(73)}$$

and we know immediatly :

$$S_1 = \alpha_1\sin\theta_1 + i\alpha_o\cos\theta_1 \qquad A_{(74)}$$

$$T_1 = \alpha_1\cos\theta_1 - i\alpha_o\sin\theta_1 \qquad A_{(75)}$$

equations leading to :

$$G_1 = -\alpha_1\tan\theta_1 \qquad A_{(76)}$$

$$H_1 = \frac{\alpha_1}{\tan\theta_1} \qquad A_{(77)}$$

$$F_1(X) = i\alpha_o(X) = i\frac{K}{h}\sqrt{X} \qquad A_{(78)}$$

The condition to be satisfied in order to have all F_k real is thus :

$$X < 0 \quad \text{or} \quad \omega < \omega_o \qquad A_{(79)}$$

From Eq. (19) and (20), we get

$$F_1(X) = -\gamma_o(X) = -\frac{K}{h}\sqrt{-X} \qquad A_{(80)}$$

This result is very important : all resonance frequencies of the structures are lower than the cut-off frequency of the infinite surrounding plate, what agrees perfectly with the energy trapping theory and the results indicated by many authors for single resonator.

We shall see now that $F_k(X)$ are increasing functions for $-1 < X < 0$: we verify in Eq. $A_{(80)}$ that $F_1(X)$ is increasing function of X, and it will be shown that, if we assume that F_{n-1} is increasing, then F_n is increasing too, recurrence demonstration leading to the required property.

We know already that G_{n-1} and H_{n-1} are real and decreasing for any X, we know that, in $(-1,0)$, $F_{n-1}(X)$ is real, and we assume it is increasing.

By derivation of Eq. $A_{(72)}$, and omitting (n-1) index of F_{n-1}, G_{n-1}, H_{n-1}, L_{n-1} and θ_{n-1}, F_n' can be written as follows :

$$F'_n(X) = -H'\left(F-\frac{GH'+G'H}{2H'}\right)^2 + F'H(H-G) + \frac{H(H-G)(\theta^2)'^2\left(\frac{1}{\theta^2}-\frac{1}{\sin^2\theta}\right)}{4L^2H'} \Big/ (F-H)^2 \qquad A_{(81)}$$

It can be verified that all three quantities appearing in numerator are positive : then $F_k(X)$ is a real increasing function of X in the interval (-1, 0).

Poles of F_k, G_k and H_k

Poles of $F_k(X)$. Function $F_k(X)$ is given by Eq.(51) which shows obviously that poles of $F_k(X)$ are roots of $T_{k-1}(X)$, thus of $W_{k-1}(X)$, and zeros of $F_k(X)$ are roots of $S_{k-1}(X)$, thus of $V_{k-1}(X)$; X_{k-1} is a zero of F_k.

Poles and zeros of $G_k(X)$, $H_k(X)$. When $X_k > 0$, G_k and $H_k(X)$ are given by Eq. $A_{(60)}$ and $A_{(61)}$, and have neither poles nor zeros in the interval (-1, 0), where they are positive.

When $X_k = 0$, $H_k(X)$ is still positive, and $G_k(X)$ is positive except for $G_k(X_k) = 0$.

When $X_k < 0$, these functions may have poles, and we shall see in what conditions. Let

$$P_{kj} = m_k^2\left\{1 + \left(j\,\frac{\pi h}{2KL_k}\right)^2\right\} - 1 \qquad A_{(82)}$$

It can be easily verified that

$$\theta_k(P_{kj}) = j\,\frac{\pi}{2} \quad (j = 1, 2, \ldots) \qquad A_{(83)}$$

showing that :
if j is odd, P_{kj} is pole of G_k and zero of H_k, while, if j is even, P_{kj} is pole of H_k and zero of G_k.

Only negative P_{kj}, thus inside the interval (-1, 0), are of interest in our problem. If we let

$$\Omega_k = \theta_k(0) = K\,\frac{L_k}{h}\left(\frac{1}{m_k^2}-1\right)^{\frac{1}{2}} \qquad A_{(84)}$$

we get the following result : if

$$j\,\frac{\pi}{2} \leq \Omega_k < (j+1)\,\frac{\pi}{2} \qquad A_{(85)}$$

then

$$P_{kj} \leq 0 < P_{k(j+1)} \qquad A_{(86)}$$

Using the notation E(N) for the integer part of a positive number N, we shall note

$$Q_k = E\left(\frac{2\Omega_k}{\pi}\right) \qquad A_{(87)}$$

and Eq. (85) and A(86) mean that the Q_k first P_{kj} are in the interval (-1, 0).

If Q_k is even, G_k and H_k have each $Q_k/2$ poles in (-1, 0) ; if Q_k is odd, G_k has $(Q_k+1)/2$ poles, and H_k, $(Q_k-1)/2$.

More generally, it can be verified that G_k has $E\left(\frac{Q_k+1}{2}\right)$ poles, and H_k, $E\left(\frac{Q_k}{2}\right)$ poles.

We get besides

$$G_k(0) = -\frac{1}{L_k}\,\Omega_k\tan\Omega_k \qquad A_{(88)}$$

$$H_k(0) = \frac{1}{L_k}\,\frac{\Omega_k}{\tan\Omega_k} \qquad A_{(89)}$$

To complete the discussion we need the sign of $V_k(-1)$ and $W_k(-1)$. We noticed that $G_k(-1)$ and $H_k(-1)$ are both negative quantities, cf. Eq. A(70) and A(71). If we show that $F_k(-1)$ is negative, $V_k(-1)$ and $W_k(-1)$ will then be negative.

From Eq. A(72)

$$F_k(X) = -H_{k-1}(X)\,\frac{V_{k-1}(X)}{W_{k-1}(X)} \qquad A_{(90)}$$

If we assume $F_{k-1}(-1) < 0$, $V_{k-1}(-1)$ and $W_{k-1}(-1)$ are negative, and since $H_{k-1}(-1)$ is positive, $F_k(-1)$ is negative. Since $F_1(-1)$ is negative, from Eq. A(80), $F_k(-1)$ is negative, then $V_k(-1)$ and $W_k(-1)$, for any k.

References

1. W. Shockley, D.R. Curran, D.J. Koneval, "Energy Trapping and Related Studies of Multiple Electrode Filter Crystals", Proc. 20 Annual Frequency Control Symposium, Atlantic City, April 1966.

2. W.G. Cady, "Piezoelectricity", Mc Graw Hill (1946)

3. W.D. Beaver, "Analysis of Elastically Coupled Piezoelectric Resonators", Jour. Acous. Soc. Am., Vol. 43, No 5 ; 1968.

4. A. Glowinski, "Etude des vibrations des lames AT en cisaillement d'épaisseur. Comparaison entre l'excitation normale et l'excitation parallèle", Centre National d'Etudes des Télécommunications, Groupement Composants, Note Technique Interne N° 1517 ; 1967.

5. W.D. Beaver, "Theory and Design of the Monolithic Crystal Filter", Proc. 21 Annual Frequency Control Symposium, Atlantic City, April ; 1967.

6. H. Yoda, Y. Nakazawa, S. Okano, N. Kobori, "High Frequency Crystal Mechanical Filters", Proc. 22 Annual Frequency Control Symposium, Atlantic City, April ; 1968.

ELECTRIC FIELD EFFECTS IN MONOLITHIC CRYSTAL FILTERS

H. F. Tiersten
Rensselaer Polytechnic Institute, Troy, New York 12181

Abstract

It is well-known that the vibrations of monolithic crystal filters are very accurately described by the equations of the thickness-shear approximation, which yield the thickness-shear branches for both the electroded and unelectroded portions of the plate. These branches indicate that, for all sections of the plate, the wave-number is real for frequencies above the thickness-shear frequency for that section and purely imaginary for frequencies below that frequency. However, the analysis predicting those thickness-shear branches completely ignores the (albeit small) electric fields in the atmosphere near the surface of an unelectroded portion of the plate. A more careful investigation, including the electric fields outside an unelectroded section of the plate, reveals that the dispersion curves for that section are shifted a small amount. Analytic expressions, relating the shift to the ratio of dielectric constants, the piezoelectric coupling factor and the pertinent elastic constants are derived. In particular, it is shown that the real branch has zero slope at a small, positive value of wave-number and that the previous purely imaginary branch has become complex. One consequence of these results is that energy trapping disappears at a frequency slightly below the thickness-shear frequency of the unelectroded section.

1. Introduction

It is well known that the vibrations of monolithic crystal filters[1,2] and trapped energy resonators[3] are very accurately described by the equations of the thickness-shear approximation[4,5]. The thickness-shear approximation is an asymptotic form of Mindlin's equations of coupled thickness-shear and flexure of elastic plates, which includes the thickness-shear branch but ignores the flexural branch. The idea of excluding the flexural branch originated with Mindlin and Forray[6]. It should be noted that while the equations of thickness-shear and flexure and the associated edge conditions constitute a system that is mathematically well set, the thickness-shear approximation does not because it is only an approximation, which is valid in a highly restricted frequency range and under highly restricted circumstances and all the edge conditions can be satisfied only in an approximate sense. It should also be noted that in the thickness-shear approximation both shear and flexural motions coexist, neither can be neglected. These comments not withstanding the restricted frequency range and restricted circumstances do not constitute a practical limitation because the monolithic crystal filter and trapped energy resonator are operated in precisely this range.

The dispersion curves obtained from the equations of the thickness-shear approximation for all sections of an elastic plate indicate that the propagation wave-number is real for frequencies above the thickness-shear frequency and imaginary for frequencies below that frequency. Consequently, for frequencies slightly below the thickness-shear frequency, the standing wave shape is hyperbolic (or exponential) and for frequencies slightly above, trigonometric. Since the solutions for standing waves in discontinuous plates may be composed of the standing wave solutions in infinite plates[7,8,9], and in the frequency range of interest only the thickness-shear branch need be considered, plates constructed with slightly different properties in adjacent sections and operated in a frequency range between the thickness frequencies of the adjacent sections will have standing wave shapes that are oscillatory in one region and decaying in the other. This behavior serves as the basis of the trapped energy resonator, as observed by Shockley, Curran, and Koneval[3], and the monolithic crystal filter, as observed by Sykes and Beaver[1] and Onoe[2].

The purely elastic- or even piezoelectric-analyses predicting the behavior of the aforementioned thickness-shear branches completely ignore the small electric fields in the atmosphere near the surface of an unelectroded portion of the plate. A more careful investigation in which the quasi-static electric boundary conditions at the dielectric-air interface and Laplace's equation in the atmosphere outside the plate are satisfied as well as the linear piezoelectric plate equations and traction-free conditions at the plate-atmosphere interface, within the framework of the thickness-shear approximation, reveals that the previously purely imaginary wave-numbers for straight-crested waves have become complex, with a small constant real part. In this paper analytic expressions for the real and complex portions of the thickness-shear dispersion curves are derived. These expressions give, among other things, the aforementioned constant real part of the complex portion of the thickness-shear branch and the frequency at which the real portion of the branch becomes complex. One consequence of these results is that energy trapping disappears at a frequency slightly below the thickness-shear frequency of the unelectroded portion of the plate. Another consequence is that no energy is radiated from the free major surfaces of the plate in the steady-state.

2. Mechanical Considerations

It is well known that the mechanical behavior of both the trapped energy resonator and the monolithic crystal filter, is very accurately described by Mindlin's equations of coupled thickness-shear and flexure for all frequencies below a frequency slightly above the thickness-shear frequency of the unelectroded portion of the plate[10]. Schematic diagrams of a trapped energy resonator and a monolithic crystal filter are shown in Figs. 1 and 2, respectively. For waves independent of x_3, Mindlin's equations of coupled thickness-shear and flexure predict the well-known dispersion curves shown in Fig. 3. Of course, at a given frequency ω, a real value of ξ corresponds to a trigonometric function of x_1 and an imaginary value of ξ to a hyperbolic (or exponential) function of x_1. Mindlin's equations may be used in obtaining eigensolutions for the structures shown in Figs. 1 and 2, by satisfying the respective plate equations in the electroded and unelectroded regions and the continuity conditions[10,11]

$$T_{12}^{(0)} = \bar{T}_{12}^{(0)},\ T_{11}^{(1)} = \bar{T}_{11}^{(1)},\ u_1^{(1)} = \bar{u}_1^{(1)},\ u_2^{(0)} = \bar{u}_2^{(0)} \qquad (2.1)$$

at each interface between the regions. The bar refers to the electroded region.

If the frequency range is confined to be close to the thickness-shear frequencies of each section of the plate, as is usually the case, only the thickness-shear branch need be included as first observed by Mindlin and Forray[6]. Under these circumstances the thickness-shear approximation[4,5] is valid and may be employed in place of the more cumbersome equations of thickness-shear and flexure. For rotated Y-cut quartz the differential equations of the thickness-shear approximation take the form[12]

$$\frac{h^2}{3}\left[\gamma_{11} + \varkappa_6^2 c_{66}\right]u_{1,11}^{(1)} + \left[\frac{h^2}{3}\rho\omega - \varkappa_6^2 c_{66}\right]u_1^{(1)} = 0, \qquad (2.2)$$

$$\frac{h^2}{3}\left[\bar{\gamma}_{11} + \bar{\varkappa}_6^2(1+2R)c_{66}\right]\bar{u}_{1,11}^{(1)} + \left[\frac{h^2}{3}\rho(1+3R)\omega^2 - \bar{\varkappa}_6^2 c_{66}\right]u_1^{(1)} = 0, \qquad (2.3)$$

where (2.2) is for the unelectroded portion of the plate and (2.3) for the electroded portion and the slopes $u_1^{(1)}$ and $\bar{u}_1^{(1)}$ are related to the displacements u_1 and $\bar{u}_1$ by

$$u_1 = x_2 u_1^{(1)},\ \bar{u}_1 = x_2 \bar{u}_1^{(1)}.$$

It has been shown that the conditions for the validity[4,5] of (2.2) and (2.3) are that the circular frequency ω be very near the thickness-frequency and that the propagation wave number ξ be much less than the plate thickness 2h, i.e.,

$$\omega = \frac{\sqrt{3}\,\varkappa_6}{h}\sqrt{\frac{c_{66}}{\rho}} + \epsilon,\ \epsilon \ll \frac{\sqrt{3}\,\varkappa_6}{h}\sqrt{\frac{c_{66}}{\rho}}, \qquad (2.4)$$

$$|\xi h| \ll 1. \qquad (2.5)$$

In Eqs. (2.2) and (2.3) $R = 2\rho' h'/\rho h$ is the mass per unit area of a set (two) of electrodes to the mass per unit area of the plate between the electrodes, c_{66} is a shear elastic stiffness, γ_{11} a flexural stiffness[4], e_{26} a piezoelectric constant and $\varkappa_6(\bar{\varkappa}_6)$ Mindlin's shear-correction factor, which is determined by equating the lowest thickness-shear frequencies obtained from the three-dimensional equations for each region with that obtained from Eqs. (2.2) and (2.3), respectively. In Fig. 4, the dispersion curve predicted by Eq. (2.2) is shown dotted and by Eq. (2.3) is shown solid. Note that both the flexural (γ_{11}) and shear constants (c_{66}) appear in equations (2.2) and (2.3), and, indeed, γ_{11} is larger than c_{66}. As a consequence, equations (2.2) and (2.3) are merely an approximation and cannot satisfy all the interface continuity conditions in (2.1). However, when the conditions for the validity of the thickness-shear approximation, $(2.4)_2$ and (2.5), are satisfied it has been shown[4,5] that the continuity conditions

$$u_1^{(1)} = \bar{u}_1^{(1)},\ u_{1,1}^{(1)} = \bar{u}_{1,1}^{(1)}, \qquad (2.6)$$

at an interface satisfy all four conditions in (2.1) approximately.

3. Piezoelectric Considerations

The purely mechanical plate equations discussed in Section 2 do not include the piezoelectric coupling. However, Mindlin's equations of coupled thickness-shear and flexure have been extended[13,14] to include the piezoelectric coupling and the electric field equations. The non-trivial electrostatic plate equation accompanying Mindlin's equations of coupled thickness-shear and flexure is

$$D_{1,1}^{(1)} + 2D_2^{(2)} = 0. \qquad (3.1)$$

where

$$D_1 = x_2 D_1^{(1)}, \tag{3.2}$$

$$D_2 = D_2^{(0)} + x_2^2 D_2^{(2)}, \tag{3.3}$$

and D_1 and D_2 are the components of electric displacement in the 1 and 2 directions, respectively; and $D_1^{(1)}$, $D_2^{(0)}$ and $D_2^{(2)}$ are plate electric displacement functions, which depend on x_1, x_3 and t in general. In the thickness-shear approximation the piezoelectric plate equations take the form[12]

$$\frac{h^2}{3}\left[\gamma_{11} + \kappa_6^2 c_{66}\right] u_{1,11}^{(1)} + \left[\frac{h^2}{3}\rho\omega^2 - \kappa_6^2 c_{66}\right] u_1^{(1)} - \kappa_6 e_{26}\frac{V}{2h} = 0, \tag{3.4}$$

$$\frac{h^2}{3}\left[\bar{\gamma}_{11} + \bar{\kappa}_6^2(1+2R)c_{66}\right] u_{1,11}^{(1)} + \left[\frac{h^2}{3}\rho(1+3R)\omega^2 - \bar{\kappa}_6^2 c_{66}\right] u_1^{(1)} - \bar{\kappa}_6 e_{26}\frac{\bar{V}}{2h} = 0, \tag{3.5}$$

where $V = \varphi(h) - \varphi(-h)$ is the voltage across the major surfaces of an unelectroded portion of the plate and $\bar{V}$, the voltage across the surfaces of an electroded portion of the plate, and φ is the electric potential. Equations (3.4) and (3.5) contain the piezoelectric coupling and replace the purely mechanical equations (2.2) and (2.3). Since, for the thickness-shear approximation to be valid, Eq. (2.5) must be valid, we have

$$|D_{1,1}^{(1)}| \ll |D_2^{(2)}|, \tag{3.6}$$

and, hence, the non-trivial plate equation of electrostatics reduces to the form

$$D_2^{(2)} = 0. \tag{3.7}$$

In the thickness-shear approximation the piezoelectric plate constitutive equations take the form

$$T_{12}^{(0)} = 2h\,\kappa_6^2 c_{66} u_1^{(1)} + \kappa_6 e_{26} V, \tag{3.8}$$

$$T_{11}^{(1)} = \frac{2}{3}h^3 \gamma_{11} u_{1,1}^{(1)}, \tag{3.9}$$

$$D_1^{(1)} = \mathring{\gamma}_{11} u_{1,1}^{(1)} - (3/2h^3)\zeta_{11}\varphi_{,1}^{(1)}, \tag{3.10}$$

$$D_2^{(0)} = \kappa_6 e_{26} u_1^{(1)} - (15/4h^3)\epsilon_{22}\varphi^{(1)} + (3/4h)\epsilon_{22} V, \tag{3.11}$$

$$D_2^{(2)} = (15/4h^5)\epsilon_{22}(3\varphi^{(1)} - h^2 V), \tag{3.12}$$

where the electric potential resultant $\varphi^{(1)}$ is defined by[13,14]

$$\varphi^{(1)} = \int_{-h}^{h} x_2 \varphi dx_2.$$

Note that if the electric fields outside an unelectroded portion of the piezoelectric plate are excluded - by setting $D_2^{(0)} = 0$ in (3.11) -, Eqs. (3.4), (3.7), and (3.12) reveal that the resulting thickness-shear branch is exactly the same as the dotted curve in Fig.4 for the purely elastic case, except for a slight shift in the thickness-shear frequency, which is a result of the piezoelectric stiffening. However, as we shall see in the next section, this is not the case when the electric field outside the piezoelectric plate is included.

4. The Effect of External Electric Fields on the Thickness-Shear Branch

In this section we determine the effect of the electric fields outside an unelectroded portion of the plate on the thickness-shear dispersion curves in the vicinity of the thickness-shear frequency. To this end we consider the electric fields outside the plate within the framework of the thickness-shear approximation. In the atmosphere outside the surface of the unelectroded portion of the plate, the electric potential satisfies the equation[15]

$$\varphi_{,11}^{e} + \varphi_{,22}^{e} = 0. \tag{4.1}$$

The electric boundary conditions at the interface between the atmosphere and the plate are

$$\varphi^{e} = \varphi^{i} \text{ at } x_2 = \pm h, \tag{4.2}$$

$$D_2^{e} = D_2^{i} \text{ at } x_2 = \pm h, \tag{4.3}$$

where

$$D_2^e = -\epsilon_o \varphi^e_{,2}, \tag{4.4}$$

and D_2^i is given by (3.3), (3.7), and (3.11). In this section it is particularly convenient to consider the solutions in the atmosphere to be of the form

$$\varphi^e = \pm\, Be^{\mp \xi x_2}\, e^{i(\xi x_1 - \omega t)}, \quad x_2 \gtrless \pm h, \tag{4.5}$$

which satisfies (4.1) and where ξ and B may be complex. Note that the first ξ in (4.5) must have a positive real part in order that φ^e be bounded as $x_2 \to \pm\infty$, but that the second ξ, which is the propagation wavenumber, can be + or - the first ξ by virtue of (4.1). From (4.4) and (4.5) we obtain

$$D_2^e = \pm\, \epsilon_o \xi \varphi \text{ at } x_2 = \pm h, \tag{4.6}$$

which, with (4.3), (3.3), and the fact that $V = \varphi(h) - \varphi(-h)$ yields

$$D_2^{(0)} + h^2 D_2^{(2)} = \epsilon_o \xi V/2. \tag{4.7}$$

From (3.7) we obtain

$$\varphi^{(1)} = Vh^2/3, \tag{4.8}$$

from which, with (4.7) and (3.7), we find

$$V = \frac{2h\kappa_6 e_{26} u_1^{(1)}}{\epsilon_{22}(1 + r\xi h)}, \tag{4.9}$$

where

$$r = \epsilon_o/\epsilon_{22}. \tag{4.10}$$

Substituting from (4.9) into (3.4) we obtain

$$\frac{h^2}{3}\left[\gamma_{11} + \kappa_6^2 c_{66}\right] u^{(1)}_{1,11} + \left[\frac{h^2}{3}\rho\omega^2 - \kappa_6^2 c_{66} - \frac{\kappa_6^2 e_{26}^2}{\epsilon_{22}(1 + r\xi h)}\right] u_1^{(1)} = 0. \tag{4.11}$$

When $\xi = 0$ we find, from Eq. (4.11)

$$\omega_o^2 = 3\kappa_6^2 \bar{c}_{66}/h^2\rho, \tag{4.12}$$

where ω_o is the thickness-shear frequency of the unelectroded region and

$$\bar{c}_{66} = c_{66} + e_{26}^2/\epsilon_{22}, \tag{4.13}$$

is the piezoelectrically stiffened elastic constant, and in this instance

$$\kappa_6 = \pi/\sqrt{12}. \tag{4.14}$$

In the thickness-shear approximation, when $\xi \neq 0$, we have

$$\omega = \omega_o - \epsilon, \quad |\epsilon| \ll \omega_o. \tag{4.15}$$

On account of (4.5) the solution for $u_1^{(1)}$ must be written

$$u_1^{(1)} = Ae^{i(\xi x_1 - \omega t)}. \tag{4.16}$$

Substituting from (4.16) and (4.15) into (4.11), rearranging terms, neglecting terms of order $(\epsilon/\omega_o)^2$ and introducing the condition for a non-trivial solution, i.e., $A \neq 0$, we obtain

$$\psi^3 + \frac{1}{r}\psi^2 + \frac{(2\delta - k^2)}{\Gamma}\psi + \frac{2\delta}{r\Gamma} = 0, \tag{4.17}$$

where

$$\psi = \xi h, \quad \delta = \epsilon/\omega_o, \tag{4.18}$$

$$k^2 = e_{26}^2/\bar{c}_{66}\epsilon_{22}, \quad \Gamma = (\gamma_{11} + \kappa_6^2 c_{66})/3\kappa_6^2 \bar{c}_{66}, \tag{4.19}$$

and k is the piezoelectric coupling factor. Equation (4.17) is a cubic in the dimensionless wave-number ψ, which is small ($|\psi| \ll 1$) when the thickness-shear approximation is valid. The constants r, Γ, and k^2 are, of course, always positive.

The roots (values of ψ for a given δ) of (4.17) can readily be determined numerically for a particular cut of a specific material such as, e.g., A-T cut quartz, for which

$$r = 0.226, \quad \Gamma = 1.418, \quad k^2 = 0.0078, \tag{4.20}$$

as calculated from Bechmann's values of the principal constants[16]. However, because of the smallness of the permissible values of $|\psi|$ and $|\delta|$ within the thickness-shear approximation and

the usual range of values of the material constants r, Γ, and k^2, analytical expressions for the physically significant dispersion curves $\psi = \psi(\delta)$ can be determined from (4.17). Obviously, such expressions are more interesting, illuminating and potentially useful than the numerical values of the dispersion curves themselves in a specific instance. Accordingly, we proceed by first observing that the cubic (4.17) must have at least one real root and, when δ is positive, the root must be negative[17] and, hence, of no physical significance because it corresponds to solutions which diverge as $x_2 \to \pm\infty$. Moreover, the nature of the numbers is usually such ($r < 1$) that the real negative root is large and, hence, determined essentially by the cubic and square terms. Keeping these observations in mind, we differentiate (4.17) totally with respect to ψ to obtain

$$3\psi^2 + \frac{2}{r}\psi - \frac{k^2}{\Gamma} + \frac{2\delta}{\Gamma} + \left(\psi + \frac{1}{r}\right)\frac{2}{\Gamma}\frac{d\delta}{d\psi} = 0,$$

which, since $r < 1$ and, within the range of validity of the thickness-shear approximation, $|\psi| \ll 1$ when real, enables us to write

$$\frac{d\delta}{d\psi} = -\frac{3r\Gamma}{2}\psi^2 - \Gamma\psi + \frac{r}{2}(k^2 - 2\delta). \quad (4.21)$$

At the thickness-shear frequency, $\delta = 0$, Eq. (4.17) yields

$$\psi(\psi^2 + \psi/r - k^2/\Gamma) = 0,$$

which yields the three real roots

$$\psi = 0, \quad (4.22)$$

$$\psi = -\frac{1}{2r} \pm \frac{1}{2r}\sqrt{1 + \frac{4k^2r^2}{\Gamma}}. \quad (4.23)$$

One of the two roots in (4.23) is large and negative and, hence, of no physical significance. The other, which is small and positive, may be very accurately approximated by

$$\psi = k^2r/\Gamma. \quad (4.24)$$

The slopes at the two physically significant values of ψ for $\delta = 0$, may be obtained by substituting, respectively, from (4.22) and (4.24) into (4.21) with $\delta = 0$, with the results

$$d\delta/d\psi = rk^2/2 \text{ at } \delta = 0, \ \psi = 0, \quad (4.25)$$

$$d\delta/d\psi = -rk^2/2 \text{ at } \delta = 0, \ \psi = rk^2/\Gamma, \quad (4.26)$$

where we have used the fact that $r\psi \ll 1$ in obtaining (4.25) and (4.26). Setting $d\delta/d\psi = 0$ in (4.21) we obtain

$$\psi = -\frac{1}{3r} \pm \frac{1}{3r}\sqrt{1 - \frac{3r^2}{\Gamma}(2\delta - k^2)},$$

and, as usual, the large negative root does not count, and the small positive root may be very accurately approximated by

$$\psi = r(k^2 - 2\delta)/2\Gamma. \quad (4.27)$$

Substituting from (4.27) into (4.17) and making use of the fact that $r\psi \ll 1$ and $r^2k^2/\Gamma \ll 1$, we find

$$\delta = r^2k^4/8\Gamma,$$

which locates the relevant point of zero slope in the real ψ-δ plane at

$$\psi_o = rk^2/2\Gamma, \quad \delta_o = r^2k^4/8\Gamma. \quad (4.28)$$

Onoe's theorem[18] now tells us that a complex branch emanates from this point.

At this point we may note, from the large, real negative root in Eq. (4.23) and the fact that this root is essentially determined from the cubic and square terms in (4.17), that the large, negative, real root may be written

$$\psi = -\frac{1}{r} + \Delta, \ |\Delta| \ll 1, \quad (4.29)$$

and using synthetic division[19] and dropping all terms of order Δ^2 and higher, we have

$$\begin{array}{r|cccc} & 1 & \frac{1}{r} & \left(\frac{2\delta - k^2}{\Gamma}\right) & \frac{2\delta}{r\Gamma} \\ -\frac{1}{r} + \Delta & & \left(-\frac{1}{r} + \Delta\right) & -\frac{2}{r^2} & -\frac{2\delta}{r\Gamma} + \frac{k^2}{r^2} + \Delta\frac{2\delta - k^2}{\Gamma} + \frac{2}{r^3} \\ \hline & 1 & \Delta & \left(\frac{2\delta - k^2}{\Gamma} - \frac{\Delta}{r}\right) & \frac{k^2}{r\Gamma} + \Delta\left(\frac{2\delta - k^2}{\Gamma} + \frac{1}{r^2}\right) \end{array}$$

and since we wish the remainder to vanish to order Δ, we find

$$\Delta = -rk^2/[(2\delta - k^2)r^2 + \Gamma],$$

and since $(2\delta - k^2)r^2 \ll \Gamma$, we have approximately

$$\Delta = -rk^2/\Gamma, \qquad (4.30)$$

and the large real root of no physical interest is given by

$$\psi \approx -(1/r)(1 + r^2k^2/\Gamma). \qquad (4.31)$$

Thus, to order Δ, the remaining quadratic is given by

$$\Gamma\psi^2 - rk^2\psi + 2\delta = 0,$$

the roots of which may be written

$$\psi = \frac{rk^2}{2\Gamma} \pm \frac{i}{2\Gamma}\sqrt{8\Gamma\delta - r^2k^4}. \qquad (4.32)$$

Clearly, both complex conjugate roots in (4.32) are physically significant since the real part is positive and corresponds to spatial decay as $x_2 \to \pm\infty$, and the real and imaginary portions of ψ are small enough, if δ is small, to be well within the range of validity of the thickness-shear approximation. Note that the roots of (4.32) become real when

$$\delta < r^2k^4/8\Gamma,$$

and that (4.32) predicts the intersection of the complex branches with the real plane, given in (4.28) by our earlier analysis of the real branches. Indeed, Eqs. (4.32) predict the corrected version of the thickness-shear branch, due to the electric fields outside the unelectroded surfaces of the plate, within the range of validity of the thickness-shear approximation.

The dispersion curves giving the dimensionless wave-number $\psi = \xi h$ as a function of $\delta \equiv e/\omega_o$ as determined from (4.32) are plotted in Fig. 5, which shows that the purely imaginary dotted dispersion curve of the mechanical solution has become complex with the real part independent of frequency. However, it turns out that the constant real part of complex ψ is quite small for all known materials because whenever k is reasonably large, r is very small. Note that the dispersion curves having Re ψ positive only have been plotted in Fig. 5. If, as is usually the case, ψ is taken to represent the dimensionless propagation wave number, the ψ values with Re ψ negative corresponding to the Re ψ positive curves shown in Fig. 5, may be plotted also by virtue of the sentence following Eq. (4.5). An important consequence of the location of the thickness-shear branch in Fig. 5 is that energy trapping disappears at a frequency slightly below the thickness-shear frequency of the unelectroded section. The small amount of frequency lowering is $\delta_o\omega_o$, and may be calculated from (4.12), (4.14), and (4.28). Moreover, since, when energy trapping is present, the thickness-shear branches are complex, rather than imaginary, the fields in the atmosphere outside an unelectroded surface of the plate correspond to standing waves, which oscillate and decay with position, rather than traveling waves. Consequently, there is no radiation of electric field energy in the steady state.

Acknowledgement

The author wishes to thank Professor J. L. Bleustein of Yale University for a very valuable comment concerning complex branches.

This work was partially supported by the Office of Naval Research under Contract N00014-67-A-0117 with Rensselaer Polytechnic Institute.

References

1. R. A. Sykes and W. D. Beaver, "High Frequency Monolithic Crystal Filters with Possible Application to Single Frequency and Single Side Band Use," Proc. 20th Ann. Symp. Freq. Control, pp. 288-308 (1966).

2. M. Onoe, H. Jumonji, and N. Kobori, "High Frequency Crystal Filters Employing Multiple Mode Resonators Vibrating in Trapped Energy Modes," Proc. 20th Ann. Symp. Freq. Control, pp. 266-287 (1966).

3. W. Shockley, D. R. Curran, and D. J. Koneval, Proc. 17th Symp. on Frequency Control, 88-126 (1963).

4. J. L. Bleustein and H. F. Tiersten, "Forced Thickness-Shear Vibrations of Discontinuously Plated Piezoelectric Plates," J. Acoust. Soc. Am., 43, 1311 (1968).

5. H. F. Tiersten, "Linear Piezoelectric Plate Vibrations," Plenum Press, New York, 1969, Chap. 16.

6. R. D. Mindlin and M. Forray, "Thickness-Shear and Flexural Vibrations of Contoured Crystal Plates," J. Appl. Phys., 25, 12 (1954).

7. R. D. Mindlin, "An Introduction to the Mathematical Theory of the Vibrations of Elastic Plates," U.S. Army Signal Corps Eng. Lab., Fort Monmouth, N.J. (1955); Signal Corps Contract DA-36-039 SC-56772.

8. R. D. Mindlin, "Investigations in the Mathematical Theory of Vibrations of Anisotropic Bodies," CU-4-56-SC-64687-CE, Final Rept. U.S. Army Signal Corps Eng. Lab., Fort Monmouth, N.J. (1956).

9. Ref. 5, Chap. 12.

10. R. D. Mindlin, "High Frequency Vibrations of Crystal Plates," Quart. Appl. Math., 19, 51 (1961).

11. R. D. Mindlin and P. C. Y. Lee, "Thickness-Shear and Flexural Vibrations of Partially Plated, Crystal Plates," Int. J. Solids Structures, 2, 125 (1966).

12. Ref. 5, Chap. 16, Sec. 6.

13. H. F. Tiersten and R. D. Mindlin, "Forced Vibrations of Piezoelectric Crystal Plates," Quart. Appl. Math., 20, 107 (1962).

14. Ref. 5, Chap. 13.

15. Ref. 5, Chap. 4, Sec. 4.

16. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 110, 1060 (1958).

17. M. G. Salvadori, "The Mathematical Solution of Engineering Problems," McGraw-Hill Book Company, Inc., New York, 1948, Sec. 3.4e.

18. R. D. Mindlin and M. Onoe, "Mathematical Theory of Vibrations of Elastic Plates," In: Proc. 11th Annual Symposium on Frequency Control, U.S. Army Signal Engineering Laboratories, Fort Monmouth, N.J., 1957, pp. 17-40.

19. Ref. 17, Sec. 3.4d.

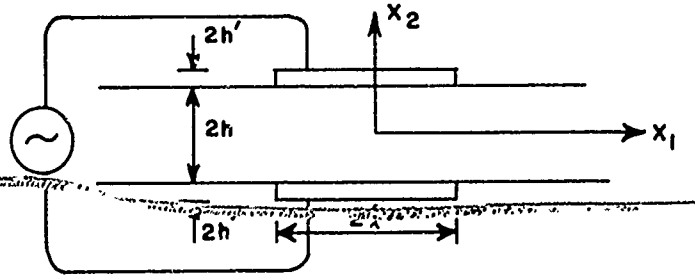


FIGURE 1

Trapped Energy Resonator

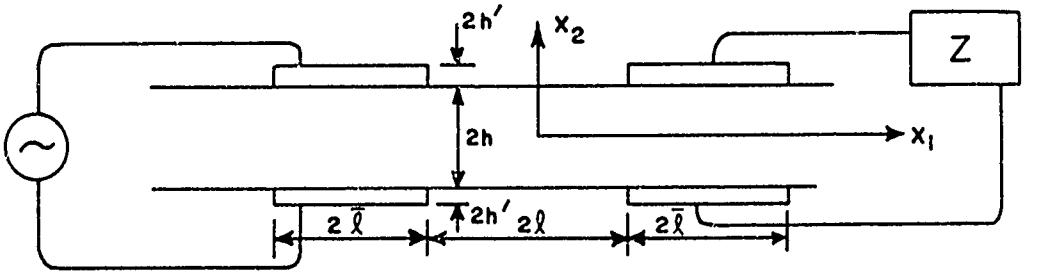


FIGURE 2

A Monolithic Structure

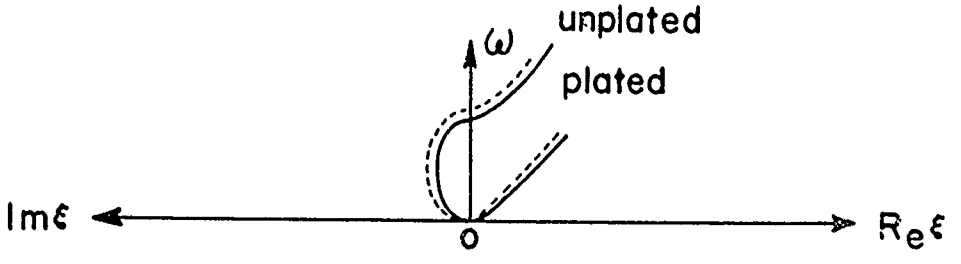


FIGURE 3

Dispersion Curves for Thickness-Shear and Flexure for Both Electroded and Unelectroded Plates for Real and Imaginary Wave-Numbers

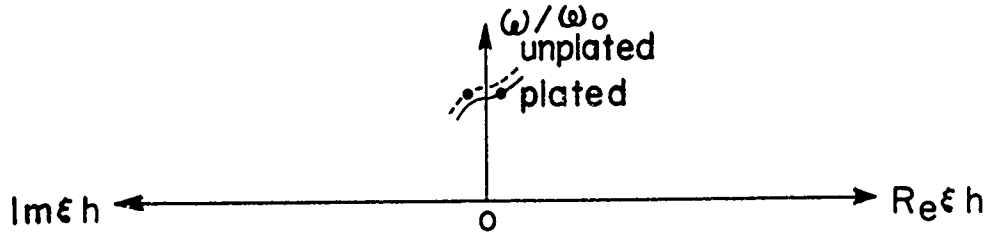


FIGURE 4

Thickness-Shear Dispersion Curves for Electroded and Unelectroded Plates for Real and Imaginary Wave-Numbers

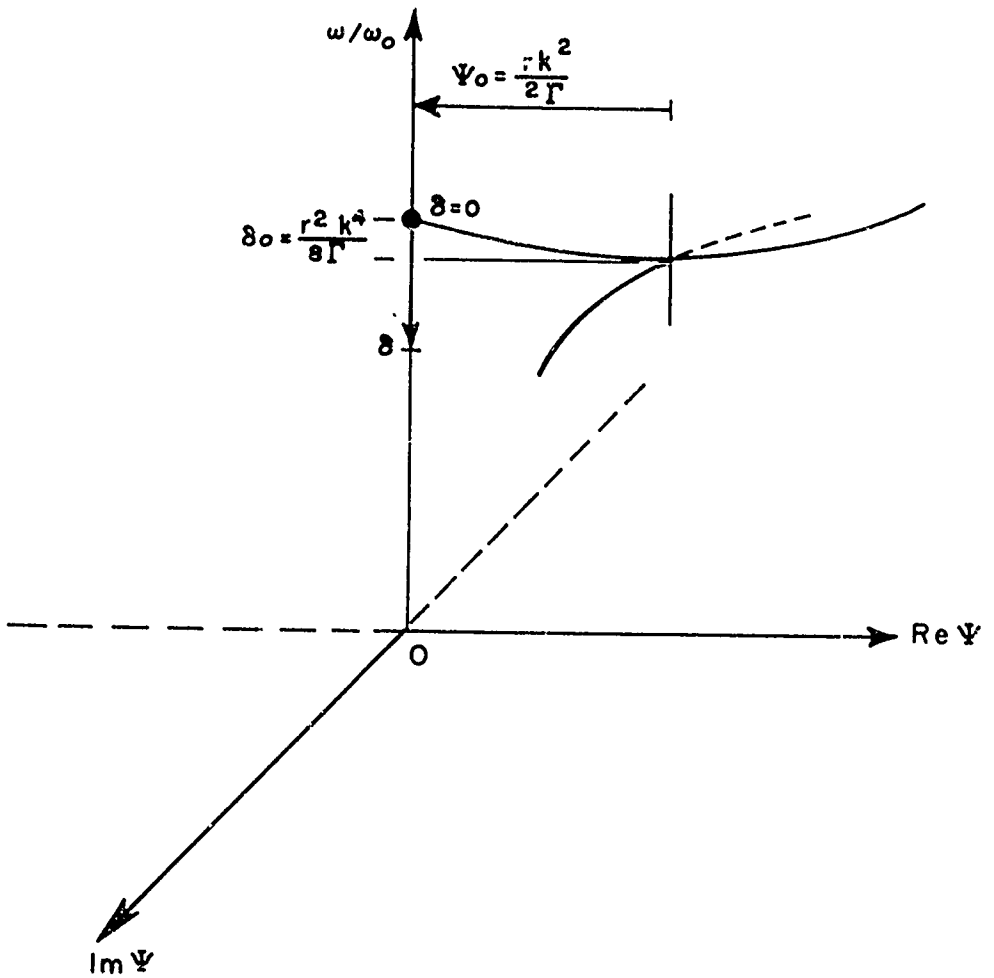


FIGURE 5

Thickness-Shear Dispersion Curves for Unelectroded Piezoelectric Plates Including the Effect of the Electric Fields Outside the Plate

THE DEVELOPMENT OF HIGH PERFORMANCE FILTERS USING ACOUSTICALLY COUPLED RESONATORS ON AT-CUT QUARTZ CRYSTALS

J.F.Werner, A.J.Dyer and J.Birch

*The General Electric Company Limited,
Central Research Laboratories,
Hirst Research Centre,
Wembley, England.

Summary

Filter response curves for experimental monolithic filters show good agreement with the theoretically expected shapes of pass band, but have a slight asymmetry which is discussed in the text.

Early filters had a large number of unwanted responses in the stop band, many of which were due to untrapped modes of vibration and were effectively suppressed by damping the edges of the crystals. However the filters still showed unwanted responses, the stop band attenuation being degraded to about 35-45 dbs. in the region of the blank frequency.

Some of these responses are caused by trapped modes and can be modified by the choice of plateback or electrode geometry in designing the filter. Also it was noted that operating monolithic filters in tandem gave a marked improvement in stop band performance, and the concept of the bilithic crystal filter was then developed. The principle is to make the filter in two halves on separate quartz wafers. On each of the wafers the coupling between resonators is acoustic and is determined by the inter-resonator spacings, but between the wafers it is electrical and is provided by a shunt capacitor. The two halves are designed so that their unwanted response spectra are displaced in frequency with respect to each other.

A six resonator filter has been developed having a band width of 10 kHz at 10.7 MHz. The device has the two wafers mounted back to back with the coupling capacitor at the upper end of the assembly and fits inside a standard HC-13/U holder (can length 0.985 in.).

The filters show a general level of stop band attenuation better than 100 dB with the unwanted responses being no worse than 85 dB at a single frequency. Curves are also shown of the effect of temperature change on the performance.

*This work has been carried out on behalf of Salford Electrical Instruments Ltd., Times Mill, Heywood, Lancs, England.

1. Introduction

In recent years theory and experiment on the monolithic crystal filter have been extensively described in the literature. One of the earliest papers to bring out the relation between the energy trapping effect and coupling between neighbouring resonators on the same quartz plate was that by Shockley, Curran and Koneval[1], although it was aimed at minimising this coupling. Later papers, such as Sykes, Smith and Spencer[2]; Onoe, Jumonji and Kobori[3] and Sykes and Beaver[4] discussed the design and fabrication of monolithic filters in which the acoustical coupling is utilised to obtain desired filter characteristics.

In this paper we describe some of our experimental work in this field and in particular the difficulties encountered in obtaining filter stop bands free of unwanted responses. It is primarily to overcome these difficulties that we have developed the bilithic filter construction which is described in the paper.

2. The Design of Filters and their Pass-band Performance

The design method we have used is that outlined by Beaver in his paper 'Theory and design of the monolithic filter'[5]. We have used his formula for the coupling between two resonators spaced along the X axis

$$k_x = \frac{3}{2}\frac{c_{11}}{c_{66}}\left(\frac{2l_e + d + \lambda}{t}\right)^{-2} e^{-2.28\Delta^{\frac{1}{2}}\frac{d}{t}}$$

as the basis for designing monolithic crystal filters having multiple resonators. In the formula c_{11}, c_{66} are elastic constants for the AT-cut, l_e and d are electrode length and resonator separation, t is thickness of the quartz plate, Δ is the plateback defined as the fractional lowering of frequency produced by the plating, and $\lambda = 2.09t$.

Dishal's tables[6] have been used to design the equivalent electrical ladder filter, and using Beaver's formula the inter-resonator spacings for a chosen electrode size and plateback have been determined. The validity of this approach has been checked by computerised calculations of the

propagation behaviour of our resulting designs. The results are in close agreement with the design bandwidth and ripple.

Typical pass band characteristics for experimental 4-pole (i.e. 4 resonator) filters are shown in figures 1 and 2. In each case the theoretical curve showing the shape of the pass band skirts for the corresponding ladder of discrete electrical elements is shown by the dotted line. It will be noticed that the monolithic filter characteristics are not symmetrical but are steeper on the low frequency than on the high frequency side. This is a characteristic of the monolithic filter. The inter-resonator coupling is a function of plateback and is therefore a fairly sensitive function of frequency in the neighbourhood of the pass band.

If the bandwidth at any chosen attenuation is designed as $2\,\delta f$ and extends from $f_o - \delta f + \alpha$ to $f_o + \delta f + \alpha$ a skew factor S may be defined as

$$S = \frac{\text{Width of pass band on high frequency side}}{\text{Width of pass band on low frequency side}}$$

$$= \frac{\delta f + \alpha}{\delta f - \alpha} \qquad \text{(See figure 3)}$$

It can be shown from Beaver's formula by taking S as

$$\frac{\text{Coupling coefficient for } \Delta = \dfrac{f_b - f_o - \delta f - \alpha}{f_b}}{\text{Coupling coefficient for } \Delta = \dfrac{f_b - f_o + \delta f - \alpha}{f_b}}$$

that approximately $\log_{10} S = 0.99 \dfrac{d}{t} \dfrac{\delta f}{f_b} \Delta_o^{-\frac{1}{2}}$ where Δ_o is the plateback for a mid band frequency f_o and blank frequency f_b i.e. $\Delta_o = \dfrac{f_b - f_o}{f_b}$. The corrections α can then be found since

$$\alpha = \delta f \frac{S - 1}{S + 1}.$$

In figures 1 and 2 this procedure has been used to correct the theoretical curves and the results, shown by the broken lines, are a better fit to the practical curves. The formula shows that for designs of equal bandwidth, those using a lighter plateback (and therefore wider inter-electrode spacing) will be more skewed. This is illustrated by the figures in which the curve in figure 1 for a filter of 0.01 plateback is more skewed than the curve in figure 2 which is for a filter of similar bandwidth but having a plateback of 0.02.

3. Factors Affecting the Stop Band

Our early experiments used quartz wafers mounted on pins. With this type of construction it was found that a large number of unwanted responses appeared in the stop band. Figure 4 shows typical behaviour of this kind for a 4-pole filter. Many of the responses occur at frequencies above the blank frequency and are due to untrapped modes of vibration which involve the whole of the quartz plate. It was found that most of these could be effectively suppressed by damping the edges of the plate and figure 5 shows the same filter after the application of a damping material to the periphery.

A suitable way of achieving this plate damping, as well as providing a convenient method of mounting, is to secure the quartz plate by its edges to a fibre glass frame. Typical stop band characteristics of monolithic filters which we have mounted in this way are shown in figures 6 and 7. It will be noticed that there are unwanted responses, particularly in the neighbourhood of the blank frequency, which severely degrade the stop band attenuation. This is not surprising as trapped overtones close to the blank frequency will be strongly coupled since they are operating at effectively low values of plateback.

Monolithic filter designs have been tried at various values of plateback without a significant improvement in this performance, the best designs giving a stop band attenuation no better than about 55 dB (see figure 7 for the stop band of a six pole monolithic crystal filter).

4. Bilithic Crystal Filters

When two monolithic crystal filters were operated in tandem, it was noticed that the level of stop band attenuation was greatly improved. This is due to breaking the mechanical link between input and output, so that the attenuations of the individual filters are additive. In a properly terminated monolithic filter the principal modes forming the pass band are matched to the external electrical circuit, but the unwanted modes in the stop band are badly mismatched so that signals at these frequencies suffer attenuation. When two such filters are connected in tandem under the correct conditions signals within the pass band are not affected but additional mismatching is introduced at the electrical junction for the unwanted stop band modes.

We then developed the concept of the bilithic crystal filter. The principle is to make the filter in two halves on separate quartz wafers. On each of the wafers the coupling between resonators is acoustic and is determined in the normal way by the inter-resonator spacings, but between the wafers it is electrical and is provided by using a shunt capacitor of the appropriate value. Thus each separate wafer does not necessarily form a useful filter on its own but together the two halves form a filter designed according to modern network theory.

The two halves are designed to operate at different levels of plateback or to have

different electrode geometries, or both, so that their unwanted response spectra are displaced with respect to each other. The limits of plateback within which such a device can be designed are set on the high side by the Bechmann number, which gives a value of plateback above which the anharmonic modes of the individual resonators become trapped, and on the low side by the necessity for the electrodes to have a low enough electrical resistance and a sufficient energy trapping effect to maintain an adequate resonator Q.

A six resonator device has been developed for use at 10.7 MHz in a 12.5 kHz channel spacing system, the band width of the filter being 10 kHz. The device, which is illustrated (figure 8) has the two wafers mounted back to back with the small coupling capacitor at the upper end of the assembly. It fits inside a standard HC-13/U holder (can length 0.985 in.).

Each of the wafers carries three resonators and can be designed individually for different electrode shapes and plateback values, a restriction being that the electrode areas should be the same on both wafers to maintain the same impedance level. A variety of combinations of these three pole wafers has been tried to obtain the best suppression of unwanted responses and examples are shown in figures 9 and 10.

In each figure the upper two curves are for the individual wafers for which the plateback and electrode shape are shown. The lowest curve shows the stop band characteristic for the filter as a whole.

In figure 9 it will be noticed that wafer B carrying rectangular electrodes with $\Delta = 0.020$ shows a strong group of responses in the region 10.855 - 10.905 MHz. This group shows up in the characteristic of the full six pole filter. In figure 10 the filter has been changed by designing the corresponding wafer B for a plateback $\Delta = 0.008$. This group of responses has now been eliminated and, although there are now minor responses in the same frequency region (10.790 - 10.820 MHz) on both wafers, the stop band is much improved. The stop band for the filter shown in figure 10 is typical of the best results obtained hitherto and shows a general level of attenuation better than 100 dB with the unwanted responses almost completely suppressed, being no worse than 85 dB at a single frequency. The problem is thus seen to be one of so choosing the wafer designs that the unwanted responses on each wafer are of minimal intensity and are not at corresponding frequencies on the two wafers.

Figures 11 and 12 show the pass band characteristics for these two filters. In both cases, as expected from the remarks in section 2, the curves are slightly skewed, being steeper on the low frequency side.

The temperature performance of several samples of these filters has been measured and the results are shown in figures 13 and 14. Figure 13 shows the envelope of the pass band characteristic for a typical filter over the temperature range -70°C to +85°C. Figure 14 shows in the upper curve the change of 3 dB bandwidth over a similar range and in the lower curve the deviation in the mid-band frequency. This curve is a good match to the standard curve for a +8′ AT-cut.

5. Conclusions and Future Work

Information contained in the existing literature provides an adequate basis for the design of monolithic crystal filters, although it is to be expected that the pass band characteristics will show some degree of asymmetry. We have concentrated on obtaining a good stop band performance and to this end have developed a bilithic form of construction. A six resonator filter of this type has been described in the paper.

Development of these units is proceeding and is being extended to an eight resonator design which can be accommodated in the same size of holder. We have currently in hand a programme of research aimed at improving our understanding of the unwanted responses in these filters at both fundamental and overtone frequencies.

References

1. W.Shockley, D.R.Curran and D.J.Koneval. Energy trapping and related studies of multiple electrode filter crystals. Proceedings of the 17th Annual Frequency Control Symposium. May 1963, p.88.
2. R.A.Sykes, W.L.Smith and W.J.Spencer. Monolithic crystal filters. I.E.E.E. International Convention Record, Part II. 1967, p.78.
3. M.Onoe, H.Jumonji and N.Kobori. High frequency crystal filters employing multiple mode resonators vibrating in trapped energy modes. Proceedings of the 20th Annual Frequency Control Symposium. April 1966, p.266.
4. R.A.Sykes and W.D.Beaver. High frequency monolithic crystal filters with possible application to single frequency and single side band use. Proceedings of the 20th Annual Frequency Control Symposium. April 1966,p.288.
5. W.D.Beaver, Theory and design of the monolithic crystal filter. Proceedings of the 21st Annual Symposium on Frequency Control, April 1967, p.179.
6. Reference data for radio engineers. International Telephone and Telegraph Corporation. 6th Edition, September 1956, Chapter 7.

List of Figures

1. Pass band of a 4-pole monolithic filter
   - - - - Theoretical curve for a 4-pole filter network
   – – – – Theoretical curve corrected for monolithic filter skew factor
2. Pass band of a 4-pole monolithic filter
   - - - - Theoretical curve for a 4-pole filter network
   – – – – Theoretical curve corrected for monolithic filter skew factor
3. Skew effect in monolithic crystal filters
4. Stop band of an experimental 4-pole monolithic filter with inadequate plate damping
5. Effect of damping the edges of the quartz plate on the stop band of the filter of figure 4
6. Stop band of a 4-pole monolithic crystal filter using edge mounting
7. Stop band of a 6-pole monolithic crystal filter using edge mounting
8. Photograph of a 6-pole bilithic crystal filter. The size of each quartz plate is 16 mm x 10 mm
9. Stop band characteristic for a 6-pole bilithic crystal filter
10. Stop band characteristic for a 6-pole bilithic crystal filter
11. Pass band characteristic of the filter of figure 9
12. Pass band characteristic of the filter of figure 10
13. Envelope of the pass band characteristics of a 6-pole bilithic crystal filter over the temperature range -70°C to +85°C
14. Variation with temperature of the band width and mid-band frequency of a 6-pole bilithic crystal filter.

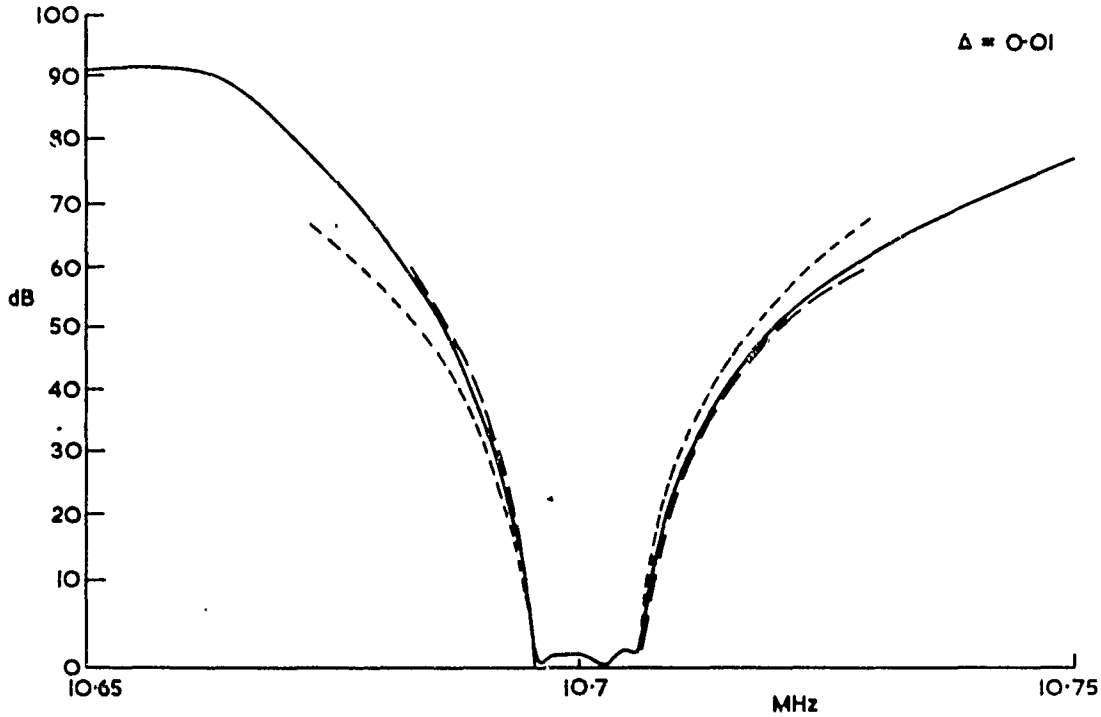


FIGURE 1

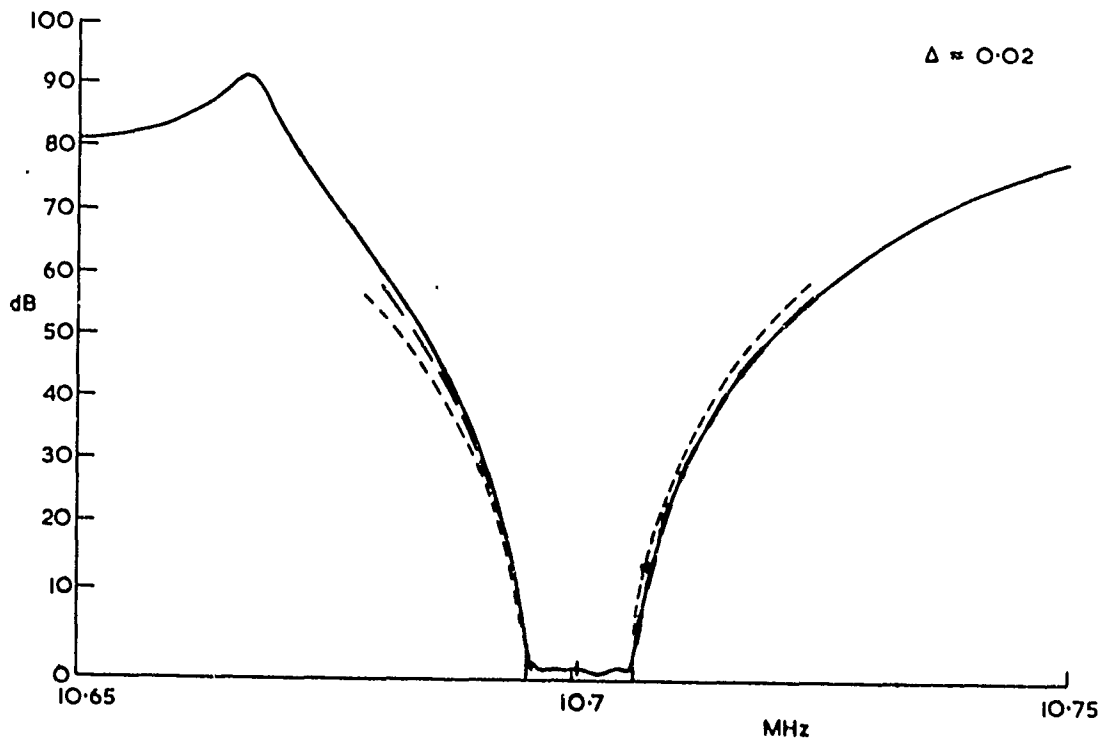
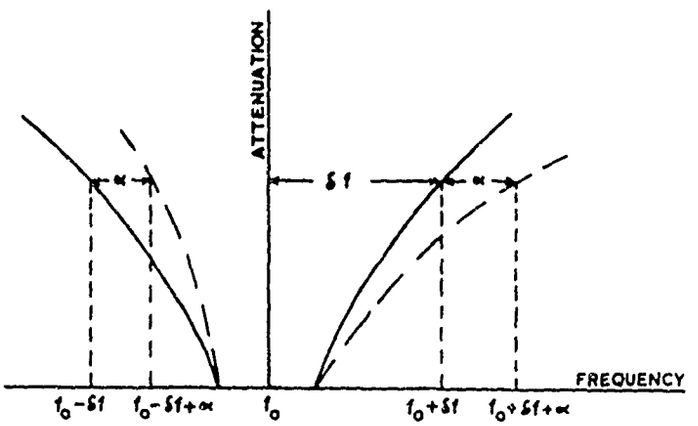


FIGURE 2

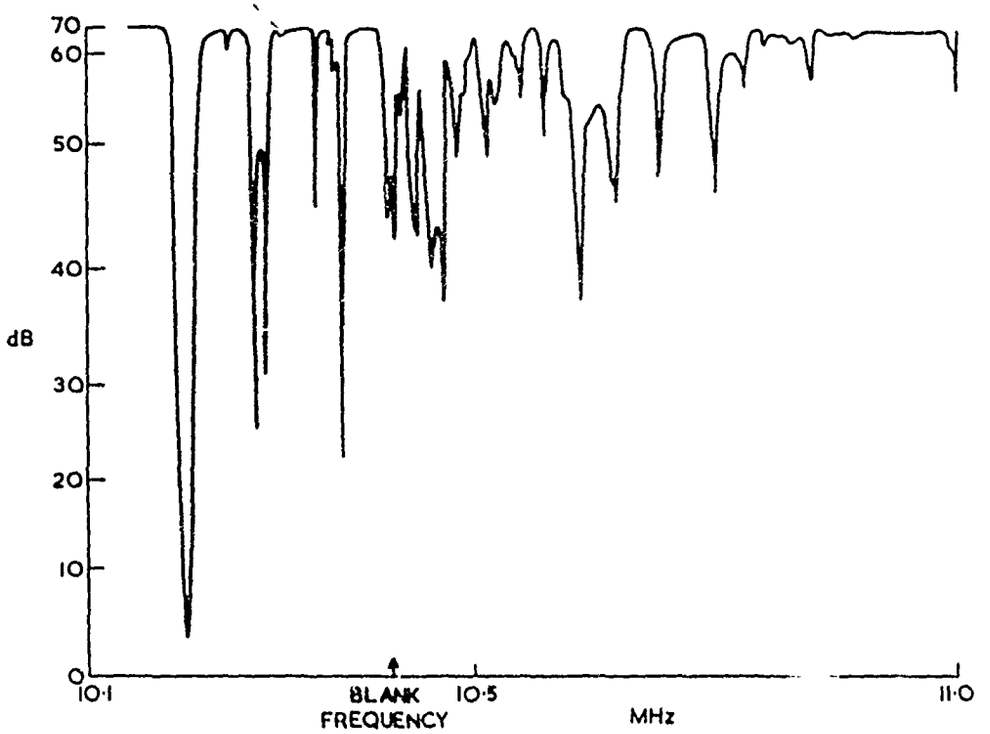


$$\text{SKEW FACTOR } S = \frac{\delta f + \alpha}{\delta f - \alpha}$$

$$\text{LOG}_{10}\, S = 0{\cdot}99 \frac{d}{t} \frac{\delta f}{f_b} \Delta_o^{-\frac{1}{2}}$$

$$\alpha = \delta f \cdot \frac{S-1}{S+1}$$

FIGURE 3



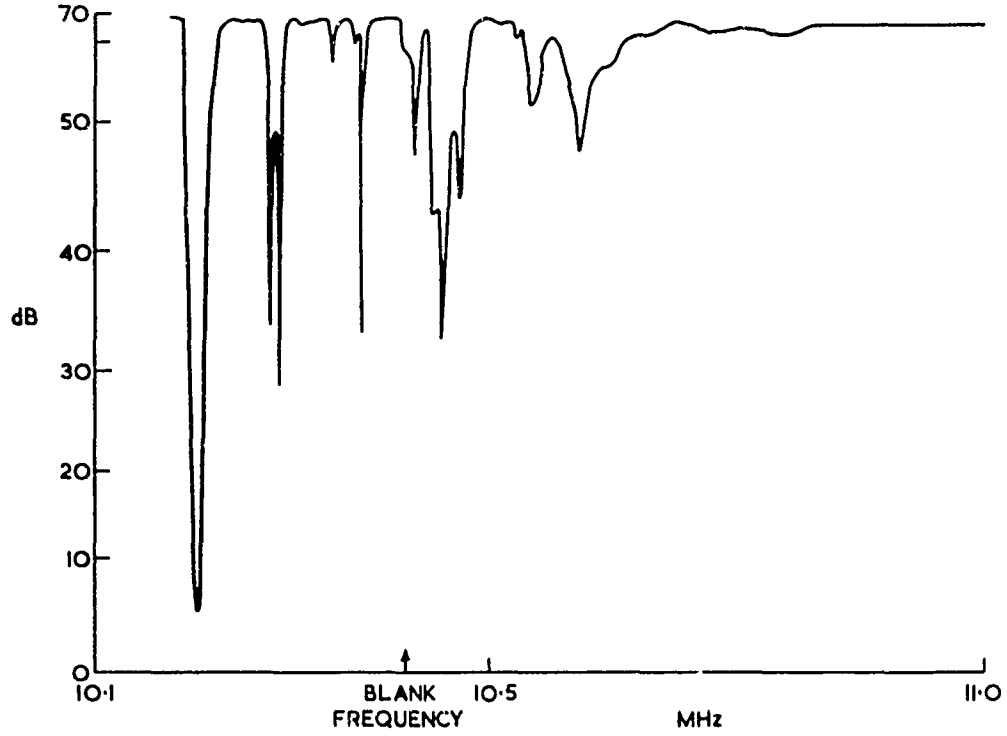


FIGURE 5

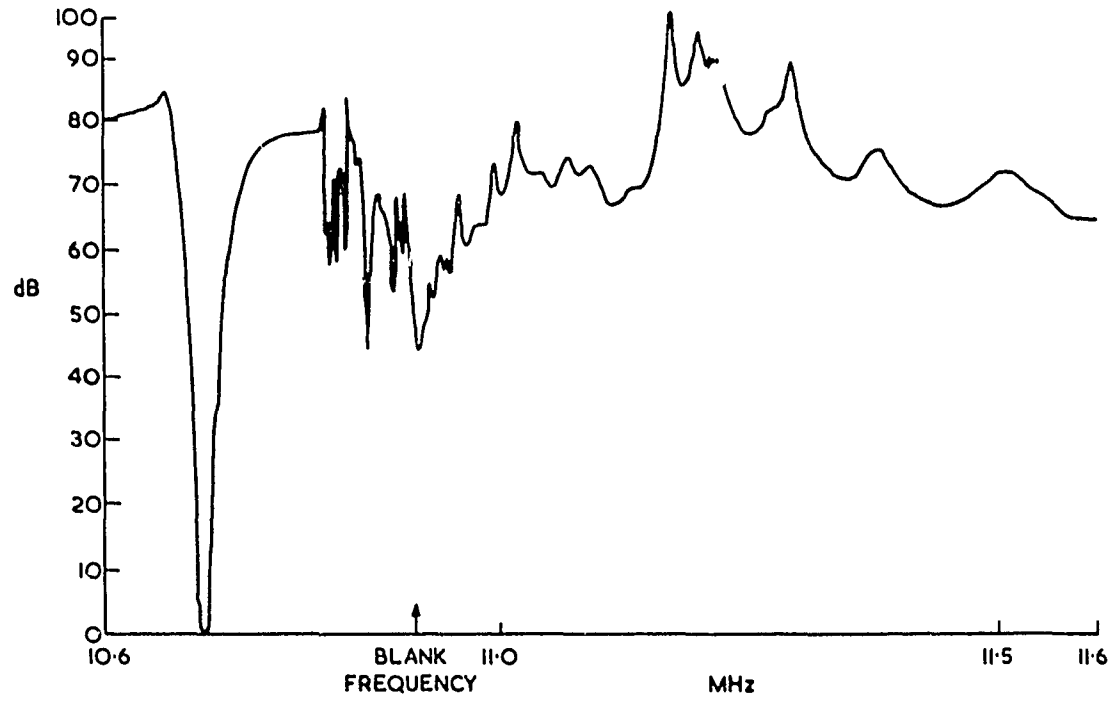
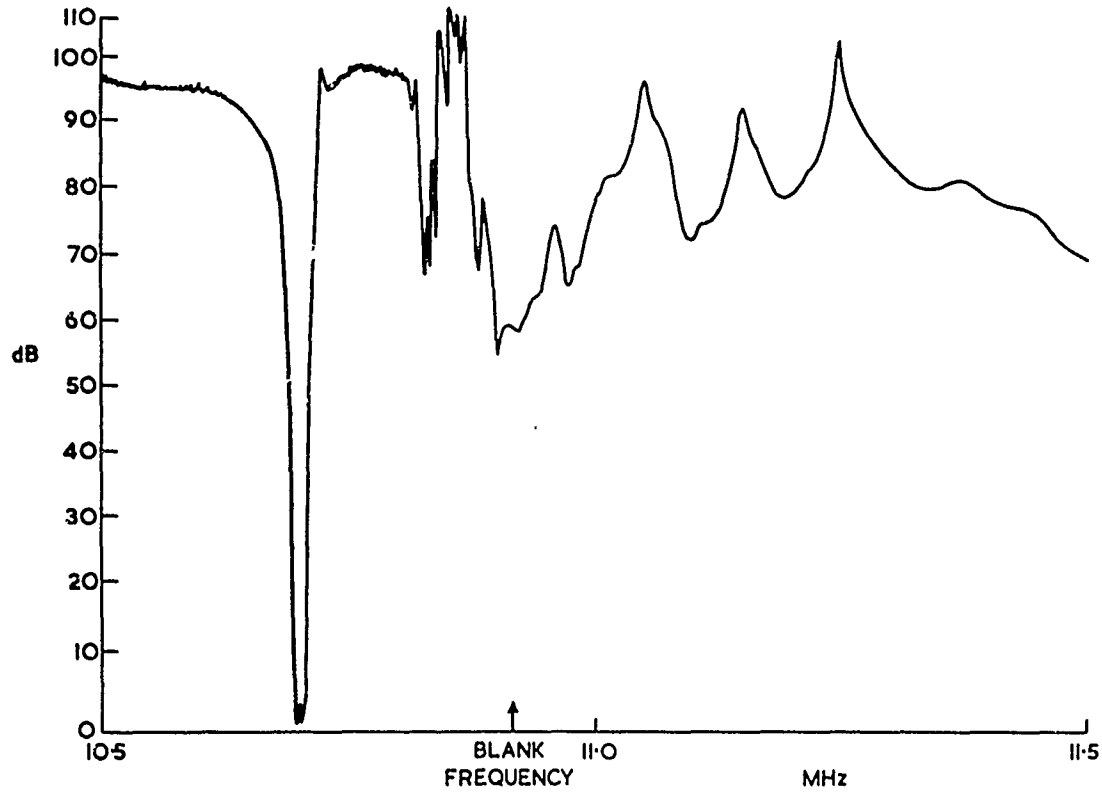
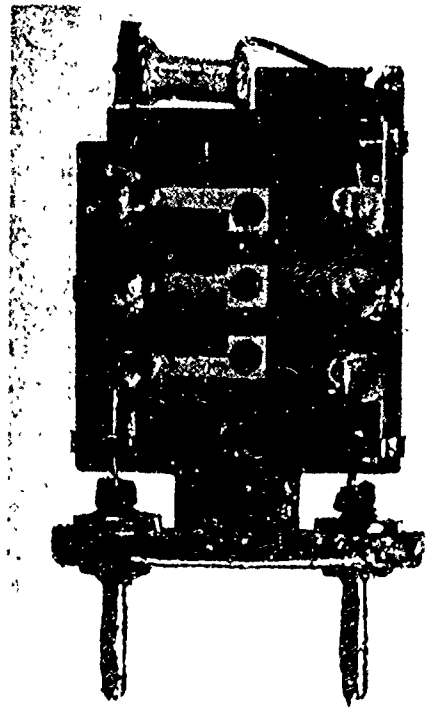


FIGURE 6





Reproduced from best available copy

FIGURE 8

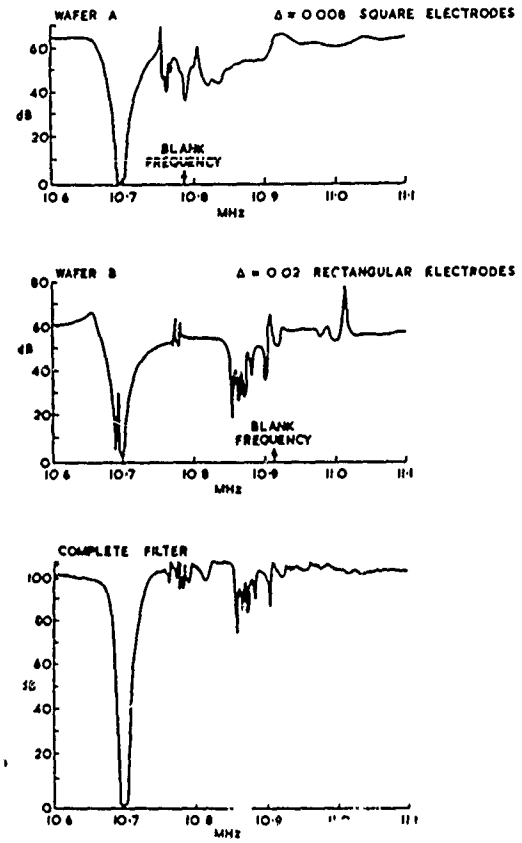


FIGURE 9

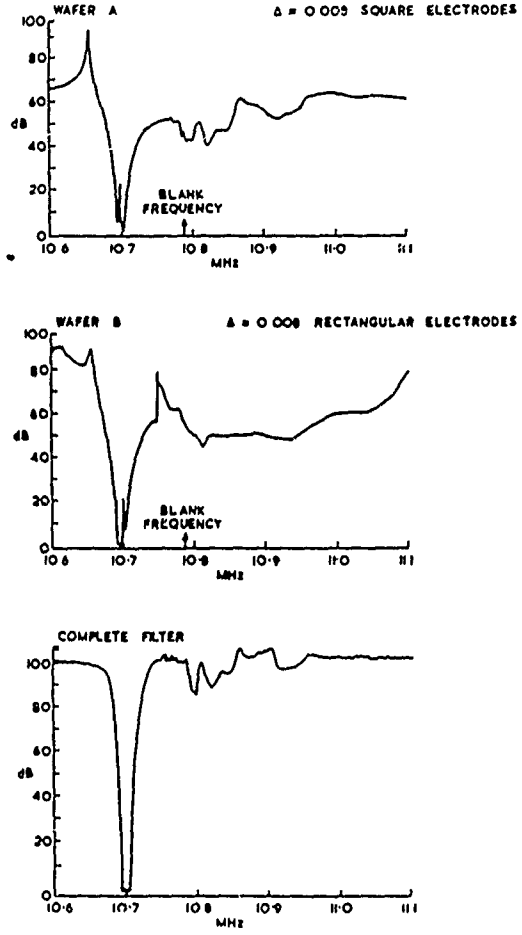


FIGURE 10

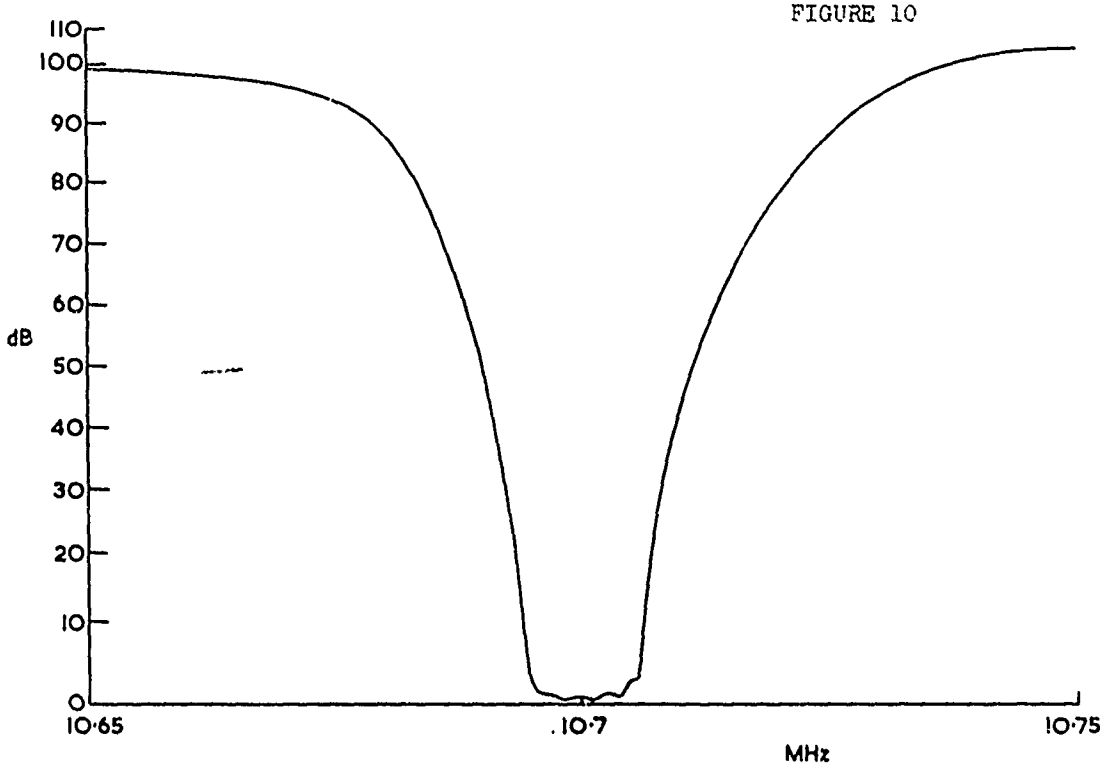


FIGURE 11

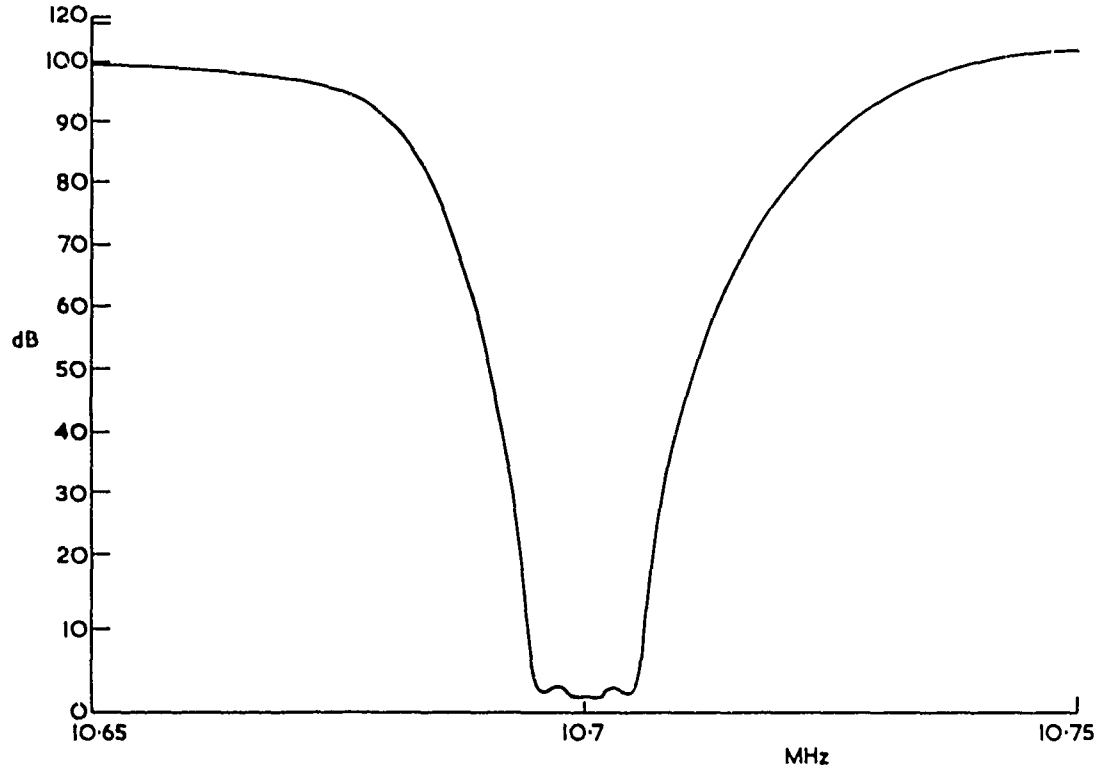


FIGURE 12

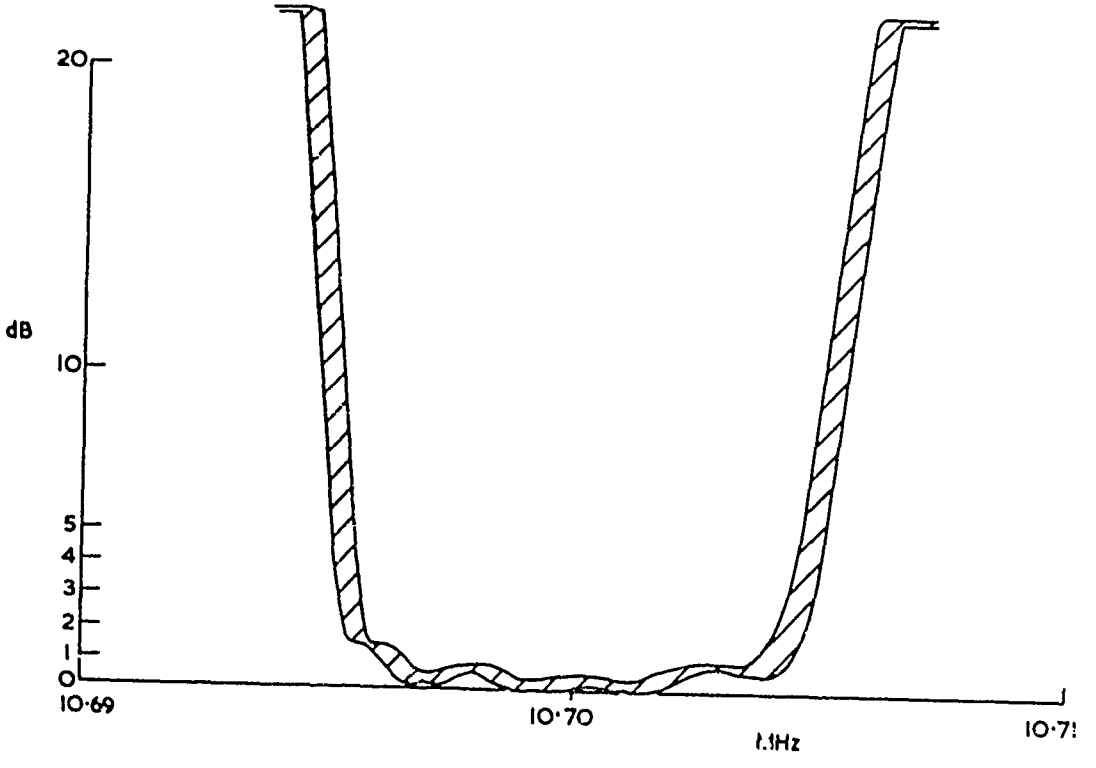


FIGURE 13

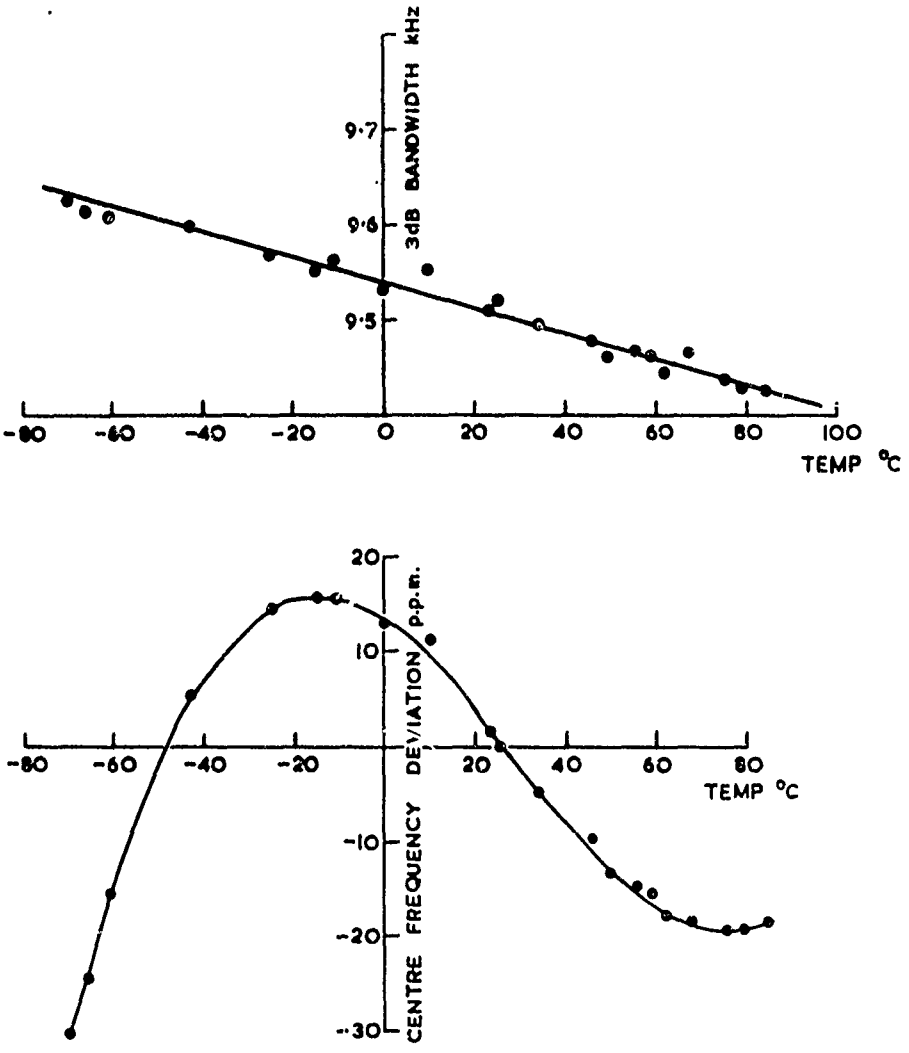

3dB BANDWIDTH kHz
9·7
9·6
9·5
-80
-60
-40
-20
0
20
40
60
80
100
TEMP °C
CENTRE FREQUENCY DEVIATION p.p.m.
20
10
-10
-20
-30
-80
-60
-40
-20
20
40
60
80
TEMP °C

FIGURE 14

HIGH FREQUENCY CRYSTAL MONOLITHIC (HCM) FILTERS

H. Yoda
Y. Nakazawa
N. Kobori
Toyo Communication Equipment Co., Japan

Introduction

Many reports have already been submitted on coupled-resonator-and cascade-connected filters, also known as high-frequency monolithic crystal filters.[1~10]

Development of this type filter has been rapid, resulting in the appearance of wide-band filters, narrow-band filters including those for SSB and channel applications, and VHF-band filters of the overtone type.

Introduced here will be filters of the attenuation peak type formed by cascading coupled resonators and narrow-band filters obtained by combining high-Q resonators of the overtone type.

HCM Filter with Attenuation Peak

Skirt characteristics of a filter with attenuation peaks at finite frequencies are sharper than those of a filter without peaks. Thus, a filter of a lower order or one with fewer elements has to be employed to obtain the required attenuation.

There are three methods of obtaining filters with attenuation peaks: inserting capacitance and by making use of acoustic properties or electrical properties. Discussing them in this order: Inserted capacitor method: Fig. 1 shows an HCM filter with the input-output terminals of a one-section bridge connected by capacitor C_p to produce attenuation peaks.[1] In filters other than the HCM, conventional mechanical or n-pole monolithic filters, for example, the capacitor is connected between two adjoining resonator elements.

Fig. 2 shows an example in which four cascade-connected-and-coupled resonators with the "n" order of the filter equalling eight are used as a filter without attenuation peaks. The drawbacks of this type of filter are that its characteristics are not symmetrical, and that in the vicinity of the upper cut-off point the skirt cuts into the pass band, pass-band ripples are increased because of the cascaded connection, and an accurate design for multiple sections can hardly be made. In Fig. 2, the broken line illustrates characteristics of the filter without attenuation peaks for the purpose of comparison.

Acoustical method: Various reports have been submitted on low-frequency filters with attenuation peaks, but few on high-frequency mechanical filters of the thickness shear or twist mode. Examples cited in such reports include the theoretical and basic construction method proposed by Börner[11] and in the patented method,[12] several resonators are connected with external couplers. However, where acoustical coupling is put to utilization, the disadvantages of these filters become evident. These are (1) construction is complicated and manufacture difficult, (2) spurious response is likely to be produced, and (3) adjustment is difficult.

Electrical method: Here, we introduce filters with attenuation peaks obtained electrically, as a result of our research. In filters of this type, the attenuation peaks are realized by using electrodes with special construction and making use of mutual cancelling of electrical charges so that the impedances of the two modes are varied.

General Discussion of Filters with Attenuation Peaks

The filter without attenuation peaks to which the pass-band-Tchebyscheff approximation is applied is subjected to K-derived conversion.

Fig. 3 shows a circuit which is converted to a lattice circuit of the n equals 5 order of the filter.

Z_{a1}, Z_{b1}, Z_{a2}, and Z_{b2} denote the arms of an attenuation-peaklers Tchebyscheff filter. The impedance Z of each of the arms can be approximated as shown here:

$$Z = \frac{1}{\omega C_0} \frac{\omega^2 - \omega_\infty^2}{\omega^2 - \omega_0^2} \simeq \frac{1}{\omega C_0} \frac{\omega - \omega_\infty}{\omega - \omega_0} \qquad (1)$$

$$\omega = \omega_0 + B_w \qquad (2)$$

ω_0 is center angular frequency
B_w is bandwidth

Substituting equation (2) into (1),

$$Z = \frac{1}{\omega C_0} \frac{\Omega - S}{\Omega - A} \qquad (3)$$

Each arm impedance is given as

$$Z_{ij} = -jK_{0_j} \frac{\Omega - S_{i_j}}{\Omega - A_{i_j}} \text{ ; where, } K_{0_j} = \frac{1}{\omega C_{0_j}} \qquad (4)$$

In the equation above, "i" denotes "a" or "b" of the series arm or lattice arm, and "j" the section number when the K-derived conversion to the impedance of each arm in the first

section, we obtain equation (5) and (6).

$$m_1 Z_{a1} - \frac{Z_{b1}}{m_1} = -J\frac{k_{o1}}{m_1}\,\frac{m_1^2(\Omega-S_{a1})(\Omega-A_{b1})-(\Omega-A_{a1})(\Omega-S_{b1})}{(\Omega-A_{a1})(\Omega-A_{b1})} \quad (5)$$

$$(m_1 Z_{a1})\cdot\frac{Z_{b1}}{m_1} = k_{o1}^2\,\frac{(\Omega-S_{a1})(\Omega-S_{b1})}{(\Omega-A_{a1})(\Omega-A_{b1})} \quad (6)$$

From these two equations, the four terminal constants are:

$$A_1 = D_1 = \frac{1}{M_1}\, m_1^2(\Omega-S_{a1})(\Omega-A_{b1})+(\Omega-A_{a1})(\Omega-S_{b1})$$

$$B_1 = \frac{1}{M_1}(2Jm_1k_{o1})(\Omega-S_{a1})(\Omega-S_{b1}) \quad (7)$$

$$C_1 = \frac{1}{M_1}\left(2J\frac{m_1}{k_{o1}}\right)(\Omega-A_{a1})(\Omega-A_{b1})$$

where

$$M_1 = m_1^2(\Omega-S_{a1})(\Omega-A_{b1})-(\Omega-A_{a1})(\Omega-S_{b1})$$

Since the attenuation peak is $\Omega_{\infty 1}$, m_1 is thus calculated.

$$m_1 = \sqrt{\frac{(\Omega_{\infty 1}-A_{a1})(\Omega_{\infty 1}-S_{b1})}{(\Omega_{\infty 1}-S_{a1})(\Omega_{\infty 1}-A_{b1})}} \quad (8)$$

Z_{a2} and Z_{b2} also can be calculated in the same manner.

Fig. 4 shows values calculated in this manner. In this example, the M derived conversion is applied to the attenuation-peakless Tchebyscheff approximation filter of 0.2 dB deviation within the pass band n = 8. The case of $m_1 = m_2 = 1.0$ is for the filter without attention peaks, that where $m_1 = 1.0$ and $m_2 = 0.844300$ is for the M-converted Z2 and Z3, and that where $m_1 = 0.938263$ and $m_2 = 0.832960$ is for obtaining Cauer response in all arms. Thus, deviation within the pass band may increase slightly. However, this has no substantial effect on the general specification. On the other hand, the attenuation range is greatly improved.

Relationship between Electrode Construction and Inductance

As already mentioned, the M derived conversion can be made in the attenuation-peakless Tchebyscheff approximation. Or, making Z_a as mZ_a and Z_b as Z_b/m, a filter with attention peaks can be obtained. Therefore, inductances of the arms also must be made mL_a and L_b/m. This can be realized employing electrodes with special construction so that electrical charges cancel each other.

Of the electrodes shown in Fig. 5 (a), $\phi 1$ and $\phi 2$ are not divided for an attenuation peakless filter. Of the coupled resonators, the equivalent of both modes are nearly equal. In this example, two pairs of electrodes are further divided to obtain main electrode ϕ_1 and sub-electrode ϕ_2. Making us of mutual cancellation of electric charges, connections are made as shown in Fig. 5 (b) so that the impedances of the series and lattice arms can be varied. Or, the electrodes are connected in such an electrode pattern that the charge produced at symmetrical mode fs is mutually multiplied with anti-symmetrical mode fa, so that the charge produced at fs is mutually cancelled.

The equivalent inductance can be varied by varying the ratio between the inductances of main electrode $\phi 1$ and sub-electrode $\phi 2$. Experimental results of this are shown in Fig. 6. Fig. 6 (a) indicates that attenuation peak points vary as the ratio of the electrode dimensions are varied at 10.7 MHz (fa - fs) = 8.7 kHz. The case of m = 1 denotes the attenuation-peakless filter without a sub-electrode. The solid line in Fig. 6 (b) shows the calculated value of the attenuation peak with respect to the inductance ratio L_b/L_a of both arms; the broken line indicates the relationship between divided electrode ratio $2\phi_1/\phi_s$ and attenuation-peak frequency f_∞, as obtained $2\phi_1/\phi_s$ by experimental measurement.

Fig. 7 shows an example of filter manufacture based on this theory. This filter has an n-value of 6, and consists of three coupled resonators. The theory just mentioned is applied to two of the three coupled resonators. At a center frequency of 10.7 MHz and bandwidth of ±7.5 kHz, so that the attenuation peaks conform to Cauer response, the filter is formed from one constant-K section and two M derived sections with $m_1 = 0.807$ and $m_2 = 0.866$, making a cascade-connection filter.

The broken line represents the attenuation-peakless filter, for the purpose of comparison.

A spurious response is produced at approximately +47 kHz. This spurious response is a result of X-direction anti-symmetrical mode (2.1.1) which is caused by complex electrodes. The problem of suppression of such spurious response is yet to be solved.

Fig. 8 shows characteristics of a filter with n equal to 8 which consists of four coupled resonators. This theory is applied to two of the four resonators. Comparison is made to the attenuation-peakless filter. As can be seen, skirt characteristics of the filter with attenuation peaks are much better than those of the filter without attenuation peaks.

Narrow Band HCM Filter

Many reports concerning high-frequency narrow-band filters have already been made by personnel at Bell Telephone Laboratories.[2, 3, 5, 6, 8] In SSB filters, channel filters, or others that require narrow bandwidth and sharp cut-off, the temperature characteristics, aging characteristics, and element tolerance are important factors. However, such filters cannot be produced unless high-Q resonators are used.

In general, it is quite difficult to manufacture crystal elements having a Q of 100,000 at 10 MHz.

If the specific bandwidth is to be as narrow as 3×10^{-4}, unless resonators of high-Q are used, pass-band deviations and insertion losses increase. Therefore, trial production of narrow-band filters was accomplished by using high-Q third-overtone resonators.

Q of Resonator and Characteristics of Filter

Let's investigate the effect of resonator Q on the cut-off characteristics of the filter. The impedance Z_1 of each arm of the foregoing circuit can be given as:

$$Z_{1j} = -j\frac{\omega_{Sij}^2}{\omega}\left(\frac{\omega_{Aij}^2}{\omega_{Sij}^2} - 1\right)\frac{\left(\frac{\omega}{\omega_{Sij}} - j\frac{1}{2Q_{ij}}\right)^2 - \left(1 - \frac{1}{4Q_{ij}^2}\right)}{\left(\frac{\omega}{\omega_{Sij}} - j\frac{1}{2Q_{ij}}\right)^2 - \left(\frac{\omega_{Aij}}{\omega_{Sij}} - \frac{1}{4Q_{ij}^2}\right)} \quad (9)$$

$$Q_{1j} = \frac{\omega_0 L_{1j}}{R_{1j}} \quad (10)$$

Taking the loss dissipation of the resonator alone into consideration, calculations were made on the assumption of an attenuation-peakless Tchebyscheff approximation, with n equalling 8, (pass-band deviation) equalling 0.2 dB, center frequency equalling 10.7 MHz, bandwidth equalling ±1.6 kHz, $Q \approx 4 \times 10^4 \sim 2 \times 10^6$. Based on these assumptions, variation or degradation of the pass-band deviation and insertion loss were calculated. The result of calculation is shown in Fig. 9. Naturally, a greater effect results and degradation is larger in filters whose specific bandwidths are narrower.

Using the overtone resonators, Q to the order of 200,000 can be realized without much difficulty at a frequency of 10 MHz. The overtone system is advantageous also from the viewpoint of plate-back frequency and electrode gap. Moreover, it is advantageous also from the viewpoints of aging characteristics and reliability.

Relationship Between Fundamental and Overtone Modes

Onoe and Jumonji and Beaver already have submitted reports on the frequency difference between symmetrical modes and anti-symmetrical modes at the fundamental frequency.[6, 10]

On the other hand, at the overtone, also, taking $n\sqrt{\Delta}$ (n: order of overtone; Δ: plate-back frequency) as a variable, analysis can be made in a fashion similar to that for the fundamental frequency irrespective of the order of overtone (n). Therefore, in an overtone resonator of the same electrode construction and same plate-back frequency, the frequency difference $(f_a - f_s)$ decreases as the order of overtone (n) becomes higher.[13] That is to say, Q increases as the electrode gap and plate-back frequency (thickness of the electrode) increase, and therefore a narrow-band filter can be obtained with thin electrodes if the overtone mode is employed.

Fig. 10 shows the relationship between the order of overtone and the frequency difference. In this example, the 3rd, 5th, and 7th overtones of a 10 MHz AT-cut crystal were employed.

Test Manufacture of Narrow-Band Filters

The 3rd overtone mode is employed in order to attain high Q. Examples of test filters are shown in Fig. 11.

The tested resonators were Vebel-processed discs having a diameter of 13.8 mm and third overtone frequency $f_3 = 12.8$ MHz. Construction and constants of the filters are shown in Fig. 11. These filters were designed for α (pass band deviation) = 0.2 dB and n = 8.

The values of Q were approximately 200,000 for both the S and A mode. Four of these resonators were connected in series. Characteristics obtained are shown in Fig. 12.

The illustration shows an enlargement of the pass-band characteristics in comparison with those of CCITT Recommendations 1 5. The insertion loss is as small as 1.2 dB, since Q of the resonators is high.

The attenuation peaks are caused by stray capacity. This is due to capacitor insertion.

Fig. 13 shows measured values of the filter designed for fo = 12.8 MHz, n = 1[illegible]5-coupled resonators in cascade), and α = 0.2 dB.

Of the narrow-band filters, the spurious response is very small as the terminated impedance is low and the many sections in cascade are employed. Attenuation of 65 dB for the filter of n = 8 and that of more than 85 dB for the filter of n = 10 can be guaranteed.

In attempting to improve pass-band deviation, we employed resonators with a high Q. Such may be realized through some method for compensating resonator loss, although this problem has not yet been solved.

Conclusion

We have introduced filters with attenuation peaks. Today the capacitor-insertion and acoustic methods are considered practical for attaining desired results. We explained in detail the electrical method. We prepared coupled resonators, the equivalent impedance of which are varied to make use of mutual cancellation

of electrical charges produced, and we verified through theoretical calculation and actual experiment that the simultaneous Tchebysheff approximation is applicable in multi-section coupled resonators.

Of the narrow-band filters of which specific bandwidths and resonators of high Q are required, in order to attain high Q, we employed the 3rd overtone mode.

Acknowledgement

We wish to acknowledge the valuable guidance given by Professor Onoe of Tokyo University.

References

(1) Y. Nakazawa; Proc. 16th Freq. Cont. Symp. (1962)
(2) R.A Sykes, W.L. Smith, and W.J. Spencer; IEEE International Convention Record (1967)
(3) R.A Sykes and W.L. Smith; Bell Lab. Record Feb. (1968)
(4) M. Oneo, H. Jumonji, and K. Kobori; Proc. 20th Freq. Cont. Symp. (1966)
(5) R.A. Sykes and W.D. Beaver; Proc. 20th Freq. Cont. Symp. (1966)
(6) W.D. Beaver; Proc. 21st Freq. Cont. Symp (1967) and JASA 43. 5. (1968)
(7) W.H. Horton and R.C. Symth; Proc. 21st Freq. Cont. Symp. (1967)
(8) D.I. McLean; Proc. 21st Freq. Cont. Symp. (1967)
(9) H. Yoda, Y. Nakazawa, S. Okano, and K. Kobori; Proc. 22nd Freq. Cont. Symp. (1968)
(10) M. Onoe and H. Jumonji; J. Inst. Elec. Comm. Eng. Japan 48 (1965)
(11) M. Börner; AEÜ 17, 103, (1963)
(12) Telefunken patent; 43-15538 Japan patent, Sept. (1964)
(13) M. Onoe and S. Kumagai; Proc. Nat. Conv. Inst. Elec. Comm. Eng., Japan S-9 (1968)

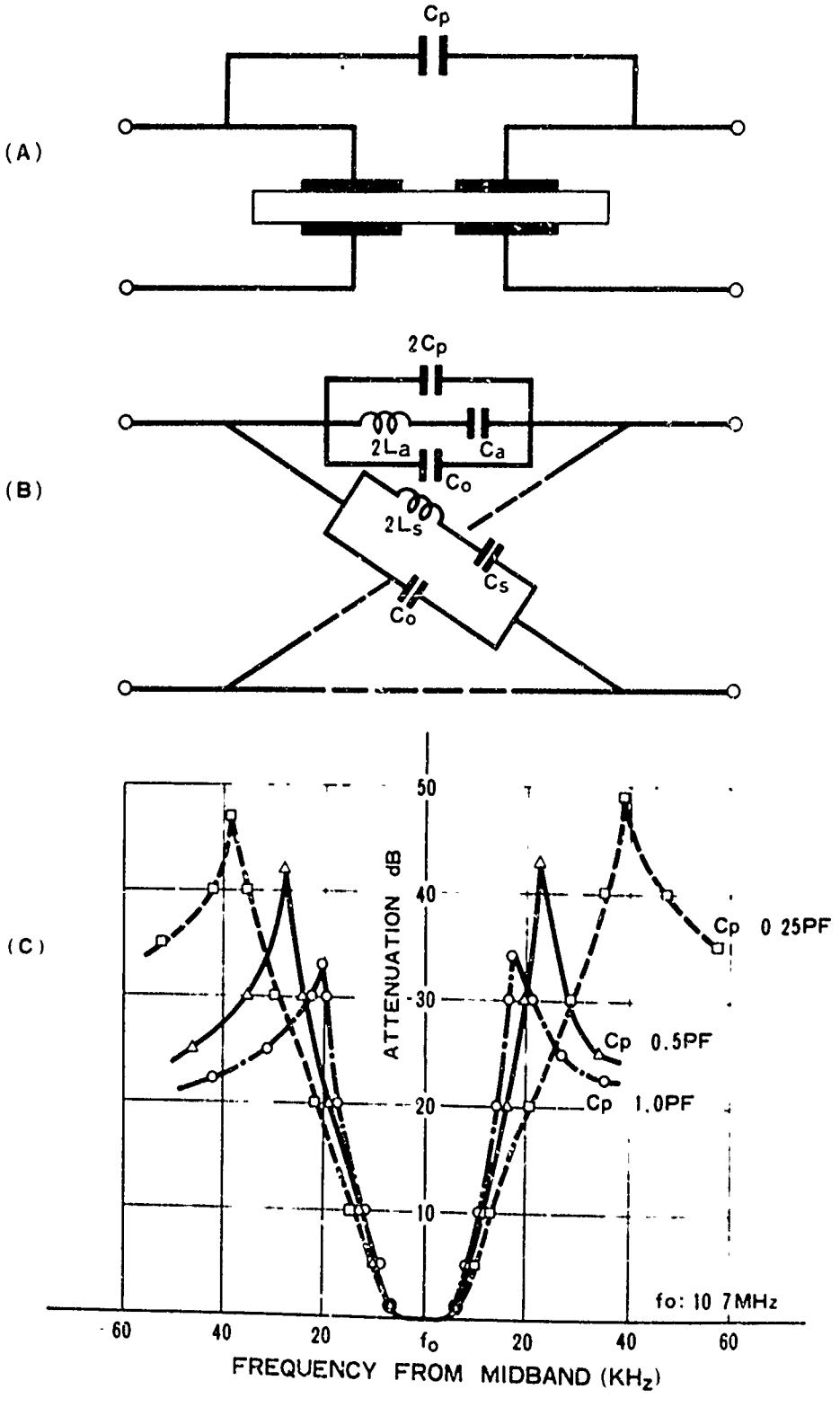
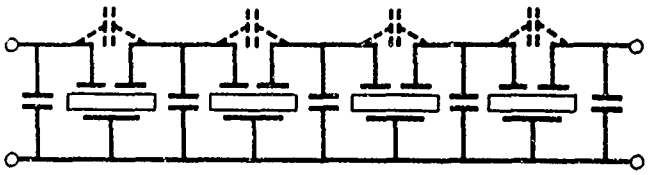


FIG 1

One section HCM filter with attenuation peak (capacitor inserted between input and output)



f_o = 10.7MHz
ORDER OF FILTER (n) = 8
TERMINATING RESISTANCE = 3 KΩ
INSERTION LOSS = 1.9dB

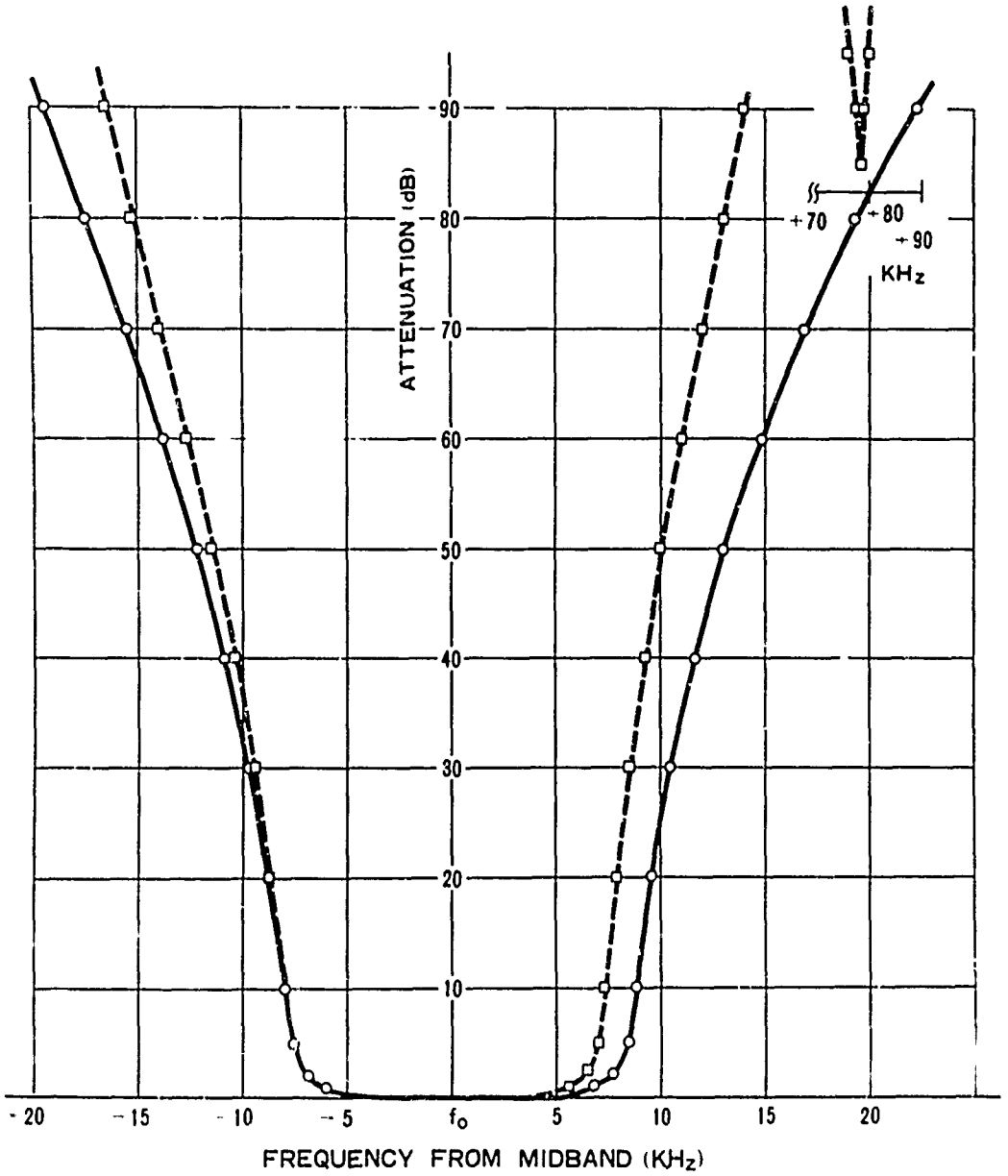
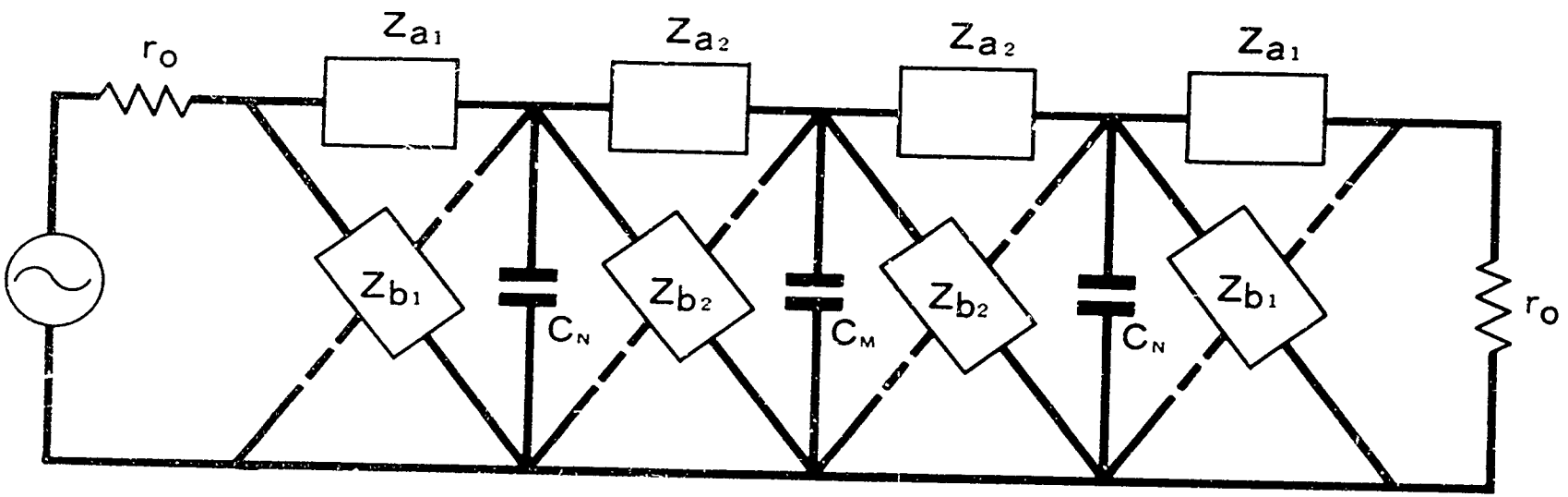


FIG. 2

Four section HCM filter with attenuator peaks, capacitor inserted



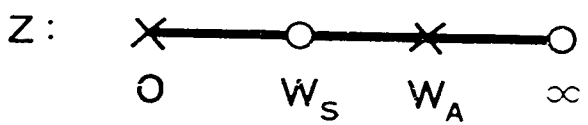


FIG. 3

Configuration of Filter (n 8) having Tchebyscheff response

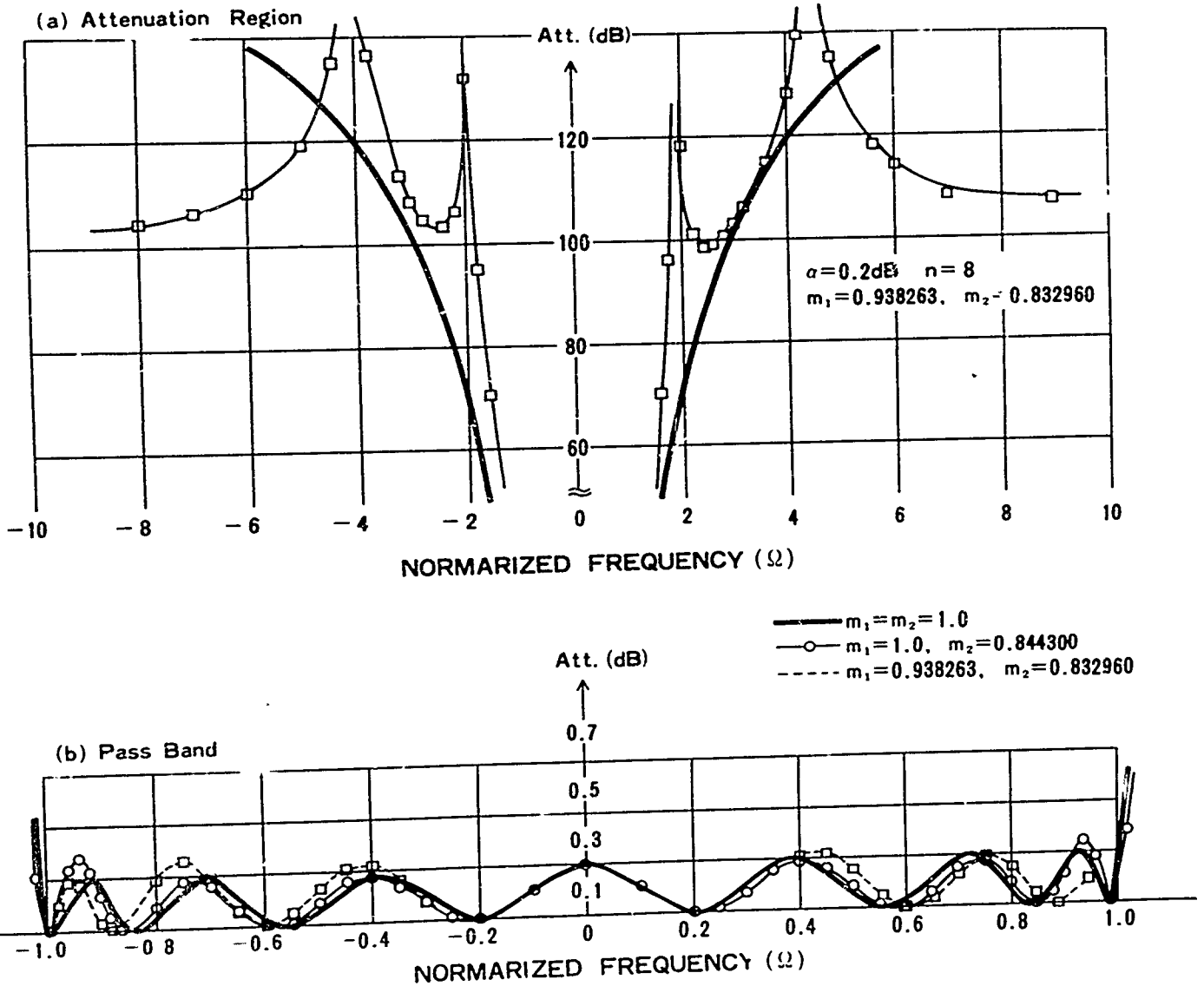
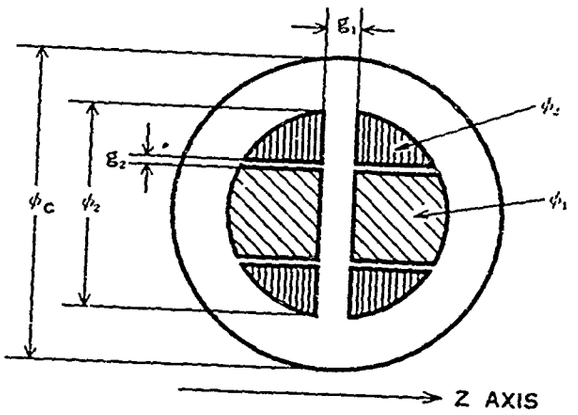
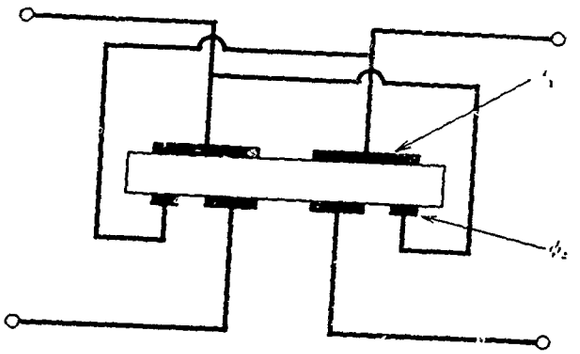


FIG. 4

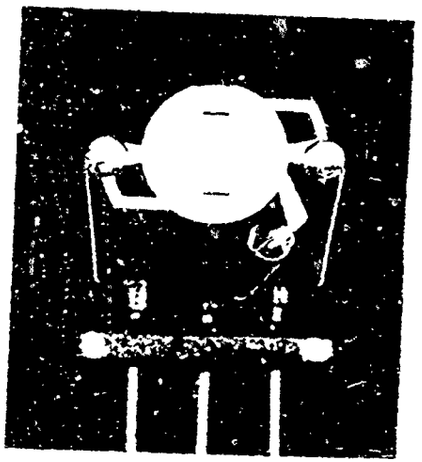
Computed values of attenuation applied derived M conversion on Tchebyscheff responses



(A)



(B)



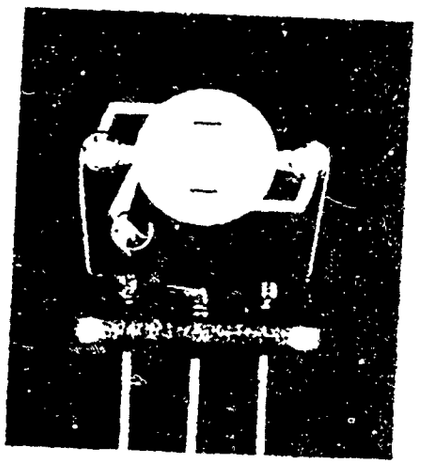


FIG. 5

A coupled resonator with attenuation peaks due to electrical method and photograph of resonator

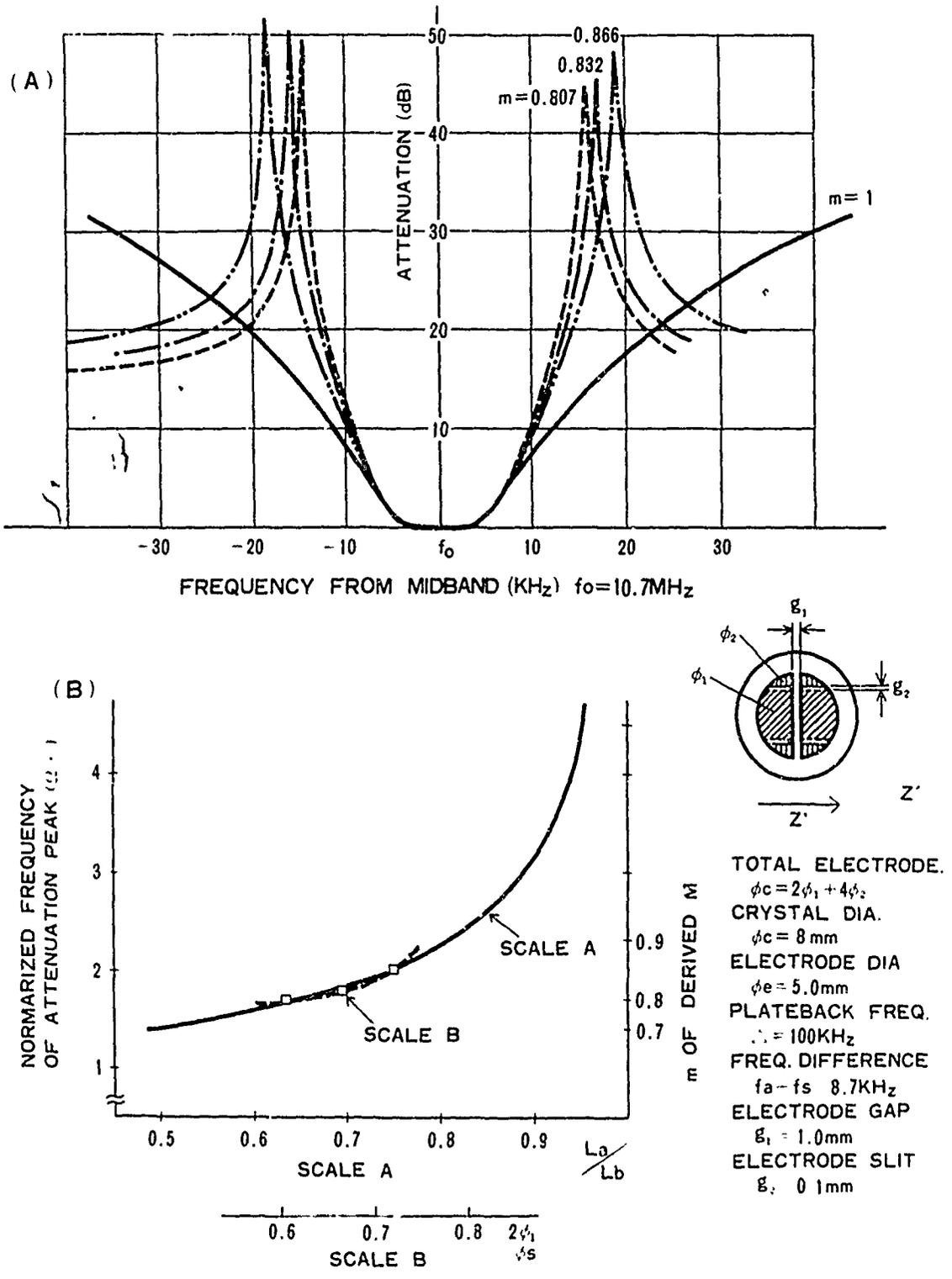
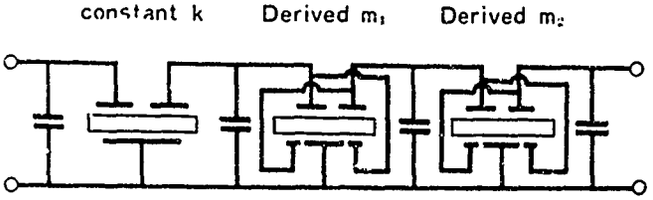


FIG. 6

Relation between divided electrode ratio $2\phi_1/\phi_s$ and attenuation peak frequency Ω_∞



ORDER OF FILTER (n_f) = 6
TERMINATING RESISTANCE = 3 KΩ
INSERTION LOSS = 1.6dB
f_o = 10.7MHz

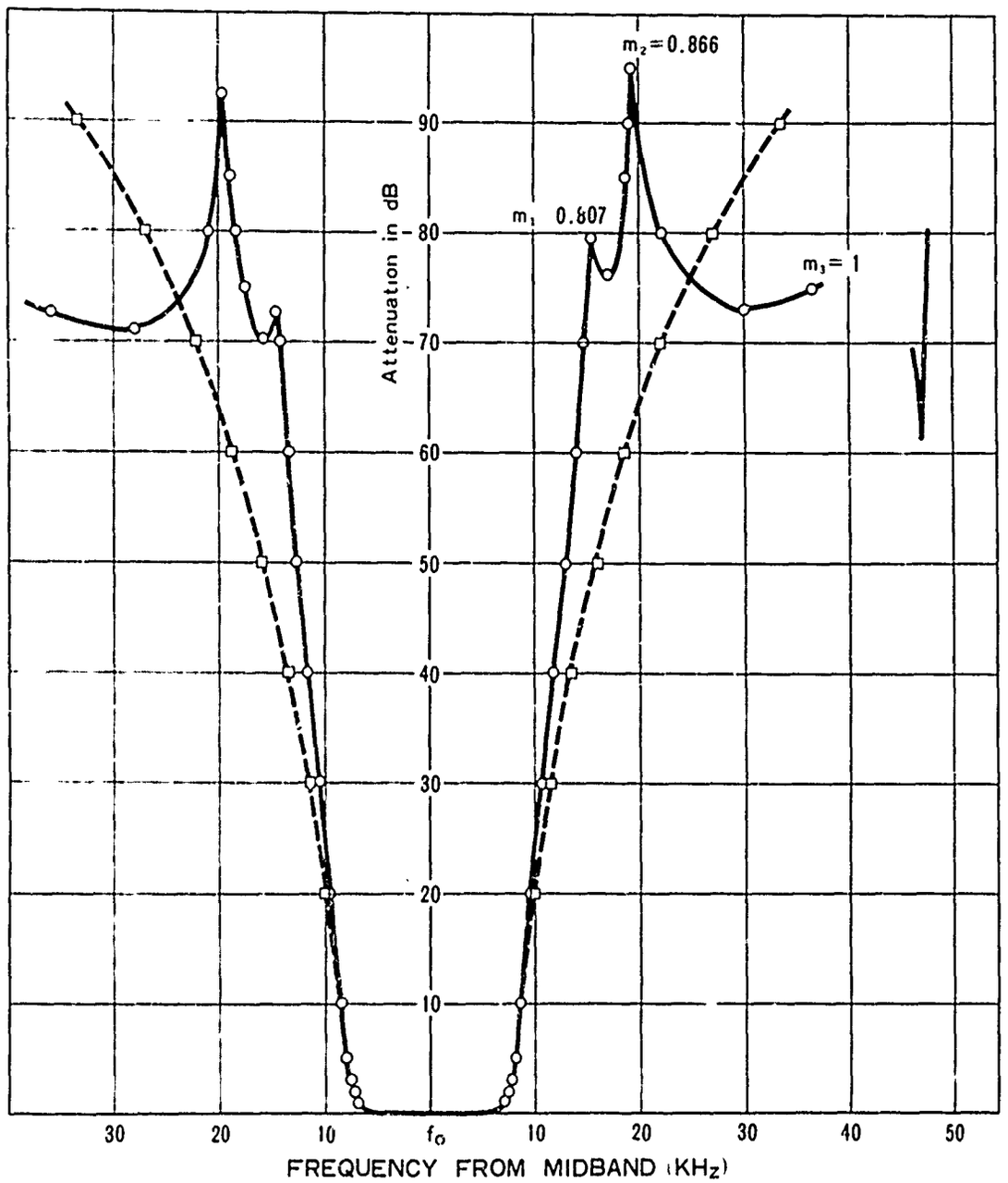


FIG. 7

Example of filter with attenuation peaks (m_1 = 0.6)

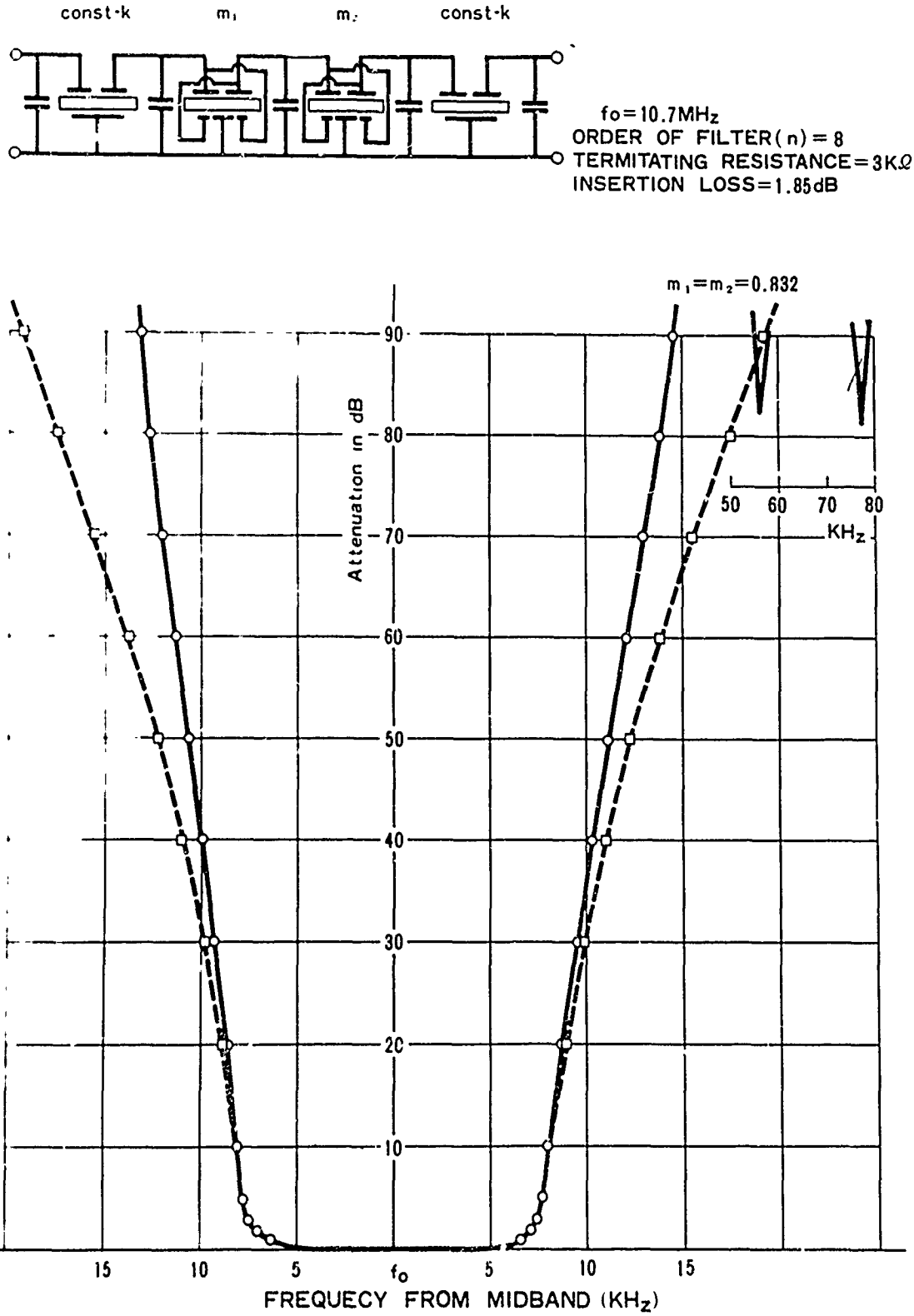


FIG. 8

Example of filter with attenuation peaks (n = 8)

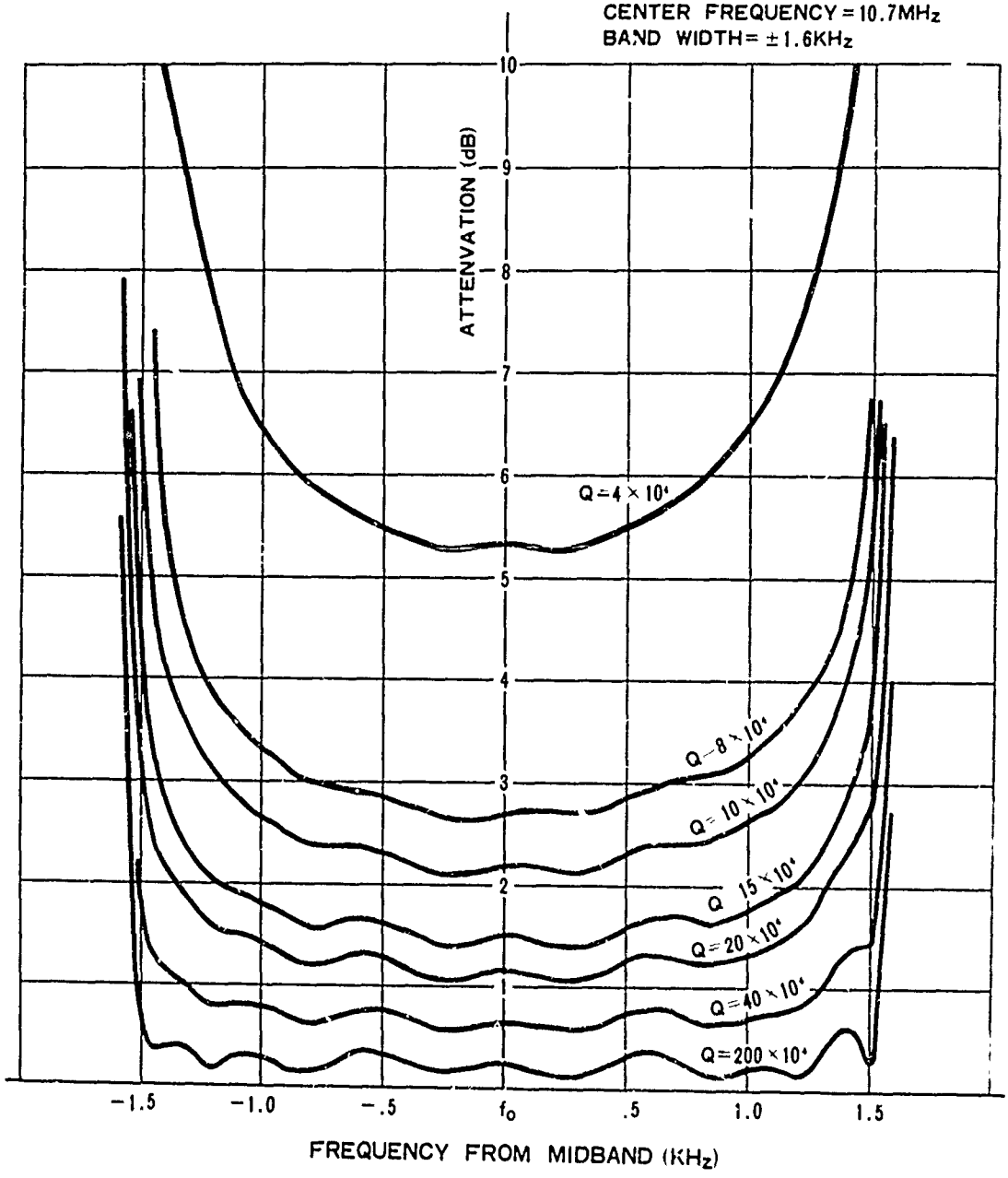
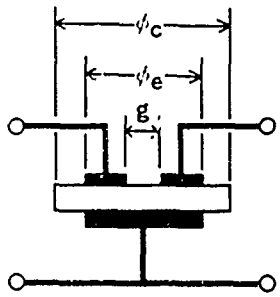


FIG. 9

Relation between Q value of the resonator and pass band characteristics, insertion loss



THICHNESS TWIST
FOUNDAMENTAL FREQUENCY f_1=10MHz
ELECTRODE GAP g=0.2mm
CRYSTAL DIA. ϕc= 8 mm
ELECTRODE DIA. ϕe= 2 mm
PLATE BACK FREQ. △=100KHz

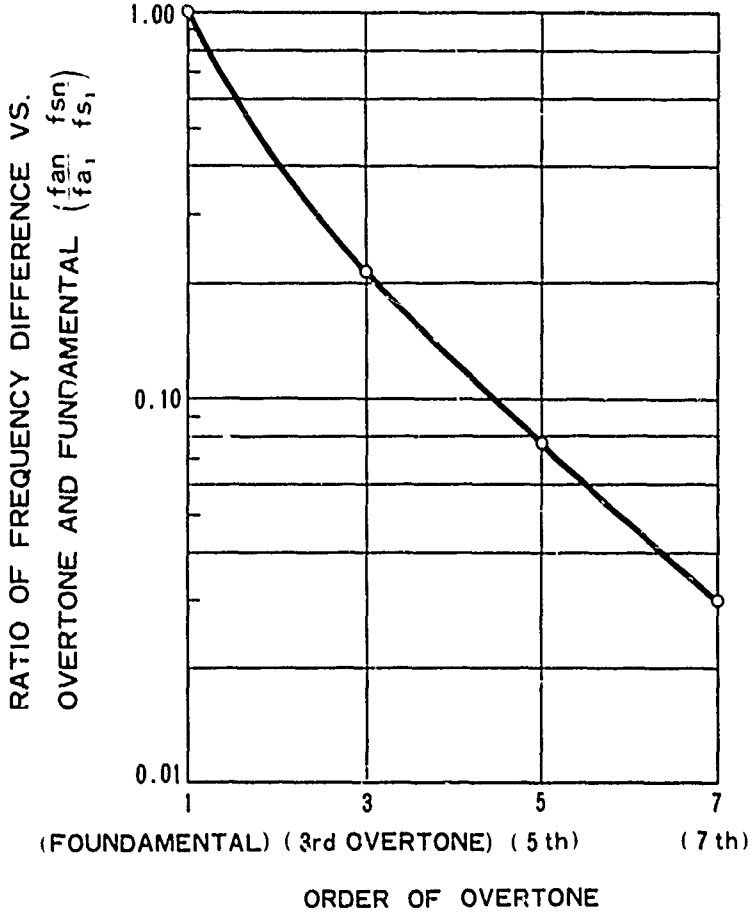
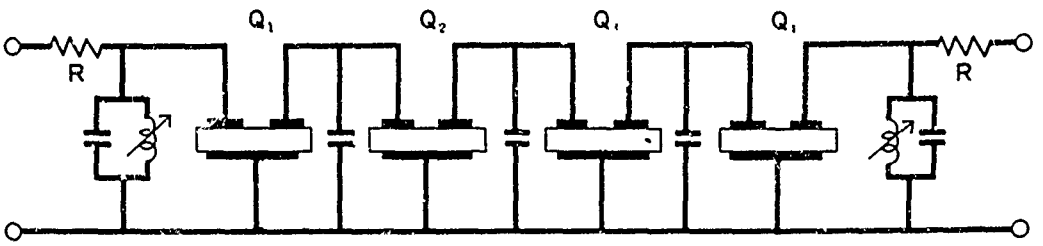


FIG. 10

Relation between order of overtone and resonant frequency difference



NO. \ ITEM / MODE		FREQUENCY (MHz)	EQUIVALENT INDUCTANCE mH	EQUIVALENT RESISTANCE Ω	Q VALUE
Q_1	S MODE	12.825 246	227	96	$19.1 \cdot 10^4$
	A MODE	12.827 990	258	75	27.8
Q_2	S MODE	12.825 597	224	112	16.4
	A MODE	12.827 706	284	150	15.6
Q_3	S MODE	12.825 505	225	84	21.6
	A MODE	12.827 585	234	95	19.9
Q_4	S MODE	12.825 243	218	84	21.0
	A MODE	12.828 023	236	87	21.9

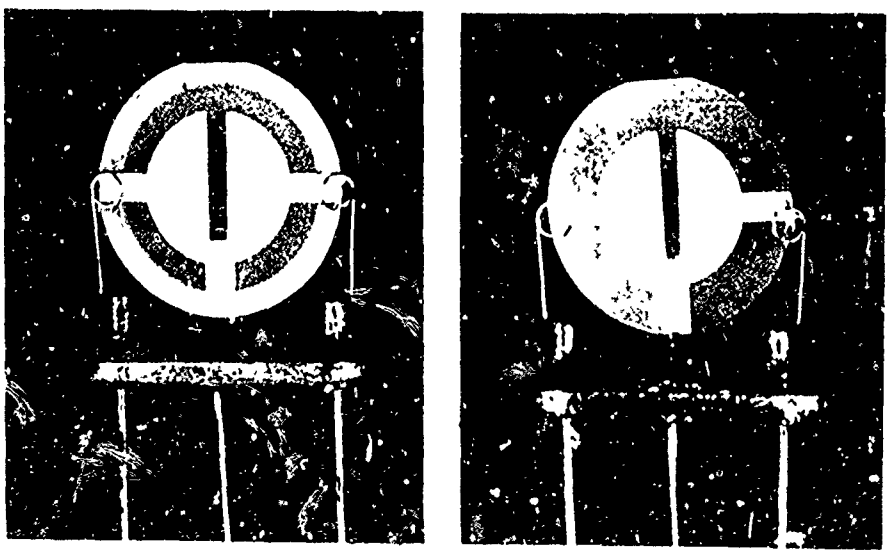


FIG. 11

Configuration of a 3rd overtone narrow-band HCM filter, electrical equivalent constant of each 3rd overtone coupled resonator, and photograph of resonators.

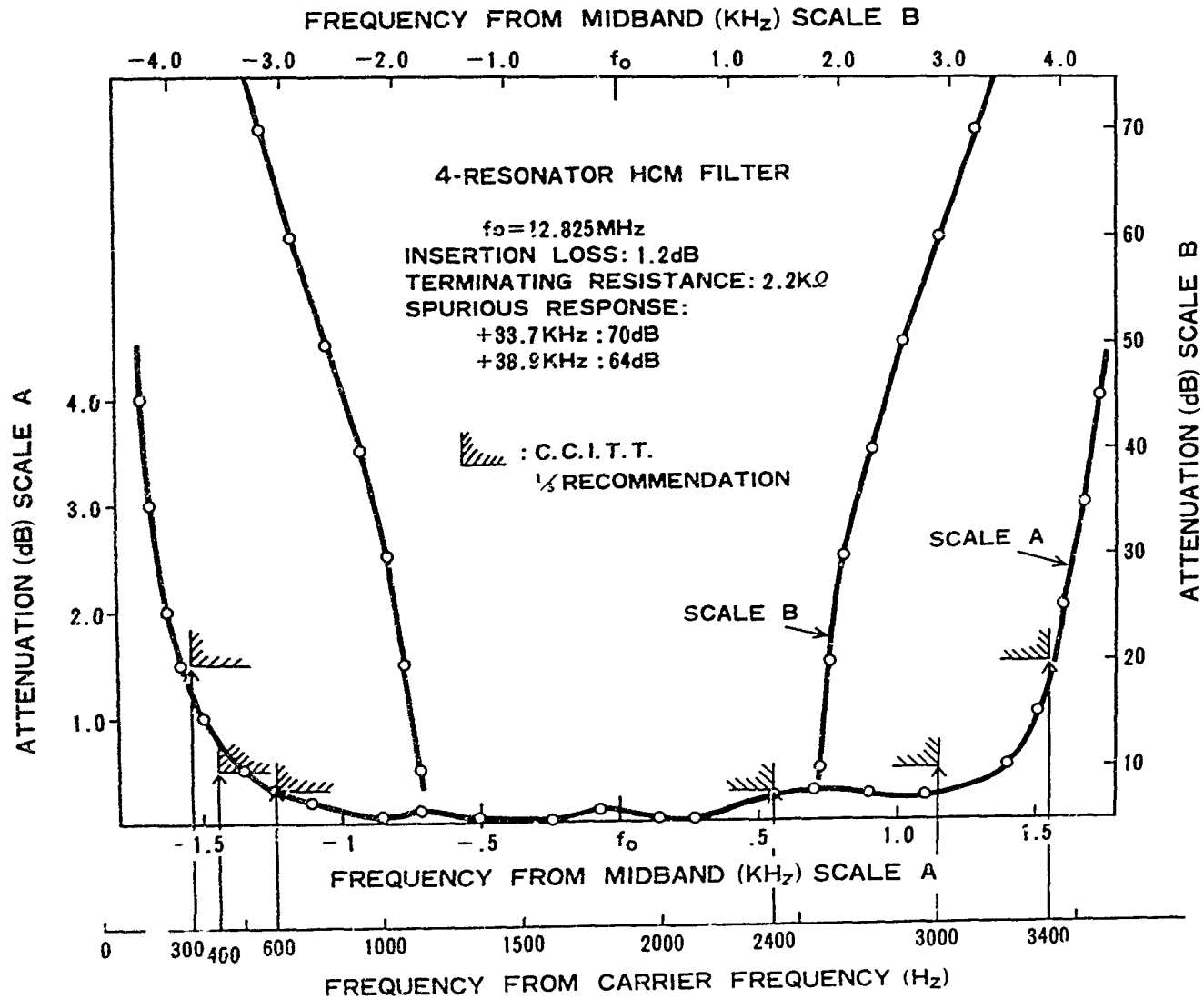


FIG. 12

3rd overtone narrow-band HCM filter (n = 8)

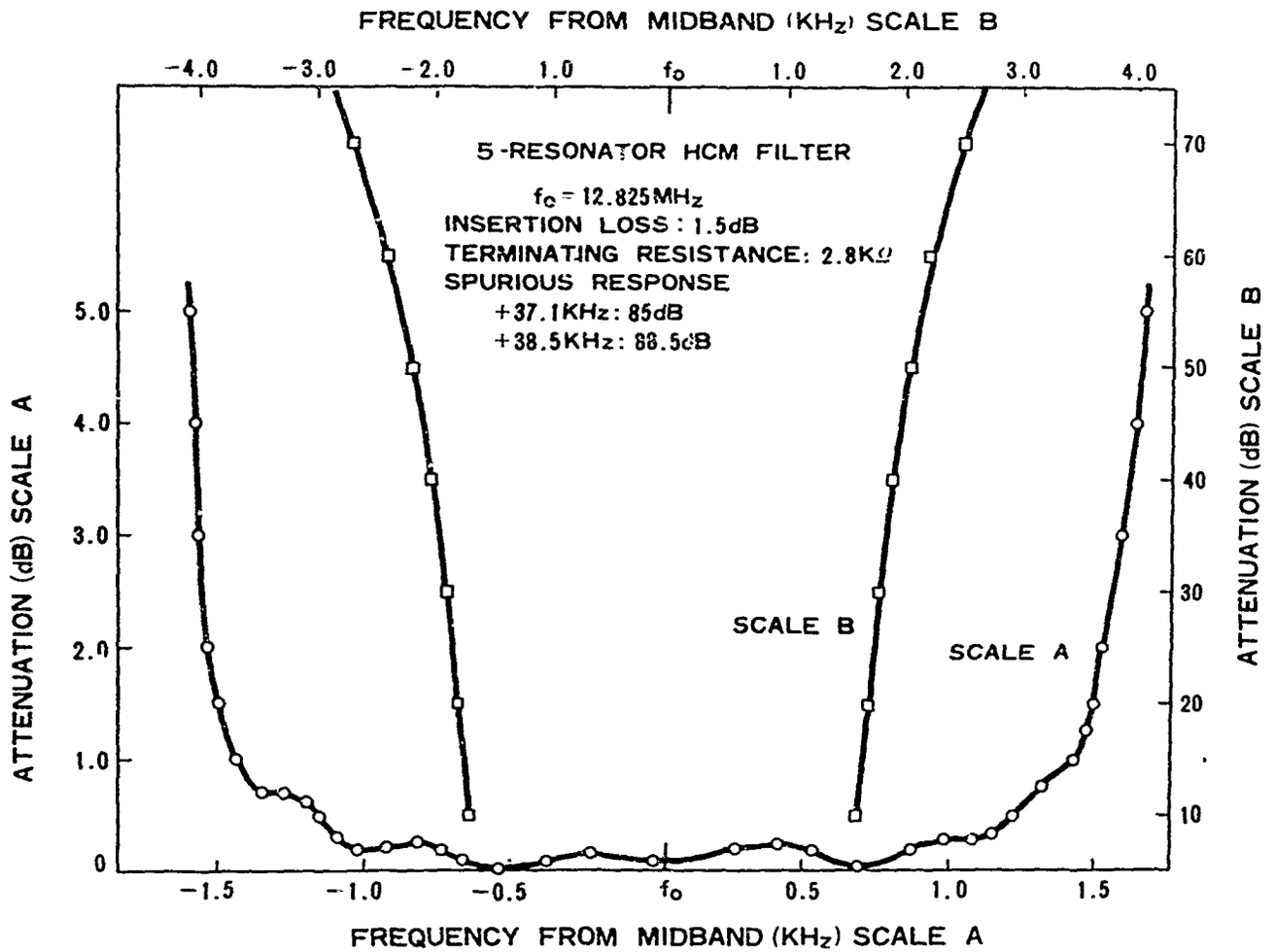


FIG. 13

3rd overtone narrow-band HCM filter (n = 10)

VECTOR VOLTMETER CRYSTAL MEASUREMENT SYSTEM

Marvin E. Frerking
Collins Radio Company
Cedar Rapids, Iowa

Summary

This paper describes a crystal measurement system which is built around an rf vector voltmeter. In this system, the crystal under test is part of an oscillating loop which generates the rf signal. The phase shift around the loop is adjusted until the phase angle across the crystal is zero, as indicated by the vector voltmeter, and the frequency is read using a counter. The voltage capability of the vector voltmeter is used to measure the voltage on either side of the crystal. The crystal resistance and drive level are determined from these readings.

Test data taken in the 3 to 20 MHz region indicates that the accuracy of the system is about 5 PP 10^8 for semiprecision crystals. The resettability is 1 or 2 parts in 10^8. The system is capable of performing measurements at drive levels as low as a few microwatts and as high as several milliwatts.

An experimental test set using these principles was built around the Hewlett-Packard 8405A Vector Voltmeter. The set covers the frequency range from 2.5 to 20 MHz and was built primarily for temperature tests of crystals at low drive level.

The oscillating loop is composed of a wide-band amplifier, an adjustable diode limiter circuit, a tunable phase shifter capable of being both mechanically and electrically adjusted, and a test fixture containing the crystal.

The vector voltmeter probes are connected across the crystal so that the phase angle across the crystal is sensed. The phase reading is available on the rear of the vector voltmeter as a recorder output. This output is amplified by an operational amplifier and applied to voltage variable capacitors in the phase shifter. This negative feedback tends to stabilize the system and also make the mechanical adjustment easy to set.

An interesting and useful feature of the system results from the fact that the crystal is operated in a π network which is located at the end of coaxial lines. It is thus convenient to operate the crystal in a temperature chamber with the π network.

The voltage reading of the vector voltmeter is available as a recorder output. Since the crystal voltage on the output side of the π network is a measure of crystal activity, it is used to record crystal resistance during temperature runs.

Introduction

The Vector Voltmeter Crystal Measurement System is a semiprecision crystal test set. It is not quite as accurate as a bridge system but it is considerably less expensive and much easier to operate. In many cases, it is easier to operate than a Q-meter, yet the accuracy is sufficient to satisfy nearly all of the semiprecision and nonprecision measurement needs. The absolute accuracy is about 5 PP 10^8 while the resettability is in the order of a part in 10^8.

An experimental unit was built to demonstrate the principles involved. It was built around the Hewlett-Packard 8405A Vector Voltmeter and covers the frequency range from 2.5 to 20 MHz.

General Description

The vector voltmeter test set is basically an oscillating loop containing an amplifier, phase shifter, and the crystal. The phase angle across the crystal is sensed by the vector voltmeter. The phase shift around the loop is adjusted until the frequency of oscillation is of the proper value to produce a zero phase angle across the crystal. The frequency is then read using a frequency counter. A photograph of the experimental unit is shown in figure 1. A block diagram is shown in figure 2. As can be seen from the diagram, the oscillating loop consists of five units which perform the following functions. The amplifier provides the necessary gain to make the system oscillate. It is broadbanded and consists of five transistors providing 45 dB of gain. The output of the amplifier is passed through a limiter whose threshold level is adjustable by means of a control on the front panel. The limiter effectively sets the crystal drive level. It consists of two diodes in series with the signal path and an appropriate biasing circuit so that the dc current through the diodes can be adjusted. The drive level, in the experimental unit, is adjustable from approximately 1 microwatt to 1 milliwatt.

The signal from the limiter passes through a variable phase shifter which, in reality, is a tunable bandpass filter. It contains three resonators and a 3-gang variable capacitor to tune them. Each variable capacitor has a small varicap across it which is used for electrical fine tuning. The purpose of the electrical fine tuning will be explained subsequently. A schematic diagram of the phase shifter is shown in figure 3. The inductors are band switched, and approximately an octave can be covered with one band.

The phase shifter, as one might expect, has a large pulling effect on the frequency of oscillation, and it is used to pull the frequency of the loop to the series resonance frequency of the crystal. The method for determining when series resonance is reached, of course, is the vector voltmeter which senses the phase angle of the voltage across the crystal.

Basically, the vector voltmeter has the capability to measure two rf voltages and the phase angle between them. It uses sampling techniques and covers the frequency range from 1 MHz to 1 GHz. Sensitivity is adequate to detect phase changes of about 0.05 degree.

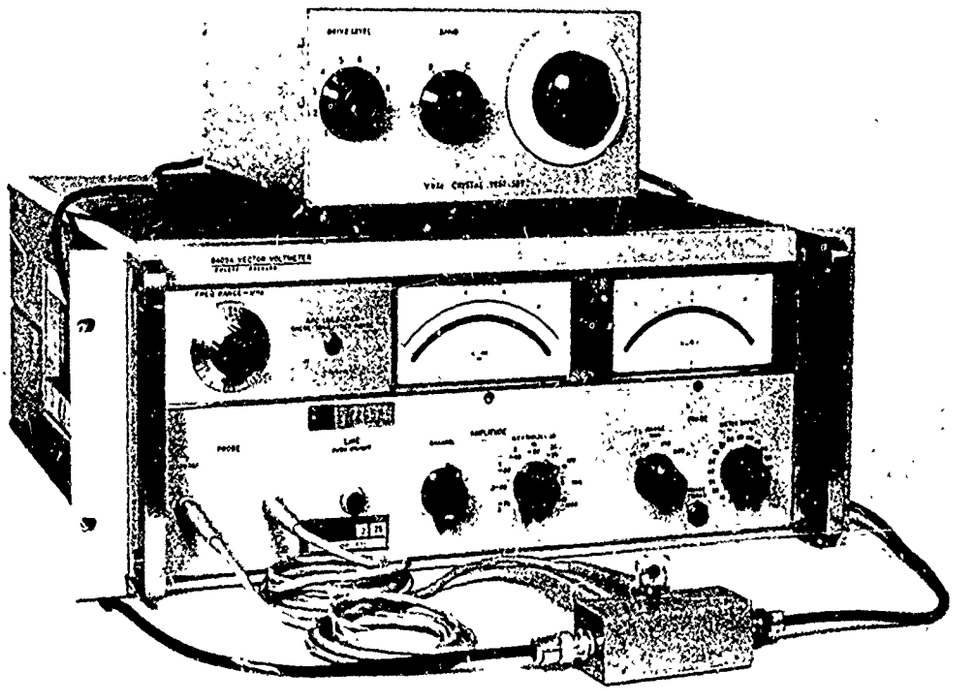


Figure 1. Experimental Unit.

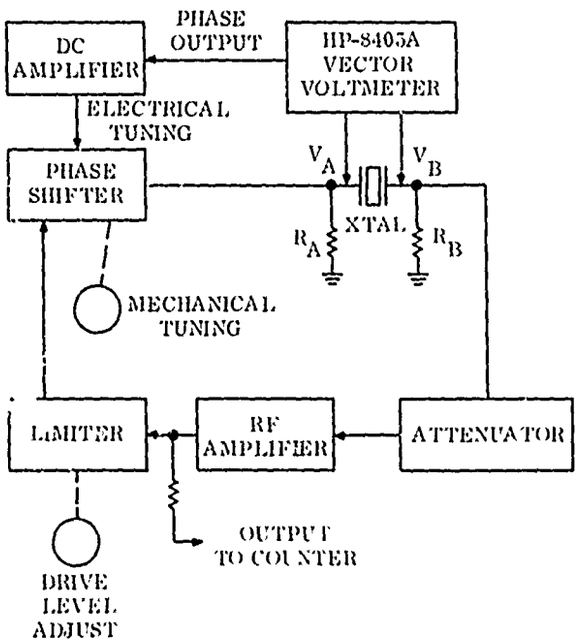


Figure 2. Block Diagram of Measurement System.

An attenuator is used between the crystal test fixture and the amplifier input to ensure that the test fixture sees a purely resistive load.

The phase reading of the vector voltmeter, in addition to being displayed on the front panel meter, is brought out to a BNC connector on the rear of the unit. This voltage is amplified in a dc amplifier and applied to the varicaps in the phase shifter. The voltage is applied to the varicaps with the proper polarity so that a positive phase angle (V_B leading V_A) will increase the varicap voltage and thus raise the frequency of oscillation; bringing it closer to series resonance. The dc feedback to the varicaps has several desirable effects. In the first place, it makes the mechanical adjustment easy to set. Secondly, it holds the frequency at series resonance over long periods of time even though the components of the loop drift somewhat, and thirdly, it makes it possible to plug in and measure crystals of the same nominal frequency without retuning. The dc amplifier has a voltage gain of 200 and uses an integrated operational amplifier.

The Hewlett-Packard vector voltmeter contains a single voltmeter which can be switched to measure the voltage to ground at either probe A or probe B. It is

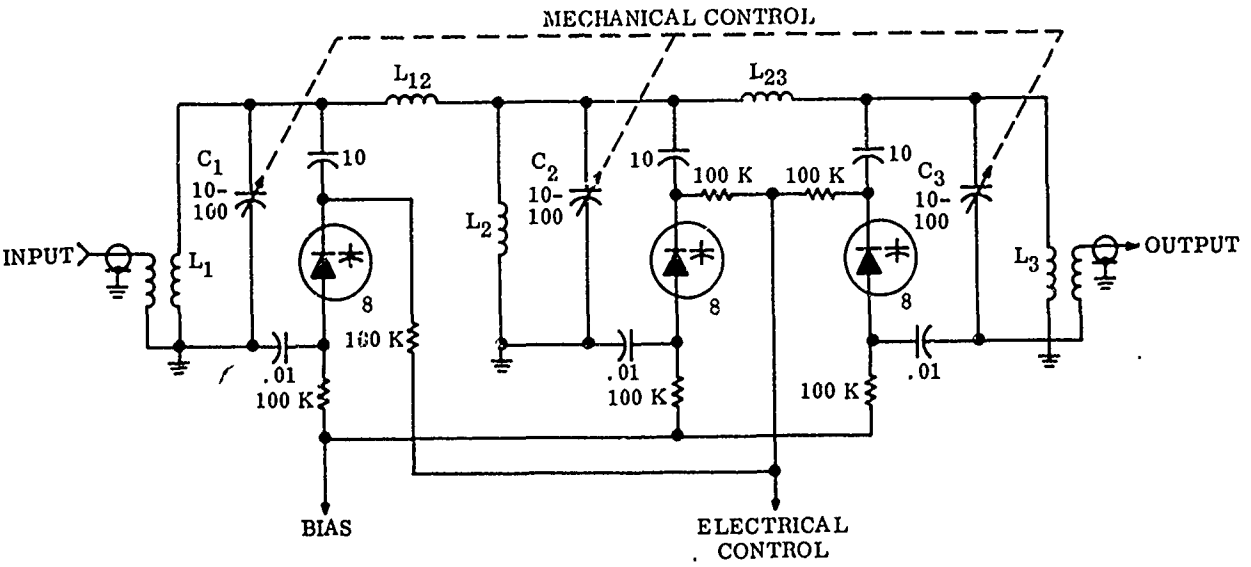


Figure 3. Schematic Diagram of Phase Shifter.

convenient to use these voltages to find the crystal resistance and drive level. This can be done graphically by plotting families of resistance and power curves. The voltage readings V_A and V_B determine the X and Y coordinates of a point which in turn determines a constant resistance line and a constant power curve. Equations for these curves can be derived from the schematic diagram of the test fixture shown in figure 4. Here it is assumed that the frequency has been adjusted to series resonance and the crystal is represented by an, as yet, unknown resistance R_c. It can be shown that the crystal resistance is given by

$$R_c = R_{eq}\left(\frac{V_A}{V_B} - 1\right) \qquad (1)$$

where R_{eq} is the equivalent resistance of the parallel series combination of R_B, R_1, R_2, and R_{in}. It is important to realize the R_{eq} is a known quantity which is fixed for the unit. For the experimental test set, R_{eq} = 28 ohms.

It can also be shown that the crystal power is given by

$$P_c = \frac{V_B}{R_{eq}}(V_A - V_B) \qquad (2)$$

Families of resistance curves can be plotted using equation 1 and families of power curves using equation 2. These curves are shown in figure 5.

To use these curves, one in general first measures V_A and V_B without regard to drive level. The crystal resistance is then read from the graph. The intersection of the desired drive level curve with the now known crystal resistance line determines a unique point corresponding to specific values of V_A and V_B. The drive level control is then adjusted until V_A or V_B reads the proper value. Only one of the voltages is observed since the other will necessarily follow.

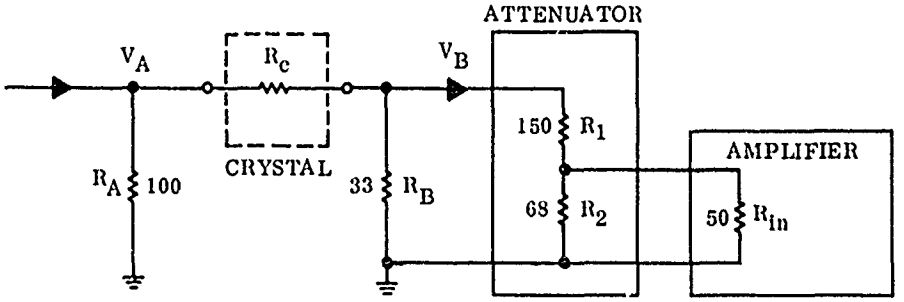


Figure 4. Schematic Diagram of Test Fixture.

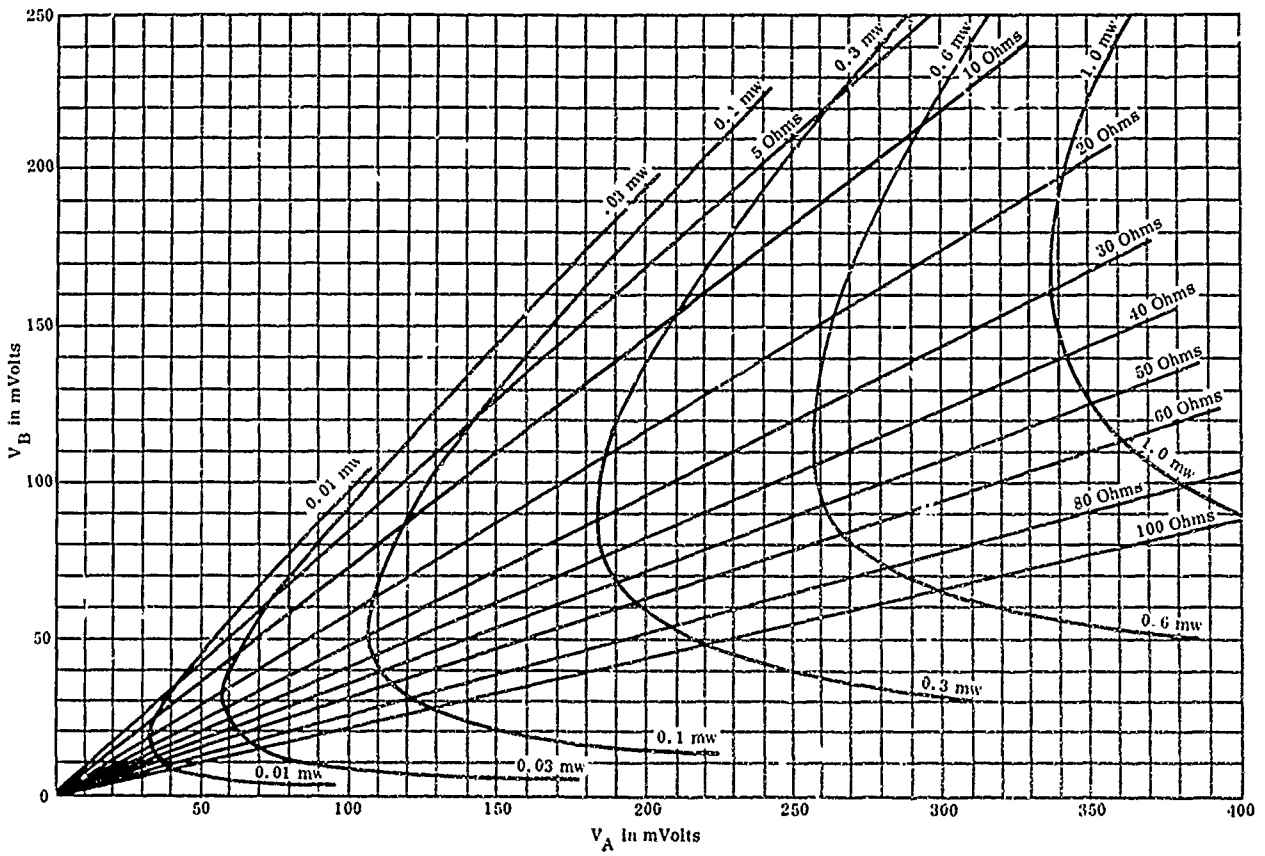


Figure 5. Resistance and Drive Level Graph.

Operating Procedures

There are several variations in the operating procedure which can be used to calibrate and operate the test set. One possible procedure is as follows:

Step 1: Place a short circuit in the test fixture in place of the crystal. With the phase shifter set to the correct frequency, as indicated on the dial, adjust the vector voltmeter to read zero in the ±6 degree range.

Step 2: Replace the short with the crystal, and adjust the phase shifter for a zero phase reading.

Step 3: Determine the crystal resistance either by reading V_A and V_B and going to the graph, or by resistance substitution until V_B reads the same for the crystal as for the resistor.

Step 4: From the graph, determine the required V_B for the desired drive level. Adjust the drive control until it is obtained. If the drive level is adjusted over a large part of its range, it may be necessary to retouch the phase shifter setting.

At this point, the frequency can be read to about a part in 10^7 for semiprecision crystals. Additional crystals of the same nominal frequency can also be tested simply by inserting them in the crystal socket and reading the frequency. The dc feedback will automatically adjust the phase shift to zero. To obtain additional accuracy, the following two steps should be added.

Step 5: Place a short in the test fixture and recheck the zero of the vector voltmeter. (For crystals greater than about 50 ohms, a resistor equal to the crystal resistance can be used to improve the accuracy, provided a pure resistance is used.)

Step 6: Remove the short and reinsert the crystal.* Read the frequency.

The test procedure is the same for 32 pF crystals except that a calibrated 32 pF capacitor is used in series with the crystal.

It is interesting to note, in step 4, that a special graph can be constructed giving V_B as a function of crystal resistance which will result in the actual drive level encounter in a specific oscillator.

---

*The vector voltmeter may lock in different modes, and depending on which mode is obtained, the zero may shift as much as about 0.4 degree. In order to prevent the occurrence of these zero shifts, the vector voltmeter bandswitch is set to the highest band for which a stable lock can be obtained. When this is done, the zero shifts are apparently eliminated.

Accuracy

The basic accuracy of the crystal test set, exclusive of crystal temperature effects, is about 3 to 5 parts in 10^8 for semiprecision crystals. The resettability error is in the order of 1 to 2 parts in 10^8. Crystal resistance can be read from the graph with an error generally less than 5 percent, or to about 1 ohm, whichever is greater. Resistance can also be determined to 1 or 2 percent by interpolating between readings obtained by substituting standard resistors. Drive level inaccuracy is in the order of 5 percent plus any error in the resistance determination.

There are several sources of frequency error in the test set. Among the more important are the following:

1. Stray inductance and capacitance in the load of the test fixture.
2. Stray capacitance and inductance in the substitution resistors, if they are used to zero the vector voltmeter.
3. Dependence of the vector voltmeter on source impedance.
4. Dependence of the vector voltmeter phase reading on signal amplitude.
5. Errors in reading the phase meter.
6. Errors in setting the crystal drive.

Each of these errors is discussed in more detail in the following material.

The first effect listed is that of stray reactance in the test fixture. This can be better understood by referring to the equivalent circuit shown in figure 6.

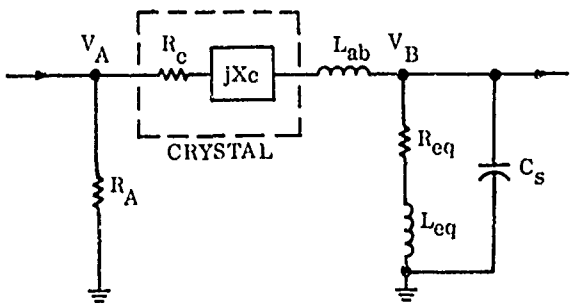


Figure 6. Equivalent Circuit of Test Fixture.

If it is assumed that the stray reactances are small and that the crystal is near series resonance, there is little interaction between the various reactances. The phase shift contributed by each effect can be examined independently. In order to get a feel for the frequency error caused by the various phase shifts, it is convenient first to determine the frequency error resulting from a given phase angle error. This can be done by examining the effect of the crystal. It can be shown that the phase angle change between V_A and V_B caused by the crystal is

$$\Delta\theta_c = -2Q\left(\frac{R_c}{R_c + R_{eq}}\right)\left(\frac{\Delta f}{f_s}\right) \qquad (3)$$

where

f_s = series resonant frequency

Q = normal crystal Q

and θ is in radians.

For a typical TCXO crystal at 5 MHz, the resistance might be 3.2 ohms and the Q = 725,000. For these values, a phase change of 0.1 degree results in a frequency change of 1.17 PP 10^8. As a rule of thumb, 0.1 degree is approximately 1 PP 10^8.

The effect of the stray capacitance C_s is given by

$$\theta_s = -T_{an}^{-1}\,\omega C_s\left(\frac{R_c R_{eq}}{R_c + R_{eq}}\right) \doteq -\omega C_s\left(\frac{R_c R_{eq}}{R_c + R_{eq}}\right) \qquad (4)$$

The effect of L_{eq} is given by

$$\theta_L = T_{an}^{-1}\,\frac{\omega L_{eq}}{R_{eq}} - T_{an}^{-1}\,\frac{\omega L_{eq}}{R_c + R_{eq}} \doteq$$

$$\omega L_{eq}\left[\frac{R_c}{R_{eq}(R_c + R_{eq})}\right] \qquad (5)$$

The effect of L_{ab} is given by

$$\theta_a = -T_{an}^{-1}\,\frac{\omega L_{ab}}{R_c + R_{eq}} \doteq \frac{-\omega L_{ab}}{R_c + R_{eq}} \qquad (6)$$

If the vector voltmeter is calibrated by placing a short in the test fixture, $R_c = X_c = 0$ and the total phase angle is given by

$$\alpha_1 = \theta_s + \theta_L + \theta_a = 0 + 0 - \frac{\omega L_{ab}}{R_{eq}} \qquad (7)$$

If the crystal is now placed in the test fixture, the phase angle indicated by the vector voltmeter is given by

$$\alpha_2 = \Delta\theta_c + \theta_s + \theta_L + \theta_a - \alpha_1 \qquad (8)$$

The phase shifter is adjusted until the phase reading, α_2, goes to zero, that is until the frequency is of the proper value to make $\Delta\theta_c$ cancel the other terms. Setting $\alpha_2 = 0$ and solving for $\Delta\theta_c$, after some manipulation, gives

$$\Delta\theta_c = \omega\left(\frac{R_c}{R_c + R_{eq}}\right)\left[C_s R_{eq} - \frac{L_{eq}}{R_{eq}} - \frac{L_{ab}}{R_{eq}}\right] \qquad (9)$$

It is interesting to note that the term in brackets is independent of frequency and crystal resistance (subject to the initial assumption that the reactances are small). By properly selecting C_s, it is theoretically

possible to make this term zero and thus eliminate the effect of the stray reactances. In practice the compensation is not perfect and errors in the order of 0.1 degree were experienced in the experimental unit. Equation 9 can be combined with equation 3 to give the following result:

$$\frac{\Delta f}{f_s} = \frac{\omega}{2Q}\left[\frac{L_{eq}}{R_{eq}} + \frac{L_{ab}}{R_{eq}} - C_s R_{eq}\right] \qquad (10)$$

which shows that the error caused by the stray reactances is proportional to frequency and inversely proportional to crystal Q as one might expect.

The third error listed results from the dependence of the vector voltmeter on the source impedance seen by the probes. If the vector voltmeter is "zeroed" using a short in place of the crystal, the samplers see a fairly low impedance. If the crystal tested then has a high resistance, the samplers see a somewhat higher impedance and the zero shifts. For this reason, the met :ero may be reset using a resistor equal to that of the crystal for large values of crystal resistance. The impedance increase over a short for a 50 ohm crystal is 6 to 8 ohms at the probes which may result in a phase error of about 0.2 degree.

A fourth error results from the dependence of the vector voltmeter phase reading on amplitude ratio. This effect amounts to about 0.2 degree for the voltage change caused by placing a 30 ohm resistor in the test fixture in place of a short. The error may be eliminated by setting the vector voltmeter to zero using a resistor equal to that of the crystal resistance. An additional small error results from changing the voltage on both probes.

A fifth error results from inaccuracy in reading the phase meter. With reasonable care, this can be held to ±0.05 degree (±5.9 PP 10^9 for the typical 5 MHz crystal). If an external meter is used on the output of the dc amplifier, the phase reading inaccuracy can be held to about ±0.01 degree.*

Finally, frequency errors result because of drive level inaccuracy. At the very low drive levels used in TCXO's (10 to 50 microwatts) this effect is negligible. At high drive levels, however, it becomes one of the major sources of error. The inaccuracy in drive level setting may be in the order of 10 percent. At a 1-milliwatt level, this could result in a frequency error of parts in 10^7.

Test Results

A number of tests were conducted to determine the accuracy of the test set. Room temperature tests were run on several 5 MHz crystals which were cut to have a flat frequency versus temperature slope at room temperature. Tests were also run on 3 MHz, 5 MHz, and 20 MHz crystals at oven temperatures. The crystals were measured both in the vector voltmeter test set and with a GR 1606-A rf bridge.

Table 1 gives the frequency deviation from the bridge readings for several of the crystals.

*For readings of this accuracy, the vector voltmeter should be allowed to warm up for several hours. It should also be operated from a constant voltage supply.

Table 1. Absolute Frequency Errors Compared to Bridge Test Set.

CRYSTAL NUMBER	NOMINAL FREQUENCY MHz	DRIVE LEVEL (MICROWATTS)	FREQ ERROR PP 10^8	RESISTANCE ERROR (OHMS)
1	3	35	4.0	0.9
2	3	6.0	4.7	0.4
3	5	7.5	-0.7	0.6
4	5	7.5	-3.0	0.1
5	5	7.5	-2.8	0.7
6	5	1.2	3.3	0.6
7	5	1.0	3.0	0.5
8	5	0.8	1.7	-0.1
9	15.8	12	1.9	0.2
10	20	6.0	5.5	0.7
11	20	7.0	-2.1	1.1

Table 2 contains information regarding the resettability of the vector voltmeter test set. Before each trial, the drive level control, the vector voltmeter zero setting, and the phase shifter position were disturbed. The crystals were also physically unplugged from the socket, and in some cases, crystals of a different frequency were tested between trials. The standard deviation of the frequency scatter is presented in the table for several of the test crystals. Data for the test crystals is given in table 3.

Table 2. Resettability Data.

CRYSTAL NUMBER	NUMBER OF READINGS	NOMINAL FREQ (MHz)	DRIVE LEVEL (MICROWATTS)	STANDARD DEVIATION OF READINGS IN PP 10^8
2	9	3	0.8	0.83
3	10	5	20	0.91
4	10	5	20	1.82
5	10	5	20	1.14
8	10	5	9.3	1.34
10	8	20	8.4	1.34

Table 3. Crystal Data.

CRYSTAL NUMBER	HOLDER	FREQ (MHz)	RESISTANCE (OHMS)	Q	MODE
1	HC-27/U	3	4.1	1,142,000	FUND
2	HC-27/U	3	4.0	1,155,000	FUND
3	HC-6/U	5	17.1	85,200	FUND
4	HC-6/U	5	16.8	89,900	FUND
5	HC-6/U	5	14.9	100,000	FUND
6	HC-27/U	5	3.2	732,000	FUND
7	HC-27/U	5	3.2	722,000	FUND
8	HC-27/U	5	3.2	724,000	FUND
9	HC-6/U	15.8	6.6	115,500	FUND
10	HC-27/U	20	5.0	346,500	3RD OT
11	HC-27/U	20	6.1	353,000	3RD OT

Temperature tests were also run on a number of TCXO crystals. It is quite convenient to use the test set for this purpose since the crystal is located in a test fixture at the end of transmission lines. A block diagram of the test setup is shown in figure 7. It should be noted that since the voltage reading of the vector voltmeter is brought out to a BNC connector, and since V_B is a measure of crystal activity, it can be used to record the crystal resistance during temperature runs. The voltage is calibrated by using the standard calibration resistors. Test results are shown in figure 8 for a 9.9 MHz third overtone crystal run at a 0.15 milliwatt drive level.

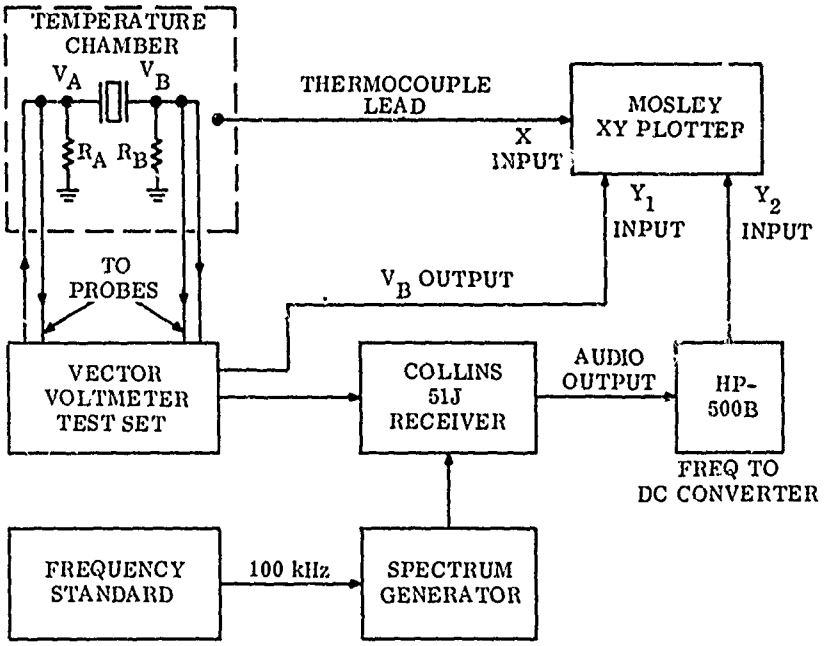


Figure 7. Block Diagram of Temperature Test Setup.

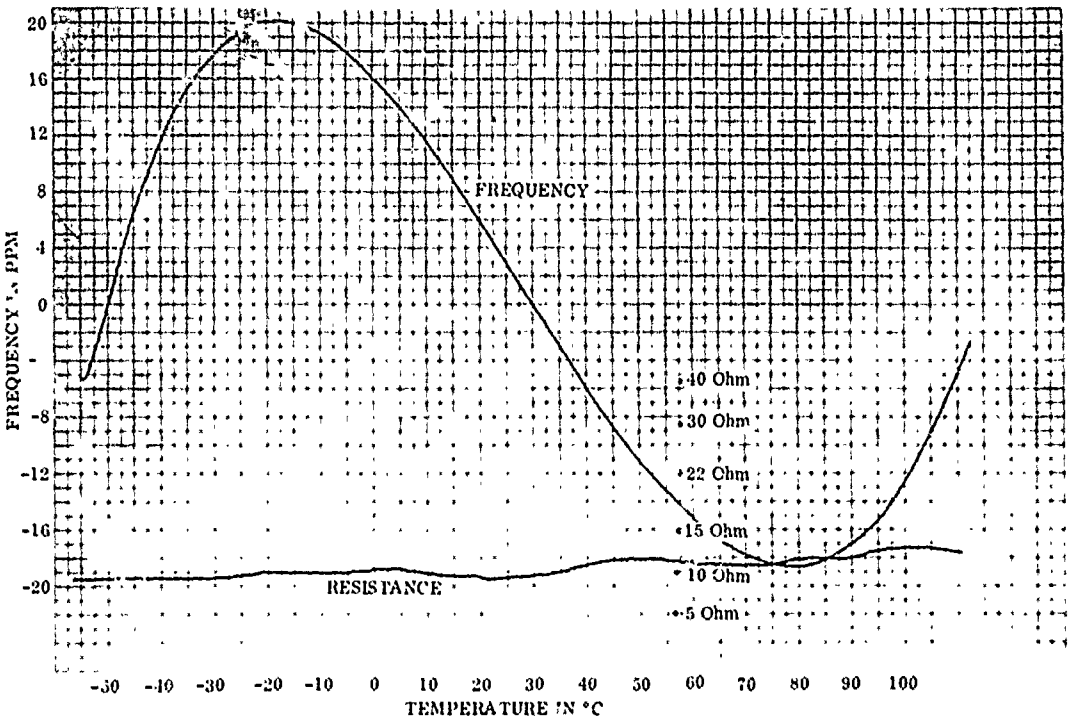


Figure 8 Frequency and Resistance Versus Temperature, 9.9 MHz Third Overtone.

Conclusions

The vector voltmeter crystal measurement technique has considerable potential for testing crystals of the "SSB quality." The experimental test set has demonstrated that the frequency accuracy is in the order of 5 parts in 10^8 while the resettability is in the range of 1 or 2 parts in 10^8. Crystal resistance and drive level can be conveniently read from a graph with reasonable accuracy. The set is well suited for making temperature runs on TCXO crystals. The experimental set covers the frequency range from 2.5 to 20 MHz; however, the technique itself is not limited to this frequency range.

Bibliography

1. "8405A Vector Voltmeter Operating and Service Manual," Hewlett-Packard Company, Palo Alto, California.

2. Grenier, R. P., "Technique for Crystal Resonance Measurements Based on Phase Detection in a Transmission Type Measurement System, Proceedings of the 22nd Annual Symposium on Frequency Control.

3. Adams, C., Kusters, J., and Benjaminson, A.; "Measurement Techniques for Quartz Crystals," Frequency Magazine, August 1968.

ON PRECISION MEASUREMENTS OF FREQUENCY AND RESISTANCE OF QUARTZ CRYSTAL UNITS

C. Franx
Crystal Department
N.V.Philips'Gloeilampenfabrieken
Eindhoven, Netherlands

Summary

The object of the investigation is to check the usefulness of the definition of resonant frequency of a quartz crystal unit laid down in IEC document 49(Secretaria')44 by measuring a set of crystals in three different ways; two bridge methods and the phase sensitive π network method. The results are compared.

Moreover a description is given of the π network set up designed for use in factories and laboratories. It will be shown that frequencies of unwanted modes and the resonant frequencies of monocrystal elements can be measured as well.

Discussion

The starting point is the formulation of an internationally agreed definition of resonant frequency. Why do we need such a definition ?

Even after a short survey of currently used specifications it appears that the most important parameter of a crystal unit, its frequency is defined as the working frequency of a specified oscillator, the test set. The specification of the test set, however, does not guarantee that sufficient correlation exists between measurements on different test sets.

A solution to the problem is to agree on such a definition of frequency that allows more than one recommended method of measurement, provided that sufficient correlation exists.

In this paper the resonant frequency is defined as the lower frequency at which the phase of the crystal impedance, measured in a reference plane 2 mm from the crystal base, is zero.

The reference plane has been introduced to keep the inductance and capacitance of the crystal terminals constant for every method of measurement.

The choice of the resonant frequency (zero phase) gives the advantage that no measurement of the elements of the equivalent circuit is necessary. The crystal unit is considered as a two terminal black box for glass enclosed units and a three terminal box for the metal enclosed units.

In cases where additional information of crystal inductance or static capacitance is required these quantities can be measured separately, but this information is not essential for the frequency measurement. This line of thought introduces a maximum frequency limit of about 100 MHz, above which the terminal inductances have an appreciable effect on the crystal impedance.

Equipment

Use has been made of 3 different set ups.

a. RX meter. Shown on photo 1 is the admittance bridge coupled to a synthesizer and a times ten frequency multiplier. To do this the internal oscillator of the bridge has been disconnected. On the bridge a coaxial adaptor with a precision GR 900 connector has been installed. On this fits the GR type 900 M component mount with the crystal unit. (Photo 2).

A plastic tag of 2 mm thickness fixes the reference plane at the specified distance from the crystal base.

The procedure starts with the calibration of the crystal drive level as a function of crystal resistance and bridge input voltage. This has been done with a selective

voltmeter connected to the GR 900 connector. In our case measurements at 50 and 25 and 16.7 Ohms are sufficient.

With the type 900-W50 standard coaxial termination the bridge is adjusted at zero capacitance and 1/50 Mho conductance. The electrical length of the coaxial system between the bridge and the reference plane of the component mount is determined with the type 900-W04 open circuit termination and with type 900-W100 100 Ohm termination. Both methods give 11.0 cm for any frequency below 100 MHz.

After adjusting the output level of the generator the crystal is inserted in the component mount on the bridge. The bridge is nulled at zero capacitance reading in the neighbourhood of f_r by adjusting frequency and conductance dials.

By means of a Smith chart the admittance at the reference plane is determined. The frequency is now changed by say 10 Hz and the admittance measured, the zero phase frequency is determined by interpolation.

b. GR type 1609 bridge (for frequencies above 40 MHz only). Although a similar procedure as with the RX meter could be followed here the electrical length between bridge and reference plane was adjusted exactly to a complete number of half wave lengths with a trombone (photo 3) using again type 900-W04 open circuit termination.

In order to obtain a sufficiently high level of drive the generator was connected to detector bridge terminal and visa versa. However, the generator voltage had to be lowered slightly to obtain an acceptable drive level, resulting in a very low detector voltage, just enough for the Rohde und Schwarz selective voltmeter type ESU.

c. Phase sensitive π network. The set up consists of 4 parts(Photo 4).

1. π network
2. Generator
3. Phase meter + accessories
4. Control network

The π network has 3 parts; 2 asymmetrical attenuators and the contact plates for the crystal terminals (Photo 5). The function of the attenuator is twofold:

Attenuation of unwanted reflections from source and load. If the phase error of the impedance seen by the crystal unit must be less than 0.2^0 it can be shown that the attenuation of the resistive networks must be 14.8 dB at least for reflection coefficients of source and load of 5%.

The second function is to transform the 50 Ohm line impedance to match the crystal impedance (Fig. 6).

The crystal sees a source with resistance R and a load R. The energy dissipated by the crystal at resonance frequency is

$$W_c = \frac{R}{R_{line}} \cdot \frac{R_r}{(2R + R_r)^2} \cdot V^2 \cdot 10^{-0.1A_t}$$

A_t = transducer loss of one π of the network.

Maximum dissipation occurs for $R_r = 2\,R$. In this case

$$W_{c\,max} = \frac{R}{R_{line}} \cdot \frac{V^2}{8R} \cdot 10^{-0.1A_t}$$

The maximum available power is

$$W_s = \frac{V^2}{4R_{line}}$$

hence

$$W_{c\,max} = \tfrac{1}{2}\, W_s \cdot 10^{-0.1A_t}$$

It appears that the level of drive is not very sensitive to changes of crystal resistance. If the crystal resistance changes from 2R to 2.8R or to 1.4R the drive level will be not more than 3% below nominal level of drive. If this is permitted the following series of π networks can be applied:

2R	R_r
12.5	9 - 18 Ohms
25	18 - 35 Ohms
50	35 - 70 Ohms
100	70 - 140 Ohms
200	140 - 280 Ohms

With these networks practically all types of AT cuts can be measured with simple control of drive level.

For the realised networks A_t = 14.8 dB, so $W_{cmax} = 1/60\ W_s$.

For the generator used W_s = 0.18 W so with a very efficient 3 way splitter $W_{c\,max}$ = 3 mW.

For higher power levels amplifiers have to be used.

The length of the connection between network and power splitter must be carefully chosen as the π network will give reflections for any crystal impedance $\neq 0$. If the connection with the power splitter is too long compared with $\frac{1}{4}$ wave a phase error will result in the A channel of the phase meter and, as a consequence, in the measurement. If long cables are required e.g. for measurements in ovens, then a 10 dB attenuator has to be added between π network and power splitter. This will lower the drive level.

The cables must also be carefully chosen and a double shielding is imperative in order to avoid parasitic couplings and hand effect. Cable RG-223/U gives satisfactory results.

The contact plates have a rim (Photo 7) that makes contact with the crystal terminals at the reference plane. The plates have been designed in such a way that the parasitic capacitances compensate the series inductance of the lead and plate for frequencies below 100 MHz. This can be checked by measuring the phase with a number of thin film ("zero phase") resistors of different resistances inserted in the π network. Graph 8 shows an undercompensated network and a compensated one. The compensation is performed by changing the thickness of the dielectric between the contact plates and the mass.

The frequency error resulting from a phase error is

$$\frac{\Delta f}{f} = \frac{1}{2} \frac{\Delta \varphi}{Q_{eff}} \; ; \; Q_{eff} = \frac{1}{\omega_r C_1 R_{eff}} \; ; R_{eff} = R_r + 2R.$$

Example: f_r = 61 MHz

C_1 = 0.0015 pF

R_r = 15 Ohm

R = 12.5 Ohm so Q_{eff} 30,000

$\Delta \varphi$ = phase error

The result is that

$$\frac{\Delta f}{f} = 33 \times 10^{-6} \times \Delta \varphi.$$

So that for a requirement of $\frac{\Delta f}{f} < 0.1 \times 10^{-6}$

$\Delta \varphi$ must be less than 1/330 radian = 0.2°. This can be achieved with a carefully designed π network.

d. Control network. Serves as a phase invertor, the low pass section suppressing oscillations in the servo loop (Fig.9).

A means is provided to make the $\frac{\Delta f}{f}$ versus $V_{control}$ curve of the generator practivally independent of frequency by using the frequency analeg output.

Closing switch S_2 for a short time lowers the frequency to that of the lower end of the controlled range (0.1% of f_s). Because of the capacitive character of the crystal units the phase meter indicates a positive phase angle. This results in a slow increase of frequency (search action) until the resonance frequency is found where the phase locks the frequency.

Pushing S_1 interrups the control loop and gives a small negative charge on the capacitor bank. This unlocks the frequency which will now slowly increase until a second mode is found. This is demonstrated in photo 4 where a dual resonator is inserted and the symmetrical mode measured. After pushing S_1 the antisymmetrical mode is excited (photo 10). The method is also useful for the search and measurement of unwanted modes in oscillator crystals.

Measuring results

Three groups of all glass crystals (HC-27/U) have been used;

87 MHz fifth overtone
61 MHz third overtone
10 MHz third overtone

All crystals have low temperature coefficients at room temperature and have been selected to avoid small activity dips and frequency jumps.

The measuring procedure starts with the π network, transfer the crystal to the GR bridge → RX meter → GR bridge → π network in order to compensate for slow temperature changes. The 10 MHz crystals have not been measured in the GR bridge because of the low frequency.

Use has been made of the π network with a 2R value of 25 Ohms, serial number 2. The measured frequency differences with respect to the π network are

Frequency	25 Ohm nr 2 π	Relative frequency differences	
		GR	RX
10 MHz	reference	-	$-0.3\pm0.7\times10^{-7}$
61 MHz	reference	$-2\ \pm1.5$	$0.5\pm2\ \times10^{-7}$
87 MHz	reference	$+0.7\pm1.6$	$+1.1\pm2\ \times10^{-7}$

Apparently no significant systematic differences have been found.

Then 3 different π networks have been compared.

Frequency	25 Ohm nr 2	Relative frequency differences	
		25 Ohm nr 1	50 Ohm nr 1
10 MHz	reference	-0.25 ± 0.5	$-0.18\pm0.4\times10^{-7}$
61 MHz	reference	-1.25 ± 1.4	$-1.6\ \pm2\ \times10^{-7}$
87 MHz	reference	$-0.3\ \pm0.3$	$-1.1\ \pm0.7\times10^{-7}$

Again no significant differences. The differences in spread are mainly produced by the different values of effective Q (290,000 21,000 resp. 107,000).

Note: The mention of the type of equipment used serves to demonstrate the flexibility of choice of measuring set up and is not intended as an endorsement of any particular product.

Conclusion

Based on the new definition of crystal impedance the differences in resonant frequency measured in 3 different π networks and in 2 different bridges appear to be smaller than 4×10^{-7}.

No significant systematic differences have been found, provided that the π networks have been designed carefully.

The resetability of the networks appears to be $\pm0.7 \times10^{-7}$ for the worst case (61 MHz).

The advantages of this set up in comparison with the bridge methods are:

a. Ease of measurement, especially with regard to adjustment of level of drive. No corrections are needed for the length of the lead between crystal and network and as a consequence the measuring time is a fraction of that needed when using the bridge methods.

b. Automatic search of resonant frequency of main mode and of other modes is possible.

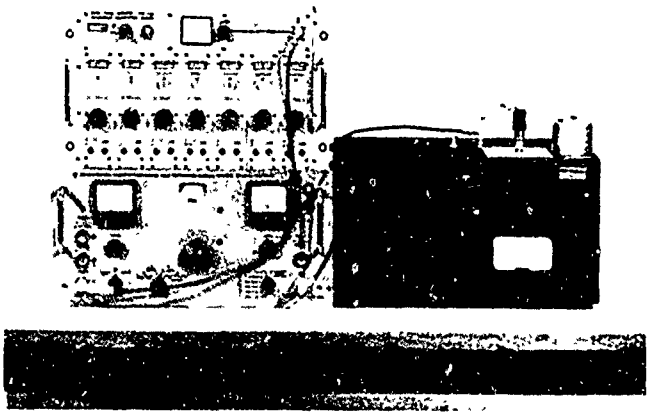


PHOTO 1

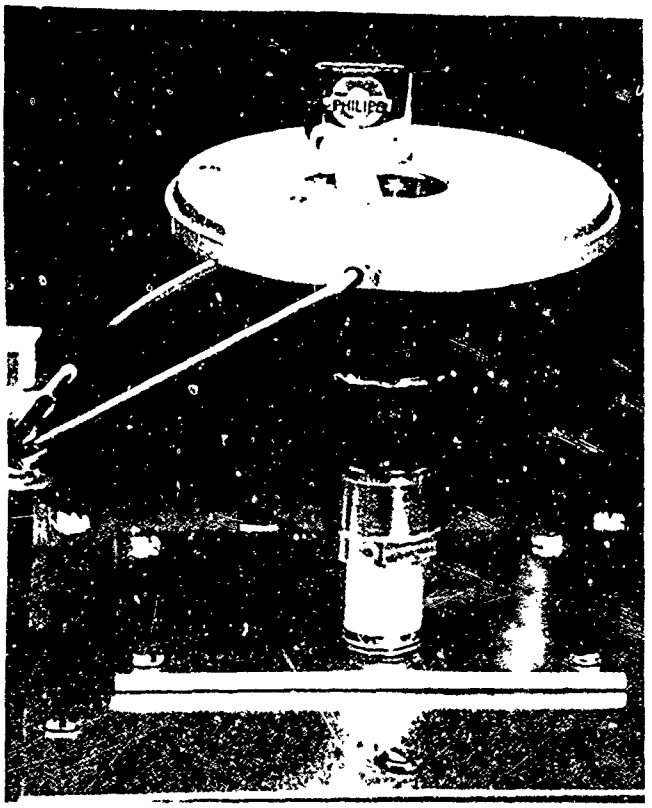


PHOTO 2

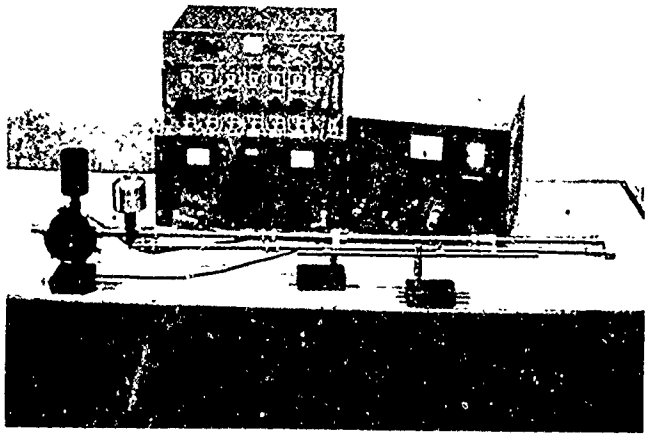


PHOTO 3

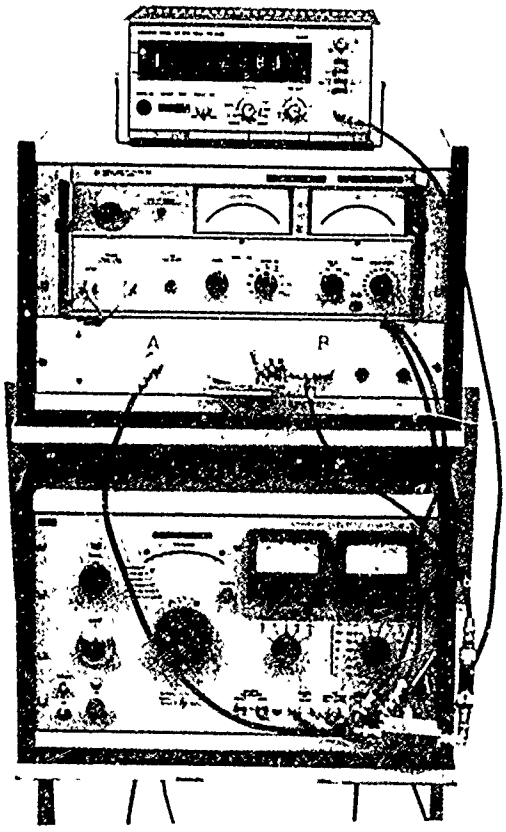


PHOTO 4



Reproduced from best available copy.

PHOTO 5

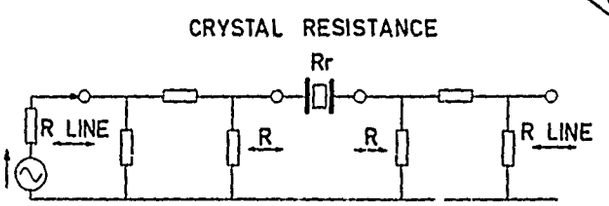


FIG. 6

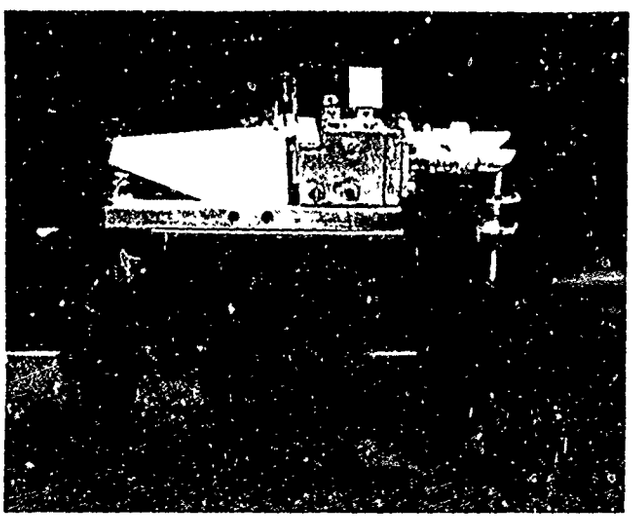
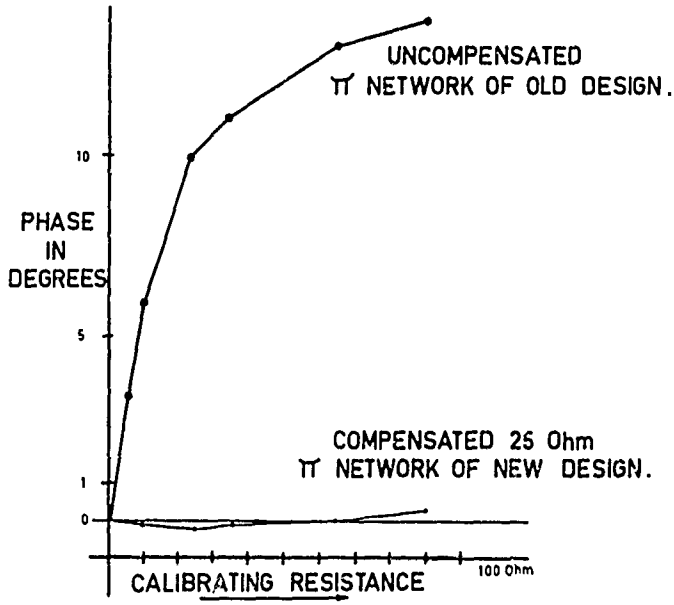


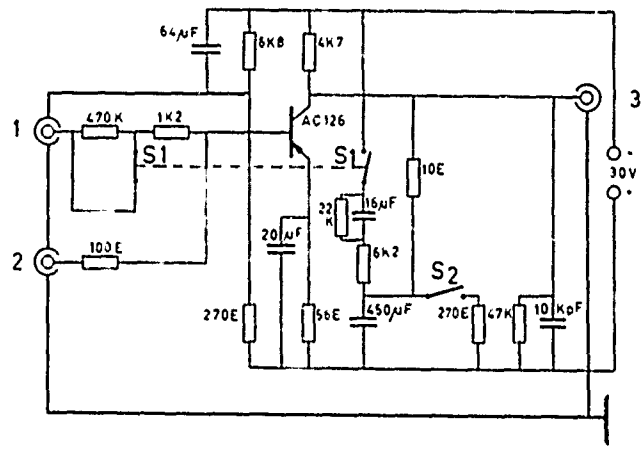
PHOTO 7

87MHz



INFLUENCE OF DESIGN OF π NETWORK ON PHASE ERRORS.

FIG. 8



TERMINAL 1 TO OUTPUT PHASE METER
TERMINAL 2 TO CONNECTOR 13 OF SIGN. GEN.
TERMINAL 3 TO CONNECTOR 12 OF SIGN. GEN.

FIG. 9

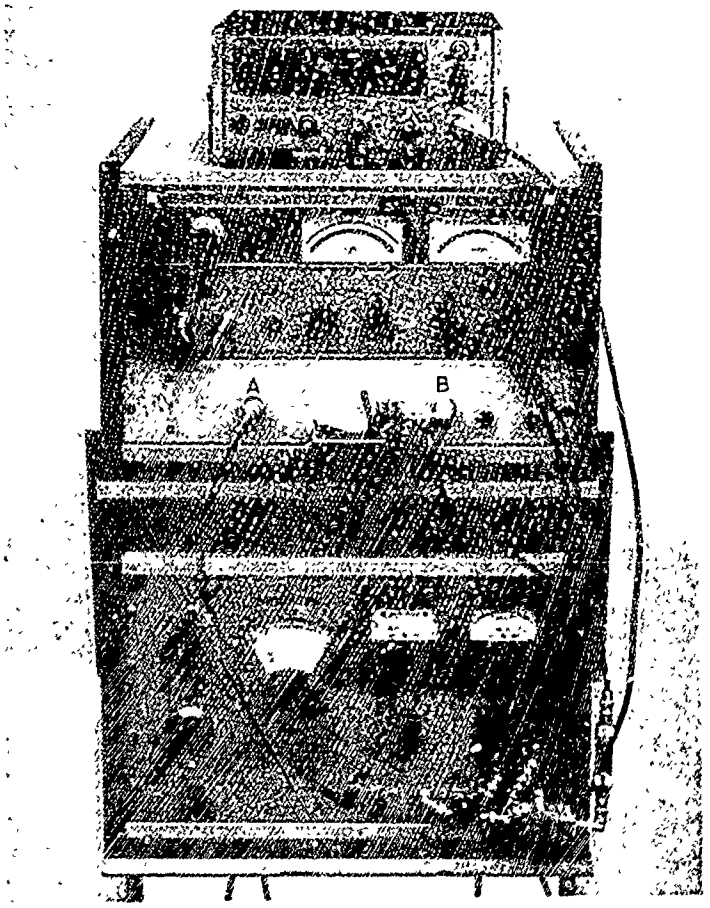


PHOTO 10

ANOTHER LOOK AT SPECIFYING A CRYSTAL

Donald W. Nelson
RCA-Camden, N.J.

Summary

Frequently, the failure to communicate problem areas by way of the drawing leads to improper price bidding and inferior products. Although there is a great tendency among engineers to pass the blame to the purchasing agent or a crystal manufacturer, the engineer is nonetheless responsible for the first line of communication - the crystal specification. Three levels of crystal specification are described with suggestions which should be helpful in eliminating misguided bids and hard feelings between manufacturer and user.

Discussion

When can you use a MIL-SPEC? How costly is Hi Rel Testing? Should you gear the specification to a specific vendor? Does it cost more to overspecify than to underspecify? Is it possible to completely controla a product by a drawing?

Perhaps you have experienced the situation where many hours were spent with a crystal manufacturer to develop a satisfactory part - only to have purchasing award the second order to a lower bidder who is completely unfamiliar with the true requirements of the circuit. While our first instinct is to blame the purchasing agent, the true responsibility belongs to the man who wrote the specification. A great deal can be placed on your drawing to better describe the crystal and protect you and the second-source bidder from certain failure.

What determines the degree of testing to be performed on a crystal? Unfortunately, the answer is usually cost. Our profit-motivated society demands a justification for additional testing. We should not, therefore, make expensive requirements unless they are necessary to the circuit; e.g., there is no reason to make requirements of Q when equivalent resistance is sufficiently meaningful. If, on the other hand, requirements such as a phase-locked loop demands a control on C_1 or Q, it would be foolish to omit control of the parameter.

Possibly, the simplest specification we shall find is in fig. 1. This is conceivably a $1 unit in quantity with a minimum of testing. Such a drawing could be expanded in accordance with requirements. Not only should the engineer change his drawing in accordance with new requirements, but the manufacturer should protect himself by informing the customer if the specification can be tightened without increased cost in the product. It might be found, as expected, that this is a poor specification, that the true requirement is for .005% from 32°F to 100°F, and that by virtue of your good design and high quality in manufacturing, your crystal fits the circuit perfectly. By all means, encourage the customer to update the drawing! He may find a less virtuous vendor on the next run.

COMMERCIAL LEVEL SPECIFICATION

H.A.M. CO.
ASHLAND, N. J.

SPECIFICATION FOR QUARTZ CRYSTALS

1. FREQUENCY - 7142KHz
2. TOLERANCE - ± .01%
3. TEMPERATURE - 50°F TO 85°F
4. LOAD CAPACITANCE = 32PF
5. HOLDER - HC-6
6. VENDOR - P.A. DELIGHT CO.
   PUNXATAUNY, PA.

FIGURE 1

It should be obvious that a crystal shipped to the above drawing has no promise of reliability other than the reputation of its manufacturer. Despite this conclusion, inadequate specifications such as figure 1 have been the greatest cause of added expense. Once a bad part has been received, there is a good chance you will pay the manufacturer. In addition, you will probably pay several times the crystal cost while conducting an investigation.

The Military crystal specification system, based on MIL-C-3098, provides a uniform series of specifications with known environmental qualifications. By virtue of the established system of testing, a MIL crystal has a knwon environmental capability and some reliability. Unfortunately, the demise of the MIL crystal lies in a manufacturer's interpretation of MIL-C-3098. Since there are many places where the MIL-SPEC is too loose, the writer suggests supplementing existing drawings with testing appropriate to your problem areas. Figure 2 displays an augmented military specification.

Some problem areas dealt with in the drawing are expanded as follows:

1. An activity dip at some temperature may cause your system to fail. Paragraph 3.1.3 not only limits the maximum equivalent series resistance, but limits the change in R over the temperature range to 20%. Furthermore, the quality assurance provisions require a temperature run chart be submitted with each crystal. The user is now assured of the quality level received.

2. It is not uncommon for a crystal manufacturer to increase the electrode size to meet the requirements of the preceding paragraph. The CR-55A/U specification allows 7pf shunt capacitance which is well above the typical value of 4.3pf. Should the electrode size be increased, the circuit will probably operate at a different frequency than when using a 4.3pf crystal.

Paragraph 3.1.4 covers this requirement.

3. Paragraph 3.1.5 shows the operating drive level which is considerably less than the normal 2mW test level. Here, again, is a poorly defined area in the standard part which can cause problems on the production line.

4. A fourth change to the CR55A/U specification which might be desirable is the addition of a Seal Test on all units. The requirement in our specification is found in para. 4.1.2.

AUGMENTED MIL-C-3098 SPECIFICATION

RCA Corp.
Camden, N.J.

Specification for Quartz Crystals

1.0 SCOPE:

1.1 THIS SPECIFICATION INCLUDES THE GENERAL AND DETAILED REQUIREMENTS AND TESTS OF RELIABLE QUARTZ CRYSTAL UNITS DESIGNED FOR PRECISE FREQUENCY CONTROL.

2.0 APPLICABLE DOCUMENTS:

2.1 THE FOLLOWING SPECIFICATION, OF THE ISSUE IN EFFECT ON DATE OF INVITATION FOR BIDS, FORM A PART OF THIS STANDARD TO THE EXTENT SPECIFIED HEREIN.

SPECIFICATIONS
MILITARY

MIL-C-3098D	CRYSTAL UNIT, GENERAL SPECIFICATION FOR
MIL-C-3098/33D	CRYSTAL UNIT, QUARTZ, CR-55A/U
MIL-H-10056/9	HOLDER, CRYSTAL, TYPE HC-18/U

STANDARDS
MILITARY

MIL-STD-202	TEST METHODS FOR ELECTRONIC AND ELECTRICAL COMPONENT PARTS.

3.0 REQUIREMENTS:

3.1 ELECTRICAL

3.1.1 THE CRYSTAL UNIT SHALL MEET ALL ELECTRICAL REQUIREMENTS OF MILITARY STYLE CR55A/U PER MIL-C-3098 AS MODIFIED HEREIN.

3.1.2 THE FREQUENCY SHALL BE 48.50000 MEGAHERTZ $\pm$ .005% OVER THE OPERATING TEMPERATURE RANGE PER PARAGRAPH 3.3.1.

3.1.3 THE MAXIMUM EQUIVALENT RESISTANCE SHALL BE 30 OHMS OVER THE OPERATING TEMPERATURE RANGE. THE CRYSTALS EQUIVALENT RESISTANCE SHALL NOT CHANGE MORE THAN $\pm$ 20% FROM THE VALUE MEASURED AT 25°C.

3.1.4 THE SHUNT CAPACITANCE SHALL BE 4.3 $\pm$ 0.5PF.

3.1.5 THE AVERAGE OPERATING DRIVE LEVEL IS 0.1MW.

3.2 MECHANICAL - CRYSTAL SHALL BE IDENTICAL TO CR-55A/U.

3.3 ENVIRONMENTAL

3.3.1 TEMPERATURE -
OPERATING -10°C TO 70°C
STORAGE -55°C TO 105°C

4.0 QUALITY ASSURANCE REQUIREMENTS

4.1 IN ADDITION TO QUALIFICATION TESTS PERFORMED IN ACCORDANCE WITH MIL-C-3098, THE FOLLOWING TESTS SHALL BE PERFORMED ON ALL CRYSTALS BEFORE THEY ARE SUBMITTED TO THE BUYER. DATA FROM THESE TESTS SHALL BE SUBMITTED WITH THE CRYSTALS.

4.1.1 TEMPERATURE RUN, OPERATING RANGE, IN ACCORDANCE WITH MIL-C-3098, PARA. 4.7.4.3.2.

4.1.2 SEAL - ALL UNITS SHALL BE TESTED IN ACCORDANCE WITH MIL-STD-202, METHOD 112, CONDITION C.

5.0 PREPARATION FOR DELIVERY

5.1 PARTS SHALL BE PACKAGED AND PREPARED FOR DELIVERY IN ACCORDANCE WITH MIL-C-3098, PARAGRAPH 5.1.2, LEVEL C.

6.0 SUGGESTED SOURCE:
Q.P.L. CORP,
INDEPENDENCE, MO.

FIGURE 2

Augmented areas of specification are meant to show typical problems which might exist. By studying your own needs, the most economical drawing may be evolved. A worst-case analysis of the circuit is most helpful and frequently inexpensive in that it tends to find problem areas before parts are purchased. The "ounce of prevention" philosophy of an augmented MIL-SPEC may well turn into a "penny saved". If a CR 55A/U costs $3.50, the additional testing could double the price. In several cases known to the author, such an expenditure would have been a bargain.

The third level of crystal specification exists in a world apart from the confines of a $1 - or $10 - price tag. High Reliability Crystals are needed where parts are inaccessible, or special environments create peculiar demands. There are seldom production runs on such parts and the testing required is extensive.

One of the important additions to Hi Rel parts is documentation; another addition is the screening tests on every part shipped. Environmental requirements are usually more stringent than those denoted by MIL-C-3098. Test procedures are detailed because of wide variations from MIL-STD-202. The resulting drawing may appear unusually large; however, it does control a component more completely and flexibly. Figure 3 outlines these requirements.

ESSENTIALS OF A HIGH RELIABILITY SPECIFICATION.

1. Mechanical & Electrical Requirements
2. Serialization & Traceability
3. Environmental Requirements
4. Quality Assurance Provisions
   - A. Qualification Tests
   - B. Group C (Shock, Vibration, Life)
   - C. Screening (100% of Shipment) data shipped with crystals
5. Detailed Test Procedures
6. Preparation for Delivery
7. Notes
   - e.g. Notification of Process Change, Failure Rate Goal

Figure 3

Documentation of a High Rel part includes serialization and traceability of the part through all its processes. All test information is filed for a number of years (usually 5) by the manufacturer. The user receives a copy of acceptance test data with each crystal, and must be notified in writing of any change in process.

Typical Screening tests invoked on all crystals include Insulation Resistance, Temperature cycling and Seal testing. In addition, a temperature run and aging are performed as part of quality assurance.

One Hi Rel Crystal Specification currently in use is eight (8) pages long. It is shown here (Fig. 4) for those who care to study the techniques more assiduously. For the small quantities used, the cost of this crystal is greater than $150. One third of this cost represents production; two-third testing and documentation. A sobering note is that we still have some problems with Hi-Rel crystals; however, the problems are much more sophisticated.

SUMMARY:

1. It is doubtful that a crystal can be completely controlled by any specification; therefore -
2. The engineer must be alert to problem areas and modify the specification when necessary.
3. The manufacturer can protect himself and the user by advising the user of desirable changes in the drawing.
4. The least expensive crystal is the one tailored to the problems of the user. It might cost $1 - or $100-. By showing the manufacturer a true picture, there are fewer chances of incorrect bidding. It is not necessary, or even desirable, to tie the specification to one vendor. He might have problems and not be able to deliver.

It is not the intent of this paper to motivate a user toward a higher level of specification. Rather, the user should be aware of techniques and practices of delineation. Always remember, a specification is a living document. Take another look at it - and another, and another.

RCA CORP.
SPECIFICATION CRYSTAL UNIT

1. Scope:

1.. This individual part specification covers the general and detailed requirements high reliability quartz crystal for use in space electronic equipment.

1.2 This specification, together with documents referenced herein, forms a complete procurement specification.

2. Applicable documents:

2.1 The following specifications and standards, of the issue in effect on the date of invitation for bid, form a part of this specification to the extent specified herein.

SPECIFICATIONS:

Military

MIL-C-3098 Crystal Units, Quartz

STANDARDS

MIL-STD-130 Identification Marking U.S. Military Equipment
MIL-STD-105 Sampling Procedures and Tables for Inspection by Attributes
MIL-STD-202 Test Methods for Electronic and Electrical Component Parts

OTHER PUBLICATIONS

National Aeronautics and Space Administration (NASA)

NPC 200-3 Inspection System Provisions for Suppliers of Space Materials, Components and Services
EIA RS-192-A Holder Outlines and Pin Connections for Quartz Crystal Units
Radio Corporation of America - 8533293 Reporting Requirements, Apollo High Reliability Parts

3. Requirements

3.1 Mechanical Requirements

3.1.1 Design and Construction - Fifth Harmonic Crystal in a Hermetically Sealed Holder.

3.1.2 Material - Quartz element, metal supports and leads encased in TO-39 holder as outlined in RS-192-A, and Figure 1.

3.1.3 Metals - metals shall be of corrosion-resistant type or shall be plated or treated to resist corrosion.

3.1.4 Fungus resistance material - external parts shall be inherently non-nutrient to fungus.

3.1.5 Leads - leads shall be solderable. Solderability requirements are specified in paragraph 3.3.1.6

3.1.6 Marking - units shall be marked in accordance with Standard MIL-STD-130 with the RCA part number. The frequency, the manufacturer's name or symbol, the manufacturer's part number and serial number. The frequency can be abbreviated to nearest kilocycle.

3.1.7 Serialization - each serial number may be alpha, numerical, or alphanumeric, and may have a number of digits up to and including 6 digits in length. The serial number must be physically on the part unless the size of the part prohibits marking. If the part itself cannot be marked, then a tag shall be attached on the part giving the serial number and all other required markings not physically on the part. Each part delivered must be individually marked or tagged. Serial numbers within a given lot need not necessarily be consecutive, but in succeeding lots, numbers shall be of higher sequence than those used in preceding lots.

3.1.8 Traceability - the vender shall maintain adequate records to identify all of the types of materials and/or types of parts and types of processes in the production of each serial numbered end item. Vendor shall upon request, within 48 hours, provide RCA with the list of materials, parts and processes for any end item, or define all end items produced from particular materials, parts or processes. All traceability information shall be retained on file by the vendor for a period fi five (5) yea. ter the parts are delivered to RCA. The information shall be forwarded to RCA upon request.

3.1.9 Workmanship - units shall be manufactured and processed in a careful and workmanlike manner in accordance with good design and sound practice.

3.1.10 Visual and Mechanical Inspection - the units shall be inspected in accordance with Paragraph 4.6.1.

3.2 Electrical Requirements----

3.2.1 Absolute Maximum Ratings - above maximum ratings are those which, if exceeded, may damage the serviceability or life of the unit.

3.2.1.1 Rated Drive Level - 0.5 mw, max.

3.2.1.2 Shunt Capacitance - 4 pf ± 1 pf

3.2.1.3 Ambient Temperature Range - 10°C to + 75°C operating
-55°C to +105°C storage

3.3.2 Characteristics:

Part No.	Frequency (MHZ)	Tolerance %		Equivalent Resistance (Ohms)
		25°C	-10°C to +75°C	
-1	86.56666	±.0005	±.002	60 max
-2	93.93333	±.0005	±.002	60 max
-3	88.93333	±.0005	±.0015	55 max.
-4	114.85000	±.0005	±.0015	55 ma :

3.2.2.2 Equivalent Resistance - over the operating temperature range the equivalent resistance shall not vary more than ±15 % as measured.

3.3 Environmental Requirements

3.3.1 The units shall be subjected to the conditions specified in the applicable test paragraph listed below subsequent to those specified. There shall be no evidence of mechanical damage, no harmful corrosion, no open-or-short circuiting, no intermittent contact. In addition, the units shall conform to the post test end points specified below.

3.3.1.1 Vibration. The frequency shall not change from the measurement made prior to test by more than ±0.0005 and the effective resistance by more than ±10% when crystal units are tested in accordance with 4.6.4. There shall be no evidence of mechanical damage.

3.3.1.2 Shock. The frequency shall not change from the measurement made prior to test by more than ±0.0005 and the effective resistance by more than ±10% when crystal units are tested in accordance with 4.6.5. There shall be no evidence of mechanical damage.

3.3.1.3 Temperature cycling: The frequency shall not change from the measurement made prior to test by more than ±0.0005% and the effective resistance by more than ±10% when the crystal units are tested in accordance with paragraph 4.6.6.

3.3.1.4 Salt Spray: There shall be no evidence of corrosion, and the frequency and equivalent resistance at ±30°C shall be as specified in paragraph 3.2.2.1 when the crystal units are tested in acc dance with 4.6.7.

3.3.1.5 Leakage: Case leakage shall not exceed 10^{-8} CC/SEC when the crystal units are tested in accordance with 4.6.9.

3.3.1.6 Life: After 24 hours soaking period at +65°C the difference frequency between any two readings shall not exceed 0.001% (10 PPM), when the crystal units are subjected to testing specified in paragraph 4.6.10.

3.3.1.7 Aging: At +65°C the aging rate slall be no greater than 14 parts in 10^{-7} per week, when the crystal units are subjected to testing specified in paragraph 4.6.11.

3.3.1.8 Solderability: The crystals shall be tested in accordance with Method 208, MIL-STD-202.

4. Quality Assurance Provisions

4.1 Classification of inspection - The examination and testing of the items shall be classified as follows:
(A) Qualification Inspection (B) Acceptance Inspection

4.2 Responsibility for Inspection

4.2.1 Supplier - The supplier is responsible for the performance of all inspection requirements as specified herein, except as otherwise specified. The supplier shall utilize his own or other inpsection facilities or services approved by RCA. Inspection records of the examinations and tests shall be kept complete and available, as specified in the purchase order. RCA reserves the right to perform any of the inspections set forth in the specification where such inspections are deemed necessary to assure that supplies and services conform to the specified requirements.

4.2.2 Test Equipment and Inspection Facilities - Test equipment and inspection facilities shall be of sufficient accuracy, quality, and quantity to permit performance of the required inspection. The supplier shall establish calibration of inspection equipment to the satisfaction of RCA.

4.2.3 Inspection System - The inspection system will be subject to evaluation for compliance with NASA Quality Publication NPC 200-3, Inspection Systen. Provisions for Suppliers of Space Materials, Components and Services; whenever the term "NASA" appears in NPC 200.3, it shall be interpreted as meaning RCA

4.3 Inspection Conditions and Methods.

4.3.1 Standard Test Conditions - Unless otherwise specified herein, all measurements and tests shall be made at an ambient temperature of 25 + 10, -5°C, atmospheric pressure, 90 percent maximum relative humidity. The tolerance on all other temperatures shall be ±3°C. When test temperatures higher than 35°C are specified adequate circulation of air shall be provided to maintain the temperatures within their specified tolerances. During all electrical measurements, the units shall be in temperature equilibrium at the specified temperatures.

4.3.2 Alternate Test Methods - Other test methods may be substituted for those specified provided the manufacturer demonstrates that the results obtained corrective with those obtained with the methods specified herein.

4.4 Qualification Inspection:

Qualification tests shall be conducted by the vendor at his facility or any other facility acceptable to the buyer in accordance with Table I.

TABLE I

TEST	TEST PARA.	CONDITIONS	ACCEPTANCE CRITERIA	DEFECTS ALLOWED
Group 1 - All Units (30)				
Visual and Mechanical (External)	4.6.1			
Insulation Resistance	4.6.8		≥ 500 Megohms	
Leakage	4.6.9		≤ 10^{-8} CC/SEC	
Frequency				
Equivalent Resistance	4.6.2.1	at -20°C, +25°C & +75°C	F ±.0005 at +25°C Δ F: See para. 3.2.2.1 Δ R: See para. 3.2.2.2	1

TABLE I (Continued)

TEST	TEST PARA.	CONDITIONS	ACCEPTANCE CRITERIA	DEFECTS ALLOWED
Group 1 - All Units (30) (Continued)				
Frequency and Equivalent Resistance	4.6.2.2	Over temp. range	ΔF: See para. 3.2.2.1 Δ R: See para. 3.2.2.2	
Unwanted Modes	4.6.3			
Group 2 - 10 Units				
Shock	4.6.5		Δ F: ≤ 0.0005% Δ R: ≤ 10% No Mechanical Damage	1
Vibration	4.6.4	In addition to Standard Test Conditions as shown in Para. 4.3.1 the temperature before and after a given test shall not vary by more than 3°C	Δ F ≤ 0.0005% Δ R ≤ 10% No Mechanical Damage	
Temperature Cycling	4.6.6		Δ F ≤ 0.0005% Δ R ≤ 10% No Harmful Corrosion	
Salt Spray	4.6.7			
Group 3 - 20 Units				
** Aging	4.6.11	At +65°C	Δ F ≤ 14 x 10^{-7}/WK	
Life (Non-Operating)	4.6.10	3000 hours at +65°C for qualification plus an additional 2000 for reliability data	Δ F ≤ 0.0015%	

*Total Δ resulting from this group of tests shall not exceed .0015% (15 PPM) and R ≤ 60 Ohms parts 1/2
** Aging concurrent with life. 45 Ohms parts 3/4

4.4.1 Prior qualification - Part type shall be considered to be qualified provided manufacturer demonstrates that the type involved, or a closely allied type, has been manufactured and qualified for other high-reliability programs under conditions consistent with the requirements of the end equipment in which the part will be used. See Note 6.5.

4.5 Acceptance Inspection:

4.5.1 Acceptance Tests

4.5.1.1 Group A Inspection. Not applicable.

4.5.1.2 Group B Inspection: Not applicable.

4.5.1.3 Group C Inspection. (16 Units) The test specified in Table II shall be performed on a random sample. Units which have been subjected to Group 1 Table II Test shall not be shipped to RCA. Units which successfully pass Group 2 Table II tests shall be shipped to RCA against purchase order.

TABLE II

TEST	TEST PARA.	CONDITIONS	ACCEPTANCE CRITERIA	DEFECTS ALLOWED
Group 1 (8 Units)				
Shock	4.6.5	In addition to standard test conditions as shown in para. 4.3.1 the temperature before and after a given test shall not vary by more than 3°C	Δ F ≤ 0.0005% Δ R ≤ 10% No Mechanical Damage	1
Vibration	4.6.4	at the turning point	Δ F ≤ 0.0005 Δ R ≤ 10 No Mechanical Damage	
Group 2 (8 Units)				
Life (Non-Operating)	4.6.10	1000 hours at +65°C	F 0.001	1

4.5.2 Screening Test: The following tests in Table III are to be performed by the vendor on all units prior to submission to the buyer:

TABLE III

TEST	PARA.	CONDITIONS	ACCEPTABLE CRITERIA
Visual and Mechanical (External)	4.6.1		
Insulation Resistance	4.6.8		$\geq$500 Megohms
Temperature Cycling	4.6.6		Δ F.: Not applicable Δ R: Not applicable No physical damage
Leakage	4.6.9 4.6.9.2		No Bubbles 10^{-8} CC/SEC
Frequency and Equivalent Resistance	4.6.2.1	At -20°C, +30°C & +75°C	F: See Para. 3.2.2 Δ F: See Para. 3.2.2.1 Δ R: See Para. 3.2.2.2
Frequency and Equivalent Resistance	4.6.2.2	Over Temp. Range	Δ F: See Para. 3.2.2.1 Δ R. See Para. 3.2.2.2
Aging (Concurrent with life Group 2, Table II)	4.6.11	+65°C	Δ F: $\leq 14 \times 10^{-7}$/WK

4.5.3 Acceptance Test Data: The data obtained from the acceptance tests shall be supplied at the time of shipment. Such data shall be shipped with the lot.

4.5.4 Lot - A lot shall consist of units which are submitted for acceptance and which have been manufactured on the same production line or lines during a period of not more than one week or a longer period of time provided the manufacturer demonstrates that a homogeneous lot has resulted from use of the same materials, processes, personnel, and assembly equipment. Where a lot contains more than one part value, sampling shall be proportionate to the lot composition.

4.6 Test Procedures

4.6.1 Visual and Mechanical - The crystals shall be inspected to verify that the materials, design, construction, physical dimensions, marking and workmanship are in accordance with the applicable requirements.

4.6.2 Frequency and Equivalent Resistance:

4.6.2.2 Frequency and Equivalent Resistance over the Operating Temperature Range - Measurements shall be performed continuously in accordance with Method A per MIL-C-3098, suitable recording devices are required to provide continuous measurements as a function of temperature.

4.6.3 Unwanted Modes - Unwanted Mode measurement shall be in accordance with para. 4.7.5 of MIL-C-3098.

4.6.4 Vibration: The crystal units shall be tested in accordance with Method 204, MIL-STD-202, Test Condition D. The following shall apply:

1 - Method of Mounting: Rigidly in each of 3 mutually perpendicular planes.
2 - Electrical Load conditions: Non-operating.
3 - Measurements: Measure frequency and effective resistance before and after testing at the turning-point temperature of each crystal unit.

4.6.5 Shock: The crystal units shall be tested in accordance with Method 202, MIL-STD-202. The following details shall apply:

1 - Mounting: Shall be successively mounted in each of its 3 principal axes.
2 - Electrical Load Conditions: Non-operating.
3 - Acceleration: 100 G.
4 - Number of Blows: One in each of the 3 mutually perpendicular planes.
5 - Measurements: Frequency and effective resistance before and after testing at the turning-point temperature of each crystal unit.

4.6.6 Temperature Cycling: The Crystal units shall be subjected to 5 cycles of temperature cycling specified Method 102, MIL-STD-202, Test Condition D, except the step 3 temperature shall be +85°C. If specified frequency and resistance measurements shall be made before and after testing at the turning-point temperature of each crystal unit.

4.6.7 Salt Spray - The crystal units shall be tested in accordance with Method 101, MIL-STD-202, Test Condition B. After this test the units shall be visually examined for evidence of corrosion, which would be detrimental to performance requirements of the crystal. If specified, frequency and resistance measurements shall be made before and after testing at +30°C.

4.6.8 Insulation Resistance - The crystal units shall be tested in accordance with Method 302, MIL-STD-202. The following details shall apply:

1. Test Potential 50 to 100V
2. Points of measurement between the pins and each pin and the holder

4.6.9 Leakage - The units shall be subjected to gross and fine leakage test in accordance with MIL-STD-202, Method 112, Condition C, Procedure III B. The equipment sensitivity shall be 10^{-9} CC/SEC., minimum.

4.6.10 Life - The crystal units shall be maintained at a temperature of +65°C ± 2°C for the specified period, but the long term temperature stability of the aging oven shall be ±0.1°C. The initial frequency measurement shall be made at 24 hours after the start of the test. Thereafter, measurements shall be made twice a week at intervals of not less than 2 days nor more than 4 days throughout the remaining period. Test oscillator approved by RCA shall be utilized for the duration of the test. The frequency measurement accuracy shall be at least ± 1 part in 10^8. In the event of any condition which will bring the temperature of the units below the aging temperature for the intervals of more than 1 hour. No measurement shall be made until 24 hours after temperature restoration, and the test period shall be lengthened by the length of time that the temperaturefailed.

4.6.11 Aging:

4.6.11.1 Test Method A:

4.6.11.1.1 Soak period: Crystals shall be maintained at +65°C for a period of two weeks, during which no measurements are to be taken. Soak period may be performed in the aging oven at the manufacturer's option.

4.6.11.1.2 Measurements: Crystals shall be transferred to a +65°C +2°C aging oven within two hours after removal from soak oven. Long term temperature stability at aging oven shall be ±0.1°C, and frequency measurements shall be accurate to at least ±1 part in 10^8. Test oscillator approved by RCA shall be used for the duration of the test. After a minimum of 36 hours for stabilization the reference frequency shall be determined . On the seventh day after the reference reading is taken a reading shall be taken from which aging rate is calculated, based on the reference reading. Crystals exceeding the specified aging rate shall be measured on the eighth day and the aging rate calculated from that reading and the better of the first two readings (the reference of the one taken one day thereafter).

4.6.11.1.3 Reference Reading: The reference rading is established by taking an initial reading and daily readings thereafter until the average difference in daily readings for the majority of crystals within the same oven is two times the daily aging limit or less. The next to last daily reading in the sequence shall be called the reference reading.

4.6.11.2 Test Method B (to step up test cycle):

4.6.11.2.1 Soak Period. None

4.6.11.2.2 Measurements: Crystal units shall be maintained at +65°C ± 2°C in an aging oven with a long term temperature stability of ±0.1°C. A transistorized test oscillator approved by RCA shall be utilized for the duration of the test. The frequency measurement accuracy shall be at least ± 1 part in 10^8. After a minimum of 36 hours for stabilization, the reference reading shall be determined. On the seventh day after the reference reading is taken a reading shall be taken from which the aging rate shall be calculated, based on the reference reading. Crystal units exceeding the specified aging rate shall be measured on the eighth day and the aging rate calculated from that reading and the better of the first two readings (the reference or the one taken one day thereafter). Crystal units meeting the specified aging limit may be removed from the aging oven. Additional aging test cycles may be performed on the remaining units in accordance with paragraph 4.6.11.1 or 4.6.11.2.

4.6.11.2.3 Reference Reading: See paragraph 4.6.11.1.3

4.7 Failure Reporting, Analysis and Corrective Action - The manufacturer shall report to the cognizant buyer on any failure which occurs during any phase of testing or conditioning within 24 hours of occurrence of such failures. A report of the cause of failure and corrective action initiated to prevent recurrence of such failure shall be submitted to the buyer within two weeks after the failure. In the event of a field failure, RCA will return the failed unit to the manufacturer, who will conduct a failure analysis on the unit and submit a report of cause of failure and corrective action initiated to prevent recurrence of such failure. The report shall be submitted within two weeks after receipt of the failed unit from RCA.

4.8 Documentation - The documentation shall be in accordance with RCA drawing 8533293.

5. Preparation for delivery

5.1 Packaging - the units shall be packaged in such a manner that they will be protected during shipment and storage. Packages shall be suitable to provide maximum physical protected in a minimum-size package without undue distortion of the leads of the device.

5.2 Packing - Units, packaged as specified, shall be packed in containers of the type, size and kind commonly used for the purpose, in a manner that will insure acceptance by common carrier and safe delivery at destination.

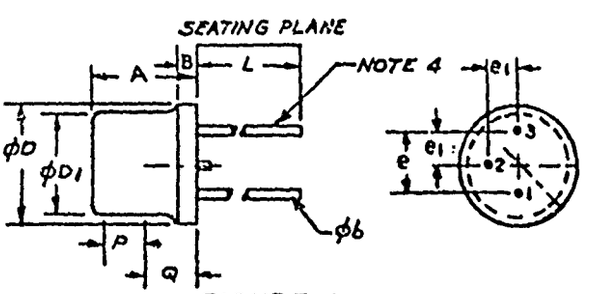
6. Notes

6.1 Design or Process Change - No change shall be made in the design or process of manufacture of units to be shipped to the customer from that of units subjected to qualification inspection without prior approval by the Radio Corporation of America.

6.2 Failure Rate: The failure rate goal for the items specified, when operated in an ambient of 25°C and subjected to 50% of specified maximum electrical ratings, shall be .005%/1000 hours.

6.3 The supplier shall exercise caution to avoid physical damage, particularly to plating of lead or terminals resulting from the abnormal amount of handling necessary for high reliability processings, such as repeated insertion and withdrawls from test fixtures.

6.4 Suggested Source: McCoy Electronics Co., (H4 Code 09136)
Mt. Holly Springs, Pa.



Symbol	Inches Min.	Inches Max	Notes
A	.240	.265	
B		.025	
ϕ b	.016	.021	
ϕD	.335	.370	
ϕD1	.305	.335	
e	.700 T.P.		2
e1	.100T.P.		
L	.500		
P	.100		1
Q			3

FIGURE 1
MECHANICAL DIMENSIONING FOR CRYSTAL UNIT USING TO-39 CASE

NOTES:
1. This zone is controlled for automatic handling. The variation in actual diameter within the zone shall not exceed .010.
2. The device may be measured by direct methods or by the gage and gaging procedure described on gage drawing GS-1.
3. Details of outline in this zone optional.
4. Active leads are numbers 1 and 3. Lead 2 is not connected, and cut off at header.

TEMPERATURE RUN, MIL-C-3098E:

AN ANALYSIS

R. Pompeo and F. Wolf
McCoy Electronics Company
Mt. Holly Springs, Pennsylvania

Summary

The advent of Mil-C-3098E, Amendment 1, July 31, 1968 in paragraph 4.7.4.3.1, specified a piece of test equipment by manufacturer and model for determining crystal performance over the temperature range. The procedure as described in paragraph 4.7.4.3.1 has presented some problems in implimentation. Many confronted with the procefure do not agree on how the test is to be conducted.

The paper will deal with the interpretation of paragraph 4.7.4.3.1, Mil-C-3098E, Amendment 1, July 31, 1968. An analysis will be made of the paragraph presenting a step by step procedure conducting the temperature run.

The test equirment required is examined. Calibration of the equipment will be discussed and versatility of the equipment for all test sets required in Mil-C-3098 is examined. Standard times for conducting the test are presented.

Recommendations to improve the test setup are presented.

Introduction

Probably since the first observer ascertained crystal frequency was contingent on temperature, measuring crystal performance over a temperature range has been the cause of disagreement between seller and customer. Standard procedures that all ir the industry could or would employ have been lacking.

Recently, in an effort to resolve this dilemma, a procedure was drafted and became paragraph 4.7.4.3.1, amendment 1, Mil-C-3098E. This paragraph describes the test procedure and specifies the equipment by manufacturer and model that shall be used. The paragraph is far from being a detailed step-wise procedure; its language is subject to interpretation. We first review the paragraph and then present our interpretation.

Test Procedure

The procedure is really two paragraphs: "4.7.4.3 Temperature Run, Operating Range, Non-controlled types. Measurements of frequency and equivalent resistance of a crystal unit, designed for operation under noncontrolled temperature conditions, shall be performed at rated drive level (see 4.7.4.5). The unit shall not be disassembled, and indirect means shall be used for determining the temperature of the resonator." and paragraph "4.7.4.3.1 Test Procedures. The temperature of the crystal unit shall be varied so as to traverse the entire operating range from low temperature to high temperature. For the operating temperature range of-55°C to 105°C, the temperature range shall be traversed in eight ± one minutes, unless otherwise specified. For the other operating temperature ranges, the time shall be proportional. Measurements of frequency and equivalent resistance shall be recorded continuously, or at intervals of not over 3°C, to ascertain that tolerances are not exceeded at any instant. The temperature of the end points shall be accurate to within ± 1°C of specified temperatures. The end-point frequencies shall be within ± 5% of the specified overall frequency tolerance, when compared to the equilibrium frequency at the end-point temperatures. For example, if the specified frequency tolerance is ± 0.005% (over-all 100ppm) then the end-point tolerance is 5 ppm. The temperature run shall be performed automatically from the low temperature to the high temperature using a T/C analyzer, Winslow Tel-Tronics, Inc. Model TCA-1070, or equal. No manual adjustments shall be made to the test setup once the temperature run has begun. Note: this type of temperature run may cause some distortion of the frequency temperature characteristics"[1]

It should be noted that the first paragraph refers to the unit as an assembly containing a resonator. The second paragraph uses only the phase crystal unit.

Analysis

How can the procedure of this paragraph be conducted? If analyzed, it is found that it can be restated in three steps.

Recall that paragraph 4.7.4.3.1 states, "The temperature of the end points shall be accurate to within ± 1°C of the specified temperatures. The end point frequencies shall

be within ± 5% of the specified overall frequency tolerance, when compared to the equilibrium frequency at these end-point temperatures". An example from the family of generalized frequency curves is represented in figure 1. [2] This is a "worst case" presentation. The solid areas represent the 5% value for a 0.005% tolerance unit. If a nominal frequency of 10.000 MHz is used as an example, the solid area at the bottom of the graph is 50 Hz in width and represents a minimum frequency of 9.999450 MHz.

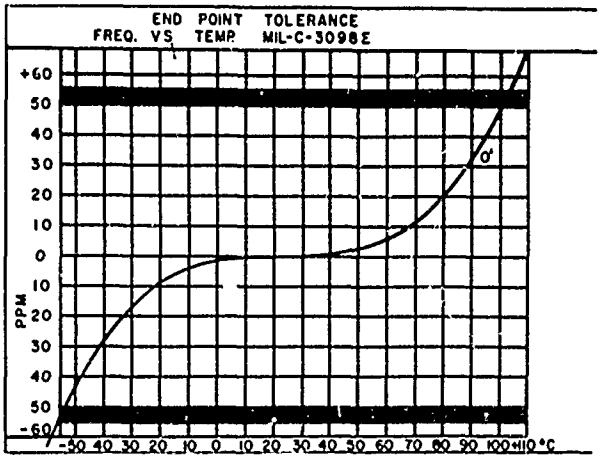


Figure 1

Thus, Step 1: Stabilize the crystal unit to be tested at an ambient of -55 + 1°C or at specified lower end point temperature, and measure the equilibrium frequency.

Step 2: Increase the temperature to 105± 1°C in a period of 8 ± 1 minutes. This shall be accomplished with the Winslow Tele-tronics, Inc. T/C analyzer, model TCA-1070 or equivalent equipment in conjunction with the appropriate CI meter as specified. During this period, the frequency and equivalent resistance shall be recorded continuously or at intervals not exceeding 3°C. No manual adjustments shall be made to the test equipment once the temperature run has begun.

Step 3 begins upon attaining 105°± 1°C, or the upper specified end point temperature: Stabilize the crystal unit and determine the frequency.

Reference is now made to the solid area at the top of the graph (Figure 1.) To use the previous example, the frequency limit is 10.000 50 MHz.

Test Equipment

What is the TC analyzer specified in the procedure? It consists of a proportionally controlled chamber for housing the crystal unit undergoing test. The assembly is represented in figure 2; the chamber accomodated an HC-6/U or HC-27/U holder. Inserts are provided for the HC-18/U, as illustrated in Figure 2, HC-32/U, and the compression welded HC-35/U, (TO-5). The adaptors supplied with the analyzer will not accomodate the HC-26/U and HC-29/U. The nominal thickness of these glass enclosed types is 0.183 inches, too large for the adaptor.

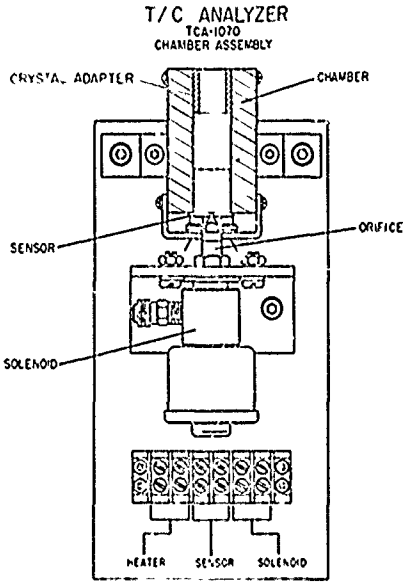


Figure 2

The chamber is provided with a solenoid operated valve for controlling carbon dioxide used to cool the specimen to the lower end point temperature. It is vitally important that a scrubber be installed on the supply line to reduce clogging of the small orfice.

This bit of knowledge was harvested as bitter fruit; an average of two temperature runs was the most that could be performed prior to disassembly of the chamber to clean the orfice. When connected to a 1000 pound carbon dioxide source, a considerable leak occured at the junction of the orfice and solenoid valve. Until repaired, as much CO_2 entered the room as the interior of the chamber.

A complete test position with the analyzer, crystal impedance meter, frequency counter with printer, and graphic recorder is shown in figure 3. The counter is adjusted so that the frequency will be printed on a two-second average. The graphic recorder senses the equivalent resistance continuously.

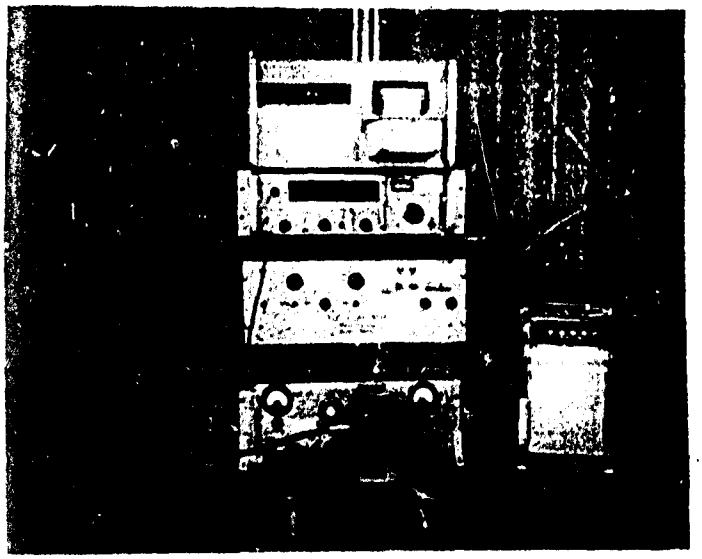


Figure 3

Utility

The TC analyzer is supposedly designed to operate with all crystal impedance meters specified in paragraph 4.2.1.1, Mil-C-3098E. It falls somewhat short of meeting this criterion.

Shown in Figure 4 is the T/C analyzer and TS330/TSM, requiring an adaptor to provide clearance of CI meter controls. This may impose no problem at low frequencies, but at the higher frequencies, correlation problems appear likely with narrow tolerance crystal types. It was found necessary to replace the original guide tracks with longer ones, increasing the travel of the carriage by 2.5 inches, to minimize the length of such adaptors, the longer guide tracks make it possible to connect an HC6/U holder to the CI meter without adaptors, if the attractive cover is removed.

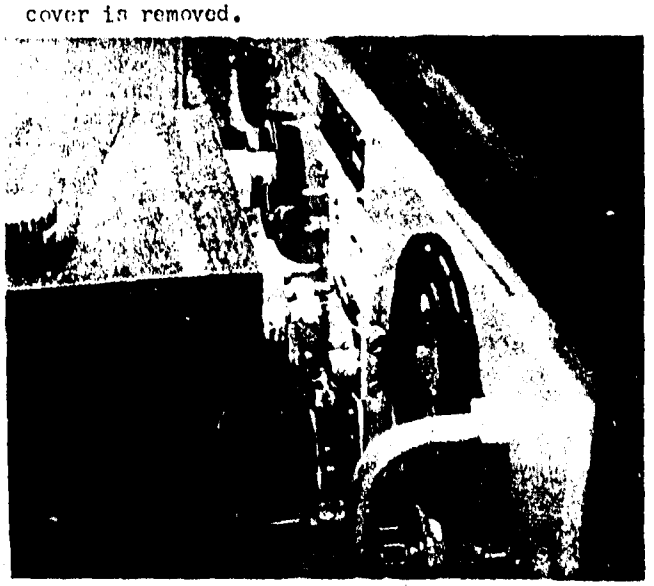


Figure 4

When the T/C analyzer is used with the TS-683/TSM, an alinement problem is encountered, see Figure 5. This may promote a crack in the glass seal. This shortcoming can be corrected by adjusting the height of the carriage platform.

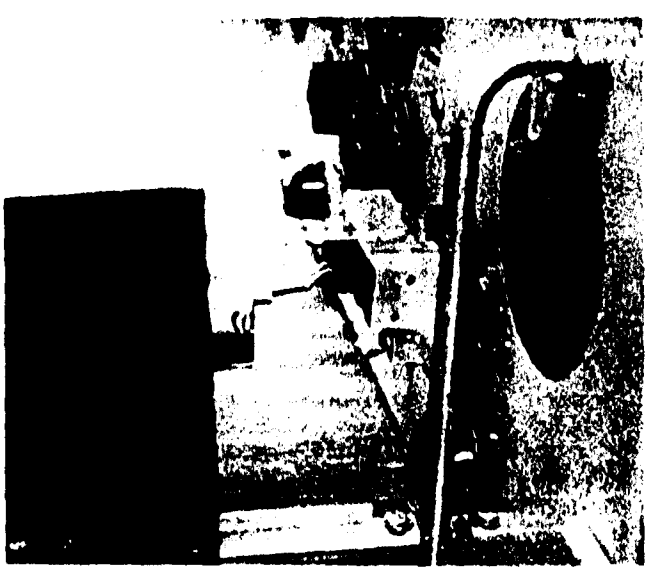


Figure 5

Many applications today require the AN/TSM-1C. However, this circuit was never considered in the design of the analyzer. As shown in Figure 6, there is no position which will allow the chamber to engage crystal units in the receptacles of the CI meter.

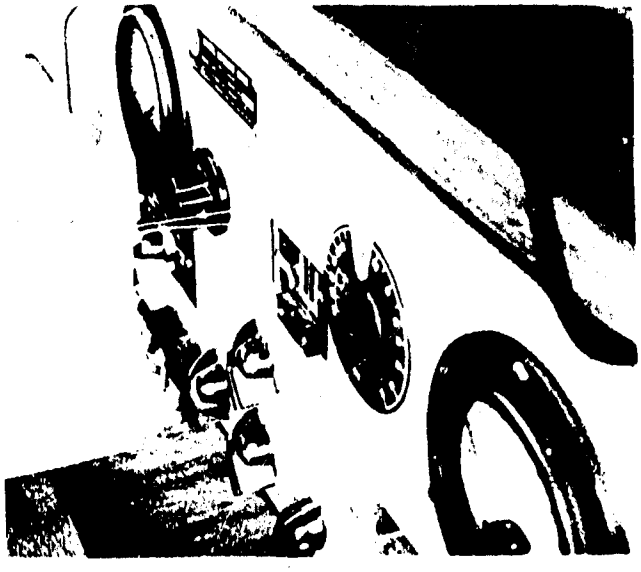


Figure 6

Chamber Design

As previously stated, and shown in figure 2, the chamber will accomodate the HC-6/U holder. A symmetry exists between the unit being tested and the sensor. However, if an HC-13/U were to be tested, this symmetry would probably be upset, causing erroneous temperature characteristics to be observed.

The adaptor as illustrated in figure 2, required the milling of a socket to engage the ball on the inside wall of the chamber to prevent its sliding the complete depth of the chamber, again destroying the symmetry.

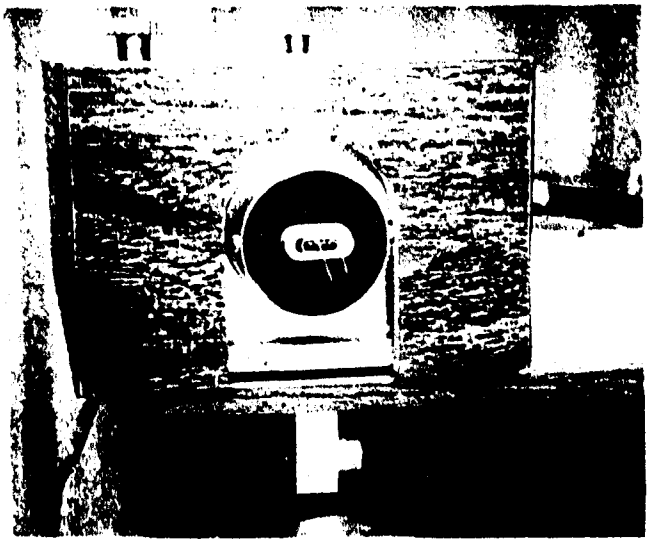


Figure 7

Thermal conductivity might be enhanced if the adaptor was open only on one end and more closely fitted the crystal being tested, see Figure 7.

Calibration

As with all test equipment employed for inspection, the T/C analyzer must be maintained in accordance with Mil-C-45662. Temperature evaluation of the analyzer was accomplished by placing a platinum resistance probe, model 542, United Systems Corp., with model 522 platinum thermometer, in the chamber cavity and the cooling circuit energized. After about fifteen minutes to allow for stabilization, the temperature was recorded at about -35°C. At the extreme upper temperature setting (105°C), the analyzer was also found to attain a temperature of about 85°C.

Contacts with other owners of T/C analyzers, and government representatives in the field, disclosed ours was not an isolated case. All who were contacted experienced similar situations.

Adjusting Analyzer

The temperature within the analyzer chamber cavity was corrected by indirect means.

The frequency of a 10.000 MHz, CR-19/U was determined at the end point temperatures. The correlation specimen, with platinum resistance probe taped to it, was placed inside a temperature chamber, model RH 1100, Associated Testing Laboratories, using a coaxial lead, type RG58U, 1½' long, to connect the unit to a crystal impedance meter, TS-330/TSM. After a period of about one hour of stabilization at -55°C, the frequency was recorded with a digital recorder, model 5050, Hewlett-Packard. The procedure was repeated at +105°C. The correlation unit was then transferred to the chamber of the analyzer and the temperature reduced. The resistor, R-28, controlling the solenoid, was then adjusted until the frequency equivalent to that observed at -55°C was reproduced and maintained. About eight hours were necessary to accomplish this adjustment.

Again at the 105°C setting, the same procedure was followed, adjusting the resistor R29 controlling the heaters until the previously recorded frequency was duplicated. As before, a minimum of 15 minutes was allowed to verify each adjustment. The entire procedure was conducted for each of the ten intermediate end point temperature settings (-40, -30, -20, 0, 10, 20, 40, 60, 85, 95°C) possible with the analyzer.

After several correlation specimens were evaluated in this manner, the test position satisfied compliance to the requirements of Mil-C-3098E and was approved by the government inspector.

Subsequent calibration as requred for maintenance of the equipment has been conducted. No adjustment to the analyzer circuitry has been required.

Operation

The T/C analyzer has been used only in what its manufacturer calls the recording mode; many require a recorder of performance characteristics.

The instruction manual for the analyzer describes the procedure on page 12.[3] The instruction states, "When the low temp. lamp goes out, allow 30 seconds for crystal stabilization". This is supposed to satisfy the requirement in paragraph 4.7.4.3.1, Mil-C-3098E for end point frequency equilibrium requirements. In our efforts, we have found the crystal exhibits a frequency equivalent to a temperature below -55°. A minimum of two minutes (120 seconds) is required before the crystal exhibits an equilibrium frequency. When allowing only 30 seconds for equilibrium, repeatability is unattainable. Likewise, at the upper end point temperature, the crystal has not yet attained an equilibrium frequency when sensor is at +105°C. A minimum of two minutes is required to achieve equilibrium.

The equipment was used to test crystals from 1 to 75 MHz. Time required to perform the tests is shown in Table I.

No doubt the validity of a 2 minute stabilization period is subject to question; however, these mean times were derived from observation for considerably longer periods. Each type to be tested should be evaluated accordingly.

Ascertain from these observations, the time required to test a crystal over the range required in Mil-C-3098E, -55°C to +105°C, averages 14 minutes, not including the time required to place the unit in the CI meter and the chamber in position over it.

Conclusion

The principles of the test procedure of Amendment 1, Mil-C-3098E, offer a practical solution to a vexing problem, with a few modifications, it will suffice for qualification and acceptance testing.

If a specific piece of test equipment is to be included in a specification, its utility must be carefully analyzed. The T/C analyzer when properly calibrated, electronically complies with the requirements of paragraph 4.7.4.3.1, of Mil-C-3098E. However, the model presently available is not adaptable to all the CI meters included in paragraph 4.2.1.1 of Mil-C-3098E. As indicated, it exhibits shortcomings in this area. The mechanical design of this equipment should be reviewed.

TEMPERATURE RUN MEAN TIME VALUES

TYPE	FREQ	COOL CHAMBER TO -55°C	LOW TEMP EQUILIBRIUM	TEMP RUN	HI TEMP EQUILIBRIUM
CR-18 A/U	1.00000 MHz	1.3	1.8	8.7	2.0
CR-19 A/U	5.00000 MHz	1.2	2.6	8.7	3.0
CR-19 A/U	10.000 MHz	1.2	2.0	8.72	2.0
CR-56A/U	74.000 MHz	1.2	1.9	8.75	1.9
		1.25	2.0	8.75	2.0
		14 MIN. TOTAL TIME			

Table I

The following modifications are suggested for consideration:

1. Longer guide tracks to permit the chamber to encapsulate a crystal unit without the need for lead extension.

2. The adaptors for small crystal holders be more carefully machined and their cavity be only as large as the maximum illustrated dimensions of the drawing.

3. The carriage for the chamber be modified so that it can be adjusted to aline with the crystal receptacle of all CI meters. The cover should be designed so as not to interfere with this function.

To the government agencies responsible for the specification, the following recommendations are offered:

1. An operating procedure be supplied to all field representatives responsible for certifying these tests.

2. Include in the procedure minimum times for stabilization at the end point temperatures; 2 minutes are suggested at each temperature.

3. Require equipment manufacturers whose products are referenced in military specifications, to comply with the same calibration and quality control procedures required of crystal manufacturers.

It is sincerely hoped this analysis will contribute to clarification of paragraph 4.7.4.3.1, Mil-C-3098E, and a standard procedure for testing crystals over a temperature range.

Bibliography

1. Military Specification, Crystal Units, Quartz, General Specification for, Mil-C-3098E, amendment 1, 31 July, 1968., p.1.

2. Instruction and Operating Manual for Model TCA 1070, T/C Analyzer, Winslow Tele-tronics, Inc. Asbury Park, N.J., p. 12.

3. Bennet, Roger E., Manufacturing Guide for "AT" Type Units, Union Thermoelectric Division Comptometer Corporation, Niles, Illinois, 196 p. 91.

ANOMALOUS VIBRATIONS IN AT-CUT PLATES

Issac Koga

Research Laboratory, Kokusai Denshin Denwa Co.
(Japanis Overseas Radio and Cable System)
Nsba-meguro, Tokyo, Japan

Summary

Some of the foreign vibrations in an AT-Cut quartz plate absorb the energy of the principal vibration in such a way that the higher the drive level of the principal vibration, the greater the absorption by the foreign vibration when the frequencies of both vibrations coincide at a certain temperature. These phenomena, called activity dips in the United States, were investigated in the following way. In a rectangular shape AT-Cut plate, of which one side was made perpendicular to an electric axis of quartz, one of the overtone frequencies of the longitudinal vibration in the direction of electric axis (say, X-vibration) was adjusted to become practically equal to that of the principal thickness-shear-flexural vibration. Then, although a coupling between the two vibrations seemed to be almost zero, an activity dip was clearly observed in the principal vibration at a certain temperature when the latter was driven at a certain level, and the higher the drive level of the principal vibration beyond that level, the greater the activity dip. Thus the so called activity dips were artificially realized.

The mechanism of the activity dips is understood as follows. In case when the drive level of the principal vibration is small, the friction loss for the X-vibration is large enough to prevent the starting of X-vibration, but when the drive level of the principal vibration is beyond a certain level, the friction loss for the X-vibration becomes so small that the X-vibration can start by absorbing the energy of the principal vibration. A series of experiments were performed to verify this interpretation.

Introduction

As is well known, the principal vibration of a crystal is coupled loosely or closely with many other foreign vibrations, and whenever the frequency of a foreign vibration is coincident with that of the principal vibration at a certain temperature, the energy of the latter is absorbed by the former. Therefore in a vacuum tube oscillator, consisting of such a plate as a frequency control element, the output of the oscillator dips and the frequency changes in a complicated way. However, sometimes a foreign vibration absorbs the energy of the principal vibration only when the amplitude of the latter is beyond a certain level. These phenomena, which are called activity dips in the United States, were found in 1950, nearly twenty years ago, by one of the collaborators of the writer. He found that the frequency-temperature characteristics of a crystal oscillator were not necessarily similar when D.C. anode voltages of the oscillator were considerably different.

Artificial Activity Dips

After a series of experiments thereafter the writer eventually understood that these activity dips were due to the fact that the larger the amplitude of the principal vibration, the smaller the friction loss aginst the foreign vibrations. To verify this interpretation the writer prepared a rectangular shape AT-Cut plate, the thickness and frequency of the principal (thickness-shear) vibration being about 0.53 mm and 3.1458 MHz respectively, and the size of the plate being about 19.06×26.00 (mm), and the shorter sides of the plate being made parallel to one of the electrical axes of quartz (Fig. 1). This shorter sides were finely adjusted so that the 21st harmonic overtone frequency of the longitudinal vibration was

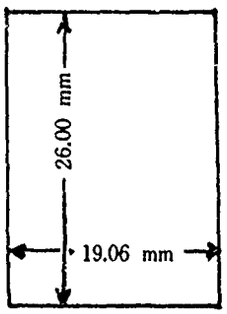


FIG. 1 - Dimensions of crystal plate under investigation

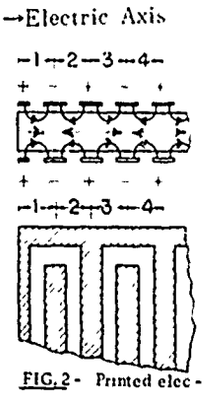


FIG. 2 - Printed electrode for the excitation of X-vibration.

equal to the frequency of the principal vibration. This frequency adjustment was performed in the following procedure. A pair of printed electrodes were prepared (Fig. 2), so that an electric field in the direction of electric axis was applied alternately in the same and opposite senses respectively to every 1/21 part of the plate. The response frequency, that is, the 21st harmonic overtone frequency of the longitudinal vibration along the shorter sides (say "X-vibration" briefly) was adjusted to become very nearly equal to the frequency of principal vibration by carefully reducing the length of the shorter sides. According to the

general understanding the X-vibration seems to have no coupling with the principal vibration, but a remarkable activity dip and a complicated frequency change were observed at a certain temperature when the plate was driven beyond a certain level as the frequency control element in an oscillator. In case when the drive level was made still higher, hysteresses were observed both in activity dip and frequency change (Fig. 3).

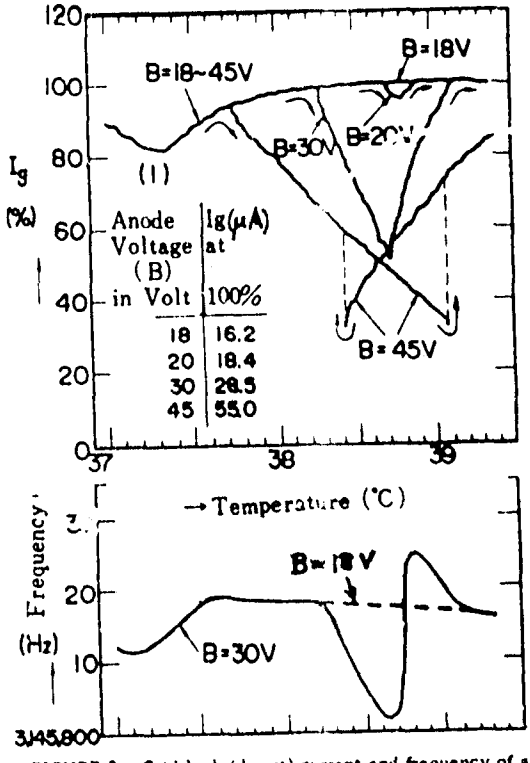


FIGURE 3- Grid leak (direct) current and frequency of a crystal oscillator vs ambient temperature of crystal at its various drive level

In order to confirm that this energy absorbing vibration was certainly the X-vibration, the frequency-temperature coefficient of this foreign vibration was measured and compared with that of the X-vibration which was determined directly by means of the arrangement already shown in Fig. 2. To measure the frequency-temperature coefficient of an energy absorbing vibration, the oscillator frequency was shifted to some extent by adjusting the circuit constants, and found anew the temperature where the activity dip occurred again (Fig. 4).

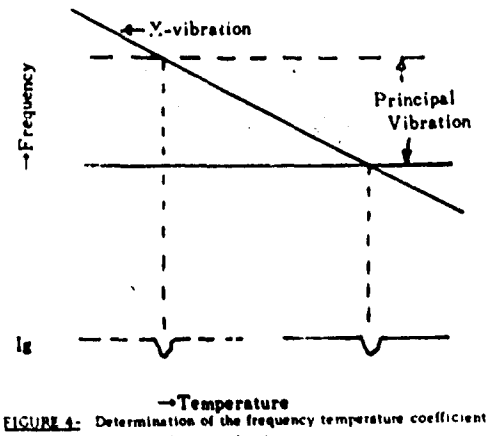


FIGURE 4- Determination of the frequency temperature coefficient of an energy absorbing foreign vibration

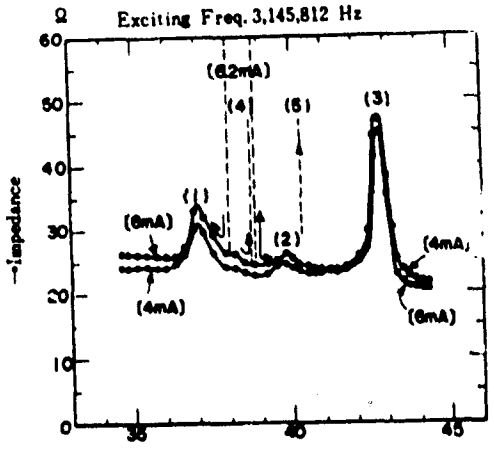


FIGURE 5- Crystal impedance vs ambient temperature at various drive level

Impedance vs Temperature at Various Drive Level

Next, in order to know the mechanism how and why the X-vibration would play a role of energy absorbing source, the impedance vs temperature at a fixed frequency (3,145,812 Hz, or the natural frequency of the principal vibration at 35°c) was observed by taking crystal current as parameter (Fig. 5). It was noticed that, while the impedance was always less than 30 Ohms in the temperature range from 37°C to 40°C at a crystal

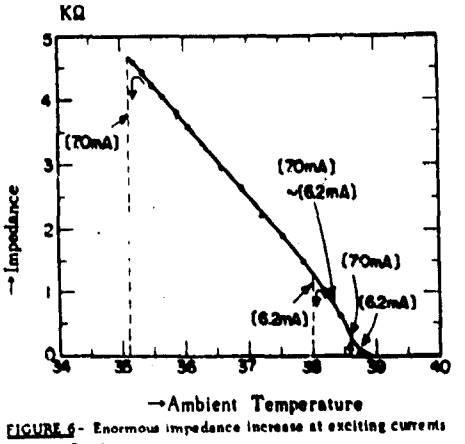


FIGURE 6- Enormous impedance increase at exciting currents above 6.0 mA

current less than 6.0 mA, it increased up to about 1 Kilo-Ohms at 6.2 mA and to about 4.5 Kilo-Ohms at 7.0 mA (Fig. 6). This means that the increase of the excitation of principal vibration significantly causes the reduction of frictional loss for the X-vibration.

Hysteresis was again observed in impedance vs ambient temperature. This relation can be explained as follows. Since the relation between the principal and the X-vibration may be illustrated by means of an equivalent circuit shown in Fig. 7, and actually the frequency of the principal vibration is independent of temperature while the frequency vs temperature is linear for X-vibration, the impedance regarded from the electrodes for crystal changes as shown by Curve (1) in Fig 8. Therefore if the crystal is driven by a constant current, the ohmic loss or the heat generation in the crystal vs crystal temperature is similar to Curve (1). Accordingly, the input impedance vs ambient temperature will incline leftwards as shown by Curve (2) in Fig. 8. Thus the impedance will actually change as indicated by arrows.

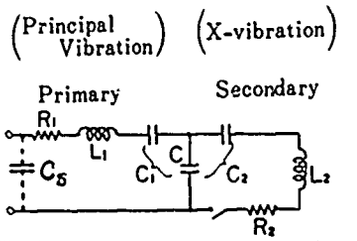


FIGURE 7 - Equivalent electrical circuit of a crystal representing anomalous vibration

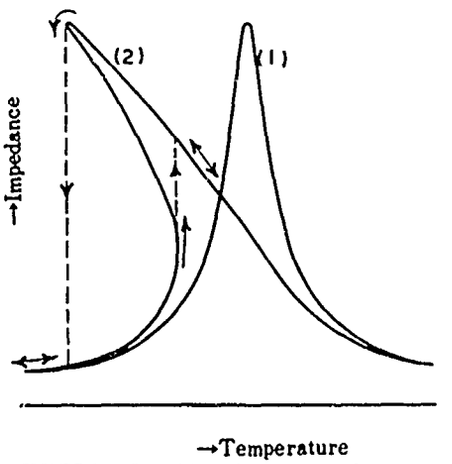


FIGURE 8 - Crystal impedance vs crystal temperature and ambient temperature

Reduction of Friction Loss for X-vibration

The reduction of friction loss for the X-vibration seems to be due to the following reasons. As has already been explained in the previous paper,[1] the principal (thickness-shear) vibration of a rectangular AT-Cut plate is accompanied by displacements in the thickness direction (Fig. 9). Therefore, if the plate, placed between two horizontal plane electrodes, is excited, it is tossed up from the lower electrode. This fact was easily recognized by observing the shifting of interference fringes produced by a projected light on a transparent electrode, which is used for another electrode and placed over the plate leaving a very small gap between them.

Thus it is understood that, if the amplitude of the principal vibration is smaller than a critical value, the friction between the electrode and the plate is large enough to clamp the motion of the X-vibration, but if the amplitude of the principal vibration is beyond a certain critical value, the X-vibration will start at a certain temperature, and the larger the amplitude of the former, the larger will be that of the latter.

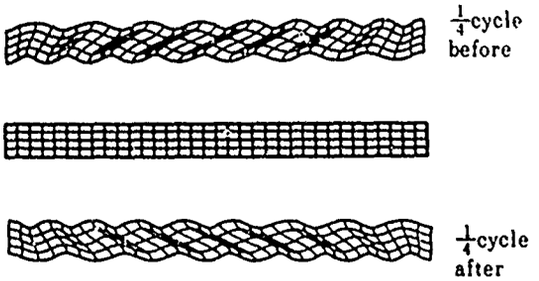


FIGURE 9 - Vertical displacements of the crystal associated with its thickness-shear vibration

Behavior of Plato in an Oscillator Circuit

When this crystal is used as a frequency control unit in an oscillator circuit, it vibrates at a frequency that makes itself zero-reactance at its terminals. Since the equivalent electric circuit can be represented as shown in Fig. 7, the zero-reactance frequency will be displayed as shown in Fig. 10. It is, therefore, easily seen that in case when the resistance R_2 in the secondary circuit, or the friction for the X-vibration is small to a certain extent, a hysteresis will occur in the frequency vs temperature as indicated by arrows.

Figure 11 shows the total resistance of crystal at its zero-reactance frequency. As is clearly seen from this figure, the smaller the resistance R_2 of the secondary circuit, the larger the total resistance at the terminals. In practical cases, if the temperature is changed of a crystal driven at a certain amplitude in an oscillator circuit, the total resistance of the crystal will start to increase at a certain temperature. This increase of resistance will cause the amplitude to decrease, in other words, will give activity dip. However, on the other hand, this decrease of amplitude will prevent from the increase of total resistance, so the amplitude of vibration will come to an equilibrium point. Thus the amplitude of vibration, represented by grid leak D.C., changes as shown in Fig. 3.

In conclusion, it is almost needless to say that actually the similar phenomena are also very often observed in case when crystal plates are wire-mounted. The behavier, though, is naturally more complicated.

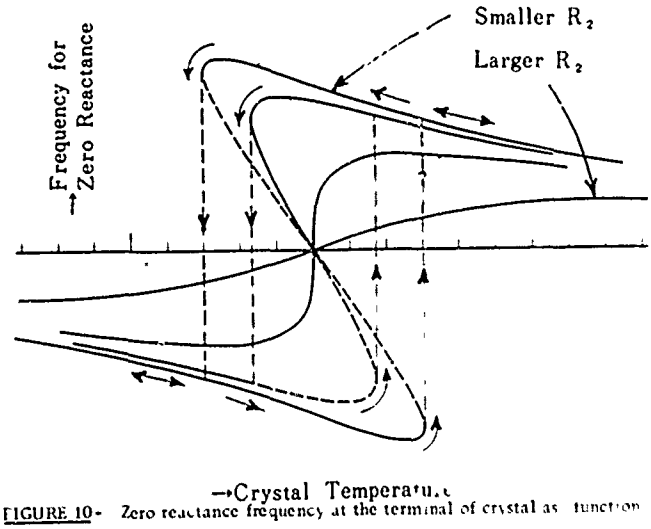


FIGURE 10 - Zero reactance frequency at the terminal of crystal as function of temperature

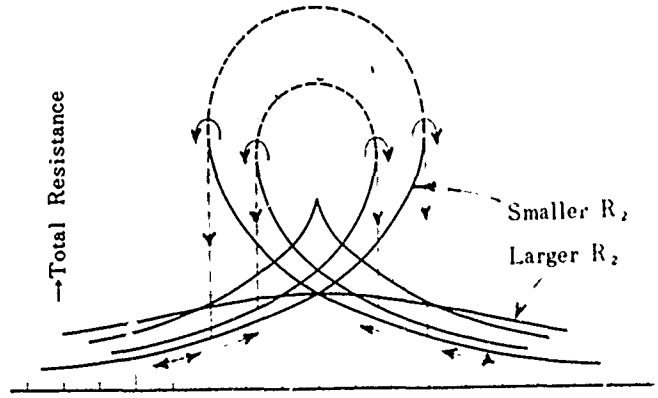


FIGURE 11 - Total resistance of crystal at its zero reactance frequency as function of temperature

Reference

(1) I. Koga, Radio Frequency Vibration of Rectangular AT-Cut Quartz Plates, JAP Vol.34 No.8, pp. 2357-65, August, 1963.

COMPARISON OF AGING PERFORMANCE OF 5 MHz RESONATORS PLATED WITH VARIOUS ELECTRODE METALS

Richard B. Belser and Walter H. Hicklin
Georgia Institute of Technology

Summary

AT-cut plano-convex resonators of 5 MHz frequency have been plated with various electrode materials including standard configurations of gold, silver, copper, aluminum, and selected bimetal layers, or with annular configurations. These resonators have been compared as to aging performance over periods of 180 to 730 days. Plating was performed in an oil-free system at pressures in the range 10^{-7} to 10^{-8} Torr. Bonding was by the ultrasonic method. Sealing was performed in cold-weld containers, HC-6/U or Bliley E7-1 types, after extended pumping and bakeout times at selected temperatures up to 450° C and at pressures in the range 10^{-6} to 10^{-7} Torr.

Resonators of excellent aging performance have been obtained with electrodes of copper, silver, aluminum, and gold. Resonators of all plating types have displayed an initial aging rate $> 1 \times 10^{-8}$/week. Within 3 weeks the rate may decrease to $< 1 \times 10^{-9}$/week and may approach 5×10^{-10}/week within 60 days with a high yield of acceptable units. Bakeout periods up to 24 hours at 400° C followed by sealing at this temperature have not necessarily improved either initial aging or the subsequent aging performance. Even for units with gold electrodes or annular electrodes, after extended bakeouts, strong positive initial aging vectors are present; and it appears that some negative vectors which are normally present have been removed or else new positive ones have been added. Resonators baked for shorter periods at lower temperatures frequently gave better total aging performance. Temperature cycling experiments for aging retrace data by oven cut-off for 48 hours and subsequent restart have not pointed to adsorption-desorption phenomena as being the principal aging vectors encountered here. Mounting strain has also been essentially eliminated. Effects within the quartz or related to the quartz film interface are pointed to by the process of elimination.

Investigations by an electron microprobe x-ray analyzer and neutron activation analysis indicate diffusion of gold into the quartz to a depth of one micron or more under clean, high-temperature processing conditions. Strain and mass transfer thus occur in a layer near the quartz surface. Copper and aluminum were found to have penetrated to lesser depths. Reduction in the metal transfer at the interface is expected to give improved aging performance with a high yield of units aging $< 2 \times 10^{-10}$/week.

Introduction

The wafers used for the experiments to be described were 5 MHz fundamental of natural quartz one-half inch in diameter. The shape was plano-convex with a 2.5 diopter curvature.

All electrode plating operations were done by evaporation in an oil-free vacuum system using a combination of ion and titanium sublimation pumping. Pressures in the 10^{-8} Torr range were easily obtained without extended bakeout.

The HC-6 type cold-weld holders were evacuated and baked before sealing using a similar vacuum system. Pressures in the sealing chamber of about 10^{-6} Torr were obtained during bakeout initially. A subsequent modification of the chamber and system lowered the pressure to about 1×10^{-7} Torr. The bakeout temperature was monitored by means of a thermocouple in contact with the cold-weld die. The thermocouple voltage was used as a reference to control the temperature of the bakeout oven which was external to the vacuum chamber to maintain it at the desired temperature. Fail-safe controls made possible extended bakeout periods during which the equipment could operate unattended. Sealing was normally done while the units were at the bakeout temperature. All of the aging measurements were made at 85° C.

Typical Aging Characteristics

1. Aluminum-Plated Units

Two characteristic aging trends were observed. Figure 1 shows a typical aging curve for a unit baked 60 minutes at 125° C. A unit of the same group baked 120 minutes at 125° C aged as shown in Figure 2. However, increasing the bakeout time and temperature to 22 hours at 350° C either activated the alumina surface or caused a reaction at the quartz-aluminum interface resulting in the aging behavior shown in Figure 3.

Early in the program a weakness was discovered in aluminum-plated resonators. They aged very poorly at low temperatures. Units which aged < 1 ppl10^{-9}/week at 85° C aged > 1 ppl10^{8}/week in a negative direction at 45° C. This observed aging behavior thus supported the theory that the principal aging mechanism of aluminum-plated units is due to gas adsorption by the electrodes since adsorption is known to be inversely related to temperature.

2. Copper-Plated Units

Some of the best units with respect to aging have been ones plated with copper. Figure 4 displays data for a good copper-plated resonator. This unit was evacuated and baked 22 hours at 250° C before sealing. The aging rate of this unit was essentially zero for a 23 week period. Copper thermocouple wire of high purity was used to plate the unit data for which are shown in Figure 4. A later group of units was plated with copper from another source and the aging data shown in Figure 5 were obtained. A spectrographic analysis showed the second copper specimen to be of lower purity than the first and that it contained about 100 ppm each of iron and aluminum.

3. Silver-Plated Units

All of the silver-plated units studies had an undercoat of evaporated chromium. Thus, the aging rates were probably not those which would have been obtained by silver alone.

Resonators plated with the chromium plus silver combination failed to develop a consistent and predictable aging pattern. For example, unit 32-2, the data for which are shown in Figure 6, was an excellent crystal. Unit 32-4 on the other hand aged as shown in Figure 7. The reason for the difference in the behavior of the two resonators cannot be explained satisfactorily at this time; the two crystal wafers were plated at the same time and baked out for the same time and temperature before sealing. The negative aging pattern for 32-4 appeared to predominate among the units of group No. 32, however.

4. Units Final Plated with Electroplated-Nickel

Resonators plated with Cr + Au and with Cr + Ag were electroplated with nickel as the final plating step. A study of the effect of this action was made since frequency adjustment by electroplating nickel is commonly used in the industry. Unit 29-2 base-plated with Cr + Au was plated with nickel using a Watts plating solution. The plate back was 3277 Hz. The aging of the unit after bakeout and sealing is shown in Figure 8. The very high initial aging is typical of units plated with electroplated-nickel and is due to relaxation of the strain inherent in such films.[1] Unit 31-2, base-plated with Cr + Ag, was nickel-plated in the same manner as unit 29-2. The resulting aging data are shown in Figure 9. The addition of nickel-plating to Cr + Ag base-plating improved the performance of the units over those plated with Cr + Ag alone. The aging behavior was consistent and predictable with aging rates after 60 days in the range of a few parts in 10^9/week fairly common.

5. Gold-Plated Units

Much of the work in the research discussed has been devoted to improving the aging performance of gold-plated units.

Units plated with single films of gold and baked out for only one hour at 125° C usually aged in a manner similar to that shown in Figure 10. Increasing the bakeout temperature from 125° C to 250° C gave no improvement in the aging rate as shown in Figure 11, but rather increased it.

Since gold has poor intrinsic adherence to quartz a number of units were base-plated with Cr + Au. Aging as shown in Figure 12 was typical of these units when baked out for no more than 2 hours at 200° C.

The gold-plated resonators discussed thus far were baked out in the unmodified vacuum baking system. Pressures were about 1×10^{-6} Torr during bakeout.

The earlier experiments with gold-plated units were repeated after the modifications of the system were made to give lower bakeout pressures and long bakeout times. Figure 13 shows data for a typical Au plated unit having Cr + Au bonding tabs and single layer gold electrodes. Here the bakeout time was 22 hours at 350° C. The Cr + Au bonding tabs were used to improve the ultrasonic bonding properties and were well removed from the most active region of the quartz. Thus they are not believed to contribute in any large degree to the observed aging. Unit 34-1 base-plated with Cr + Au gave the data shown in Figure 14. Here again the initial aging was highly positive.

Unit 41-3 was evacuated for 70 hours, over one weekend, and then sealed without a bakeout. The aging rate of this unit after 120 days, as shown in Figure 15, was only slightly greater than units of the same series which were baked in vacuum before sealing.

As we improved our fabrication capability and techniques to include a lower bakeout pressure at longer times and higher temperatures the initial positive aging of gold-plated units became greater although the final rates were sometimes improved. This action could either be due to the removal of negative aging effects or to the addition of ones causing positive aging or some combination thereof.

All of the units discussed thus far were mounted in tab clips and ultrasonically bonded using aluminum foil spot-welded to the clip. A series of experiments were undertaken in which the gold-plated wafer was supported at only one point, the bottom; this was accomplished by means of a slot (or hook) mount spot-welded to the HC-6 base to which the wafer, positioned in the slot, was cemented with a very small portion of Duramic cement (non-conducting). The foil leads were then welded to the prepared bonding sites of Cr + Au. Thus the wafer is very nearly strain-free, its weight being supported by the single slot mount. The possibility of flexing or bending due to poor tab clip alignment was removed and other undesirable forces reduced to a minimum. Typical aging of a unit of this type is shown in Figure 16. Since

the aging was not improved, the positive aging of gold-plated units must not be caused by mounting strain to any appreciable degree.

After the research discussed it appeared that a reduction of the initial high positive aging rate of gold-plated units and the subsequent moderately high one was not to be found by improving the vacuum, increasing bakeout temperature, or reducing mounting strain. In fact, each change designed to reduce the aging rate seemed to increase the initial rate without particular benefit at the end of a three-month period.

Frequency Retrace Studies

Frequency "retrace" studies were made by first establishing an aging rate for a selected unit at a selected temperature. The oven controlling the crystal temperature was then turned off for a period of time. The frequency "retrace" was the subsequent aging performance of the unit after the oven had been returned to the original operating temperature.

Byrne and Hokanson[2] have shown that resonators baked out at low temperatures essentially "retraced" the original aging curve. Those baked out at high temperatures exhibited only small frequency shifts and after restart returned to the previously established aging slope very quickly.

The time for a crystal unit to return to its original aging rate after oven restart is considered to be an indication of the quality of the fabrication process employed. The shorter the time, the better the process.

Frequency retrace studies were made in this work by turning the ovens off at 1600 hours, waiting 48 hours, and then turning them back on. The oven was then allowed 16 hours to return to the operating temperature of 85° C before reaging measurements were started.

Typical retrace aging for a unit plated with Cr + Au electrodes is shown in Figure 17. The first frequency measurement after oven restart was exactly on the projected aging curve. Subsequent measurements deviated but slightly from the projected track due probably to small oven temperature fluctuations.

The most obvious retrace aging obtained for any of the units studied was for unit 38-5, a copper-plated unit with Ti + Au bonding tabs. The retrace data are illustrated in Figure 18. Note that the unit was given a 21 hour bakeout at 250°C then backfilled to 100 μm with nitrogen while still hot before sealing.

Studies of the Diffusion of the Metal Electrodes into the Quartz

An electron microprobe x-ray analyzer* was available at Georgia Tech and studies were begun about 1 February 1969 to determine if the electrode metal diffused into the quartz during plating, bakeout and subsequent operation at the aging temperature.

The electron-microprobe is a device designed to analyze the chemical composition of a solid material by using the solid as the target of a precisely focused electron beam. The x-rays from the solid are diffracted from a suitable crystal and analyzed according to wavelength and count rate. The detector is a gas-flow proportional counter. The count rate at an established voltage for a specific wavelength is proportional to the amount of a given element in the material under examination. The electron spot size can be varied. However, a spot 50 μm in diameter was used for our work to give a more representative sampling of the material. The penetration of the beam into the quartz plates was about 3 μm at 20 kv. Thus the composition of the surface zone, rather than that of the bulk material, was revealed. The electron beam parameters used were 20 kv at 0.3 μa. Counts at selected sites on the quartz wafer were taken for 10 seconds; heating effects with the large beam cross-section were minimal.

The first units selected for study were about 4 years old. They had been aged at 85° C for at least 2 years and then stored at room temperature. The wafers were removed from the HC-27/U holders and the gold electrodes stripped using ultrasonically agitated aqua regia. No visible traces of the electrodes were seen on the polished quartz wafers after stripping. The wafers were then examined using the microprobe as follows: (1) an average background count at the edge of the wafer where no gold had been deposited was obtained (5 sites were counted); (2) an average gold count at 10 randomly selected sites in the previously electroded area was made. The wafer was then etched 10 minutes and re-examined with the probe in the manner described. The examination and etch steps were continued until the gold count in the electrode region reached the background level, the limit of detection. The frequency of the wafer was measured after stripping the electrode and after each etching step. The etch rate was calculated to be about 200 Å/minute using a saturated solution of ammonium bifluoride as the etchant. Figure 19 shows the result of the experiment. Ten thousand angstroms of quartz were removed from the evaporated gold-plated unit (No. 2-5) before the gold count reached the background. The amount of quartz removed from the sputtered gold unit was 8000 Å.

A unit plated with aluminum was similarly studied after removing the electrode with HCl. The data obtained are shown in Figure 20. Here 7000 Å of quartz was removed before the background level was reached. The initial count before etching was about three times the background count, approximately the same as for the gold-plated units. The data presented above and in the following figures are semi-quantitative in that the relative

*Acton Model MS-64.

proportion of the material present in the bombarded zone is reflected by the count for the material at constant electron beam parameters. More precise quantitative calculations were not attempted in the initial work. There is no direct relation between the amount of gold and aluminum in the quartz matrix based upon the count rate. However, quantities of these materials in the basic quartz were determined by neutron activation analyses of several specimens and are discussed near the end of this paper.

Two copper-plated units were treated with HNO_3 to remove the copper and examined with the microprobe for the presence of copper before and after successive etching. The penetration depth of copper into quartz was found to be higher for the unit baked 20 hours at 350° C than for the unit sealed without bakeout as illustrated by Figure 21.

As a final experiment two wafers of polished quartz were cleaned and plated on one side only with 1500 Å of evaporated gold while at a substrate temperature of 200° C. One of the wafers (Au-2) was then vacuum-baked 24 hours at 300° C. Both wafers were stripped and microprobe studies made. The results of this study are illustrated in Figure 22. The depth at which gold was found in the quartz was very similar for the two units, about 5000 Å, indicating that a considerable penetration of the quartz occurred during the plating.

Neutron activation analyses of the wafers, the electrode diffusion data which were presented in Figures 19 and 20 were made after the etching had been completed. In addition, two unplated and unetched wafers were also selected for study. The results are shown in Table 1. The bulk rather than the surface analysis of the quartz is presented by this method. Note especially the relatively large amount of gold in the wafer previously coated with sputtered gold. The total aging and final aging rate of sputtered gold wafers was higher than that obtained for evaporated gold.

Table 1

Neutron Activation Analysis of Quartz Samples

Wafer No.	Plating	Elements Present (ppm)		
		Au	Al	Cu
No.-1	none	0.01	101	<0.10
No.-2	none	<0.01	62	<0.10
8-13	Sp. Au	0.11	92	<0.10
2-5	Evap. Au	0.008	94	<0.10
1-1	Evap. Al	0.05	90	0.53

It appears that the positive aging of gold-plated units may be caused in part by the relief of the stress created by the diffusion of the gold into the quartz. Whether a shift of mass or a change in the elastic constants at or near the quartz surface is likewise partially responsible has not been determined. It would be desirable to deposit a barrier on the quartz surface to block the diffusion of the electrode metal into the quartz. However, a suitable barrier material must first be selected and carefully evaluated to prove its merit.

We have barely "scratched the surface" with these studies of the diffusion of the electrode metal into the quartz wafer. However, the pursuit of similar studies may furnish information that will allow resonators to be fabricated by predictable methods based on knowledge of all important factors involved rather than by a recipe derived from trial and error which permits a reasonable yield of acceptable units but which may result in resonators of poor and unexplicable aging behavior.

References

1. R. B. Belser and W. H. Hicklin, "Quartz Crystal Aging Effects," Interim Report No. 8 on Contract No. DA 36-039 AMC-02251(E), U.S. Army Electronics Laboratory, Fort Monmouth, New Jersey, 15 February 1965.

2. R. J. Byrne and J. L. Hokanson, "Effect of High-Temperature Processing on the Aging Behavior of Precision 5 MHz Quartz Crystal Units," IEEE Transactions on Instrumentation and Measurement IM-17, No. 1, pp. 76-79 (1968).

Conclusions

Quartz plates fabricated into resonators under clean conditions and subjected to high temperatures during plating, subsequent bakeout, or aging are penetrated by atoms of the metal electrodes to depths of 0.5 to 1 micron under procedures we have employed. Even plating with gold or copper at a substrate temperature of 200°C appeared to result in penetration to depths of 0.4 micron. The degree of penetration of the quartz by the metal atoms and the accommodation of the atoms by the lattice or imperfections in it are specific to the metal atom-lattice parameters and to the degree of perfection of the quartz. The direction of the resulting frequency shift is dependent on the atom accommodation method, any strain effects resulting, elastic property changes, and moment of inertia changes of the plate due to lattice growth, shrinkage, or mass transfer within the affected zones. A mechanism of the nature described will explain many current aging phenomena considered anomalous. Its consideration and proper control will allow improvement in the ultimate stabilities of quartz resonators and the attainment of projected aging requirements of 2×10^{-11} per day for resonators of 5 MHz fundamental frequency.

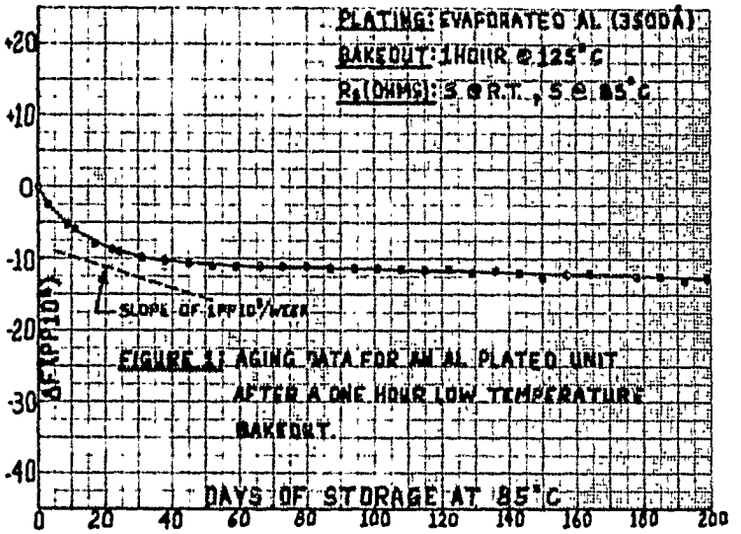


FIGURE 1: AGING DATA FOR AN AL PLATED UNIT AFTER A ONE HOUR LOW TEMPERATURE BAKEOUT.

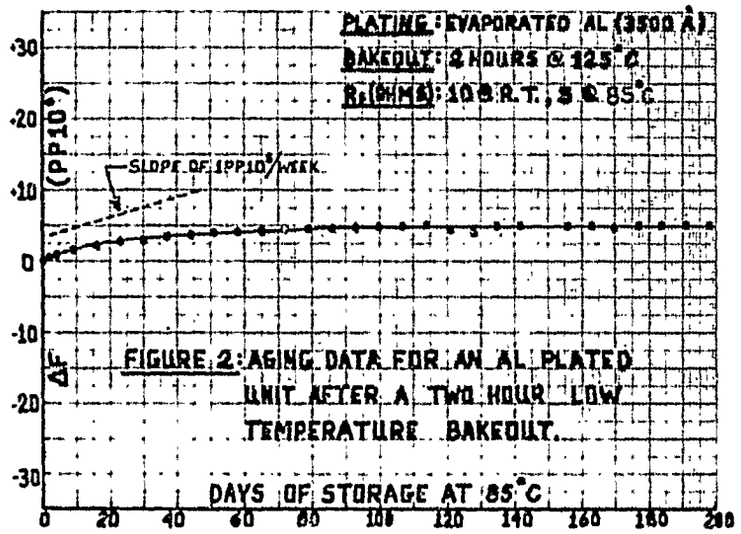


FIGURE 2: AGING DATA FOR AN AL PLATED UNIT AFTER A TWO HOUR LOW TEMPERATURE BAKEOUT.

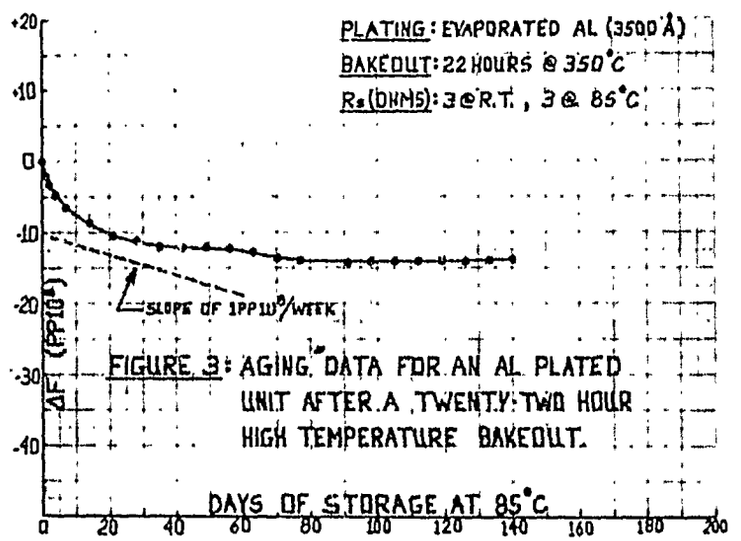


FIGURE 3: AGING DATA FOR AN AL PLATED UNIT AFTER A TWENTY-TWO HOUR HIGH TEMPERATURE BAKEOUT.

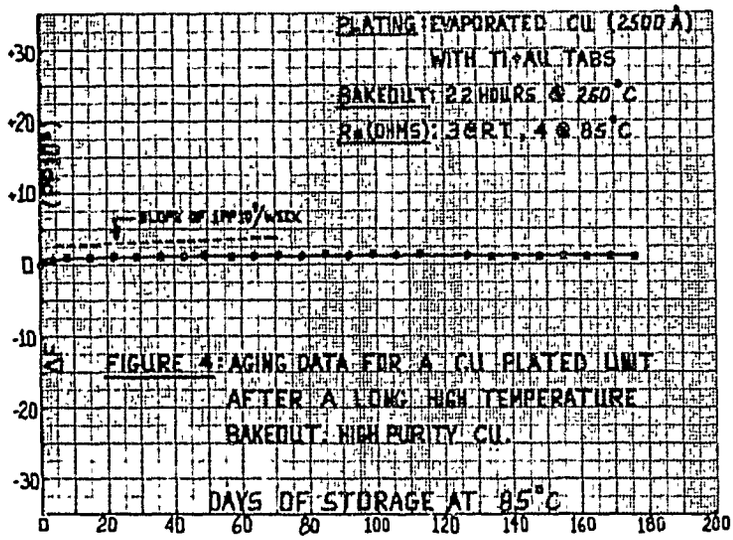


FIGURE 4: AGING DATA FOR A CU PLATED UNIT AFTER A LONG HIGH TEMPERATURE BAKEOUT: HIGH PURITY CU.

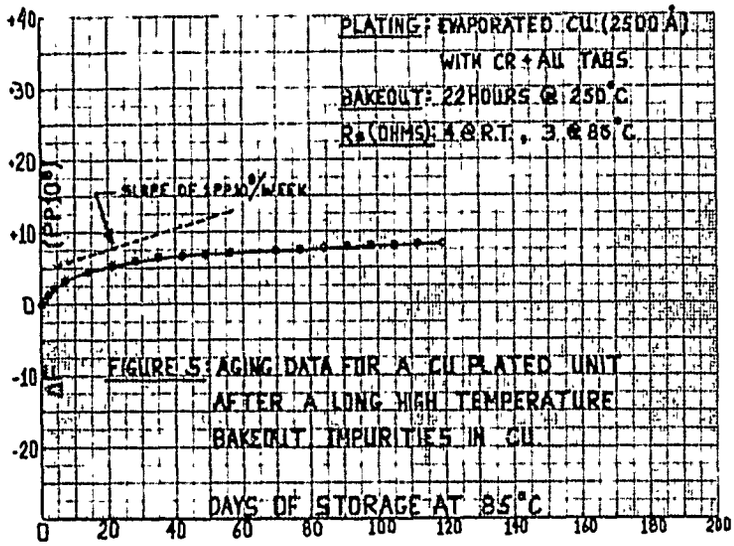


FIGURE 5: AGING DATA FOR A CU PLATED UNIT AFTER A LONG HIGH TEMPERATURE BAKEOUT IMPURITIES IN CU.

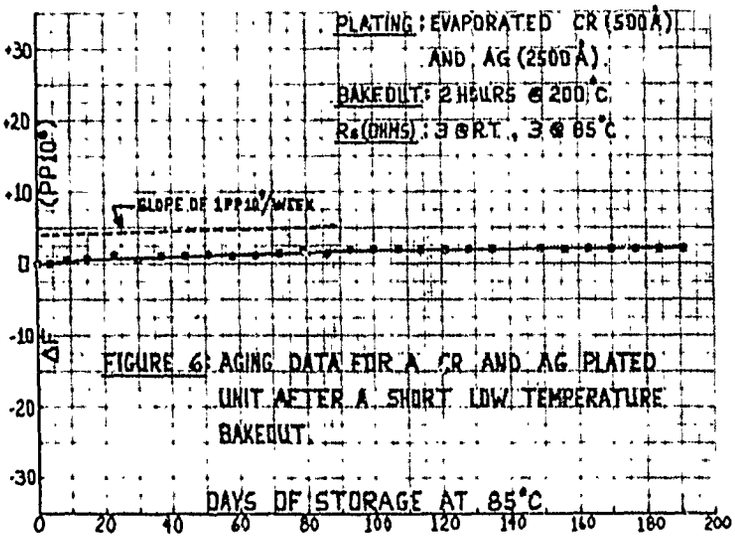


FIGURE 6: AGING DATA FOR A CR AND AG PLATED UNIT AFTER A SHORT LOW TEMPERATURE BAKEOUT.

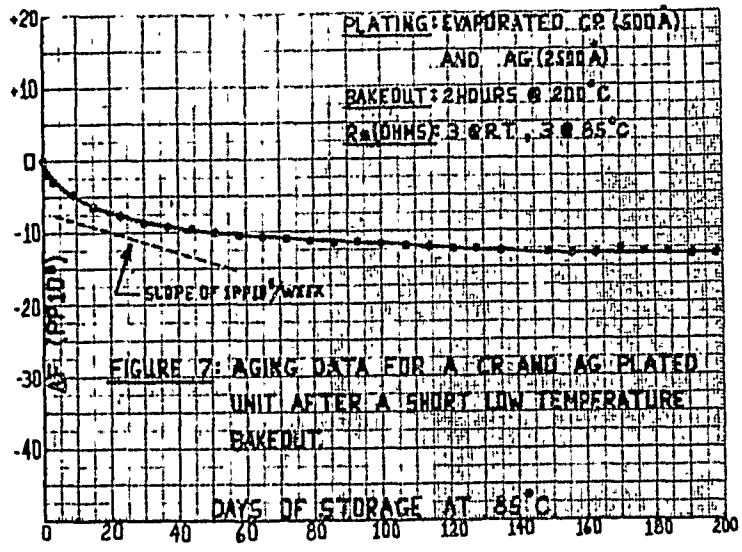


FIGURE 7: AGING DATA FOR A CR AND AG PLATED UNIT AFTER A SHORT LOW TEMPERATURE BAKEOUT.

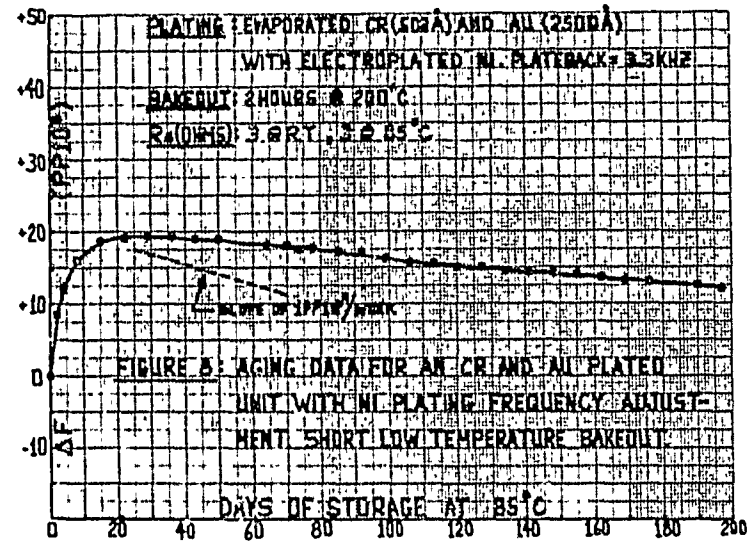


FIGURE 8: AGING DATA FOR AN CR AND AU PLATED UNIT WITH NI PLATING FREQUENCY ADJUSTMENT SHORT LOW TEMPERATURE BAKEOUT.

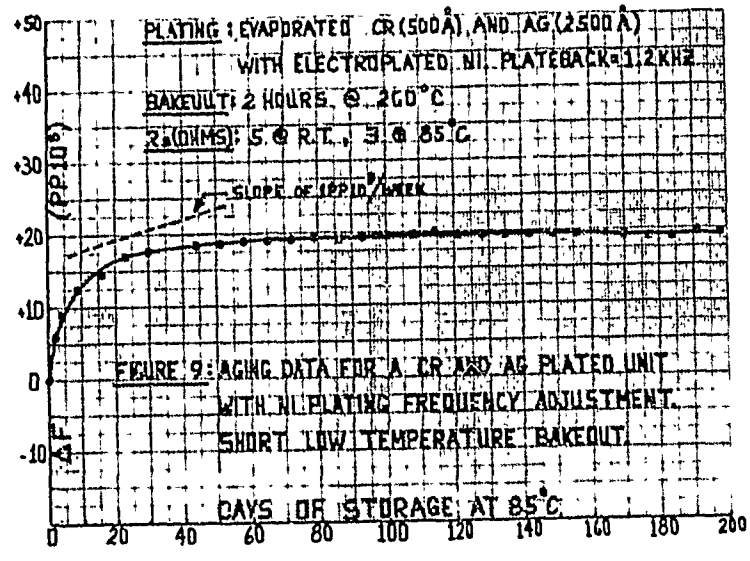


FIGURE 9: AGING DATA FOR A CR AND AG PLATED UNIT WITH NI PLATING FREQUENCY ADJUSTMENT. SHORT LOW TEMPERATURE BAKEOUT.

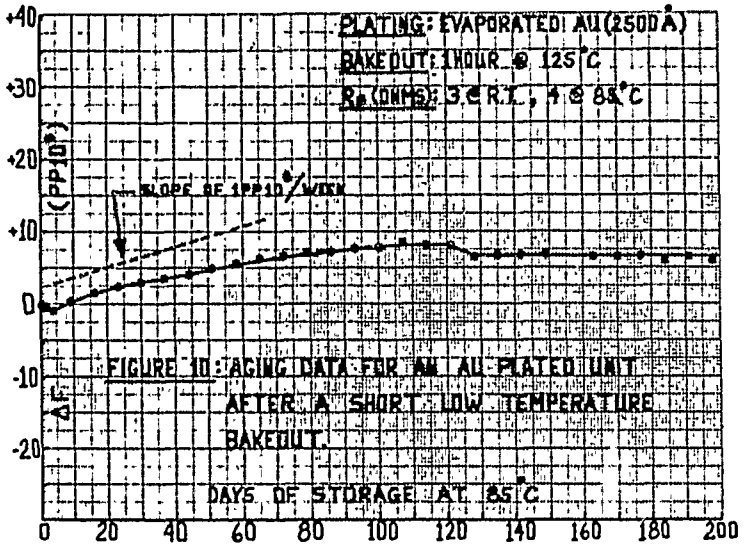


FIGURE 10: AGING DATA FOR AN AU PLATED UNIT AFTER A SHORT LOW TEMPERATURE BAKEOUT.

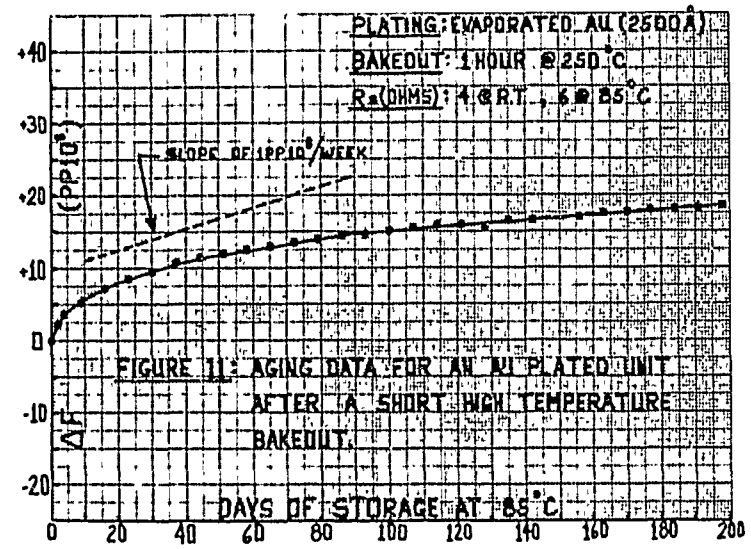


FIGURE 11: AGING DATA FOR AN AU PLATED UNIT AFTER A SHORT HIGH TEMPERATURE BAKEOUT.

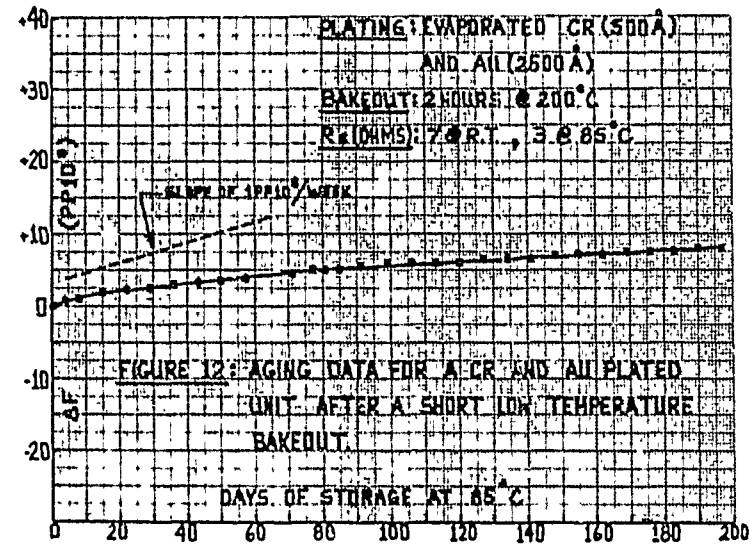


FIGURE 12: AGING DATA FOR A CR AND AU PLATED UNIT AFTER A SHORT LOW TEMPERATURE BAKEOUT.

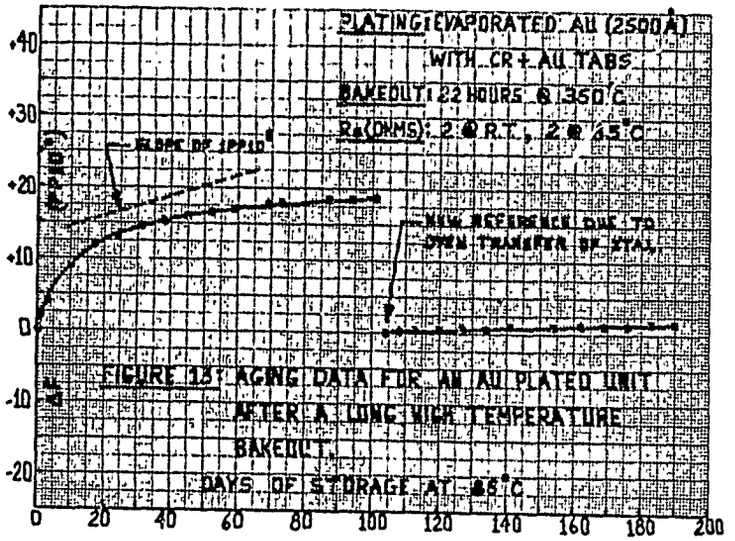


FIGURE 13: AGING DATA FOR AN AU PLATED UNIT AFTER A LONG HIGH TEMPERATURE BAKEOUT.

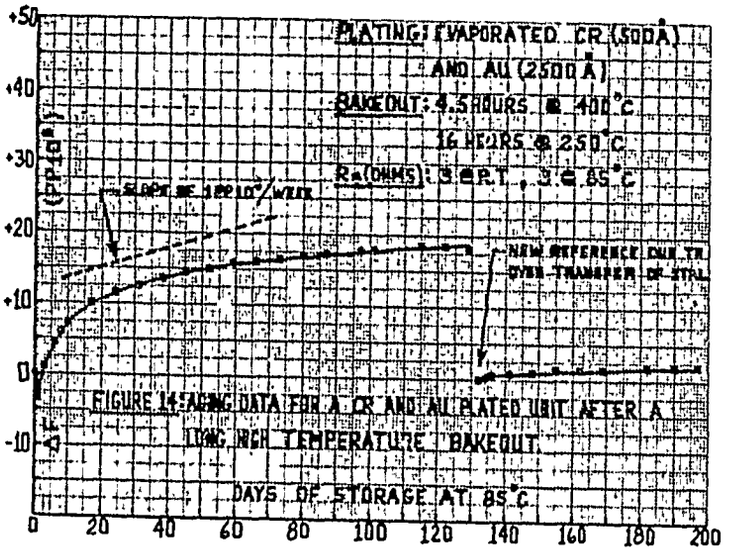


FIGURE 14: AGING DATA FOR A CR AND AU PLATED UNIT AFTER A LONG HIGH TEMPERATURE BAKEOUT.

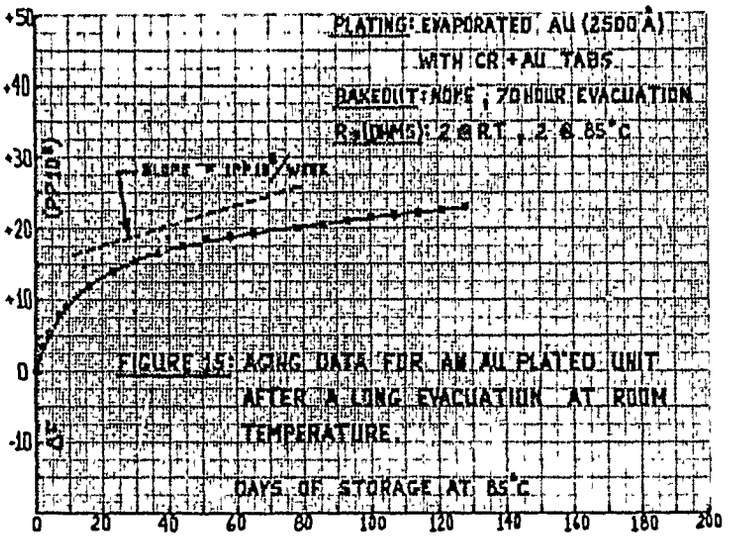


FIGURE 15: AGING DATA FOR AN AU PLATED UNIT AFTER A LONG EVACUATION AT ROOM TEMPERATURE.

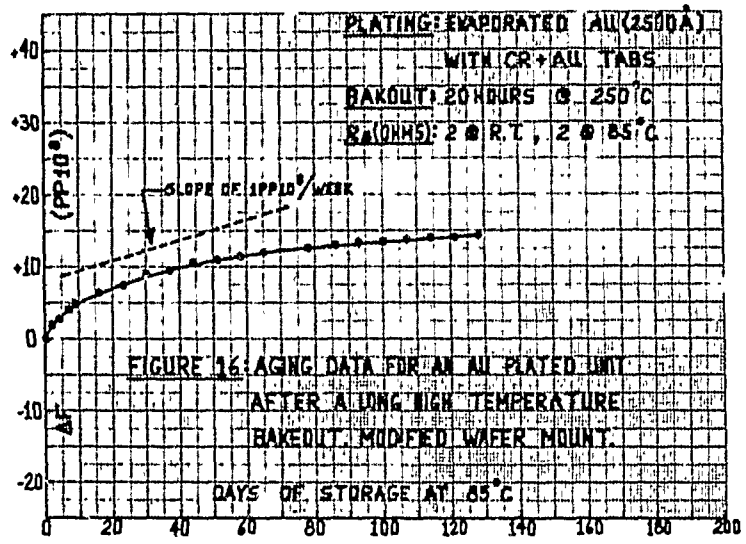


FIGURE 16: AGING DATA FOR AN AU PLATED UNIT AFTER A LONG HIGH TEMPERATURE BAKEOUT. MODIFIED WAFER MOUNT.

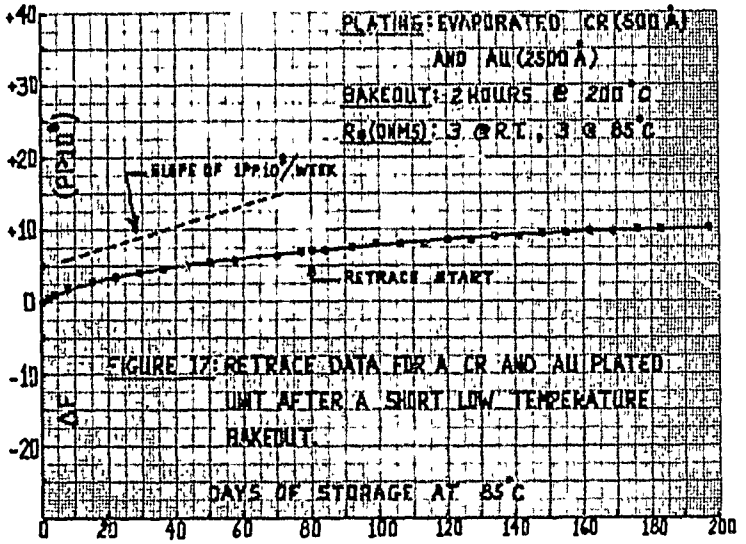


FIGURE 17: RETRACE DATA FOR A CR AND AU PLATED UNIT AFTER A SHORT LOW TEMPERATURE BAKEOUT.

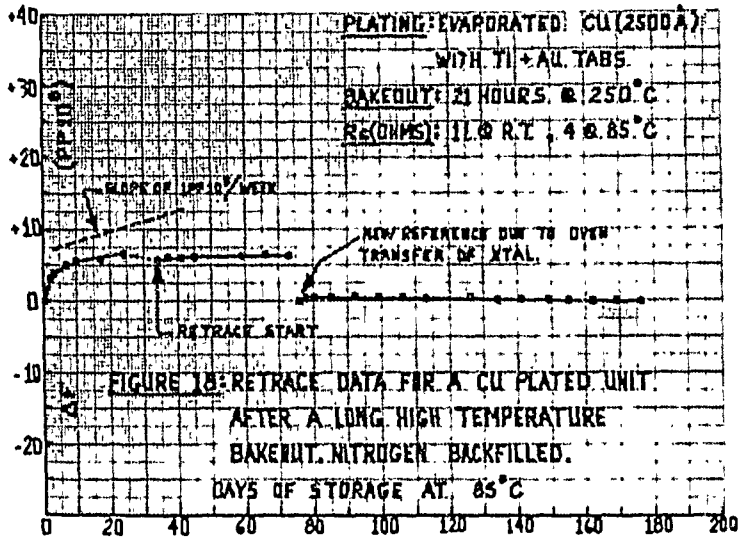


FIGURE 18: RETRACE DATA FOR A CU PLATED UNIT AFTER A LONG HIGH TEMPERATURE BAKEOUT. NITROGEN BACKFILLED.

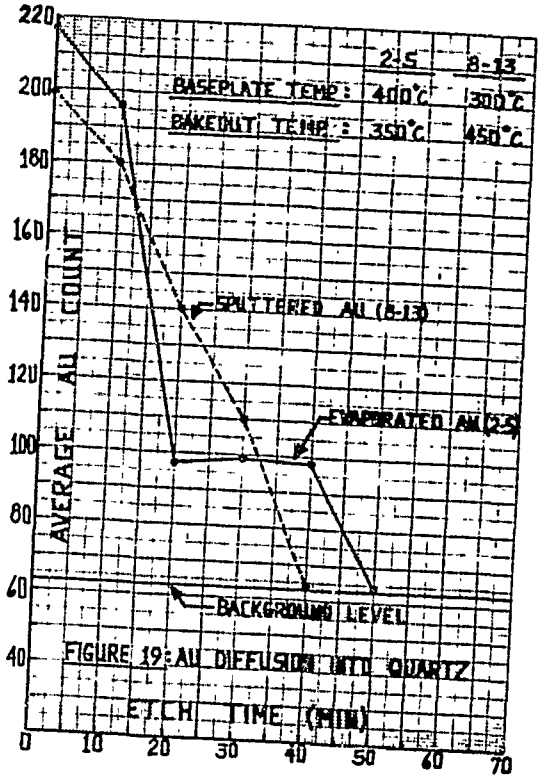


FIGURE 19: Al DIFFUSION INTO QUARTZ

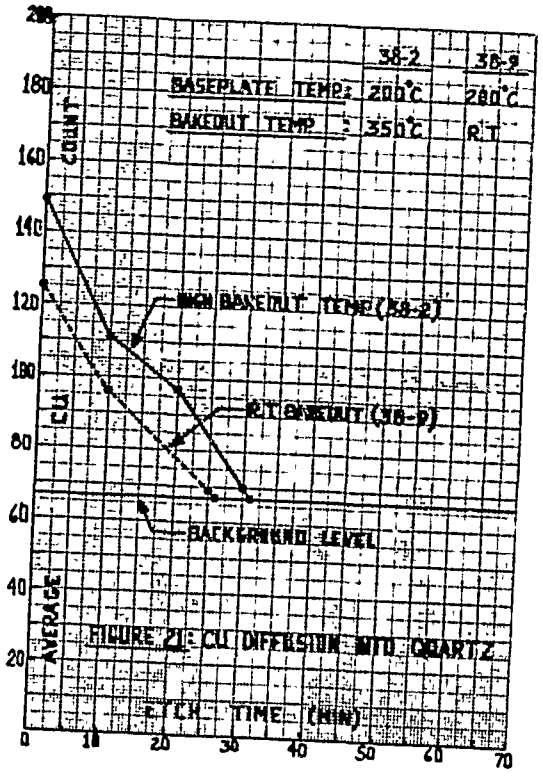


FIGURE 21: Cu DIFFUSION INTO QUARTZ

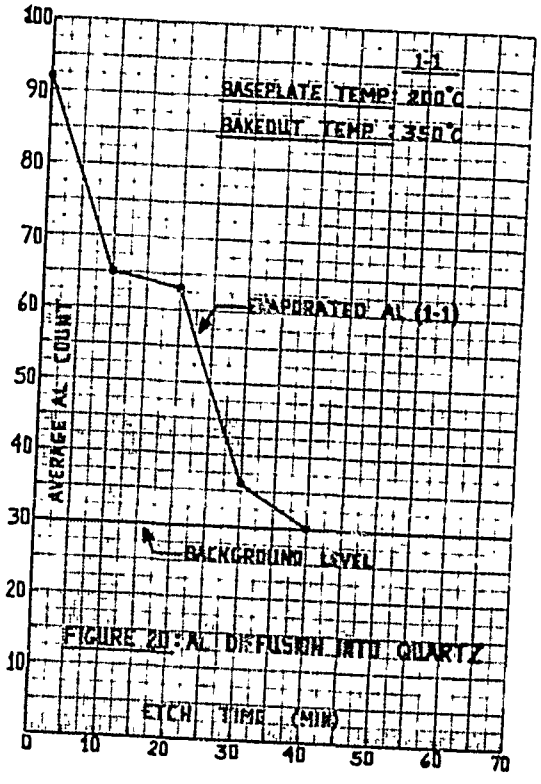


FIGURE 20: Al DIFFUSION INTO QUARTZ

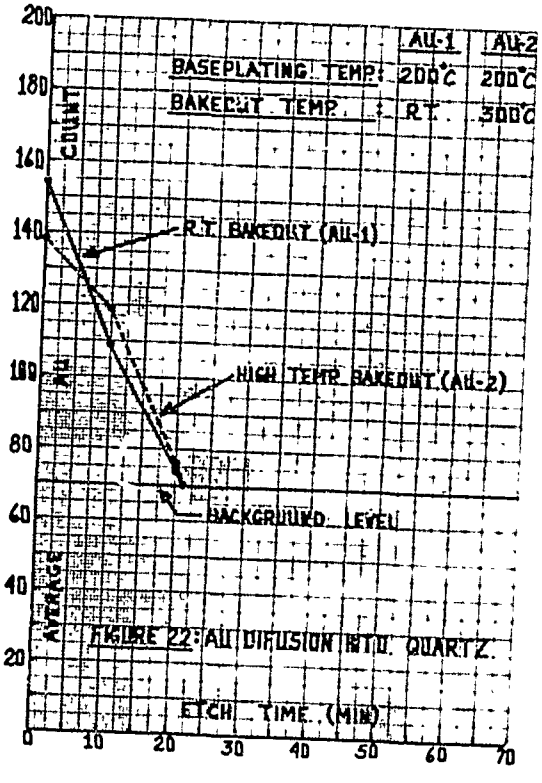


FIGURE 22: Au DIFFUSION INTO QUARTZ

A NOVEL ALGORITHM FOR THE DESIGN OF THE ELECTRODES OF SINGLE-MODE AT-CUT RESONATORS

John H. Sherman, Jr.

General Electric Company
Mobile Radio Department
Lynchburg, Virginia

Summary

The paper reviews the basic concepts of energy trapping from the standpoint of physical intuition. The equation of Onoe and Sykes, derived to calculate the responses of a simple rectangular AT-cut plate, is used to generate relationships expressing the effect of simultaneously bounding the X and Z' extensions of the electrode, leading directly to the evaluation of rectangular electrode configurations.

Criteria for optimizing rectangular electrode configurations are considered, resulting in an X:Z' ratio of 1.1226, the square root of the value proposed elsewhere in the literature. The design equations derived calculate actual resonator frequencies as well as the characteristic frequencies. The plateback constant calculated is twenty to fifty percent larger than that found by other investigators. The derivation presents a previously unexplicated criterion for minimum plateback.

Electrode thickness is considered per se. This results in a means to determine requirements for absolute parallelism of the blank. The stringency of parallelism requirements for realistic design objectives is illustrated.

A 13 step design algorithm for single-response AT resonators is stated.

The mathematics required is algebra.

Introduction

Six years ago, during the 1963 Symposium, we all listened, deeply impressed because we sensed that what we were hearing was of the greatest importance, as Dr. Shockley[1] offered the first explanation most of us had ever heard why some electroded AT-cut resonators display only one resonance while others display very complex spectra.

His approach was to solve the wave equation for thickness dependent plane waves in an isotropic medium with a discontinuity in velocity parallel to the wave front and discover the laws of reflection at the discontinuity. Altogether it was rather heady stuff and we all took notes so we could try it out as soon as we got home. And, of course, it didn't work quite as advertised. In the intervening years we have had substantial clarification and explanation and some very vigorous exploitation of the ideas has occurred.[2,3,4,5,6] It is doubtful that the coupled-mode monolithic filter would have become the important device it is without this work.

Among the significant refinements soon upcoming were modification of the results to account for the anisotropic medium[7] and the clarification of the concepts of symmetric/antisymmetric unwanted responses. It had been long recognized that there was a close relationship between the unwanted modes of the electroded resonator and the enharmonic modes of the whole-plate resonator investigated by Onoe and by Sykes some thirty years ago. The mathematical approach employed has not made it explicit that the ideas are identical, but equating the concepts made possible the rapid exploitation of the ideas in the Toyo[8] series of filters of cascaded coupled-mode dual quartz resonators. Equating the concepts made it possible for Beaver[9] to evaluate the coupling coefficient between two resonators in the design of the general monolithic filter. Equating the concepts made it possible for Mortley to design single response resonators for frequency modulated oscillators in the middle 1950's.

One of the most humbling of the things brought forward was that essentially the same ideas had been not only sorted out in successful detail by Mortley[10,11] but also published seven years earlier, only to be ignored by the entire frequency control fraternity. This kind of oversight has its precedents, of course, in dismaying number. Among them are the demonstration of wireless communication[12] in 1866 by Mahlon Loomis and the invention of the junction transistor[13] and the back-biased diode varactor in 1925 by Julius Lilienfeld.

Failure to make secure identification of the well-understood enharmonic mode with the unwanted mode has allowed some significant amounts of effort to be less than ideally effective. What is to follow is the consequence of making the absolute identification of the mode spectrum of the electroded resonator with that of the unelectroded AT-cut blank of the older literature. The approach is very similar to Mortley's, but is carried through to the form of design equations which Mortley considered an impractical hope. Some of the resulting magic numbers are unfamiliar. What is reassuring is that in form the equations are as they ought to be, that some problems inadequately handled previously yield readily to this approach, and that the approach requires only a reasonably sophisticated physical intuition and simple algebra to derive its results. There isn't a derivative or a matrix in the whole bit.

Enharmonic Frequency Spectrum

Before writing the equation of Onoe and Sykes certain peculiarities of notation will be explained. The thickness of the crystal, in the y' direction, will be designated as lower case t; the extension of the resonator in the x- direction will be designated by upper case X while the direction will be mentioned in lower case; similarly the extension in z' will be expressed in upper case and the direction itself in lower case. Furthermore, after this point no axis will be identified as primed, z' will be called z and the extension will similarly be unprimed. It will be convenient to use the subscript lower case m in three senses, (1) thickness mode index, (2) modifier to signify minimum value with upper case signifying maximum, and (3) modifier of the density symbol ρ (rho) to signify metal density in contrast with quartz density which is not subscripted.

The equation of Onoe and Sykes is:

$$f_{mnp} = \frac{1}{2\sqrt{\rho}} \sqrt{\frac{C_{66m}^2}{t^2} + \frac{C_{11n}^2}{X^2} + \frac{C_{55p}^2}{Z^2}} \qquad \text{Eq. 1}$$

This is an approximate* solution with distinct limitations well discussed and acknowledged in the literature. It is most nearly valid at low orders of the lateral (enharmonic) overtones as it is employed in this paper.

If we extract from this the frequency

$$f_{moo} = \frac{m\sqrt{C_{66}}}{2t\sqrt{\rho}} \qquad \text{Eq. 2}$$

we recognize the familiar form

$$f_{moo} = \frac{m\,K_f}{t} \qquad \text{Eq. 2a}$$

where $$K_f = \frac{1}{2}\sqrt{\frac{C_{66}}{\rho}} \qquad \text{Eq. 2b}$$

When the standard values of C_{66} and ρ are substituted a numerical value of K_f is found:

$$K_f = 165.1 \times 10^3 \text{ Hz cm.}$$

$$= 65.0 \times 10^3 \text{ Hz in.}$$

Physically f_{moo} is identified as the frequency of an infinite plate and is measured as the frequency of a plate immersed in a parallel electric field formed by electrodes larger than the plate. Since the dielectric constant of quartz is reasonably high the approximation is quite good.

The equation can be rewritten by factoring under the radical:

$$f_{mnp} = \frac{1}{2\sqrt{\rho}} \sqrt{\frac{C_{66m}^2}{t^2} + \frac{C_{11n}^2}{X^2} + \frac{C_{55p}^2}{Z^2}} \qquad \text{Eq. 1}$$

*Epilogue 2

$$= \frac{m}{2t}\sqrt{\frac{c_{66}}{\rho}}\sqrt{1 + \frac{t^2}{c_{66m}^2}\left(\frac{c_{11n}^2}{x^2} + \frac{c_{55p}^2}{z^2}\right)} \qquad \text{Eq. 1a}$$

$$= f_{moo}\sqrt{1 + \frac{t^2}{c_{66m}^2}\left(\frac{c_{11n}^2}{x^2} + \frac{c_{55p}^2}{z^2}\right)} \qquad \text{Eq. 1b}$$

This form of the equation makes it explicit that the frequency of the resonator is higher than the frequency we normally consider to be the "blank frequency." As a matter of fact, there is a family tree of frequencies all standing above fmoo coming from the possible values of n and p. These are the enharmonics.

Another valuable manipulation of the equation is made by factoring in the parentheses under the radical:

$$f_{mnp} = f_{moo}\sqrt{1 + \frac{t^2}{c_{66m}^2}\left(\frac{c_{11n}^2}{x^2} + \frac{c_{55p}^2}{z^2}\right)} \qquad \text{Eq. 1b}$$

$$= f_{moo}\sqrt{1 + \frac{t^2\, c_{55}}{z^2 c_{66m}^2}\left(\frac{c_{11}\, z^2 n^2}{c_{55}\, x^2} + p^2\right)} \qquad \text{Eq. 1c}$$

Now, if the policy decision is made to proportion the blank appropriately the enharmonic spectrum is substantially simplified. Let the resonator be proportioned so that:

$$\frac{c_{11}\, z^2}{c_{55}\, x^2} = 1 \qquad \text{Eq. 3}$$

or $$x = z\sqrt{\frac{c_{11}}{c_{55}}} \qquad \text{Eq. 3a}$$

The full equation then becomes:

$$f_{mnp} = f_{mpn} = f_{moo}\sqrt{1 + \frac{c_{55}\, t^2\, (n^2 + p^2)}{c_{66}\, z^2\, m^2}} \qquad \text{Eq. 4}$$

Again, this simplification is convenient and has some very desirable implications. It is not necessary, and perfectly adequate resonators can be designed with other proportions. In the language of the textbooks, if other ratios are desired the details are left as an exercise for the reader.

As is well known, if the plate is electroded then all three indices must be odd numbers. The sequence** of resonances, then, becomes

$$f_{m11} = f_{moo}\sqrt{1 + \frac{2\, c_{55}\, t^2}{m^2\, c_{66}\, z^2}} \qquad \text{Eq. 5a}$$

$$f_{m13} = f_{moo}\sqrt{1 + \frac{10\, c_{55}\, t^2}{m^2\, c_{66}\, z^2}} \qquad \text{Eq. 5b}$$

$$f_{m33} = f_{moo}\sqrt{1 + \frac{18\, c_{55}\, t^2}{m^2\, c_{66}\, z^2}} \qquad \text{Eq. 5c}$$

$$f_{m15} = f_{moo}\sqrt{1 + \frac{26\, c_{55}\, t^2}{m^2\, c_{66}\, z^2}} \qquad \text{Eq. 5d}$$

$$f_{m35} = f_{moo}\sqrt{1 + \frac{34\, c_{55}\, t^2}{m^2\, c_{66}\, z^2}} \qquad \text{Eq. 5e}$$

$$f_{m55} = f_{moo}\sqrt{1 + \frac{50\, c_{55}\, t^2}{m^2\, c_{66}\, z^2}} \qquad \text{Eq. 5f}$$

etc.

So far, nothing is novel. This is all standard in the literature and in the art. It is, however, the necessary prelude to the concept of energy trapping as considered in this exposition.

Stored and Trapped Energy

Energy is stored in all excited resonant structures. This storage is not perfect in any ordinary resonator; there is a steady leakage of energy from the resonator. This leakage is called dissipation and its reciprocal is the quality factor Q of the resonator. Q is mentioned here specifically to point out that nowhere else in this discussion is the Q concept to be offered as an explanation of any phenomenon considered.

Energy is said to be trapped if, due to some peculiarity of a resonant system, the site of the storage is only a part of the entire system. The distinction between energy storage

**Epilogue 3

and energy trapping becomes of great significance in thickness mode quartz crystals with deposited electrodes or with thickness discontinuities. Trapped energy is stored energy, but energy stored only in regions defined by electroding or by thickness discontinuities. Other modes of energy storage exist which involve the remainder of the plate as well as the electroded region. We term these modes of storage "whole plate modes." This exposition will avoid the attempt to discuss whole plate modes in detail. Understanding of modes of energy trapping will provide the background of better understanding of whole plate modes as well.

We have described the resonant modes of an isolated rectangular AT plate. No discussion or means of supporting this isolated plate is made or implied in the equations. We will now propose one. We will postulate that the plate is supported by a web of crystalline quartz which is thinner than the resonator itself. This web has a characteristic frequency

$$f'moo = \frac{m}{2t'}\sqrt{\frac{c_{66}}{\rho}} \qquad \text{Eq. 2}$$

which, because the web thickness is postulated to be less than that of the resonator, is greater than fmoo, the characteristic frequency of the resonator. The web is bounded in some way, hence has its own set of resonances relating to its x and z extensions. All of these resonances are higher in frequency than f'moo.

If the imbedded resonator is excited by a complex signal -- if, for instance, it is swept in frequency -- all of the resonances of the resonator up to f'moo will be observed to be essentially the same as if the resonator were totally isolated. Above f'moo exciting the imbedded resonator couples into the web so that the resonances of the web can be excited. In terms of our definition of trapped energy given above, all resonant modes lower than f'moo in frequency are trapped energy modes trapped in the resonator while those higher than f'moo are generally associated with the entire plate including the web, hence not trapped.

This structure consisting of a region of quartz surrounded by a boundary of higher characteristic frequency is recognized to be the mesa structure discussed by the Clevite group. The derivation of the concept to this point in terms of thickness will now be generalized by avoidance hereafter of the implication of thickness or thickness difference. We will have occasion to speak explicitly of thickness when we discuss electroding and parallelism. The ratios Z/t and X/t are of course retained as parameters in the enharmonic equation. From this point on, the discussion will be based upon the two frequencies fmoo and f'moo alone. We will continue to assume that the resonator is rectangular in shape, oriented along the x and z axes and proportioned so that:

$$X = Z\sqrt{\frac{c_{11}}{c_{55}}} \qquad \text{Eq. 3a}$$

We will at this point add some terminology in order to clarify the discussion. We will designate our two characteristic frequencies with the terminology commonly used for them since Shockley's paper:

$$fmoo \triangleq fe, \qquad f'moo \triangleq fs$$

The Clevite group also used the symbol f_{te} for trapped energy frequency. There can be several frequencies of trapped energy in accordance with the following condensation of the discussion preceding:

$$fs > f_{te} = fe\sqrt{1 + \frac{n^2 + p^2}{m^2}\,\frac{c_{55}}{c_{66}}\,\frac{t^2}{Z^2}} > fe \qquad \text{Eq. 6}$$

The inequality on the right has been discussed already. The inequality on the left will concern us from here on out. It is obvious that fs can conceivably be sufficiently greater than fe that many values of n and p can be combined to satisfy the conditions, resulting in many values of f_{te}. We are concerned with the problem of defining conditions which will permit one and only one value of f_{te}, the "single mode criterion."

Single Mode Criterion

Earlier we tabulated in Equation 5 the formulas of several possible values of f_{te}. Reference to that table shows that the lowest of these, which we will designate fd, the designed frequency is Eq. 5a, which becomes:

$$fd = fm11 = fe\sqrt{1 + \frac{2\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}} \qquad \text{Eq. 7}$$

If the resonator is to display a single trapped response, then the next frequency Eq. 5b must be an unwanted frequency.

$$fu = fm13 = fm31 = fe\sqrt{1 + \frac{10\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}} > fs \qquad \text{Eq. 8}$$

The inequality on the right indicates that a signal at this frequency results in exciting the entire plate and not the electroded region alone. This frequency, therefore, would not be trapped. Also all higher frequencies would not be trapped as they would fail to satisfy the inequality and would excite the entire plate. Since, in general, the resonator region is only a small fraction of the entire plate, the electromechanical coupling of the electrode to a whole plate mode will tend to be small and the whole plate will usually, though not always, couple only small impedance fluctuations into the electrodes.

Since fs must be greater than f_{te} for f_{te} to exist, fs must be greater than fd. This is written:

$$fs = fe + \Delta f > fd = fe\sqrt{1 + \frac{2\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}} \qquad \text{Eq.9}$$

which becomes:

$$1 + \frac{\Delta f}{fe} > \sqrt{1 + \frac{2\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}} \qquad \text{Eq. 9a}$$

which in turn, by the standard approximation becomes:

$$1 + \frac{\Delta f}{fe} > 1 + \frac{C_{55}\ t^2}{m^2\ C_{66}\ Z^2} \qquad \text{Eq. 9b}$$

From this we conclude that

$$\frac{\Delta mf}{fe} = \frac{C_{55}\ t^2}{m^2\ C_{66}\ Z^2} \qquad \text{Eq. 10}$$

that is, that there is a minimum plateback which must be exceeded before energy trapping can occur. This can be expressed as a minimum electrode thickness which must be deposited. A sample calculation of this kind will be made later to illustrate the concept. A survey of the earlier literature of energy trapping has failed to disclose a discussion of this phenomenon.

For only a single mode to exist, the first unwanted mode fm13 must be greater than fs. This is written:

$$fu = fm13 = fm31 = fe\sqrt{1 + \frac{10\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}} > fs \qquad \text{Eq. 8}$$

Whence:

$$1 + \frac{\Delta f}{fe} < \sqrt{1 + \frac{10\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}} \qquad \text{Eq. 8a}$$

or: $$\frac{\Delta f}{fe} < \frac{5\ C_{55}\ t^2}{m^2 C_{66}\ Z^2} \qquad \text{Eq. 8b}$$

This inequality can be recast:

$$\left(\frac{\Delta f}{fe}\right)\frac{Z^2}{t^2} < \frac{5\ C_{55}}{m^2\ C_{66}} \qquad \text{Eq. 8d}$$

whence $$\left(\frac{\Delta Mf}{fe}\right)^{\frac{1}{2}}\frac{Z}{t} = \frac{\sqrt{5}}{m}\sqrt{\frac{C_{55}}{C_{66}}} \qquad \text{Eq. 11a}$$

or $$\left(\frac{\Delta Mf}{fe}\right)^{\frac{1}{2}}\frac{X}{t} = \frac{\sqrt{5}}{m}\sqrt{\frac{C11}{C_{66}}} \qquad \text{Eq. 11b}$$

These equations are of familiar form, being unfamiliar only in the proposed numerical values of the constants on the right hand side. The two forms of Equation 11 are equivalent under the dimensional transformation we are using:

$$X = Z\sqrt{\frac{C_{11}}{C_{55}}} \qquad \text{Eq. 3a}$$

Resonater Resonant Frequency

Unfortunately the equations derived above are of no particular value, as they are all expressed in terms of the unmeasureable frequency fe. It becomes our next objective to recast Equation 11 in terms of the measureable frequency fd, the resonator resonant frequency. The effect of this transformation will illuminate some of the empirical corrections we have all had to make in attempting to reference plateback to fe.

Write again Equation 8

$$fu = fe \sqrt{1 + \frac{10\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}} > fs \qquad \text{Eq.8}$$

If fs be written

$$fs = fd + \Delta f < fe \sqrt{1 + \frac{10\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}} \qquad \text{Eq.8f}$$

And the inequality Eq. 8f be divided by Equation 7:

$$\frac{fd + \Delta f}{fd} = 1 + \frac{\Delta f}{fd} < \sqrt{\frac{1 + \frac{10\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}}{1 + \frac{2\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}}} \qquad \text{Eq. 12}$$

This can be simplified by noting that

$$\frac{1 + 5X}{1 + X} = 1 + 4\ (X - X^2 + X^3 - X^4 + \ldots)$$

So Equation 12 becomes:

$$1 + \frac{\Delta f}{fd} < \sqrt{1 + \frac{8\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}} - \ldots \qquad \text{Eq. 12a}$$

Whence:

$$\frac{\Delta f}{fd} < \frac{C_{55}\ t^2}{m^2\ C_{66}\ Z^2} \qquad \text{Eq. 12b}$$

This can be recast into the form:

$$\left(\frac{\Delta Mf}{fd}\right)^{\frac{1}{2}} \frac{Z}{t} = \frac{2}{m} \sqrt{\frac{C_{55}}{C_{66}}} \qquad \text{Eq. 13a}$$

or $$\left(\frac{\Delta Mf}{fd}\right)^{\frac{1}{2}} \frac{X}{t} \quad \frac{2}{m} \sqrt{\frac{C_{11}}{C_{66}}} \qquad \text{Eq. 13b}$$

Again, the two versions of Eq. 13 are equivalent by virtue of the transformation Eq. 3a.

It is important to note that the expression ΔMf appears in Equations 11 and 13 with different meanings in each. For this reason ΔMf must be handled most carefully. In this paper it is always associated with the appropriate resonator frequency fs or fd. It turns out that the form of Equations 11 involving $\Delta f/fs$ is of significance in determining electrode thickness, estimating the consequences of deviation from parallelism of the blanks, etc. The form of Equations 13 is of significance in determining measured frequencies in the design of practical resonators. Equation 13 corrects a calculation which has previously required empirical adjustment.

Electrode Design Formulas

So far, except for the confirmation of the frequency constant, the various formulas have not been reduced to simple numerical coefficient form. This will be done next in order to compare the numbers here derived with those proposed by other sources.

Eq. 3a becomes:

$$X = Z \sqrt{\frac{C_{11}}{C_{55}}} = 1.1262$$

which is to be compared with the value generally proposed in the literature:

$$X = 1.26Z = \frac{C_{11}}{C_{55}} Z$$

The value proposed as Eq. 3a in this paper is preferred as it is the ratio of velocities, which has the proper physical significance.

Equations 11 become

$$\left(\frac{\Delta Mf}{fe}\right)^{\frac{1}{2}} \frac{Z}{t} = \frac{\sqrt{5}}{m}\sqrt{\frac{C_{55}}{C_{66}}} = \frac{3.446}{m} \qquad \text{Eq. 11c}$$

$$\left(\frac{\Delta Mf}{fe}\right)^{\frac{1}{2}} \frac{X}{t} = \frac{\sqrt{5}}{m}\sqrt{\frac{C_{11}}{C_{66}}} = \frac{3.868}{m} \qquad \text{Eq. 11d}$$

Equations 13 become

$$\left(\frac{\Delta Mf}{fd}\right)^{\frac{1}{2}} \frac{Z}{t} = \frac{2}{m}\sqrt{\frac{C_{11}}{C_{66}}} = \frac{3.0818}{m} \qquad \text{Eq. 13c}$$

$$\left(\frac{\Delta Mi}{fd}\right)^{\frac{1}{2}} \frac{X}{t} = \frac{2}{m}\sqrt{\frac{C_{11}}{C_{66}}} = \frac{3.4596}{m} \qquad \text{Eq. 13d}$$

Lest the point be forgotten, we again call attention to the fact that these numbers pertain to an oriented rectangular electrode of previously specified proportions. But, regardless of their referents the magic numbers appear large compared with previously published values. It is instructive to compute the value of the magic number discussed extensively by the Clevite group from this approach and compare values.

Semi-Infinite Strip Electrodes

Consider the semi-infinite electrode of finite extension X and no field variation along Z.

This has resonances:

$$fd = fm10 = fe\sqrt{1 + \frac{C_{11}\,t^2}{m^2\,C_{66}\,X^2}} \qquad \text{Eq. 14}$$

$$fu = fm30 = fe\sqrt{1 + \frac{9\;C_{11}\,t^2}{m^2\,C_{66}\,X^2}} > fs = fe + \Delta f \qquad \text{Eq. 15}$$

The latter becomes:

$$1 + \frac{\Delta f}{fe} < \sqrt{1 + \frac{9\;C_{11}\,t^2}{m^2\,C_{66}\,X^2}} \qquad \text{Eq. 15a}$$

or $$\frac{\Delta f}{fe} < \frac{9\;C_{11}\,t^2}{2m^2\,C_{66}\,X^2} \qquad \text{Eq. 16}$$

This can be recast as:

$$\left(\frac{\Delta f}{fe}\right)^{\frac{1}{2}} \frac{X}{t} < \frac{3}{m\sqrt{2}}\sqrt{\frac{C_{11}}{C_{66}}} = \frac{3.661}{m} \qquad \text{Eq. 16a}$$

This magic number is twenty percent over the familiar one determined by much elaborate measurement. As satisfactory answer to the discrepancy between the numbers is not available; we did not perform the experiments. Three possible explanations come to mind, that the blanks upon which the measurements were made were substantially convex that the subscripted C values used (which are elastic moduli) do not apply to electroded plates and that the fringing effects are large. "Substantially convex" as will be illustrated later, can be used to describe parallelism of better than one fringe of visible light. No estimate of fringing effect comes out of the equations, but if it is large, the effective lateral dimensions will be substantially larger than the extent of the metalization. Experiments in our laboratory on crystals of known convexity are so far consistent with the fringing hypothesis. If compensated in the constant, both effects would reduce the constant. For the sake of understanding of concept we proceed anyway.

It is possibly equally instructive to determine the resonator frequency under these conditions. Again we write:

$$fu = fm30 = fe\sqrt{1 + \frac{9\;C_{11}\,t^2}{m^2\;C_{66}\,X^2}} > fs = fd + \Delta f \qquad \text{Eq. 15}$$

and dividing by fd

$$1 + \frac{\Delta f}{fd} \sqrt{\frac{1 + 9\;\frac{C_{11}\,t^2}{m^2\,C_{66}\,X^2}}{1 + \frac{C_{11}\,t^2}{m^2\,C_{66}\,X^2}}} = \sqrt{1 + 8\;\frac{C_{11}\,t^2}{m^2\;C_{66}\,X^2}} \cdots$$

This reduces to

$$\frac{\Delta f}{fd} < \frac{4\ C_{11}\ t^2}{m^2\ C_{66}\ X^2} \qquad \text{Eq. 16}$$

which is the value calculated for the rectangular electrode. If the same procedure is followed, calculating fm01 and fm03 the equation for $\Delta f/fd$ in terms of Z will come out

$$\frac{\Delta f}{fd} < \frac{4\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2} \qquad \text{Eq. 12b}$$

Again the value calculated for the rectangular electrode. Different values for the limit of fe are calculated but the expressions for the resonator frequency are the same. If the two expressions for the limit of $\Delta f/fd$ are equated then we find that

$$\frac{X^2}{C_{11}} = \frac{Z^2}{C_{55}}$$

or $$X = Z\sqrt{\frac{C_{11}}{C_{55}}}$$

which is our previously postulated dimensional ratio. From this it follows that our proportions have yet a further significance. This electrode shape results in the maximum electrode area for a given thickness or the maximum electrode thickness for a given area. This is an important design consideration for single response crystals.

Rectangular Electrodes

This might have been anticipated by examining the values of x and z extension for a given electrode area assuming circular, elliptical and rectangular electrodes.

circle $$A = \frac{\pi\ Dc^2}{4} = \frac{\pi\ Xc^2}{4} = \frac{\pi\ Zc^2}{4}$$

ellipse $$A = \frac{\pi\ XeZe}{4} = \frac{\pi\ Xe^2}{4}\sqrt{\frac{C_{55}}{C_{11}}} = \frac{\pi\ Ze^2}{4}\sqrt{\frac{C_{11}}{C_{55}}}$$

rectangle $$A = XrZr = Xr^2\sqrt{\frac{C_{55}}{C_{11}}} = Zr^2\sqrt{\frac{C_{11}}{C_{55}}}$$

By equating the areas it follows that:

$$Xr^2 = \frac{\pi}{4}\sqrt{\frac{C_{11}}{C_{55}}}\,Xc^2 = .883Xc^2$$

$$Xr^2 = \frac{\pi}{4}\,Xe^2 = .785Xe^2$$

$$Zr^2 = \frac{\pi}{4}\sqrt{\frac{C_{55}}{C_{11}}}\,Zc^2 = .699Zc^2$$

$$Zr^2 = \frac{\pi}{4}\,Ze^2 = .785Ze^2$$

If the shape of the electrode were to affect the value of the electromechanical coupling the effect would be a significant source of a criterion for electrode shape. Experience in the use of rectangular electrodes as elements of monolithic filters has shown that any effect on the inductance of an electrode ranging from square to 4:1 is below our ability to measure.

Rectangular electrodes are adequately simple to fabricate and thoroughly satisfactory to use. We are convinced that there is no electrode configuration superior to rectangular for critical crystals in either filter or overtone oscillator application.

Minimum Plateback Criterion

Earlier in this paper an expression was derived for the minimum plateback for energy to be trapped. If the approach here is valid this equation should yield results consistent with experience in predicting the frequency at which special treatment of contour or bevel must be applied to override the deficiencies of electrodes alone. In what follows after some estimates are made concerning the usual practices in applying electrodes, frequencies will be calculated below which special blank treatment is required, for several electrode materials.

We will start by repeating equation 2a

$$fe = f = \frac{mKf}{t} \qquad \text{Eq. 2a}$$

unorthodox electrode patterns and assume an extremely high order of surface parallelism of the quartz plates.

The indication is that substantially thicker electrodes can be used than is normal practice with advantage. There is little point in looking for a denser electrode material; Osmium, the densest metal, has less than an 8% advantage over Gold, is rare and expensive, and is difficult to evaporate. Gold, on the other hand, is 2.67 times as effective as Aluminum. The four metals listed are, or should be, the important electrode metals.

Parallelism

Several times we have had occasion to note that parallelism of crystal plates has a severe effect on the applicability of the design principles being discussed. It was stated rather categorically that deviation from parallelism will result in decreased values of what we have been calling "the magic numbers." To justify this assertion it is necessary to discuss quantitatively the effect of crystal surface deviations on electrode thickness. We will discuss in terms of yellow light and polished blanks so that interference fringes can be used as a basis of discussion. Any other means capable of measuring deviations of the magnitudes resulting from this discussion will be equally applicable.

Common usage refers to a total difference in optical path length of one wavelength as one fringe. For yellow light this is a length of some 5890Å in space. Quartz has two indices of refraction which vary by less than 1/4% around 1.55, so in quartz the wavelength is 3800Å. For a fringe to be seen the light must pass through the quartz twice, so the thickness variation must be one half wavelength or 1900Å (7.5 microinches).

Now, if we examine fs and fe we will lay the basis for understanding the maximum deviation from parallelism permissable.

$$fs = \frac{m\ Kf}{t'}$$

$$fe = \frac{m\ Kf}{t}$$

$$fs - fe = \frac{m\ Kf}{t'} - \frac{m\ Kf}{t} = \frac{m\ Kf\ (t-t')}{t\ t'}$$

or $$t-t' = \Delta t = \frac{m\ Kf\ (fs-fe)}{fs\ fe}$$

We have derived this as quartz thickness making Δt be the increment of mesa thickness. This has an equivalent in metal thickness. We know that there is a minimum plateback which must be applied for energy to be trapped given by

$$\frac{\Delta mf}{fe} = \frac{C_{55}\ t^2}{m^2\ C_{66}\ Z^2}$$

and a maximum for a single response which is given by

$$\frac{\Delta Mf}{fe} = \frac{5\ C_{55}\ t^2}{m^2\ C_{66}\ Z^2}$$

Surely the minimum must apply at the edge of the mesa (electrode) and the maximum at least at some inner point. The full range is only 5 to 1. To determine the maximum permissable deviation from parallelism, the minimum equivalent quartz thickness increment is computed. The thickness variation of the plate in the region to be covered by the electrode must be less than 4 times this minimum increment. For this calculation the frequency constant is useful in unorthodox units:

$$Kf = 1.651 \times 10^{13}\ Hz\text{Å}$$

If the deviation from parallelism exceeds four times the minimum then the curvature of the plate will dominate the energy trapping and multiple responses will occur with any electrode thickness whatever.

A sample calculation will be made to show how the system works. For this illustration we will choose the geometry of the system[14] described by Guttwein, Lukaszek and Ballato for an improved CR-67 at 47 Mhz. They chose elliptical electrodes. We will evaluate the curvature for the elliptical case and then consider the redesign with rectangular electrodes.

We transcribe:

Plate diameter = 8.5 mm
Parallelism = 0.27 μm max across diameter
(1.4 fringes)
Electrode: elliptical shape
Z = 3.0 mm
X = 4.0 mm

Plateback = 30 KHz
$\Delta f/fd = 3 \times 10^4/4.7 \times 10^7 = 6.38 \times 10^{-4}$
Mode index, m = 3

Since

$$\frac{X}{Z} = \frac{4}{3} > 1.1226$$

the first unwanted mode will occur due to a phase reversal along x. Under these circumstances the forms of equations 8, 10, and 12 required will be written in terms of X. Hence:

$$\frac{\Delta mf}{fe} = \frac{C_{11}\ t^2}{m^2\ C_{66}\ X^2} \qquad \text{Eq. 10}$$

$$\frac{\Delta Mf}{fe} = \frac{5\ C_{11}\ t^2}{m^2\ C_{66}\ X^2} \qquad \text{Eq. 8c}$$

$$\frac{\Delta Mf}{fd} = \frac{4\ C_{11}\ t^2}{m^2\ C_{66}\ X^2} \qquad \text{Eq. 12}$$

Since fs and fd and fe vary from each other only by parts in 10^4, wherever the conceptual differences do not require separation of concept the nominal frequency will be used for calculation. Separation of the concepts is significant at several points, however.

$$\frac{\Delta mf}{fe} = \frac{\Delta mt}{t'} = \frac{C_{11}\ t^2}{m^2\ C_{66}\ X^2}$$

$$\frac{\Delta Mf}{fe} = \frac{\Delta Mt}{t'} = \frac{5\ C_{11}\ t^2}{m^2\ C_{66}\ X^2}$$

$$t = \frac{m\ Kf}{fe} = \frac{3(1.663)(10)^5}{4.7(10)^7}$$

$$= 1.0615 \times 10^{-2}\ \text{cm}$$

$$\frac{X}{t} = \frac{.4}{1.0615 \times 10^{-2}} = 37.68; \left(\frac{X}{t}\right)^2 = 1420.0$$

$$\Delta mt = \frac{Kf\ C_{11}}{(1420)\ f_s m\ C_{66}}$$

$$= \frac{1.663(10)^{13}\ (2.9923)}{3(1420)(47)(10)^6} \quad \text{Å}$$

$$= \frac{4.976 \times 10^{13}}{2.428 \times 10^{11}} = 205.\ \text{Å}$$

$$\Delta Mt = 5\Delta mt = 1025\text{Å}$$

The maximum permissable curvature would be the difference of these two, or

$$\Delta Mt - \Delta mt = 4\Delta mt = 820\text{Å}$$

The actual equivalent electrode thickness must also lie within these limits; in fact, the sum of the equivalent electrode thickness and the contour variation inside the electrode boundaries must lie inside $4\Delta mt$. We should, therefore, determine the actual electrode equivalent thickness.

$$\frac{\Delta f}{fe} = \frac{\Delta t}{t'}$$

$$\Delta t = 6.77 \times 10^{-6}\ \text{cm} = 677\text{Å}$$

The allowed deviation of parallelism, then, is:

$$4\Delta mt - \Delta t = 820 - 677 = 143\text{Å}$$

This is somewhat under a tenth of a fringe while the actual deviation from parallelism was 2700Å. It is possible but improbably that only this tiny fraction of the total curvature would occur in the central half of the plate diameter. The reported experience was that all plates had detectable unwanted responses associated with the electrode which were not excited by the oscillator circuit.

Deviations from parallelism are not such as to cause the surfaces to be sectors of spheres. The principal cause of deviations from parallelism is the distribution of particle size in the abrasive slurry; the second most significant is systematic deviation from flatness of the lap surfaces. The former cause results in a thinned edge but a center sector of lesser curvature than the edge while the second tends toward a uniform curvature which extends in to the center of the plate.

A redesign of the plate merely requires the computation of the dimensions of the rectangular electrode of the same area and the recalculation of the plateback range

$$Ae = \frac{\pi(.3)(.4)}{4} = .03\pi = \sqrt{\frac{C_{11}}{C_{55}}}\, Z^2 = 1.1226Z^2$$

$$Z = \sqrt{\frac{.03\pi}{1.1226}} = .290 \text{ cm}$$

$$X = 1.1226Z = .325 \text{ cm}$$

$$\frac{X}{t} = 30.6$$

$$\left(\frac{X}{t}\right)^2 = 936$$

$$\Delta mt = \frac{205(1420)}{936} = 312\text{Å}$$

$$\Delta Mt = 5\Delta mt = 1560\text{Å}$$

$$\Delta Mt - \Delta mt = 4\Delta mt = 1248\text{Å}$$

$$4\Delta mt - \Delta t = 1248 - 677 = 571\text{Å}$$

This is four times the value for the elliptical electrode, or a third of a fringe spread over the diagonal of the electrode, 54.5% of the blank diameter. It is probable that most of the units would have had more than one detectable response with the rectangular electrode of the same impedance as that used, but the prediction of improvement is distinctly illustrated.

In the original exercise using oval electrodes the electrode material used was Aluminum. The density of Aluminum is only 1 1/2% greater than quartz, 2.70 vs. 2.66. When a mesa thickness increment is computed to be 2000Å or less the metal of choice will probably be Aluminum. The electrode thickness will be half the mesa increment of aluminum applied to each side.

Epilogue 1

Design Algorithm

1) Approximate the equivalent thickness of the resonator

$$t \approx \frac{m}{fd}\sqrt{\frac{C_{66}}{\rho}} = \frac{165.1(10)^3 m}{fd} \text{ cm}$$

2) Approximate the electrode area from the specified parameter L or Ce

$$A \approx \frac{1.9m^3}{L\, fd^3} \quad \text{or} \quad A \approx \frac{m\, Ce \times 10^{18}}{2.42fd}$$

3) Compute the dimensions of the electrode

$$Z = \sqrt{A\sqrt{\frac{C_{55}}{C_{11}}}} \qquad X = \sqrt{A\sqrt{\frac{C_{11}}{C_{55}}}}$$

4) Calculate the minimum plateback

$$\frac{\Delta mf}{fe} = \frac{C_{55}\, t^2}{m^2\, C_{66}\, Z^2} = \frac{C_{11}\, t^2}{m^2\, C_{66}\, X^2}$$

5) Calculate the maximum plateback

$$\frac{\Delta Mf}{fe} = \frac{5\Delta mf}{fe} = \frac{5\, C_{55}\, t^2}{m^2\, C_{66}\, Z^2} = \frac{5\, C_{11}\, t^2}{m^2\, C_{66}\, X^2}$$

6) Calculate fe

$$fe = fd\Big/1 + \frac{C_{55}\, t^2}{m^2\, C_{66}\, Z^2} = fd\Big/1 + \frac{C_{11}\, t^2}{m^2\, C_{66}\, X^2}$$

7) Choose a value of plateback

$$\Delta mf < \Delta f \qquad 5\Delta mf = \Delta Mf$$

8) Calculate the blank frequency, fs

$$fs = fe + \Delta f = fd + \Delta f - \Delta mf$$

9) Compute the equivalent thickness difference

$$\Delta t = \frac{m}{2}\sqrt{\frac{C_{66}}{\rho}}\left(\frac{fs - fe}{fs\, fe}\right) = m\, Kf\left(\frac{fs - fe}{fs\, fe}\right)$$

10) Determine the electrode thickness

$$tm = \frac{\Delta t}{2}\Big/\frac{\rho}{\rho_m}$$

Note: tm must exceed 100Å. By a choice of metal the electrode thickness can be kept in the desirable range of 500 to 2000Å down to a frequency below 5 MHz.

11) Compute the maximum equivalent thickness difference

$$\Delta Mt = m\ Kf \left[\frac{\Delta Mf}{fs\ (fs - \Delta Mf)} \right]$$

12) The difference $\Delta Mt - \Delta t$ must exceed the deviation from parallelism in the blank subtended by the electrode (diagonal).

13) Mask electrodes of dimensions and orientation (3) onto blank of frequency (8) and parallelism (12) and adjust resonator to resonator frequency fd.

Epilogue 2

The equation of Onoe and Sykes was said to be only an approximation. It is proper, however, to explain the basis for the assertion that it is only approximate.

All single dimensioned resonators have resonant frequencies given by an equation of the form

$$f = nv/2\ l$$

where n is integer, v is the velocity of the wave motion, and l is the bounded length of the resonator. In some complex cases analysis indicates that the velocity of sound varies within the resonator as a function of position resulting in a very complex formula for the frequency. The law remains of this form. Thus the equation for the frequencies of a taut flexible string is:

$$f = \frac{n}{2\ l} \sqrt{\frac{t}{\rho}}$$

where t is tension and ρ is mass per unit length; the equation for the resonant frequencies of a segment of transmission line is:

$$f = \frac{n}{2\ l \sqrt{\mu\epsilon}}$$

where μ is the permeability and ϵ the dielectric constant of the medium surrounding the line; the equation for the resonant frequencies of an organ pipe or flute is

$$f = \frac{n}{2\ l} \sqrt{\frac{P\gamma}{\rho}}$$

where P is pressure, γ is the ratio of the specific heats and ρ is the volume density of air.

In these relatively simple cases the velocity is given as the square root of the ratio of the elastic modulus appropriate to the wave by the density of the medium.

Similarly the resonances of a rectangular microwave cavity are given by the formula:

$$f = \frac{1}{2\sqrt{\mu\epsilon}} \sqrt{\frac{m^2}{l^2} + \frac{n^2}{w^2} + \frac{p^2}{d^2}}$$

and the acoustic resonances of a room by the formula:

$$f = \frac{1}{2} \sqrt{\frac{P\gamma}{\rho}} \sqrt{\frac{m^2}{l^2} + \frac{n^2}{w^2} + \frac{p^2}{d^2}}$$

It is tempting to try to write the equation for the resonances of an isotropic solid rectangular prism by the similar formula:

$$f = \frac{1}{2} \sqrt{\frac{E}{\rho}} \sqrt{\frac{m^2}{l^2} + \frac{n^2}{w^2} + \frac{p^2}{d^2}}$$

Such an assumption is generally incorrect. The assumption that a single elastic coefficient applies in a solid is a gross oversimplification. If the wave motion is simple dilation in all directions the assumption is good. It can be made correct for such a simple mode of oscillation in an isotropic prism by an appropriate adjustment of the modulus E. For any other mode of distortion the appropriate modulus of elasticity may be different in different directions. An improved approximation will be found in the equation:

$$f = \frac{1}{2\sqrt{\rho}} \sqrt{\frac{m^2 E_1}{l^2} + \frac{n^2 E_2}{w^2} + \frac{p^2 E_3}{d^3}}$$

This approximation is an improvement but requires that the mode of distortion be known in order that the proper elastic moduli be chosen.

Certainly if the material is not isotropic the elastic moduli in the three directions will not, in general, be the same. Any adequate approximate solution would require the recognition of this in the expression for the modull.

Additionally in an anisotropic medium the very anisotropy causes a driven mechanical distortion of one kind to be coupled into simultaneous distortion of another kind. In quartz, especially in rotated cuts, a pure body shear or a pure face shear or a pure extension is not excited by the field between the electrodes, but simultaneously a whole complex of kinds of distortions that the desirable temperature coefficients of certain cuts are generated. Each of the distortions has its appropriate modulus so a complete expression would have to be much more complex than this one. It is possible to direct the analysis of a particular resonant state to a more accurate conclusion than this formula yields. No other presentation is more directly accessable to physical intuition or displays the correlations of experience more clearly.

Epilogue 3

The dimensional ratio chosen as optimum for the rectangular electrode has the mathematical property that the formula frequencies of f_{mij} and f_{mji} are equal. The ratio, first chosen for this reason, was further justified by the argument that this ratio yields the largest unit capacitance for a given plate-back with only a single response. The latter criterion is frequently of great practical importance.

It is, however, true that the frequencies can never actually be equal. There will be in actuality the same number of responses in the electroded plate due to the electrode as there are with square, round, elliptical or any other shape of electrode. Every attempt to cause a pair of responses to coalesce will be frustrated due to the couplings between all the various modes of motion.

When resonators on the same frequency are coupled, the resonance is split and maxima of response appear spaced symmetrically above and below the isolated resonant frequency of either. This well-known pattern is uniform in all instances, causing alike the double-humped response of the ordinary superhetrodyne radio receiver and the fine structure of atomic spectra.

The response does split and the separation of the two peaks resulting is a measure of the coefficient of coupling between the two modes. During our investigation, still incomplete, of the discrepancy highlighted in the body of the paper between the magic numbers derived algebraically and empirically, a value for this coupling of approximately 6×10^{-3} has been indicated.

Bibliography

1. Shockley, W., Curran, D. W., and Koneval, D. J. "Energy Trapping and Related Studies of Multiple Electrode Filter Crystals" Proc. 17th A.F.C.S., 1963
2. Curran, D. W., and Koneval, D. J., "Energy Trapping and the Design of Single and Multi-Electrode Filter Crystals," Proc. 18th A.F.C.S., 1964
3. Curran, D. W., and Koneval, D. J., "Factors in the Design of VHF Filter Crystals," Proc. 19th A.F.C.S., 1965
4. Lukaszek, T. J., "Improvements of Quartz Filter Crystals," Proc. 19th A.F.C.S., 1965
5. Waren, A. D., Gerber, W. J., and Curran, D.R., "Trapped Energy Modes, Network Synthesis and the Design of Quartz Filters" Proc. 19th A.F.C.S., 1965
6. Ballato, A., Lukaszek, T., Wasshauser, H., and Chalak, E.,"Design and Fabrication of Modern Filter Crystals, Proc. 20th A.F.C.S., 1966
7. Mindlin, R.D., "Studies in the Mathematical Theory of Vibrations of Crystal Plates," Proc. 20th A.F.C.S., 1966.
8. Yoda, H., Nakazawa, Y., Okano, S., and Kobori, N., "High Frequency Crystal Mechanical Filters," Proc. 22 A.F.C.S., 1968.
9. Beaver, W.D., "Theory and Design Principles of the Monolithic Crystal Filter," Doctoral Dissertation, Lehigh University 1967, University Microfilm, 1967.
10. Mortley, W.S., "Circuit Giving Linear Frequency Modulation of a Quartz Crystal Oscillator," Wireless World 57, October 1951.
11. Mortley, W.S., "Frequency-Modulated Quartz Oscillators for Broadcasting Equipment," Proc. IEEE, Vol. 104B, May, 1957.
12. Appleby, T., "Mahlon Loomis, Inventor of Radio," Loomis Publications, 1967.
13. U.S. Patent No. 1745170 Filed October 8, 1926, Issued January 28, 1930.
14. Guttwein, G.K., Lukaszek, T.J., and Ballato, A.D., "Practical Consequence of Model Parameter Control in Crystal Resonators," Pro. 21st A.F.C.S., 1967.

A FLEXURE-MODE QUARTZ FOR AN ELECTRONIC WRIST-WATCH

Max P. Forrer
Centre Electronique Horloger S.A.
Neuchâtel, Switzerland

Summary

Design considerations of a low-frequency quartz resonator to be used in an electronic wrist-watch are discussed. Among the most difficult problems are the low-loss and reasonably shock-proof suspension as well as the encapsulation within the very rigid dimensional requirements. Other considerations involve the choice of the operating frequency, the electrode configuration, the crystal cut and the ensuing temperature behavior. The principal characteristics of a crystal unit of 1.070 x 0.260 x 0.110 inches overall external dimension will be discussed. In conclusion, the other essential elements constituting the quartz wrist-watch are mentioned.

Introduction

The advent of integrated circuits has opened up new possibilities for miniaturization of electronics in the dimensional range conforming to wrist-watches. The question is therefore open whether or not such integrated circuits, together with other elements, can contribute towards a wrist-watch of improved characteristics as compared to existing mechanical, electrical or electronic watches.

Quartz Frequency

One is tempted by the excellent properties of the quartz crystal oscillator to consider it as a regulating device for such a wrist-watch. The oscillator frequency will have to be counted down to a value suitable to drive an electromechanical transducer actuating the hands. To simplify the integrated counter circuitry and also to minimize its power consumption, we are interested in an oscillator frequency as low as possible. This leads us to consider flexure-mode quartz resonators.

Dimensional requirements for the quartz on the other hand suggest the use of high frequencies, so that we must search for a reasonable compromise.

For a man's wrist-watch the length of the quartz bar cannot exceed 25 mm. A minimum allowable thickness of 0.8 mm has been assumed to assure adequate soldering connections of the suspension wires. The frequency formula for flexure-mode resonators then yields

$$f = 5800 \frac{h}{l^2} = 7400 \text{ Hz}$$

as the lowest admissible resonator frequency. Since our counter circuits under development are based on binary logic elements and the possibility of a 1 Hz counter output frequency seems important the binary frequency value of $2^{13} = 8192$ Hz was chosen.

Oscillating Mode

Among the three wellknown methods for flexure-mode excitation the NT type was excluded because of its higher frequency for a given length. The Curie-strip, which would be attractive because of its high coupling coefficient, does not appear practical due to the very small thickness dimension. Additionally, the solder joint between the two quartz halves would unduly derate the resonator performance.
A practical solution is offered by the XY mode, whose electrode geometry is shown in Fig. 1. Isolation between X and Y surface metalisations is obtained, where required, by a mechanical angling operation with a diamond file. The four transverse isolating bands are produced by appropriate masking during evaporation.

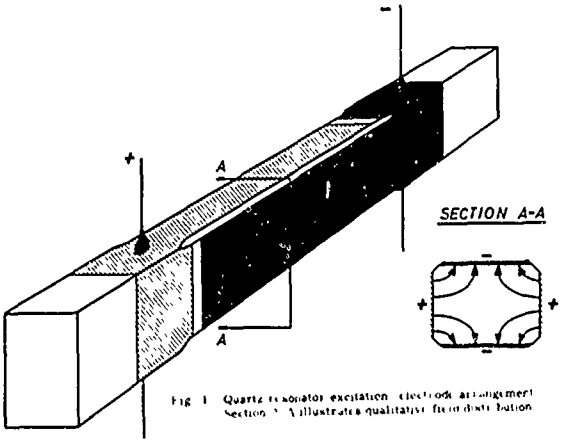


Fig. 1 Quartz resonator excitation electrode arrangement. Section A-A illustrates qualitative field distribution.

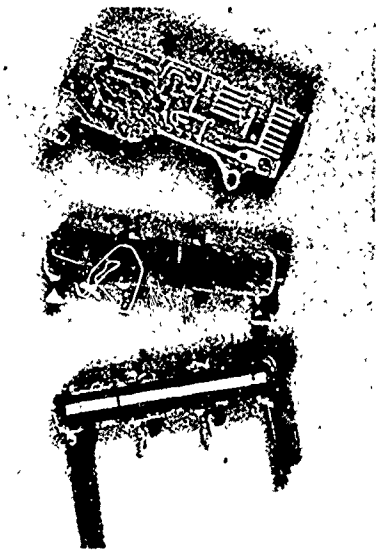


Fig. 2 Experimental quartz resonators, left: open, right: with cover and welded encapsulation. Encapsulated quartz resonator with pinch-off tube.

Mounting and Encapsulation

The major problems involved in the development of a high quality crystal resonator of such small dimensions lie in its suspension and encapsulation.
Figures 2 and 3 illustrate a laboratory

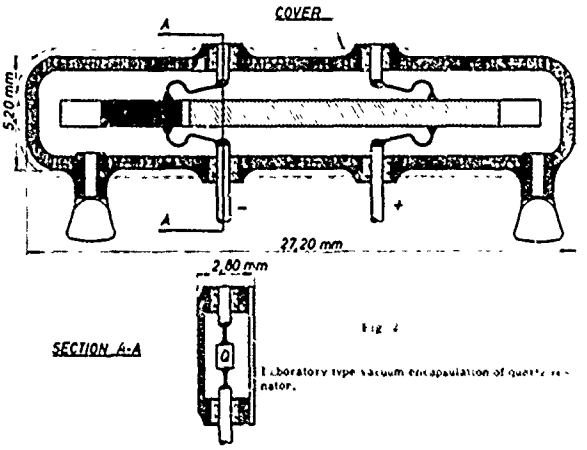


Fig. 3 Laboratory type vacuum encapsulation of quartz resonator.

type encapsulation, which has given good results. It utilizes a copper housing, containing four miniature glas feed-throughs. Two copper tubings are attached to allow circulation of liquid cleaning agents after mounting and encapsulation and to allow evacuation and sealing by cold-weld pinch-off.

One of the most important requirements for a practical watch application is a satisfactory shock resistance. Present tests show that a free fall through 1 meter on a hardwood floor does not cause destructive effects. Prolonged tests on a centrifuge at 450 g acceleration show frequency changes smaller than 10^{-6} due to static deformation of the quartz suspension. What is more important, however, is the fact that to an accuracy of 10^{-7} no permanent frequency changes are observed. This indicates that the quartz is reasonably shock-proof for normal watch use, where the statistic[illegible]ow that practically all shocks produce accelerations below 10 g.

Improvement of the shock resistance calls for a more rigid quartz suspension. But this in turn has a degrading effect upon the resonator performance. Due to the miniaturization of the quartz slab, the suspension wires become a non-negligeable part of the resonating system. With larger diameter suspension wires a higher fraction of the total stored energy of the resonator is in the suspension system bringing about an increased influence of the metal wires and solder joints upon resonator performance. Affected primarily are the Q-factor, the temperature characteristic and the aging rate.

Temperature Characteristics

The wellknown temperature characteristic of the flexure-mode quartz is a parabolic behaviour of the form

$$\frac{\Delta f}{f} = -\beta (T - T_0)^2$$

where the quadratic temperature coefficient is approximately $\beta = 3.7 \times 10^{-8}/°C^2$. The inversion temperature T_o is chosen by the appropriate crystal cut of X + 4° to correspond to the average operating temperature of the watch on the wrist, i.e. approximately 28°C.

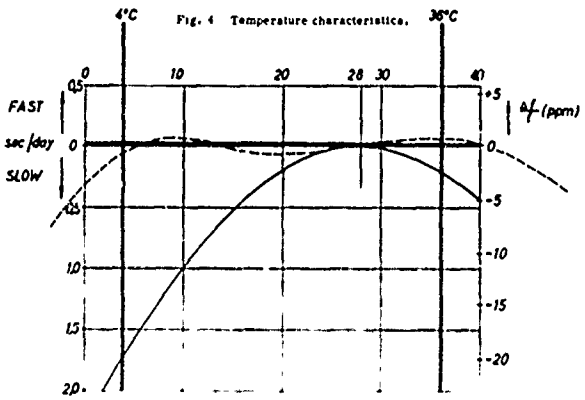


Fig. 4 Temperature characteristics.

Figure 4 shows such a temperature characteristic. For extreme precision a temperature compensation network consisting of NTC and PTC resistors and capacitors may be connected to the quartz in order to provide a phase vs temperature function having frequency compensating effect. The result of this is shown in the double humped dashed curve.

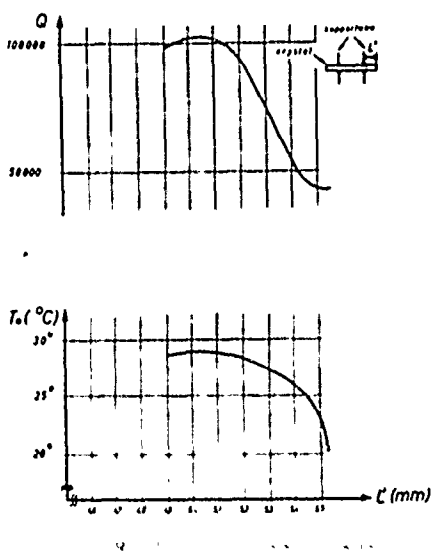


Figure 5 shows an interesting correlation between temperature behaviour (T_o), Q-factor and the location of the suspension points. A series of resonators was fabricated and the suspension points were carefully varied. A maximum Q exceeding 100'000 is obtained for L' = 5 mm indicating the situation where the quartz suffers the least amount of degradation due to suspension wire losses. We note that the analytical calculation of the nodal point yields L' = 0.224 x 23.8 = 5.3 mm.

In the optimum situation T_o comes closest to the value expected for the quartz alone. Due to the negative temperature coefficient of the metal suspension wires, the experimentaly observed decrease of T_o is understandable for those situations where more oscillation energy is contained in the wires.

These considerations offer a simple practical way for resonator optimization with regard to the suspension.

Frequency Adjustment

In a practical watch the oscillator frequency shall be slightly adjustable to permit compensation of manufacturing tolerances and aging effects. This has been realized at the prototype stage by a capacitor C_s connected in series with the quartz.

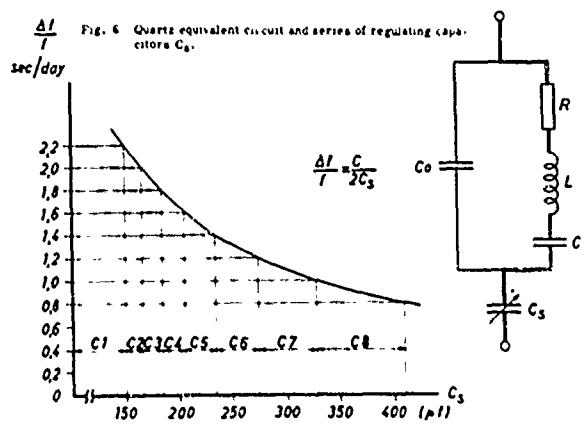


Fig. 6 Quartz equivalent circuit and series of regulating capacitors C_s.

Figure 6 shows the equivalent circuit of the quartz with the regulating capacitor C_s. The characteristic values are:

$C = 7.8\,10^{-3}$ pf

$L = 4.8\,10^{4}$ Henry

R = 25 k ohm

C_o = 8 pf

The frequency deviation is plotted vs C_s and various incremental capacitor values C_2-C_8 are chosen to correspond to equal frequency steps of 0,2 seconds/day. The miniaturized realization of these capacitors uses the integrated MOS technique.

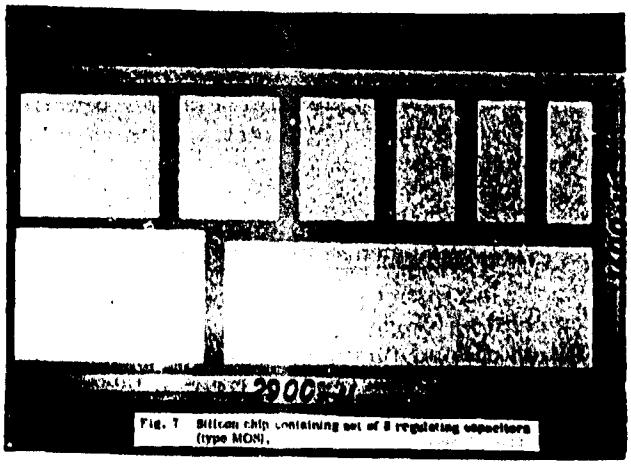


Fig. 7 Silicon chip containing set of 8 regulating capacitors (type MOS).

Figure 7 illustrates the silicon chip containing the set of capacitors with a total value of 420pf.

Performance Characteristics

Performance characteristics of the entire watch, that have not yet been mentioned and which are mainly determined by the quartz, are given by the following table

Dimensions of movement	28,2x25,0x5,3 mm
Short term stability	0,01 sec/day
Position error	0,04 sec/day
Error due to support mass	0,02 sec/day
Aging	0,2 sec/day per year
Isochronism (for V/V=10%)	0,1 sec/day
Barometric error	0
Tolerable ambient magnetic field	60 Oe
Current consumption at 1,35V	16 μa*
Battery life-time (for WH-3)	14 months*

where it should be reemphasized that apart from precision the overall dimensions and the energy consumption constitute the novel and difficult problem areas.

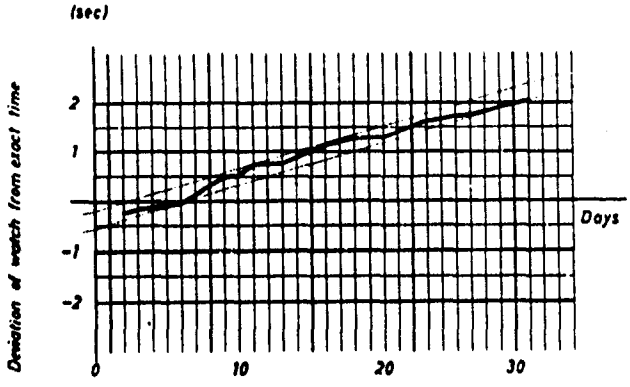


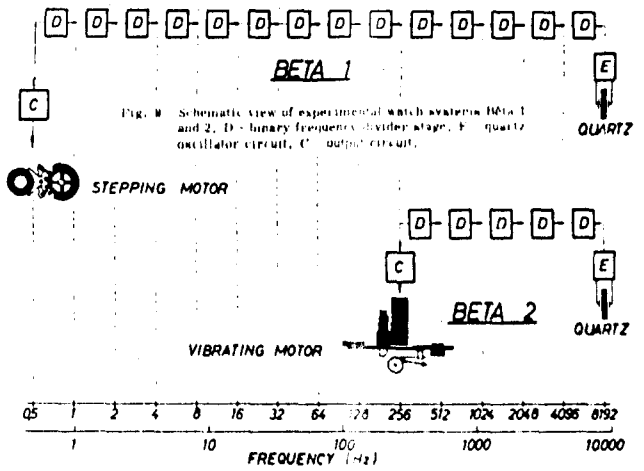
Fig. 8 Performance of watch at normal wrist wearing.

Figure 8 illustrates a practical test of a prototype watch while it was worn over a period of one month. Apart from a systematic error, which can in principle be compensated we note an uncertainty of less than one second per month.

Conclusion

In conclusion the quartz shall be placed in context with the overall watch system. Figure 10 gives a schematic diagram of two such systems that have been realized. System Bêta 1 divides the quartz frequency down to a 0,5 Hz value suitable to drive an electrodynamic stepping motor which performes one-second steps in response to bipolar current pulses.

* for Bêta 2 type watch system, as illustrated in the following paragraph



The system called Bêta 2 divides the quartz frequency to an intermediate value of 256 Hz where it is utilized to drive a resonant electro-mechanical vibrator actuating the watch movement by means of a ratchet mechanism.

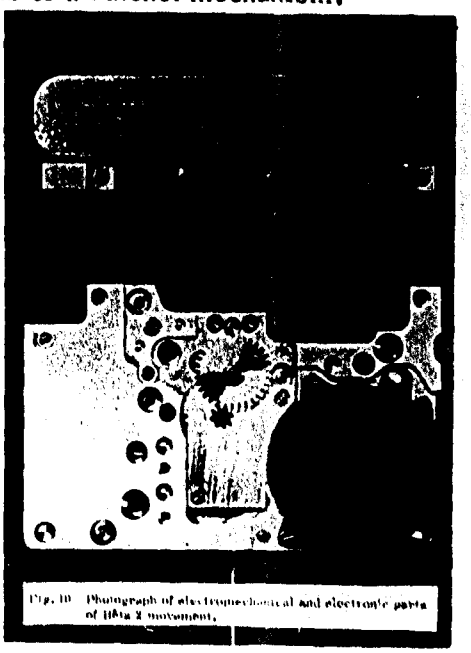
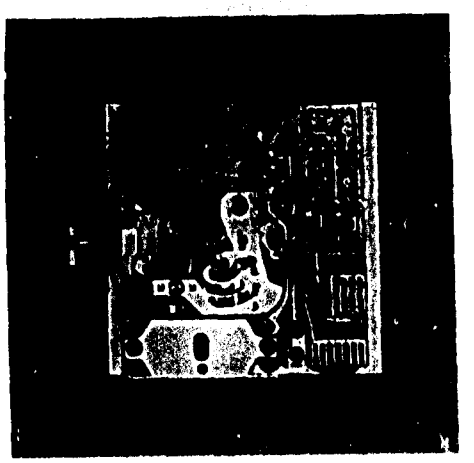


Figure 10 shows a photograph of the Bêta 2 type watch movement, in two functional parts:

The quartz and the integrated circuits mounted on a printed circuit card form the electronic part. The electromechanical part contains the battery housing, the electromagnetic transducer, the vibrator with the ratchet mechanism, the necessary gears to actuate the three hands and the hand-setting mechanism. When assembled, these two parts form the complete movement as illustrated by Figure 11.



Reproduced from best available copy.

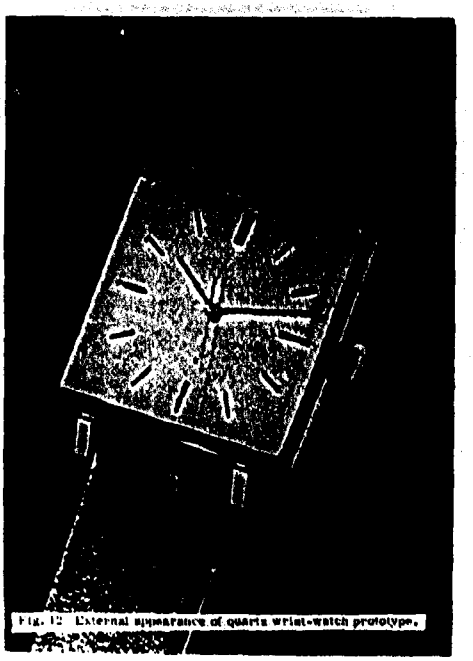


Figure 12 shows the external appearance of the Bêta 2 prototype when mounted into an experimental watch case for practical testing.

Manufacturing of this relatively complex watch at reasonable cost poses many challenging problems. However, simplifications and compromises are possible so that a competitive, high-performance quartz wrist-watch is coming into

the realm of practical reality.

The experimental quartz slabs described have been fabricated by Ebauches S.A., Neuchâtel, Switzerland.

The author wishes to give credit to Messrs. A.H. Frei, J. Hermann and L. Omlin, who have contributed with dedication and patience to the solution of this tedious and very delicate miniature quartz problem.

LASER MACHINING THIN FILM ELECTRODE ARRAYS ON QUARTZ CRYSTAL SUBSTRATES

J. L. Hokanson
Bell Telephone Laboratories, Incorporated
Allentown, Pennsylvania

Summary

The feasibility of laser machining AT-cut quartz crystal electrode arrays for performing vernier adjustments to both the frequency of individual resonators and bandwidth of monolithic crystal filters has been demonstrated. The frequency has been adjusted (upward only) by vaporing spot patterns on the electrode surface. The bandwidth has been adjusted (downward only) by removing material from the edge of one electrode and by removing bits of mass from a stripe of metal evaporated between two adjacent resonators. Laser machining has also been used to produce divided electrodes on extensional and flexural vibrators. It is a noncontact process which can be carried out in air and without any apparent long range damage to the substrate.

Introduction

Thin metallic films have traditionally been used to excite quartz resonators, and in many cases act as a vernier adjustment on the resonant frequency. In the case of monolithic crystal filters, the electrode geometry and surface mass loading serve to control the bandwidth, band shape and the activity of unwanted anharmonic mode responses. Tighter requirements on final frequencies and the maximum suppression of unwanted resonances place stringent requirements on electrode shapes and thicknesses for many present resonator designs. Heretofore, the alteration of electrodes deposited on quartz substrates without having deleterious effects to the resonator performance has been very difficult. It now appears that laser machining of these metallic electrode arrays to change resonator shape and electrode mass loading can be realized through the use of a neodymium doped yttrium-aluminum-garnet (YAG) laser without any apparent damage to the quartz itself. (The feasibility of removing or machining evaporated films has been reported in Refs. 1, 2, and 3). The machining is accomplished by (1) pulsing or Q switching the laser with a repetition rate up to a few kHz, (2) focussing the beam on the crystal electrode, and (3) traversing the sample through the laser beam. By varying the pulse repetition rate and speed with which the crystal is traversed through the laser beam, the electrode may be removed in strips or in small individual spots.

This paper will describe the use of such a laser machining arrangement to adjust the frequency of individual resonators, and the frequency and bandwidth of monolithic crystal filters by selectively removing portions of metal electrodes evaporated on the crystal faces. This technique has also been used to produce divided electrodes on extensional and flexural vibrators. The process, however, is not without its own limitations, and these also will be discussed.

Experimental Apparatus

The laser machining apparatus used in this work is shown schematically in Figure 1. The heart of the system is a continuously pumped yttrium-aluminum-garnet (YAG) laser oscillating in a transverse mode. The laser is Q-switched by means of an acoustooptic deflector mounted on the rear mirror[4,5] which permits the varying of the laser pulse repetition rate from 1 Hz up to 4 kHz. The Q-switched output consists of a train of pulses with peak power approaching 500W and a pulse duration of about 250 nsec. at the half power point. The output, at 1.06 microns in wavelength, is monitored by means of a photomultiplier tube located behind the cavity which detects the pulse shape, and is used in conjunction with a thermopile which measures mean power. From these two measurements the peak power of the laser pulse may be calculated.

The beam is focussed on the crystal electrode by means of an objective lens and the machining operation is viewed by means of a closed circuit television system. A pair of beam splitter mirrors permit both the illumination of the sample and TV viewing. A 23 mm microscope objective (10X) was used for most of the machining work. This produces a vaporized spot size in the electrode film of 13 microns (.5 mil) in diameter.

The sample is mounted on a x-y-z micropositioning table which is driven in the horizontal direction by a variable speed motor. The other two directions are used for translating the workstage between passes and for focussing the sample under the microscope objective.

The machining operation is controlled by (1) selecting the workstage travel rate and laser pulse repetition rate to produce the desired spot center-to-center distance and (2) adjusting the laser intensity be means of neutral density

filters. Thus, material may be removed by producing an array of holes or a continuous line may be generated by a series of overlapping holes. Reindexing at the end of each row and repeating this line machining permits the removal of larger areas of material. A photograph of the laser machining permits the removal of larger areas of material. A photograph of the laser machining apparatus is shown in Figure 2.

Experiments and Measurements

In order to ascertain the nature and effect of laser machining on quartz crystal electrode arrays several experiments were performed to determine the following:

A. Multiple pass stripe definition
B. Extent of damage to the substrate
C. Bandwidth adjustment feasibility
D. Frequency adjustment feasibility.

A. Stripe Definition

The first experiment was designed to ascertain the line definition that could be obtained during a multiple pass or milling type operation. The AT-cut quartz wafer used in this experiment was 15 mm in diameter with an 8 micron surface finish, and had a fundamental frequency of 8 MHz. The electrodes had the configuration shown in Figure 3a, and consisted of an evaporated metallic trilayer composed of a base layer of 300 Å of Cr, a middle layer of 6000 Å of Cu, and outer layer of 300 Å of Au. A series of three passes with a 0.3 mil translation step between passes was made removing a total of 0.9 mil from the region shown in the inset. A photomicrograph taken of this region is shown in Figure 3b. An optical comparison of the definition at the machined electrode edge with that of the as-plated edge indicates that the machined edge is more sharply defined thus giving a qualitative measure of the process control.

This line machining technique has been used at the Western Electric Company, Merrimack Valley to produce divided electrodes on some extensional and flexural vibrators.[6] The units made in this manner passed 500 VDC - 500 Mohm breakdown tests indicating that no appreciable amount of metallic debris remained in the machined area.

B. Substrate Damage

A number of experiments were conducted to determine the extent of the damage to the quartz as a result of the laser machining process. The first experiment involved the crystal units described in the previous section. The resonator frequency, inductance, and Q of the crystal units were measured and compared to values obtained prior to the machining process to see if any gros changes in crystal unit electrical parameters had occured. No changes in resonator inductance or Q could be detected. The frequency of the resonator increased in accordance with that expected from the amount of mass removed from the electrode. Thus, as far as could be determined, no adverse affects on crystal unit electrical parameters could be detected.

The second experiment involved a laser machined crystal which had polished major surfaces and had been plated in the same manner described in the previous section. This sample did reveal some irregularities on the substrate surface when examined using interference contrast microscopy. Figure 4a is a photomicrograph taken at the corner of an electrode that had been machined by cutting a series of lines in the X direction. A photomicrograph of the same area at a higher magnification taken after the remaining electrode had been removed by chemical etching is shown in Figure 4b. The irregularity in the machined path indicates that local melting of the substrate may have occured. The sample was examined by a Talysurf profilometer, but the depth of these irregularities were masked out by the surface scratches on the substrate. The depth to which the surface may be disturbed during the machining process is reported to be on the order of the film thickness.[2,3] It should be noted that quartz is transparent to radiation at 1.06 microns and is not damaged by the direct impingement of the laser beam on its surface. In this same vein, the laser machining process could be performed through a glass bulb enclosure. The damage is a result of the vaporization of the metal electrodes from the quartz surface. The extent to which this damage occurs, however, is localized and minimal.

A third experiment was undertaken to determine whether the first antisymmetric anharmonic thickness shear mode in a quartz resonator could be charge cancelled by altering the electrode configuration using laser machining.[7] This mode can be charge cancelled if the positive charge due to the shear displacement amplitude of the middle antinode is balanced by the negative charge due to the two outer antinodes. Since this experiment involved the removal of a considerable amount of electrode material, a comparison of the static X-ray diffraction topographs[8] taken before and after machining was made to see if any substrate damage could be detected.

The resonator used in this experiment is shown in Figure 5. The plate was a 15 mm square AT-cut quartz wafer at 3.3 MHz on which a strip electrode 4 mm wide had been plated along the x_1 axis. The mass loading due to the electrodes was 0.5%. Figure 6 shows the frequency scan of this unit along with the X-ray diffraction topographs of (1) the static or nonvibrating plate which will be used for comparison purposes with the machined unit, (2) the fundamental thickness shear mode and (3) the first antisymmetric anharmonic thickness shear mode, the mode we wish to charge cancel.

To charge cancel this unwanted mode the electrode strip was shortened to the midpoint of the two end antinodes by laser machining the ends of the strip. By doing this the positive charge due to the middle antinode will be balanced by the negative charge contributions from the two end antinodes. This is shown schematically in Figure 7.

A narrow tab was left along the center of the strip to maintain electrical continuity. The positions of the three antinodes of the anharmonic mode seen in Figure 6 have been superimposed on the schematic of the laser machined electrodes. The frequency scan of this crystal after laser machining in the above described manner shows that the unwanted anharmonic mode, labled "S", has nearly vanished. The frequency of the fundamental mode was raised slightly, and the strength of the other unwanted responses were altered. The topograph of the static plate does not give any indications of laser damage to the quartz.

C. Bandwidth Adjustment

A number of techniques can be proposed for the adjustment of the inter-resonator coupling (and hence bandwidth) of a monolithic crystal filter after the initial electrode array has been deposited. These include schemes for altering the thickness of the quartz such as using abrasives, and altering the geometry of the electrode array and the distribution of electrode mass.

Two methods for altering the geometry of the electrode array using laser machining are shown schematically in Figure 8. The first method involves a machining operation at the edge of an electrode, and the second involves the removal of mass from a stripe of metal evaporated between adjacent resonators.

To test the first method the edge of one top electrode was machined a distance of 3.5 mils (10 passes) as indicated in Figure 8a. This operation increases the effective separation between resonators. The inter-resonator coupling of this subsystem decreased 10%. Had the bottom electrode been machined in the same manner as the top electrode the decrease in coupling computed from our theoretical model[9] indicates that a 17% decrease in coupling would be expected. Thus, the increase in the separation of only the top electrode pair had about half the effect on coupling as that produced by increasing the separation between resonators.

The second method for bandwidth control involves a stripe of metal evaporated between two adjacent resonators and its subsequent removal as indicated in Figures 8b and 8c. The effect of this stripe is shown schematically in Figure 9.

The stripe lowers the cutoff frequency in the region between the two resonators and alters the normal mode frequencies of this coupled system. Numerical results obtained from our theoretical model[9] show that the frequency of the antisymmetric mode, f_2, is lowered slightly, while on the other hand the frequency of the symmetric mode, f_1, is lowered considerably. Thus, as the mass loading on this stripe is increased, the bandwidth of the monolithic crystal filter is increased; i.e., the frequency difference between these two normal modes has been increased. It should be noted that the measured individual resonator frequencies, f_a, also have been lowered in the process thus lowering the midband frequency of the filter. This difficulty may be overcome by technique described in Figure 10.

The procedure relies on the difference in effect of mass placed on the trimming stripe and mass placed on the electrodes. The mass placed on the stripe has the effect of raising the bandwidth about the same order of magnitude as it lowers the midband frequency. Mass placed on the electrodes, on the other hand, can lower the midband frequency a large amount while having little effect on bandwidth. By adjusting the inter-resonator coupling first, and the frequencies of the resonators second, both correct bandwidth and band location can be achieved.

The method of bandwidth adjust by laser trimming is illustrated in Figure 11. It consists simply of removing portions of mass from the stripe of metal between the resonators. The computed and measured effects of this removal process by laser machining are shown by the solid line and circles respectively.

D. Frequency Adjustment

The frequency trimming operation consists of vaporizing a regularly spaced array of holes such as the pattern shown schematically in Figure 12. The amount of frequency increase is approximately proportional to the number of holes in the array. This technique was carried out on two types of 5 MHz oscillator crystal units--one operating on the fundamental thickness shear mode, the other operating on the fifth overtone of the thickness shear mode. The plateback on the fundamental unit was 2% while that for the fifth overtone unit was 0.1%. For the particular array shown the fundamental unit increased in frequency 6 Hz per row of holes (about 150 holes) or about 4×10^{-2} Hz/hole. The overtone unit measured after generating 55 rows of holes increased in frequency 25 Hz or about 2×10^{-3} Hz/hole. A photograph of the actual hole pattern is shown in Figure 13. The inductance and Q of both units were unaffected by the machining operation.

References

1. M. I. Cohen and J. P. Epperson, Advances in Electronics and Electron Physics (Academic Press, New York, 1968), p 139.

2. M. I. Cohen, B. A. Unger, and J. F. Milkowsky, Bell System Tech. J. 47, 385 (1968).

3. B. A. Unger and M. I. Cohen, Electronics Components Conference, Washington, D. C., p 304 (1968).

4. J. E. Geusic, H. M. Marcos, and L. G. Van Uitert, Appl. Phys. Letters 4, 182 (1964).

5. J. E. Geusic, M. L. Hensel, and R. G. Smith, Appl. Phys. Letters 6, 175 (1965).

6. A. Gingrande and B. Falanga, Private Communication.

7. G. T. Pearman, J. Acoust. Soc. Am. 45, 1969, p 228.

8. W. J. Spencer, Physical Acoustics Principals and Methods (Academic Press, New York, 1968), Vol. 5, Chap. 3, p 111.

9. P. Lloyd and K. Haruta, Private Communication.

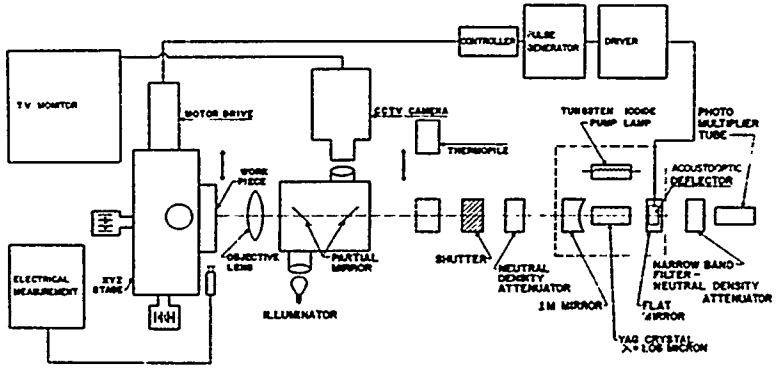
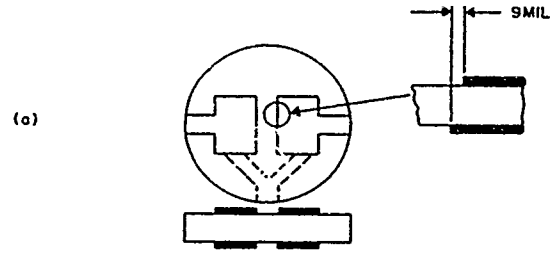
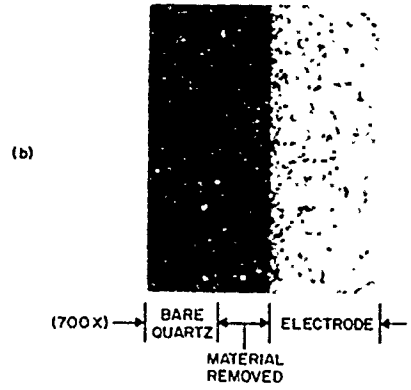


FIGURE 1 . Schematic diagram of laser machining apparatus

Reproduced from best available copy.





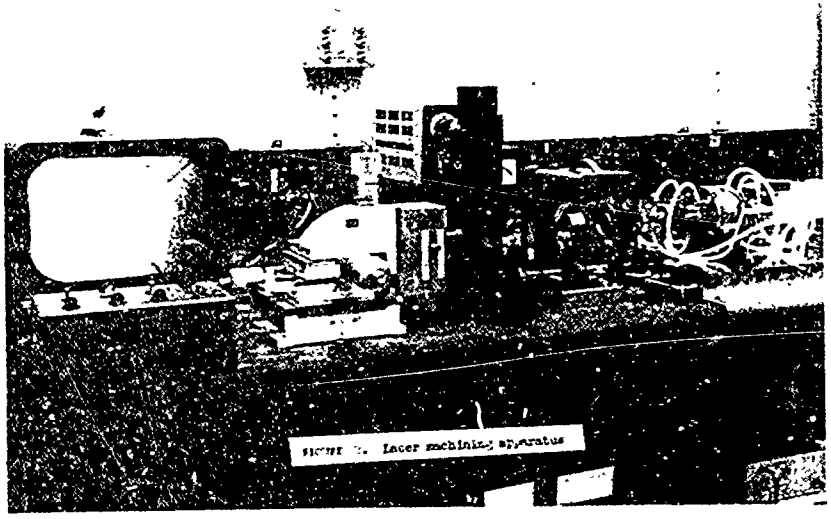
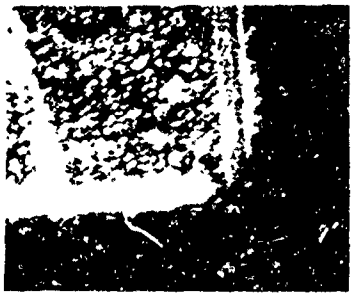


FIGURE 2. Laser machining apparatus

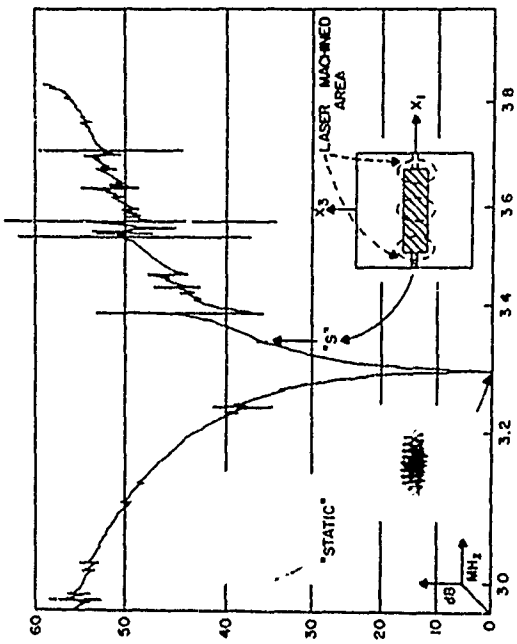


(a)

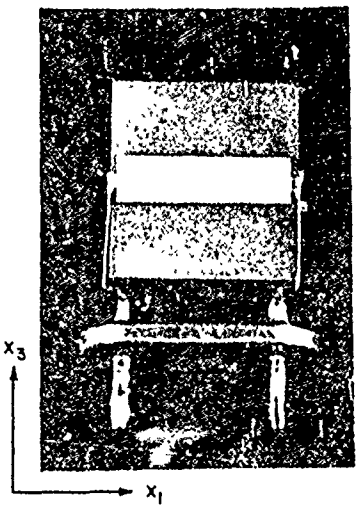


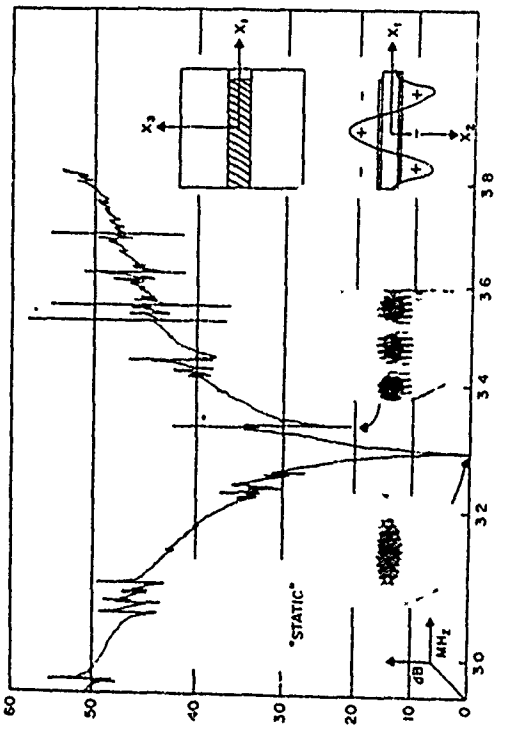
(b)

FIGURE 4 - Photograph of laser affected area



Reproduced from best available copy.





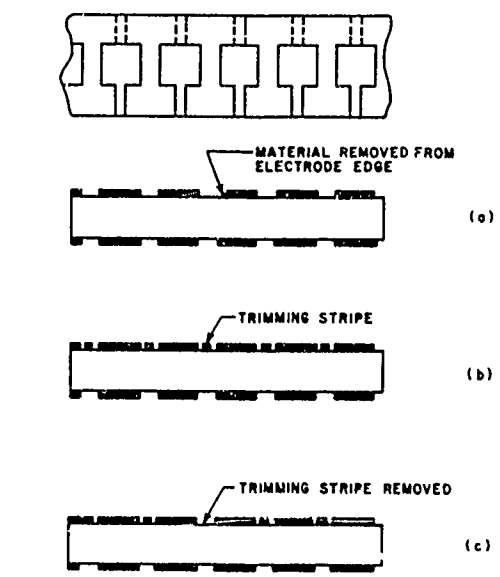


FIGURE [illegible]

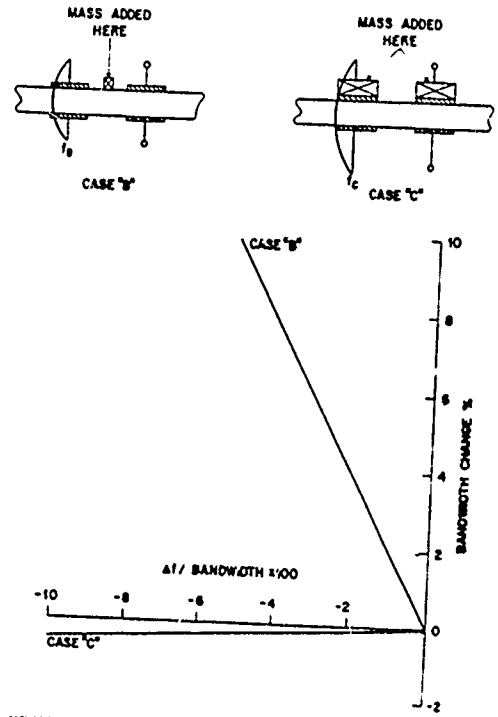


FIGURE 10. Effect of mass distribution [illegible]

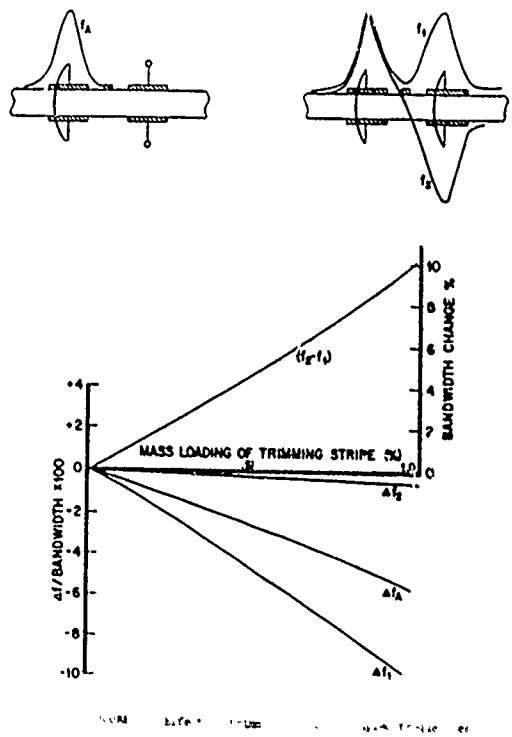


FIGURE [illegible]

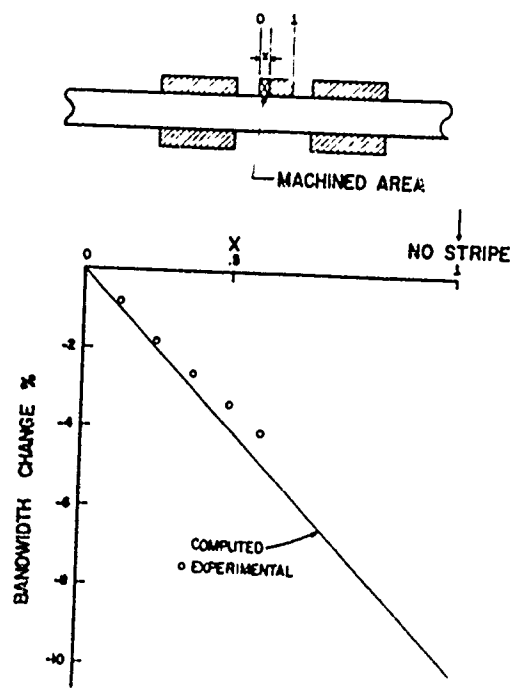


FIGURE 11 Effect of trimming stripe removal on filter bandwidth

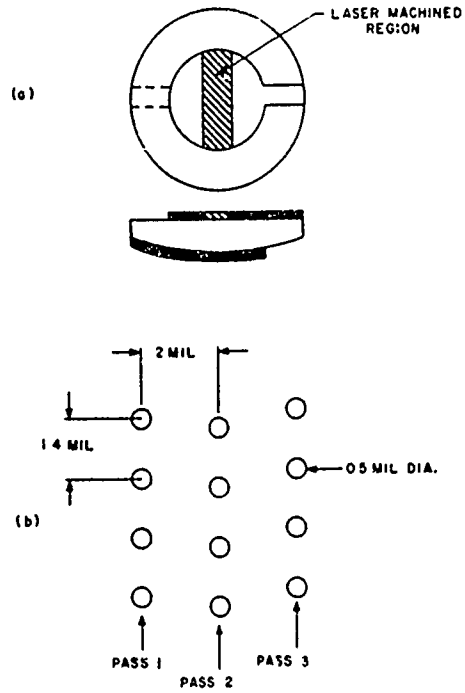
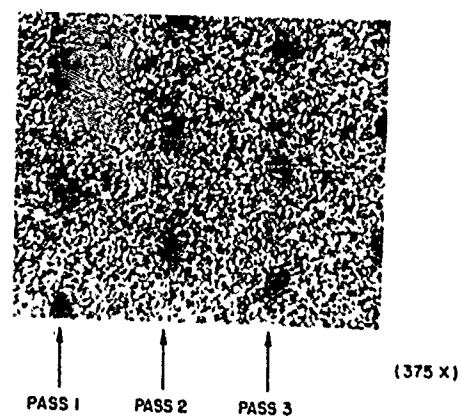


FIGURE 14 Quartz typical [illegible]



[illegible]

COMPUTERIZED PROCESS CONTROL FOR SYNTHETIC QUARTZ GROWTH

by

D. W. Rudd, A. R. Flora and N. C. Lias
Western Electric Co., North Andover, Mass.

ABSTRACT

This paper describes the computerization of a process to produce synthetic quartz by hydrothermal crystallization. In formulating the general program it was decided to use the computer not only as a control device but also as a storage facility for all engineering information concerned with this process. To accomplish this aspect of the program, all process variables were reduced to functions of growth rate and acoustic loss (Q^{-1}). Thus, the crystallization of certain types of quartz crystals may be expressed in simple equation form. These process equations are also used as the basis for the control of programmed crystal growth in which the operational variables of temperature and pressure are changed during the crystallization cycle in accordance with a suitable process equation.

A G. E. 4050 computer was chosen as the control device. Certain aspects of this machine which influenced its choice are described. A description of the software is presented and the various programs required for routine operation are discussed. Most of the control effort involved the use of a control algorithm. The development and use of this algorithm is presented. Finally, an economic appraisal of direct digital control is developed and compared to such schemes as supervisory control and the use of mechanical cams to generate a temperature profile for the crystallization cycle.

INTRODUCTION

Recent investigations have shown that it is possible to cause the uniformity of basal cut synthetic quartz crystals to be increased by changing the temperatures during the growth cycle (1). Attempts to apply this technique to several autoclaves have shown that it is necessary to have some type of automatic device to effect the temperature changes. Efforts to grow different types of quartz crystals have shown that it is often necessary to employ vessels of different designs than had previously been used (2,3). In the Western Electric Co., various types of autoclaves are in use for specific purposes (2). Use of autoclaves of different designs has shown that they cannot be controlled to the same degree of flexibility or in some cases accuracy when using a standard analog control panel. This is due to the fact that these types of facilities are usually designed to control a specific type of autoclave so that when the design of the vessel is changed the control becomes less precise. The nature of quartz growing is such that while considerable experience is necessary to completely understand the process, when the autoclaves are running, the process does not require complete engineering attention. Advantage would be obtained if it were possible to store previously developed engineering knowledge so that information could be immediately obtained when needed. If it were desired to construct an additional growing facility it would be helpful to have engineering knowledge available so that the training period of new personnel would be kept to a minimum. An economy would result by having a centrally located computer to control the growing process. With such a device variously located growing facilities could be controlled from one location. Additionally, by utilizing direct digital control with a control algorithm, various types of vessels could be conveniently and accurately controlled. Finally, because of the ability to store scientific experience it is possible to perform designed experiments based upon this previously obtained knowledge. A computer with its fast operation is well equipped to perform this task.

EXPERIMENTAL

The principal obstacle in computerizing a physical process such as the growth of quartz crystals is the reduction of the process to mathematical relationships of sufficient simplicity to be entered on a program with some facility. It was decided to attempt to characterize the process in terms of growth rate and Q, since these variables are the ones of greatest interest when considering the individual merit of a grown crystal. Because there are many other operating variables which contribute to the ultimate Q and growth properties of a given crystal, it soon developed that some techniques must be formulated to circumvent this myriad of process variables. It was found that this could be accomplished by choosing specific operating conditions that produced a crystal of interest and by building around these conditions, process equations of utility and of sufficient simplicity to be programmable. With this scheme the process may now be controlled by a process equation at least for the crystallization governed by the conditions of the process equation. Other crystals with properties of interest can be handled in a similar manner with the final program being a

a constantly changing flexible array of process equations for all sorts of crystals with various usable properties.

Two aspects of the programming are of significance. These are the development of the process equations for the various crystal types and the control algorithm for the heater power input.

As a first step in assembly of data required for programming the growth process, a curve was constructed depicting the thickness of crystal as a function of the growing time. The operating conditions were 675°F crystallization temperatures 83.5% fill and 70°F difference in temperature between dissolving and crystallization zones. These conditions were chosen after preliminary experimental work indicated that a good crystal could be produced using them. The curve, Fig. 1, was constructed by performing interrupted runs, that is, terminating the runs at specific time increments and recording the thickness of the crystals at the end of these increments. After the data was plotted, an equation of the type (1)*:

$$G = \sum_{i=0}^{n} a_i t^i \qquad \text{(1) was developed}$$

Where,

G= crystal growth (inches)
t= time (days)
a= experimentally defined coefficients
i and n are integers with
i= consecutive values from 0 to ∞

Fitting of equation (1) was performed on an IBM 1620 computer. The derivitive of this equation, dG/dt, furnishes the actual growth rate. The limitation of equation (1) is for values of $t > 0.5$ days. Prior to 0.5 days the seed is etching since the solution is still essentially NaOH.

After the growth time data has been formulated it is necessary to relate it to the Q and Q distribution. For this purpose a crystal was grown under the conditions which established equation (1). A Y cut optical flat about 0.25 inch thick was prepared from this crystal and scanned at 3500 CM^{-1}. Next a shadowgram of this flat was prepared and matched to the 3500 CM^{-1} scan so that positions within the growth region may be identified with Q (the 3500 CM^{-1} scan). This analysis is shown in Fig. 2. Knowing the magnification effect of the camera and utilizing certain landmarks (principally the seed), the 3500 CM^{-1} scan can be divided into segments (numbers ascending in chronological order on either side of the seed) proportional to thickness segments in the original crystal. Each of these thickness segments may be identified by a Q value expressed in terms of α/α_0.

* For example for the above temperature and pressure condition equation (1) becomes $G = -88.9 + .9 \ldots T - 6.25 \times 10^{-3} T^3 + \ldots$ (see Fig. 1)

The value is obtained by noting the absorption difference between the 3800 CM^{-1}, and 3500 CM^{-1}. Thus Fig. 2 relates the Q distribution of the crystal grown under the conditions of equation (1) to various thickness segments in the growth region. With this analysis, the Q as a function of growth rate throughout the crystal may be determined.

This information is obtained by solving equation (1) for t and d (thickness)/dt for each of the thickness segments of Fig. 2. After performing this operation, the Q for each thickness segment is identified with its growth rate so that the Q as a function of rate is obtained for a crystal grown under the operational conditions of equation (1). These results are shown in Table 1, Fig. 3. Table 1 gives the average growth rate, the final thickness of the crystal, the Q distribution throughout the crystal, the average Q and the range in Q. The equations used to generate Table 1 are referred to as process equations for the crystallization dictated by the temperatures and pressures fitted to equation (1). This crystallization is for the condition of fixed operating temperature. Other crystallizations of a fixed temperature variety that are of interest can be handled in the same manner so that an array of process equations are assembled each with a print out such as Table 1 associated with it.

In addition to this information storage ability it was desired to use the computer as a control device. In order to accomplish this it was necessary to arrange the machine so as to fill the role previously assumed by the analog control panel. While this was being considered, it was desired to have the computer remotely located from the growing facility. This is a desirable feature for two reasons: first, in the event of a violent autoclave failure an expensive device such as a computer would not be destroyed and second, most computers are sufficiently large to control large numbers of autoclaves and could be used to control several remotely located growing facilities. Therefore, a computer was chosen that could control the growing facility by means of telephone transmission lines. A diagram of the physical arrangement is shown in Fig. 4. Thermocouple and pressure transducer voltages are fed to a remote scanner which communicates with the computer by telephone data sets. The computer analyzes the readings and makes corrections by a reverse loop. Results are printed out on the typer when desired. Additional instructions may be entered by means of the operators console.

In order to use the computer as a control device it was necessary to arrange the machine so as to perform the task previously performed by the analog control panel. The analog controllers are continuous devices in that they compare actual and desired temperatures, then modify the power input to the autoclave to compensate for changes. In the case of the present control facility the design of the controller is centered about a

6-inch autoclave that was originally used to grow quartz. Direct digital control (DDC) is unlike an analog system in that a DDC system is discrete; that is a digital system will read temperatures at discrete time intervals which cause temperature changes recorded by the computer to be stepwise to avoid oscillations (a periodic or sine wave-like wandering of actual temperature about the set point). Care must be taken in a digital system not to over compensate for sudden changes in recorded temperature. The DDC algorithm calculates an input power based upon control parameters of the process (thermal mass of the autoclave, past temperature changes, difference between the desired and actual temperature) to change the temperature of the autoclave asymptotically to the desired temperature. There are two aspects of the control function. These consist of the algorithm itself which is a series of equations that affect input current changes and of the autoclave model, a mathematical relationship which serves as the constraint for the algorithm. The model is defined as:

$$\frac{dy}{dt} + ay = kx + b + c$$

Where

y = temperature (millivolts)
t = time (minutes)
x = input current to heaters (milliamps)
a = an anticipatory function based on the thermal lag of the vessel.
k = a gain constant
b = a model-autoclave factor relating the model with the actual autoclave response
c = a constant

The coefficients a, k and c were determined experimentally, but were found to be different for different temperature ranges and autoclave zones (bottom and top). For example, the values of these coefficients are different for 300°F than they are for 600°F, also the model is somewhat different for the bottom zone than the top. Because of the temperature changes, the model is solved for new values of current and new values of b every three minutes. New values of a and k are used three times during the entire growth cycle (twice during the warm-up cycle and once during the run itself after equillibrium is established) because of the nature of the autoclave. It was found experimentally that at full power it required 12 minutes for the effect on temperature to be observed. It was further found that there are a series of power inputs that will effect this temperature rise in a given time period. For example, a 10 degree rise in temperature applying full power can be effected in a shorter time than 1% input. A series of these temperature rise-power input curves were constructed to observe their form. A representative series is shown in Fig. 5. From this it was found that there is always a power setting to achieve a given rise in temperature in a given time.

It can be seen from Fig. 5 that if one wished to raise the temperature 10 degrees, one could put in full power for a time, record the temperature then put in a small amount of power so that the temperature would not over shoot the set point. If current changes were made in three minute intervals, the initial effect on temperature would not be felt for twelve minutes. However, by this time 4 other changes would be made so that the set point temperature is asymptotically reached and controlled by digital power inputs. It may be considered also that the algorithm is effecting a quasi-analog control over the autoclave. Two examples of control will be given to illustrate the action of the algorithm and the model.

Autoclave in run conditions

Consider the case of an autoclave in equilibrium. For simplicity, consider only the top zone of the vessel. Imagine also that the top zone is being controlled at or near a set point of 670°F. If the temperature is low, the computer senses this low temperature and by means of the algorithm instructs a current to be sent to the heaters. This amount of current is the amount required to cause an asymptotic change in 60 minutes.* Three minutes later another reading is taken and the power changed over the original setting. Imagine the temperature for some unspecified reason to be 668°F. Imagine further that based on previous and present power inputs that we would still be at 668°F in twelve minutes or one thermal time delay. This estimate is performed by the algorithm. The algorithm then predicts the power required to raise the temperature asymptotically two degrees in 60 minutes. Three minutes later the algorithm predicts, based on this power setting, what the temperature should be and makes another calculation of the power requirement to raise the temperature to the desired 670° in 60 minutes. This is repeated every three minutes during the growth cycle.

Autoclave in warm-up condition

The control problem in the warmup condition is more complicated in that the autoclave is not in a steady state condition. Here we have no set point to control about, but are attempting to reach a temperature level quite remote from room temperature (700°BT 630°TT at which point the warm-up heaters are normally shut off[5]). The autoclave is started up in the normal manner. That is, power is supplied to the bottom zone in intervals until full power is being supplied. The problem at hand is to control the power input to the top zone so that a difference of 70° between top and bottom zones is preserved. When the autoclave reaches a 50°Δ T from rest, a small amount of power (2.0 milliamps) is supplied to the top zone. This has been found to have a stabilizing effect on future power inputs. When the top temperature reaches 300°F, the bottom temperature is about 370°F. The algorithm then predicts what

* An arbitrary time interval chosen for convenience and other practical considerations.

the temperature will be in 60 minutes in three minute intervals. The top temperature will be controlled at these temperatures -70° to establish the 70°Δ T. The temperature settings also have a power input to effect the actual temperature change. With this scheme it has been found that the maximum departure from set point temperature is ± 1°F at 750°F.

It has previously been observed that the Q of fast grown basal crystals varies in quality through all of the crystallographic regions of the crystal [4,6,7]. This has been attributed to impurities in the crystal which are segregated in response to variations in growth rate during the growth process. These variations in growth rate are believed to be a consequence of the autoclave design and its relation to the growing crystal in addition to the natural anisotropic nature of the crystal. For example, during the early days of the synthesis, the size of the crystal is not sufficiently large to inhibit the efficiency of the convective flow of solvated quartz from a large nutrient area dissolving zone to the crystallization zone of the autoclave. The crystals therefore grow at a rapid rate. This effect, however, diminishes with time because as the crystals grow they occupy an increasingly larger portion of the upper autoclave zone and this decrease in rate leads in turn to a rejection of impurities with a subsequent enhancement in Q. Changing the growth rate of the crystal during the growth cycle then can lead to a more even distribution of Q throughout the crystal. By changing the temperature during the growth cycle, the rates may be changed as well, and it has been found possible to completely reverse the normal OH distribution in the Z region by this technique. Therefore it is possible to elevate the Q of the crystal in the vicinity of the seed and diminish the Q at the outer Z extremity, leading to a more even Q throughout the growth. By reducing the time interval between the temperature changes, a more even OH distribution is obtainable. The consequence of this is a higher Q for a programmed growth than an unprogrammed growth at an equivalent over all growth rate.

Several runs were completed to determine the minimal time increment that could be obtained for manual selection of desired temperature settings. This was determined to be two and one half days. At the conclusion of the runs the results were analyzed and are shown in Fig. 6. This shows a shadowgram and a Q trace of a programmed run the object of which is to make the Q of the grown crystal high and uniform. Two distinct land marks are seen, the consequence of a large temperature drop purposefully imposed to produce the land mark and recording the day on which this event took place. The shadowgram was divided into equal segments based on the thickness of the seed and the magnification effect of the camera. The magnification was determined by dividing the observed thickness of the optical flat in the Z direction by the total number of thickness segments ruled off in the shadowgram. Each thickness segment corresponds to 0.056" (actual seed thickness) curves relating the temperature of the autoclave as a function of time (days) were drawn. It can be seen that a sharp dip in top temperature occurred after 16 days and it was assumed that this adjustment produced the defect shown in the shadowgram. The rate for the first 16 days was determined by dividing the total thickness from the seed to this defect (1.00") by the time required to reach this thickness (16 days) which yields 62.6 mils/day. The rate for the remainder of the run was determined in similar manner; by subtracting 16 from the total growing time (24.9 days) and dividing this number into the remaining thickness (0.586"). This yielded 65.8 mils per day so that

Ave Growth Rate (first 16 days) = 62.6 mils/day

Ave Growth Rate (remainder) = 65.8 mils/day

It is assumed that the 3.2 mil/day difference is within the accuracy of the technique used to obtain it so that the growth rate was essentially linear.

RESULTS

Remote location of the computer was a desirable feature when initial plans were being formulated concerning the project. It was desired to keep such a complex device removed from the growing location so that in the unlikely event of an accident the machine would not be damaged. The computer chosen as a control device was a GE 4050. This machine was selected due to its having the feature of being able to communicate with the remote control site by means of existing telephone lines. Recently other manufactures supply equipment with this capability. This remote control ability is also of advantage when considering the necessity of establishing an additional growing facility.

Start up and termination of runs is accomplished in the growing area by means of the operators console, Fig. 4. Programs such as start up, termination, autoclave status each have a program number associated with them. These numbers are on one dial of the operators console and can be dialed at will. Additionally, every hour the status of each autoclave is printed out. This report lists the temperature, pressure and departure from set point and this also is printed out. Other programs such as autoclave history etc. are likewise obtainable in the same manner. The operator console also contains the alarm system so that in the event of a large departure from set point value an alarm sounds. At the present time the old analog control panel is

utilized as a back up system when it is desired to work on additional programming and as a safety precaution in the event of a computer failure. In the future this will be replaced with a simpler and much less expensive device.

Due to the ability of the computer to hold the desired temperature to close tolerance, this type of control is most desirable if other economic factors permit its use. But the ability to produce a programmed growth makes computer control exceedingly attractive. Not only are the over all growth rates linear but generally in terms of average growth rate, the Q is higher for a programmed crystal than for an unprogrammed crystal of similar average growth rates. This can be seen in Fig. 7 which shows the results of a Q analysis of two crystals having similar growth rate. While growth rate is a somewhat elusive term to correlate directly to Q, nevertheless the general effect is generally of the type shown in Fig. 7. While it is possible to obtain higher Q at growth rates of 60-70 mils per day[7], these are not always routinely reproducible while the reproducibly of a programmed growth under computer control is in the order of 90% or less than 10% deviation in Q from one batch to the next.

Other schemes were investigated to determine something of the economics of the procedure. In one scheme a mechanical cam was integrated into the existing control panel. The cam was connected to the set point mechanism and the timer. Physically, the cam is a disc ruled with lines concentrically drawn corresponding to temperature settings. A program is actually cut on the circumference of the disc. While this procedure is inexpensive it suffers from the obvious difficulty of inaccuracy in cutting and ruling the disc. Additionally, there is the further disadvantage of having the control ability limited by the accuracy of the control panel. In a second scheme a supervisory control was investigated from a cost standpoint. Here a programmable device is interfaced with the existing panel. It receives instructions from a previously designed program but requires the functioning of the existing control panel to effect the actual temperature changes. The economic picture is dictated to a great extent by the number of autoclaves that are to be controlled. This is because the cam and supervisory schemes require a control panel as the interfacing to the autoclave. The cost of a suitable control panel approaches that of the GE 4050 computer at about the control level of 25 autoclaves. Therefore, for about 25 autoclaves the computer and control panel scheme are about the same in cost. For over 25 autoclaves the computer cost remains the same but the panel cost increases. Therefore computer control becomes attractive when it is desired to control a large number of autoclaves producing a variety of quartz crystals.

CONCLUSION

Computerization of the growing process has yielded a process which can both control the process completely and store existing knowledge on the growth of quartz. The direct digital control is so devised that the controlling parameters are independent of the type and design of the autoclave to be controlled. Use of a computer can effect a more accurate temperature control than other commercially available non-computerized control facilities. Additionally, the computerized system can set up and control programmed quartz growth. Finally, economic analysis of the process indicates that computer control becomes attractive as the number of autoclaves becomes large.

ACKNOWLEDGEMENTS

The authors wish to express thanks to R. E. Dubois for his assistance in formulating much of the computer software and for securing the proper mechanical devices associated with the computer and for the computer itself. Also, thanks is expressed to J. R. Boie for experimental assistance during the course of the investigation. Thanks are also extended to R. K. Bhattacharyya, W. E. Co. - ERC for his assistance in formulating the control algorithm.

REFERENCES

1. R. E. Dubois and D. W. Rudd, Western Electric Eng. II, No. 1, 14 (1967)
2. A. A. Ballman, R. A. Laudise, D. W. Rudd et al. J. Appl. Opt. 1, No. 7, 1387 (1968)
3. A. A. Ballman and D. W. Rudd, to be published
4. D. W. Rudd, E. E. Houghton and W. J. Carroll, Western Electric Eng. 10, No. 1, 22 (1966)
5. R. A. Laudise and R. A. Sullivan, Chem. Eng. Progr., 55, 1 (1959)
6. A. A. Ballman and D. W. Rudd, Western Electric Eng. 10, 1, 3 (1965)
7. D. W. Rudd and N. C. Lias, Chem. Eng. Pro. Symposium Series, Vol. 63, No. 76, 32 (1967).

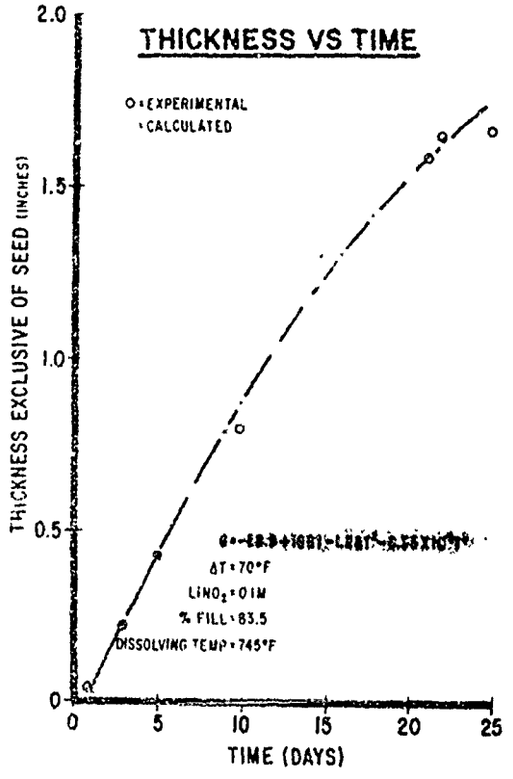


FIGURE 1

FIG 3

TABLE 1

Q vs GROWTH RATE DATA-GROWTH TEMPERATURE = 675°F ΔT=70°F % FILL = 83

REGION	TOTAL GROWTH (MILS)	TIME (DAYS)	GROWTH RATE (MILS/DAY)	$r \times 10^{-5}$	$Q_r \times 10^{-5}$
1	122	1 46	[illegible]	0 357	0 357 0 3[illegible]
2	244	2 46	113	0 432	0 464 0 401
3	366	3 51	109	0 452	0 411 0 493
4	488	4 68	99.3	0 517	0 574-0 [illegible]
5	610	5 97	90 2	0 583	0 639-0 52
6	730	7 39	81 7	0 612	0 659 0 564
7	854	8 96	74.0	0 665	0 72 0 [illegible]
8	976	10 7	67 5	0 726	0 [illegible]
9	1098	12.6	61 4	0 812	1 [illegible]
10	1220	14 7	55 v	0 85[illegible]	0 [illegible]
11	1342	17.0	50 5	1 0	1 2 [illegible]
12	1464	19 5	44 5	2 [illegible]	2 [illegible]
13	1586	22 5	36 2	> 2 2	> 2 2

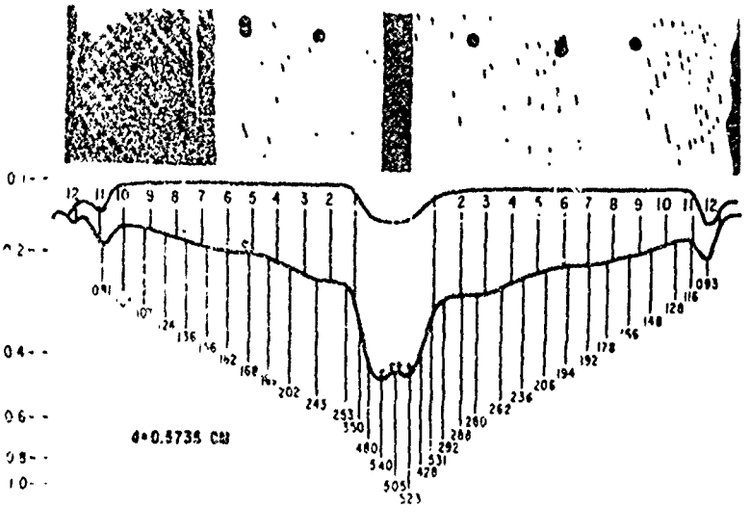
AVG GROWTH RATE = 78 mils/day

TOTAL GROWTH / TOTAL GROWING TIME = 70 mils day

OBSERVED GROWTH / TOTAL GROWING TIME = 67 mils/day

Q at $\overline{C_R}$ = 665,000

Q Range = 350,000 2,250,000



SHADOWGRAM - Q ASSIGNMENT

FIG 4

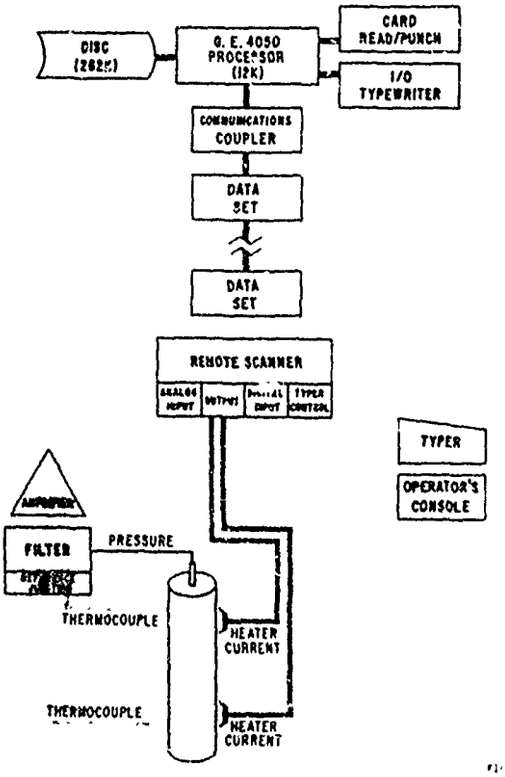
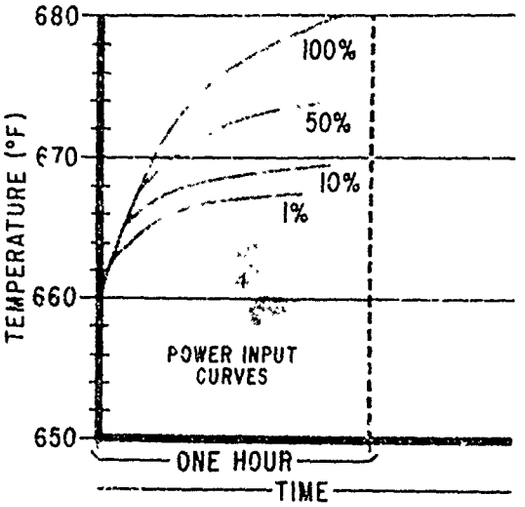
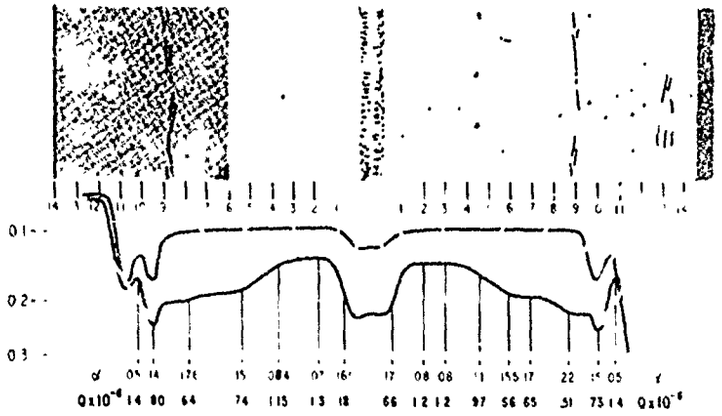


FIG 5

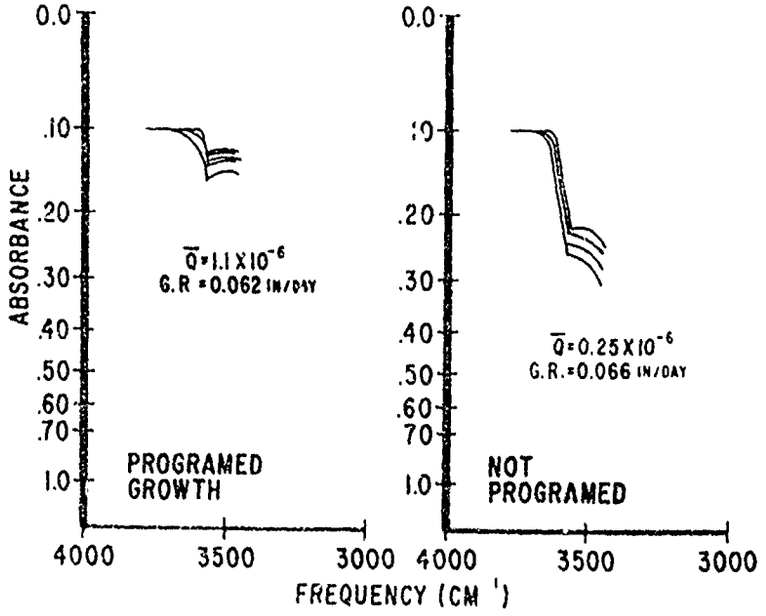
REPRESENTATIVE CURVES RELATING ASYMPTOTIC RISE IN TEMPERATURE FOR DIFFERENT VALUES OF POWER INPUT





SHADOWGRAM - Q ASSIGNMENT FOR PROGRAMMED GROWTH

FIG 5



EFFECTS OF GAMMA IRRADIATION ON FREQUENCY STABILITY OF 5TH OVERTONE CRYSTAL OSCILLATORS

Carl A. Berg - James R. Erickson
GIBBS MANUFACTURING AND RESEARCH CORPORATION
JANESVILLE, WISC.

Summary

This report presents the results of a program of subjecting a series of precision oscillators to gamma radiation. The program objective was to advance present knowledge of the effects of a radiation environment associated with nuclear weapons upon this type of oscillator. Developing an oscillator to maintain frequency stability during and after exposure was the prime objective.

Program effort concentrated on selection and use of "swept" quartz crystal resonators fabricated with precision techniques. Semiconductor device selection also constituted an important portion of the program.

A group of 3 production run oscillator units were modified with the selected components. These oscillators were subjected to a gamma radiation environment at the White Sands Missile Range nuclear test facility. Radiation doses exceeding 3×10^6 rads were used. Frequency variations of less than 14ppl0^8 without oscillator cessation were achieved. This represents an improvement factor of at least 50 to 1 over non-modified oscillators similarly tested.

Introduction

Information is presented to indicate progress made to advance present knowledge of the effects of a radiation environment associated with nuclear weapons upon quartz crystal controlled oscillators. Work on this program was accomplished for the U.S. Air Force Avionics Laboratory, Wright-Patterson Air Force Base, Ohio, under contract F33615-68-C-1589. Permission for presentation and publication of this paper has been granted by the Contracting Officer. Tests were performed on oscillators of the precision high stability type.

The severe weight and size penalties that would be imposed by the use of suitable shielding was avoided to the fullest extent. It then became necessary to evaluate, test and use components offering maximum stability during and after exposure to radiation.

Preliminary studies and evaluation indicated that semiconductors and the quartz crystal are the most susceptible to radiation. This report will be concerned primarily with the improved stability that was achieved by the use of improved transistors and crystal resonators.

Three (3) high stability type oscillators were modified for the purposes of this study program. Two (2) of the units, which had previously been subjected to gamma and fast neutron radiation, were furnished to the government. These oscillators and the third unit, supplied by Gibbs Manufacturing and Research Corporation, were modified by substitution of the improved crystal resonators and semiconductors.

Modification of Oscillators

Resonators fabricated from "swept" cultured quartz have been shown to display increased resistance to radiation. Improvement factors of three or more over natural quartz crystals have been achieved. (1) Pure Z-growth cultured quartz is "swept" by application of high voltage at elevated temperature to materially reduce the alkali and hydrogen ions within the quartz itself.(2) Type CR-71/U high stability 5th overtone crystals fabricated with "swept" quartz were substituted into the oscillators in place of the usual natural quartz CR-71/U crystal units.

Figure 1 summarizes the semiconductor complement of each oscillator as modified. No radiation resistant types were used in Oscillator #1. Prior to modification, the transistor complement of the oscillators comprised of the types 2N706, 2N1700 and MHT4501. The 2N706 was replaced with types 2N5200 and 2N5399. These are recently developed types specifically designed for radiation resistance. For comparison purposes and to gain additional information, the type 2N918 was also included as a substitute for the 2N706. The 2N918 is an improved version with greater gain and frequency response. It is also currently available as a military preferred type. Radiation resistant type transistors in the medium power category were not readily available as direct replacements for the oven heater driver transistors. Consequently, high frequency types designed for use at VHF frequencies were substituted for the types 2N1700, 2N339 and MHT4501.

Radiation Test Procedure

The oscillators were subjected to gamma radiation at the Nuclear Effects Branch, White Sands Missile Range, N.M. A linear electron accelerator (LINAC) provided the gamma radiation energy. All oscillators were kept operating nearly continuously by a minimum of transfer time from one power supply to another.

Oscillator #1 was powered by battery during transit to the test site. Prior to testing, oscillators #2 and #3 were energized and operated continuously for over two days upon arrival at the test site. Figure 2 shows a view of an oscillator mounted in front of the LINAC port. It was positioned with the crystal centered on the line of greatest gamma intensity.

Figure 3 provides a view of the LINAC port. The LINAC output consists of single intense pulses of high energy electromagnetic radiation of short time duration. The pulse energy level was established at 17 MEV and a duration of 10 microseconds. Any desired number of pulses at a rate of 10 pulses per second can be furnished by the LINAC. A shot consists of LINAC operation for one or more pulses.

Gamma Dose Calibration

Initial test setup consisted of establishing the gamma radiation level impinging on the oscillator surface. Calibration of the gamma radiation level consists of attaching dosimeters on the oscillator sides. After exposure to a test shot, the dosimeters are developed and measured to provide the radiation energy level on a per pulse basis. Only a change of distance or position of the oscillator from the LINAC port required new dosimeter measurements. Gamma dosage was changed primarily by varying the number of pulses per shot.

Frequency Measurement

Test instrumentation, as shown in Figures 4 and 5, provided frequency measurement resolution of $1ppl0^9$. Oscillator frequency and output voltage were monitored continuously. However, because of the high RFI level during the actual operation of the LINAC, the frequency counter presented erroneous readings. Consequently, actual oscillator frequency deviations during the radiation period were not obtainable with sufficient accuracy. However, frequency measurement during radiation exposure was accomplished by means of counting the beat of the lissajou pattern displayed on the oscilloscope. For each oscillator, it's frequency just prior to the initial calibration and trial shot served as the reference for subsequent gamma radiation shots.

Test Results

None of the oscillators tested experienced any cessation or output amplitude change at any time during exposure to gamma radiation.

Frequency vs. Radiation

Typical observed radiation effects are illustrated in Figure 6 through 12. These are the immediate and short term frequency changes occurring during and after each gamma shot. The dashed portion of each curve indicates actual exposure time (shot) and also shows the frequency at the start of each shot.

For oscillator #1, shots #1 and #5 in Figure 6 show the increased frequency offset occurring as the duration of the gamma dose increases. These curves show that the frequency offset, initially positive for low level radiation, becomes negative for increasingly larger doses of gamma radiation. With the completion of shot #5, the total gamma dose to which oscillator #1 was exposed had reached approximately 400K Rads.

Figure 7 further shows the effect of continuing exposure to gamma radiation. Shot #8 shows the frequency offset reached for a total gamma dose exceeding 1.2 M Rads. These curves (shots #6 and #8) also show a transient frequency increase indicating induced heating during exposure to radiation.

To further verify that a temperature change was taking place within the oscillator, the crystal oven temperature was reduced slightly below the zero T.C. point of the crystal. After a 24 hour restabilizing period, the oscillator was then subjected to a small gamma dose, shot #9, followed by a larger dose as shown in Figure 8. These curves now show a transient frequency decrease during radiation as the induced heating causes the crystal temperature to increase toward the zero T.C. point. After each shot approximately 12 minutes was required for frequency to restabilize which was essentially the same as the recovery time following other shots.

After the oven temperature of oscillator #1 was readjusted to it's original temperature, it was subjected to additional radiation at higher intensity per pulse. Figure 9 shows the further increase of frequency offset. Shot #15 shows the effect of increasing the radiation intensity to 87 Rads per pulse. Although the exposure intensity was further increased to 146 Rads per pulse, shot #19 shows that the transient frequency change is also dependent on the duration of exposure. After the completion of shot #19, the total gamma dose exposure exceeded 3.3M Rads with the permanent frequency offset reaching approximately $-12ppl0^8$.

The exposure of oscillator #2 to gamma radiation was similar to that of oscillator #1 with the exception that the crystal oven temperature was not changed. In Figure 10, two effects of increasing the intensity of radiation is shown. One, the dashed lines of each curve indicate the increase of the transient frequency change as the radiation intensity is increased. Each indicated shot was of equal duration. Secondly, the net negative frequency offset is

a function of both radiation intensity and accumulated or total dose. Total gamma dose at the completion of shot #3, #13 and #17 was 45K Rads, 1.47M Rads and 3.36M Rads respectively. The offset frequency typically continued to drift an additional 2 to 3ppl10^8 for elapsed times of approximately 24 hours between shots.

In Figure 11 the effects of two gamma doses on oscillator #2 are compared. They were of equal duration and equal intensity but separated in time by 27 hours. Total gamma dose at completion of shot #5 was 255K Rads and 1.33M Rads for shot #10. The oscillator frequency continued to drift, reaching -7.8ppl10^8 approximately 18 hours after shot #10.

Oscillator #3 was subjected to one-third the total gamma dose that oscillator #1 and #2 received. As shown in Figure 12, with the exception of the exploratory shot #1, the oscillator was subjected to one continuous dose (shot #2) which required approximately one hour to complete. The transient frequency change was not as great as that experienced by the other oscillators. This would indicate that the crystal and oven stabilized temporarily at a new operating temperature during the shot. A count of the lissajous pattern beat established a maximum frequency change of ±3ppl10^8 during the radiation period.

A plot of frequency offset versus gamma radiation is shown in Figure 13. This presents a composite view of the frequency offsets experienced by all three oscillators. The spread as indicated here represents the trend of frequency offset as a function of total gamma radiation dose. The long term effect of radiation is also included. For example, the frequency offset of oscillator #1 at the total dose of 3.3M Rads was -12.1ppl10^8 for an elapsed time of 47 minutes after exposure. The frequency offset continued to increase with time to -13.3ppl10^8 at approximately 4.6 days of elapsed time.

Long Term Stability

Oscillator #1: Frequency stability prior to and after radiation exposure is shown in Figure 14. For the pre-radiation drift rate to become less than 1ppl10^9 per day required approximately 9 days of continuous operation (Curve A).

Four days after the total gamma dose of 3.3×10^6 rads, the frequency offset was approximately 13ppl10^8 below the initial reference frequency at the start of the radiation test. Any frequency deviations occurring during transit and power interruptions are also included. At this time the oscillator frequency was reset to zero offset (Curve B). Aging in the positive direction proceeded at a rate of 3ppl10^{10} per day. Twenty-seven days later, aging had become negative at the rate of approximately 1.7ppl10^{10} per day.

The oscillator was then de-energized for a period of one day. Curve C shows the resumption of the usual frequency restabilizing characteristics of this type oscillator. Reduction of the drift rate to less than 1ppl10^9 per day required less than six days. The drift of approximately -4ppl10^{10}/day after twenty-five days was approximately 28% less than the pre-radiation drift for the same period of continuous operation. The off time of one day did not cause reversion to the gamma radiation induced frequency offset and drift. The oscillator merely repeated the usual pre-radiation long term frequency drift characteristics.

Oscillator #2: Figure 15 shows the frequency drift occurring during three different long term test periods. Curve A indicates the frequency drift prior to modification. The crystal was a natural quartz unit. Frequency drift continued to exceed 1ppl10^9 per day after fifteen days of operation. The improved frequency stability after modification is shown by Curve B. Less than six days were required to reach a drift rate of less than 1ppl10^9 per day.

Post gamma radiation dose (over 3.6×10^6 rads) versus frequency stability, Curve C, is similar to that of oscillator #1. At the elapsed time of 4.7 days after final radiation dose, the frequency offset was low by approximately 12ppl10^8 as referred to the pre-radiation reference frequency. Included in this time period is three days during which the oscillator was de-energized. Curve C begins at the end of the 4.7 day interval with the oscillator reset to zero offset. Maximum frequency drift in the positive direction was 7ppl10^{10} per day. This drift rate gradually decreased toward zero during about fifty-two days of continuous operation. Negative drift rate had commenced after 52 days and continued until the oscillator was de-energized.

Oscillator #3: This oscillator was exposed to a total gamma dose of approximately 1×10^6 rads in one continuous shot requiring one hour to complete. After an elapsed time of 5 days after the shot, the frequency offset was low by approximately 7ppl10^8 as referred to the pre-radiation reference frequency. During this period the oscillator was de-energized for 3.7 days. Figure 16 shows that the frequency drift rates are similar before and after radiation exposure. After 56 days of continuous operation, the drift rate had decreased to -3.4ppl10^{10} per day.

It is expedient at this point to review the past history of oscillators #1 and #2. Approximately two years prior to the present program, these oscillators were subjected (without any modifications) to similar radiation tests. The resonators were the type CR-71/U of natural quartz crystal construction.

Exposure of oscillator #1 to approximately 1×10^6 rads of gamma radiation caused a permanent frequency change of -6PPM. Oscillation also stopped for periods exceeding 45 minutes.

Oscillator #2, subjected to approximately 1.7×10^{12} NVT of fast neutron radiation, also stopped oscillating for periods of up to 3 minutes. Frequency change was less than $-1\text{ppl}0^8$. One transistor (low frequency type 2N1700) failed. No failures occurred among the type 2N706 transistors used at that time.

Conclusions

Salient Oscillator Improvements

1. Oscillator cessation did not occur.
2. No significant variation of output voltage occurred.
3. At these radiation levels, the high frequency transistors (ft of 500MHz and higher) performed equally as well as the radiation resistant types.
4. The "swept" quartz crystal resonators displayed significant reduced permanent frequency change. Less than $14\text{ppl}0^8$ frequency change occurred for a gamma dose exceeding 3.6×10^6 Rads. This constitutes at least a 45 to 1 reduction of frequency variation. For a dose of 1×10^6 Rads this factor increases to more than 80 to 1.
5. Substantially reduced aging rates are exhibited by "swept" quartz crystals. Less than 6 days was required to achieve drift rates of less than $1\text{ppl}0^9$ per day as compared to the natural quartz crystal which continued to drift at a rate of $4\text{ppl}0^9$ after 4 months of continuous operation after gamma radiation.

Figure 17 shows the total radiation dose to which each oscillator was subjected. The resulting permanent frequency change of each oscillator is also indicated. A comparison of the three oscillators indicates that a 3 to 1 increase of the gamma dose caused less than a doubling of the frequency change.

Short term frequency stability degradation was negligible as shown in Figure 18. The slight differences indicated for the 2 second and 20 second averaging periods could be attributed to instrumentation tolerances.

A transient frequency changes does occur during exposure to radiation. For crystals operating at the zero coefficient of temperature point; the frequency increases, temporarily over-riding other radiation induced effects. From this, it could be assumed that, this is primarily related to induced heating of the crystal blanks either directly or indirectly by a change of the oven operating temperature.

Permanent frequency change (negative offset) remains as a predominant result of gamma radiation. However, a reduction of nearly two orders of magnitude was achieved with the use of "swept" quartz resonators.

Since the purity of the swept quartz is the major factor in contributing to gamma radiation resistance, it is reasonable to assume that the holder will contribute second order effects. Recent advances in such areas as high temperature bakeout, high vacuum flameless sealing and extreme cleanliness have made possible the production of extremely stable quartz resonators. Combining these techniques with "swept" quartz fabrication should produce a quartz resonator with increased stability and immunity to a radiation environment.

References

1. J.M. Stanley - Radiation Effects in Frequency Control Devices - U.S. Army Electronics Command Proceedings of the 20th Annual Symposium on Frequency Control, PP82-102, Apr. 1966.

2. B. Sawyer - Quality in Cultured Quartz Sawyer Research Products Inc., Eastlake, Ohio

3. J.M. Wolfskill - Advancements in Production of 5MHz Fifth Overtone High Precision Crystal Units - Proceedings of the 22nd Annual Symposium on Frequency Control, PP89-117, Apr.1968.

4. E. Hafner & R.S. Blewer - Quartz Crystal Aging - Proceedings of the 22nd Annual Symposium on Frequency Control, PP136-154, Apr. 1968.

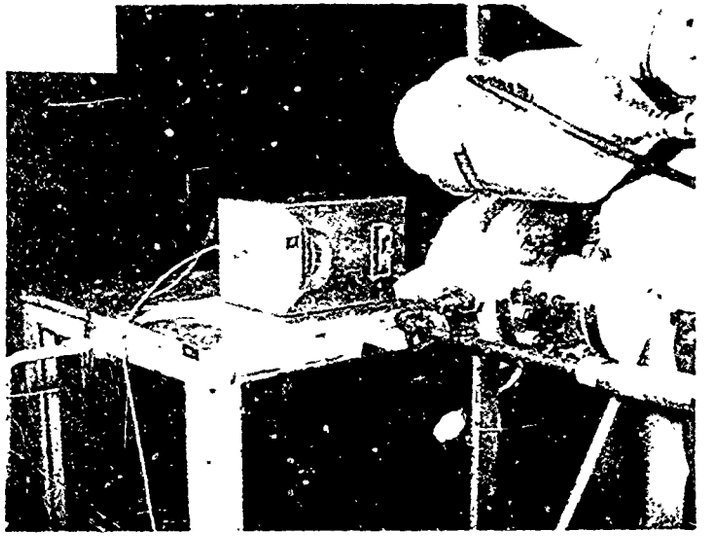
Acknowledgments

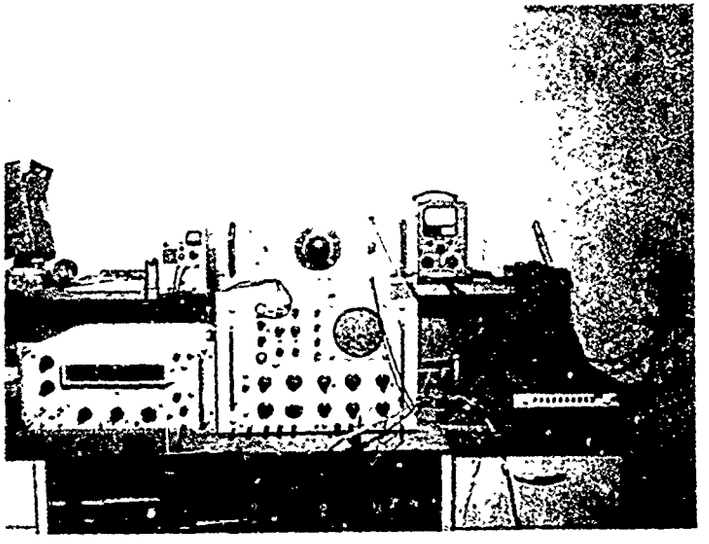
It is a pleasure to acknowledge the efforts and valuable assistance of Mr.C.W.Friend Air Force Avionics Lab., Wright-Patterson Air Force Base, Ohio.

The able assistance of Mr. Leo Flores Technical Coordinator and his group, WSMR Nuclear Effects Branch Facility, White Sands, N. Mexico, is gratefully acknowledged.

		2	3
	HI FREQ	RADIATION RESIST	RADIATION RESIST
	HI FREQ	HI FREQ	HI FREQ
	HI FREQ	HI FREQ	HI FREQ
	HI FREQ	RADIATION RESIST	RADIATION RESIST







Reproduced from best available copy.

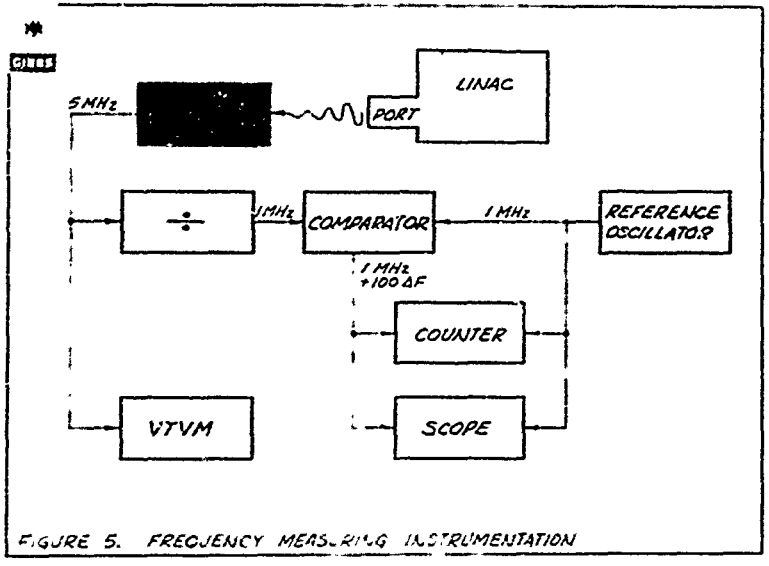


FIGURE 5. FREQUENCY MEASURING INSTRUMENTATION

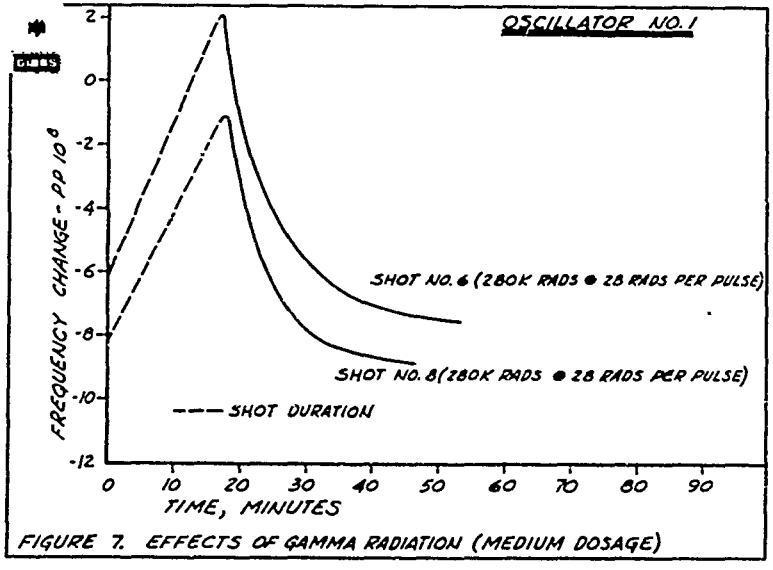


FIGURE 7. EFFECTS OF GAMMA RADIATION (MEDIUM DOSAGE)

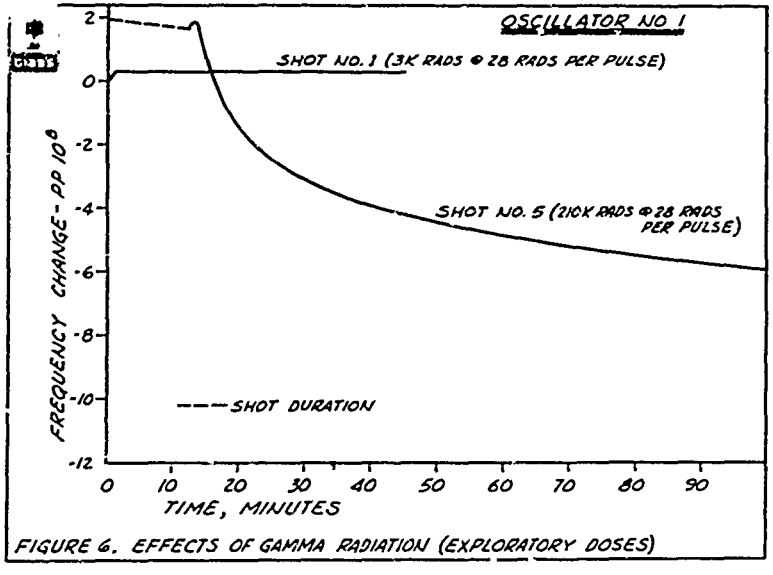


FIGURE 6. EFFECTS OF GAMMA RADIATION (EXPLORATORY DOSES)

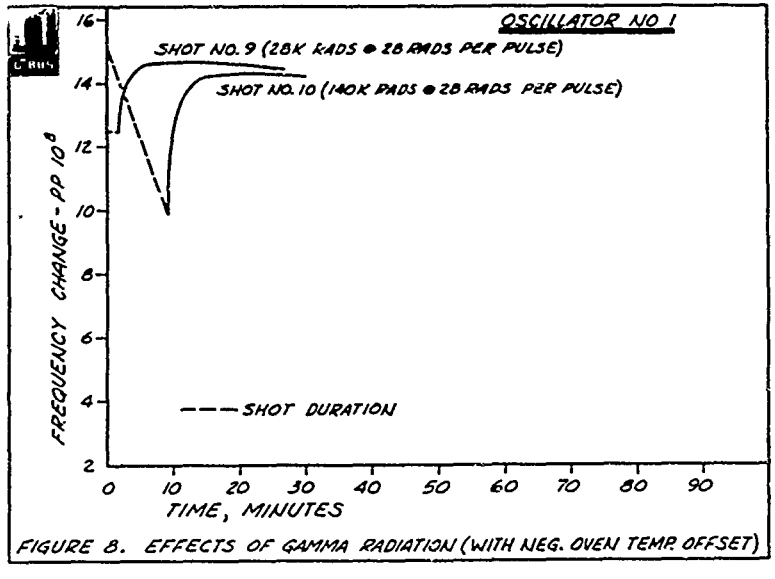


FIGURE 8. EFFECTS OF GAMMA RADIATION (WITH NEG. OVEN TEMP. OFFSET)

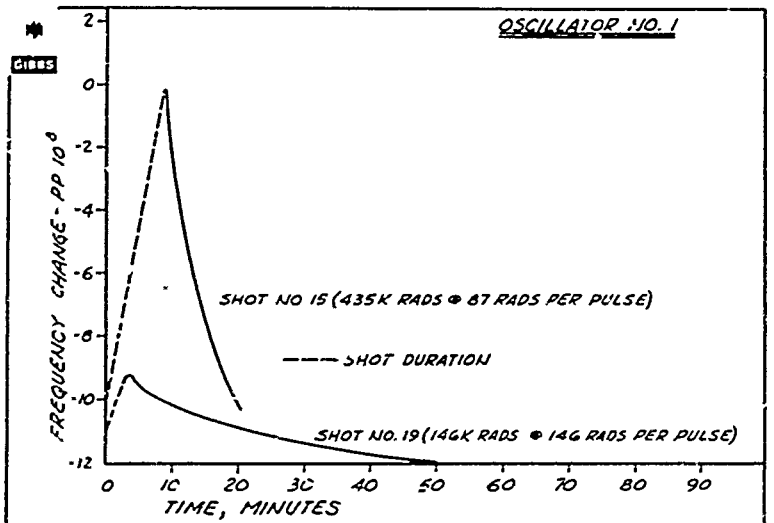


FIGURE 9. EFFECTS OF GAMMA RADIATION (OVEN ADJUSTED TO CRIGINAL TEMP)

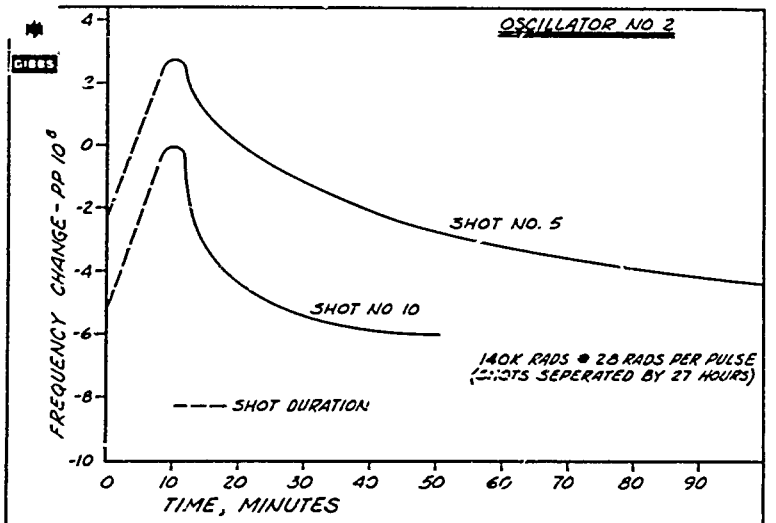


FIGURE 11. EFFECTS OF GAMMA RADIATION (SAME DOSE RATE)

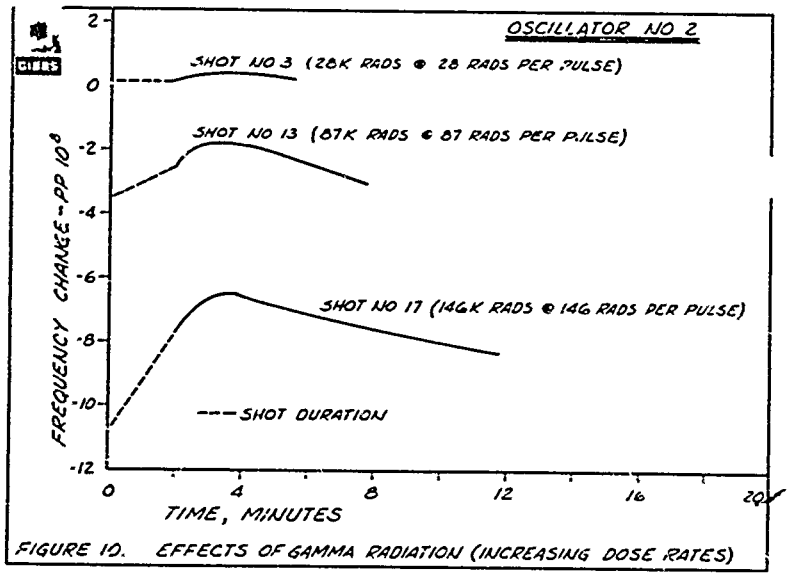


FIGURE 10. EFFECTS OF GAMMA RADIATION (INCREASING DOSE RATES)

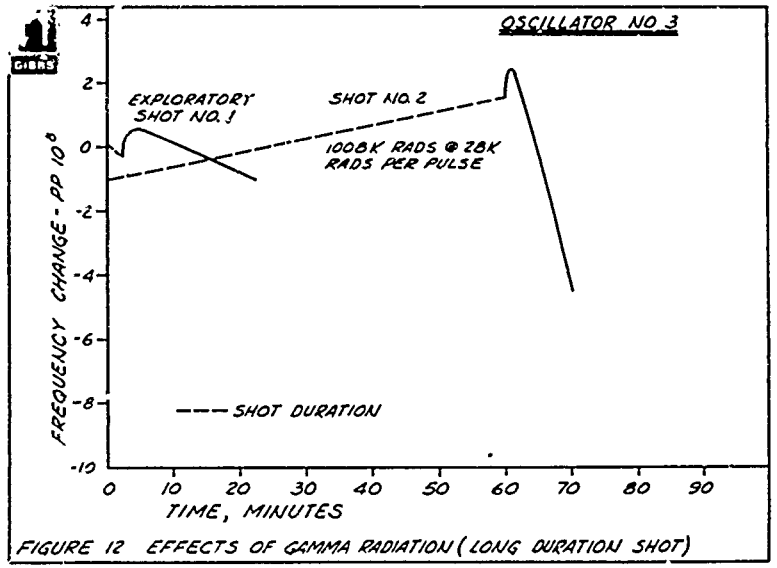


FIGURE 12 EFFECTS OF GAMMA RADIATION (LONG DURATION SHOT)

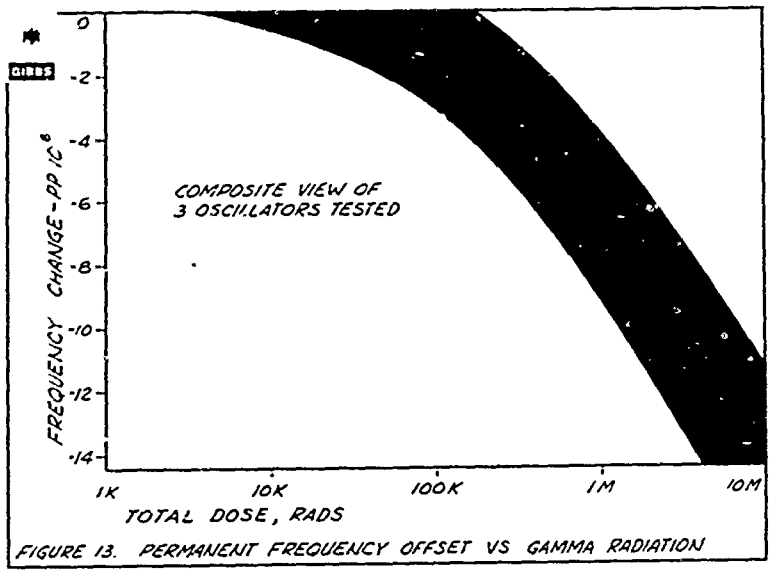


FIGURE 13. PERMANENT FREQUENCY OFFSET VS GAMMA RADIATION

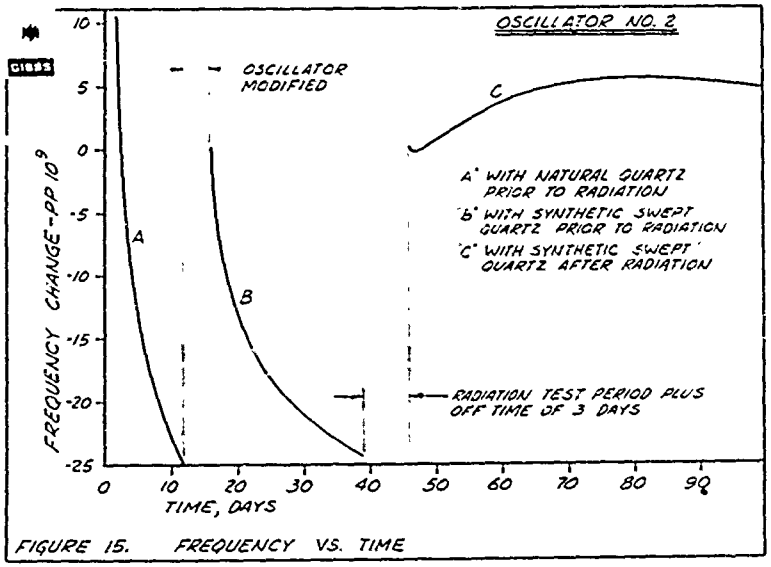


FIGURE 15. FREQUENCY VS. TIME

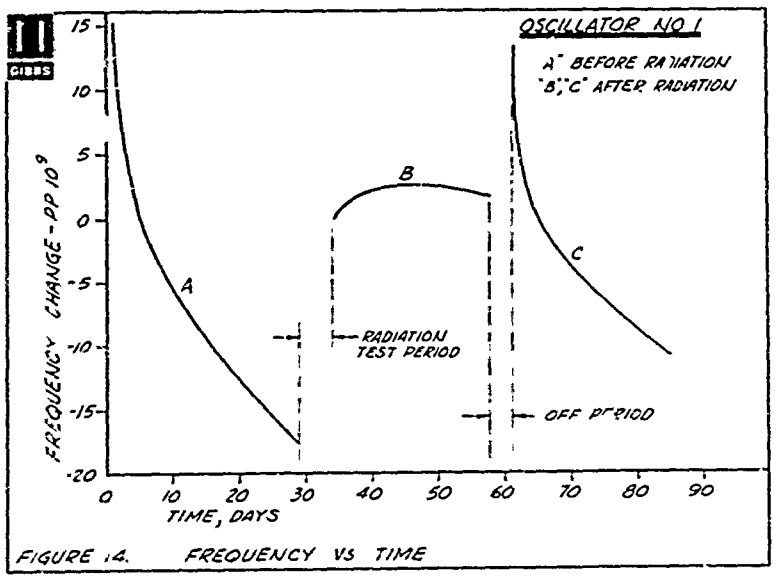


FIGURE 14. FREQUENCY VS TIME

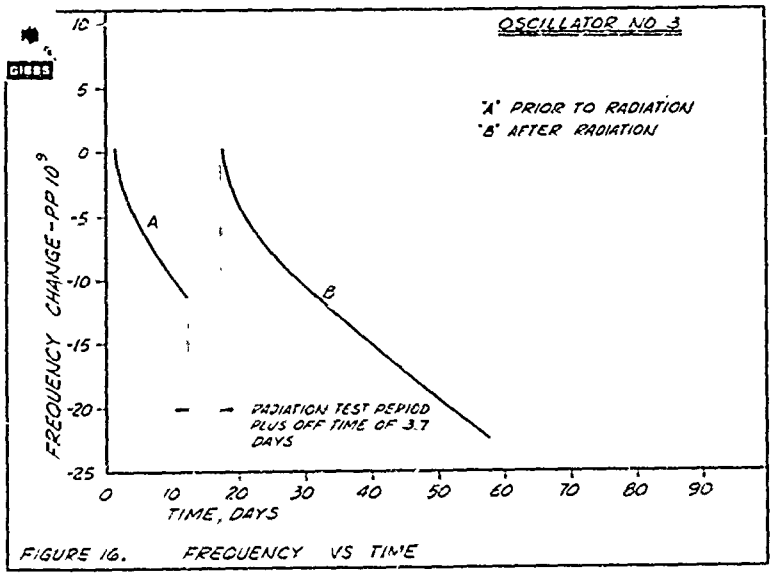


FIGURE 16. FREQUENCY VS TIME

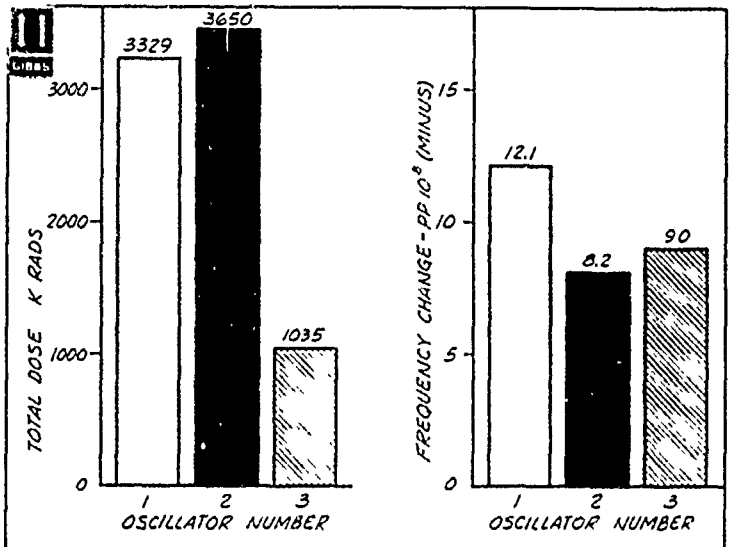


FIGURE 17. PERMANENT FREQUENCY OFFSET VS. GAMMA RADIATION

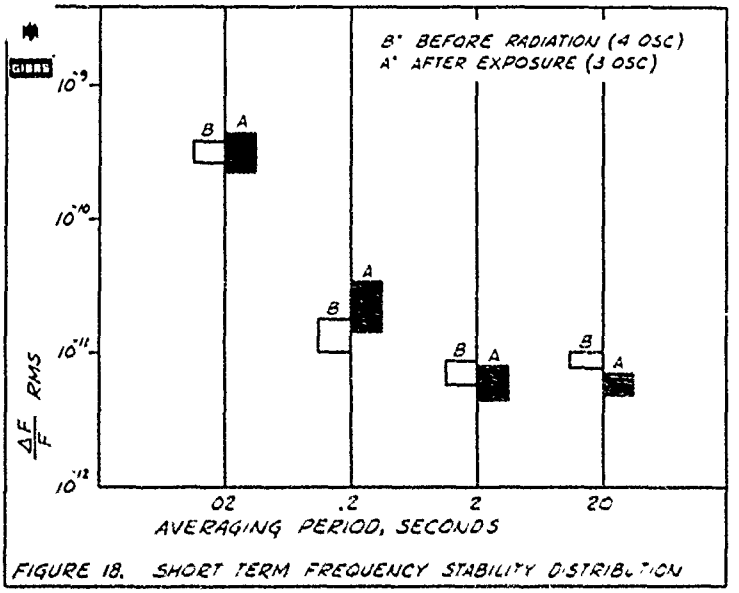


FIGURE 18. SHORT TERM FREQUENCY STABILITY DISTRIBUTION

A REPORT ON TCXO'S AND SEGMENTED COMPENSATION

D. E. Newell and H. Hinnan
CTS Knights, Inc.
Sandwich, Illinois

The first portion of the work presented in this paper was performed on the U. S. Army Electronics Research and Development Contract DAAB07-67-0433. At the 1968 Frequency Control Symposium a method of temperature compensating crystal oscillators using a segmented network was presented. This approach allows one group of components to control the frequency of the oscillator over a segment of the temperature range, and other groups to control over other segments. This approach permits the adjustment of a portion of the compensation network without affecting the remaining sections, thus setting up the boundary conditions required by automatic compensation equipment. During the past year many oscillators have been compensated utilizing the segmented approach. Perhaps the best way to discuss this method is to look at the results of several units.

Figure I shows a coarse curve and network to compensate an oscillator to $\pm$ 1 PPM over the temperature range. A second fine modulator is used to compensate for variations of the coarse network to obtain a stability of $\pm 2.10^{-7}$ over the range. Figure II shows the fine network voltage required and the network used to generate the voltage. It should be noted that the NPN transistors will always follow the maximum base voltage while the PNP's will follow the lowest base voltage. From the figure it can be seen that the fine curve becomes a very non-linear curve. Figure III shows the coarse voltage required and the network used to supply this voltage. Figure IV is the fine voltage required and the network used to generate this voltage for the oscillator shown in Figure III. Figures V and VI are fine network voltages and the networks for several other oscillators.

From the above discussion it can be seen that each fine network looks like the derivative of coarse network error and consequently is more random than the original coarse curve. However, the flexibility contained in the segmented approach permits a wide variety of variation in acceptable fine voltage curves.

It should be noted that fine curves with negative slopes are preferred at low temperatures since it is desirable not to use the PNP's until as high a temperature as possible. The PNP's make nice top temperature slope generators, but affect repeatability at the lower or medium temperature range.

Some difficulty has been encountered in being able to predict the repeatability of a resonator. It was initially our procedure to place the resonators in a controlled environment and consider them useful after they had reached an aging rate of 5.10^{-9}/day. In using these resonators we encountered repeatability difficulty and since have modified our test procedure to take the following form. Resonators and oscillators are placed in a controlled temperature environment and with the temperature set to the upper turning point the frequency is noted: the temperature is then changed to the lower turning point, allowed to stabilize, and again the frequency is noted. This cycle is repeated several times and if the various readings at each of the turning points do not deviate more than 2.10^{-8}, the resonator is accepted.

We are currently evaluating this testing procedure, at this time we believe it will solve some of the repeatability problems we have encountered.

The automatic compensation equipment development has been delayed because of repeatability problems, but some circuit development may be discussed. Figure VII shows the servo amplifier. It should be noted that by using a nulling type operational amplifier the loop will lock immediately and permit the motor to have a few seconds to reach null. This equipment has been tested and functions very well. Figure VIII is a diagram of the phase detector used to lock previously shown oscillators to the standard.

In operation the system functions as follows: The temperature controller selects a temperature and remains at the temperature until all thermal transients have diminished, a signal causes the loop to close, and current contribution from the phase detector causes the TCXO to lock. This current generates an error signal causing the network resistance to change until the current required from the phase detector is zero. At this point the resistance value is locked and temperature sequencer moves on to the next sequence.

This same procedure of operation can be repeated on ten oscillators with eight segments. While it yet remains for the equipment to be utilized for a production run of TCXO's, it appears feasible.

The unique characteristics of TCXO's make them particularly attractive for use in some special purpose oscillator requirements. Among these characteristics are instantaneous operation, low power consumption, and the ability to be electrically remotely turned on or off or controlled. Two special cases where these attributes have been utilized are described in the following paragraphs.

Case No. 1

A need existed for a signal source that would provide two square wave output signals at 8 MC with a 90° phase shift between them. It was required that the rise and fall times of these signals be less than 10 nano seconds and that the duty cycles be 50 ± 2 nano seconds. In addition this clock was required to synchronize with a 400 KC externally applied signal. Without the external synchronization signal the clock was to maintain 5.10^{-7} from -30°C. to +70°C. The unit was to be housed in a container .5" x 2-3/4" x 2-1/4". To generate this requirement an 8 MC TCXO was utilized to drive 2 RC phase shifters which in turn respectively drove individual shaping amplifiers. One of the phase shifters was used to advance the oscillator output signal by 45° and the other to retard the signal by 45°. There was some concern originally with the ability of these phase shifters to maintain the phase shift accurately over temperature. It was discovered, however, that they performed this requirement admirably. To provide the synchronization function a 400 KC signal was multiplied to 8 MC and a portion injected into a low impedance point in the oscillator loop. Locking ranges of ± 3 PPM were typical.

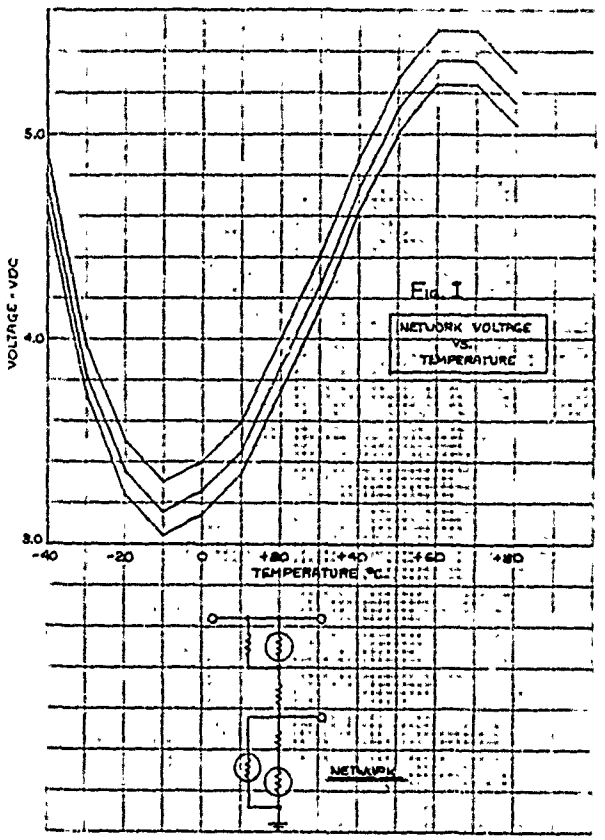
A complete schematic diagram is shown in Figure IX. The RC phase shifters consisted of a metal film resistor and a mica capacitor. The phase shift at the point was very stable and well within the output requirements. The problem was in squaring these two sine wave signals and maintaining the established phase relationship over the operating temperature range. This was squaring amplifiers. The exact configuration of these and the final output stages are shown in the schematic.

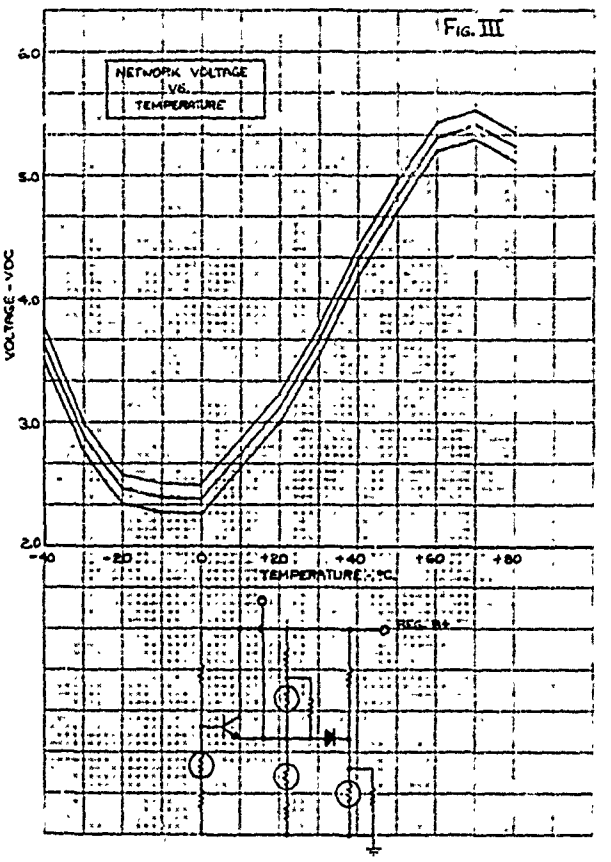
Case No. 2

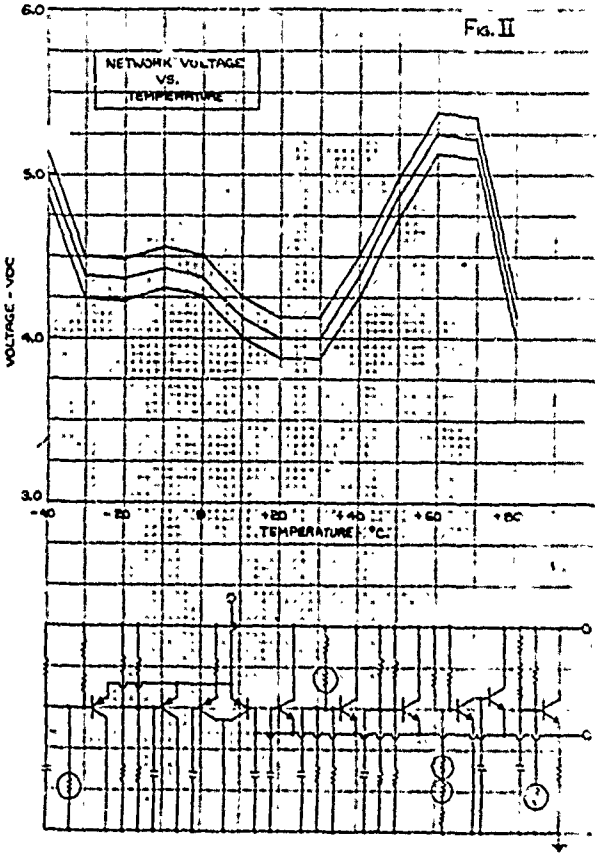
The equipment requirements for this oscillator were a 100 KC siusoidal output maintaining a stability of 1 v RMS ± .1 db over the temperature range and for 6 months. The output signal was to maintain a stability of ± 1 PPM from -40°C. to +75°C. The most difficult portion of the specification was the output requirements. In an ovenized unit this requirement can be met quite readily, whereas in a TCXO it becomes quite difficult. The reason for the difficulty is that in a TCXO all of the circuitry sees the entire temperature variation. To maintain the constant output ± .1 db requires some special attention.

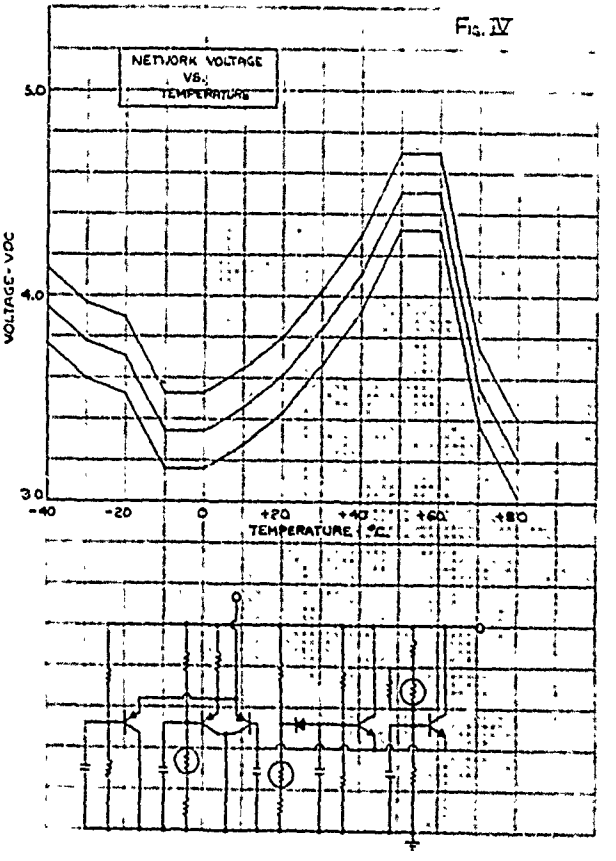
A 3.2 MHz oscillator was selected because of stability, aging, and environmental requirements. This frequency was divided to 100 KHz by 5 binary stages. The output of the last binary provides a square wave with excellent amplitude stability over the temperature range. To convert this to a very clean sine-wave, a differential tuned amplifier was utilized. In addition to the differential amplifier approach, the tank circuit of the amplifier was designed such that the temperature coefficients of the inductor and capacitor were matched. The differential portion provides the constant gain vs. temperature and the matching in the tank circuit prevents detuning. This coupled with the constant amplitude feed from the binary allows a constant output vs. temperature.

A complete schematic of the unit is given in Figure X.


Fig. I
NETWORK VOLTAGE VS. TEMPERATURE
VOLTAGE - VDC
5.0
4.0
3.0
-40
-20
0
+20
+40
+60
+80
TEMPERATURE °C.
NETWORK


Fig. III
NETWORK VOLTAGE VS. TEMPERATURE
VOLTAGE - VDC
6.0
5.0
4.0
3.0
2.0
-40
-20
0
+20
+40
+60
+80
TEMPERATURE °C.
REG. B+


Fig. II
NETWORK VOLTAGE VS. TEMPERATURE
VOLTAGE - VDC
6.0
5.0
4.0
3.0
-40
-20
0
+20
+40
+60
+80
TEMPERATURE °C.


Fig. IV
NETWORK VOLTAGE VS. TEMPERATURE
VOLTAGE - VDC
5.0
4.0
3.0
-40
-20
0
+20
+40
+60
+80
TEMPERATURE °C.

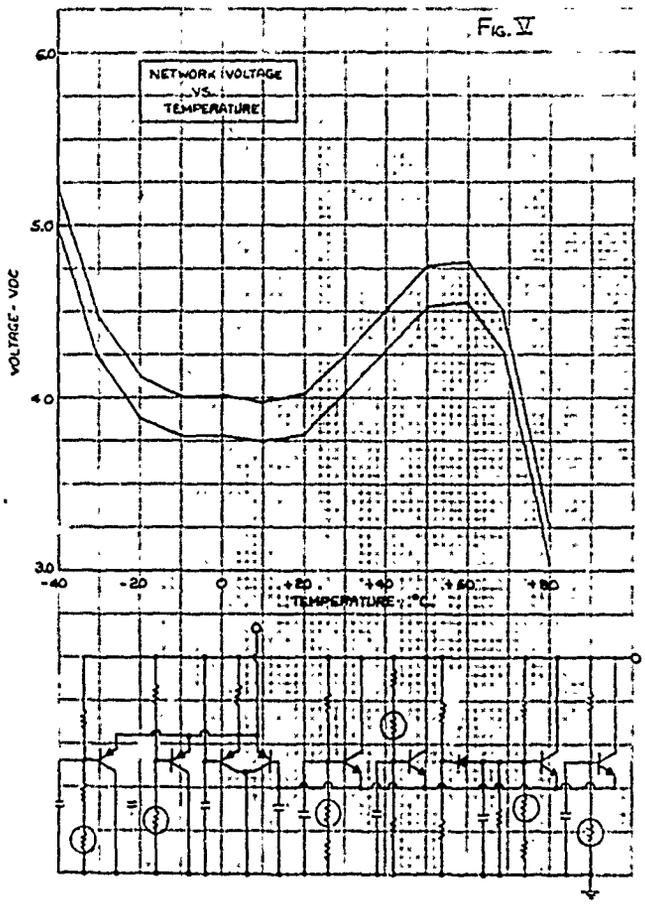


Fig. V

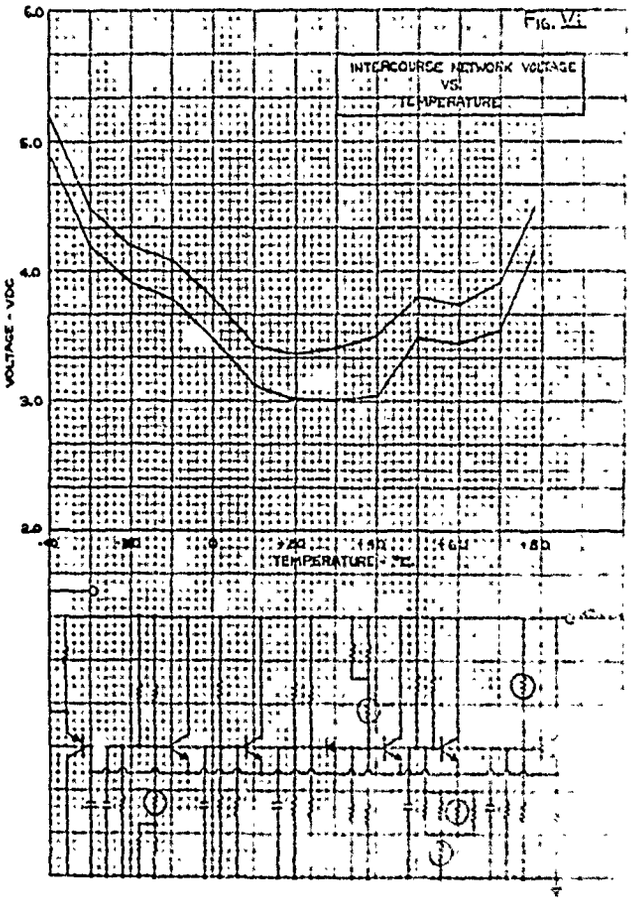


Fig. VI

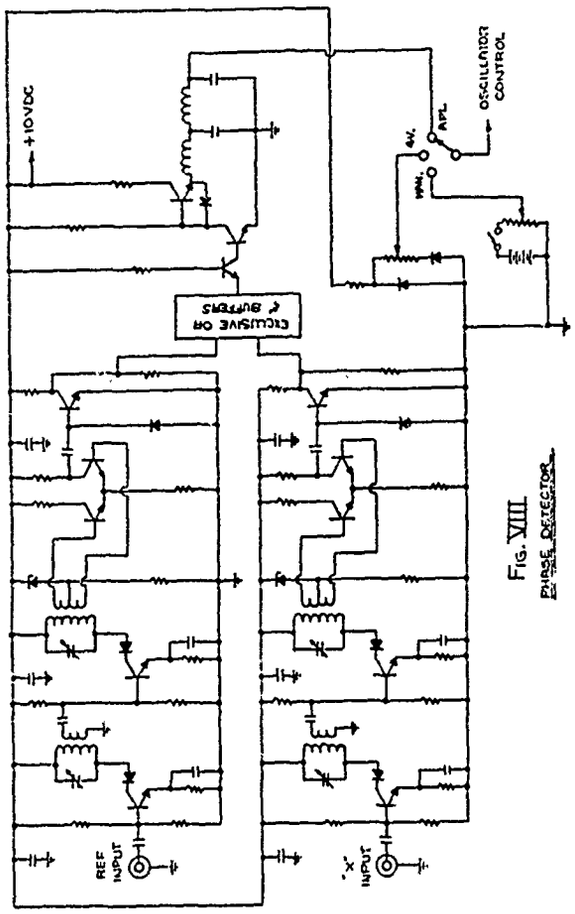


Fig. VIII
PHASE DETECTOR

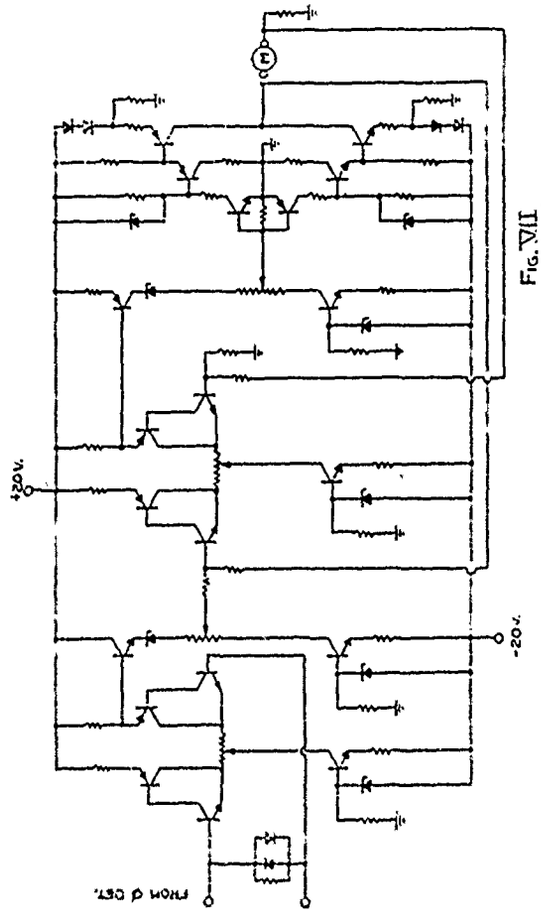
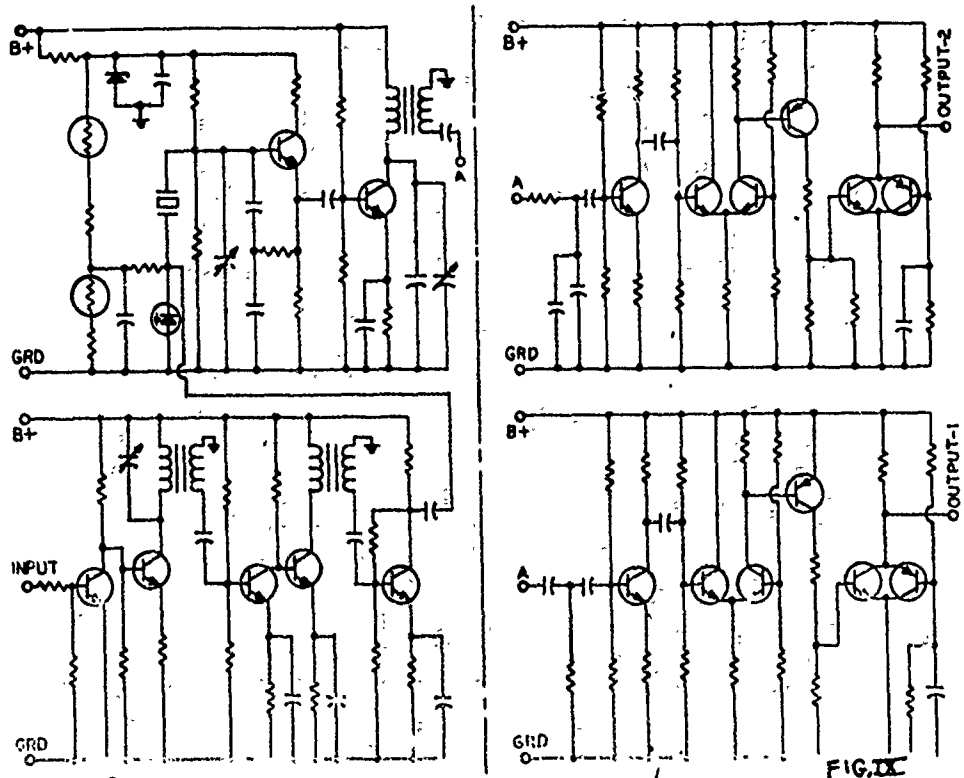


Fig. VII



8 MHz Dual Output TCXO

FIG. IX

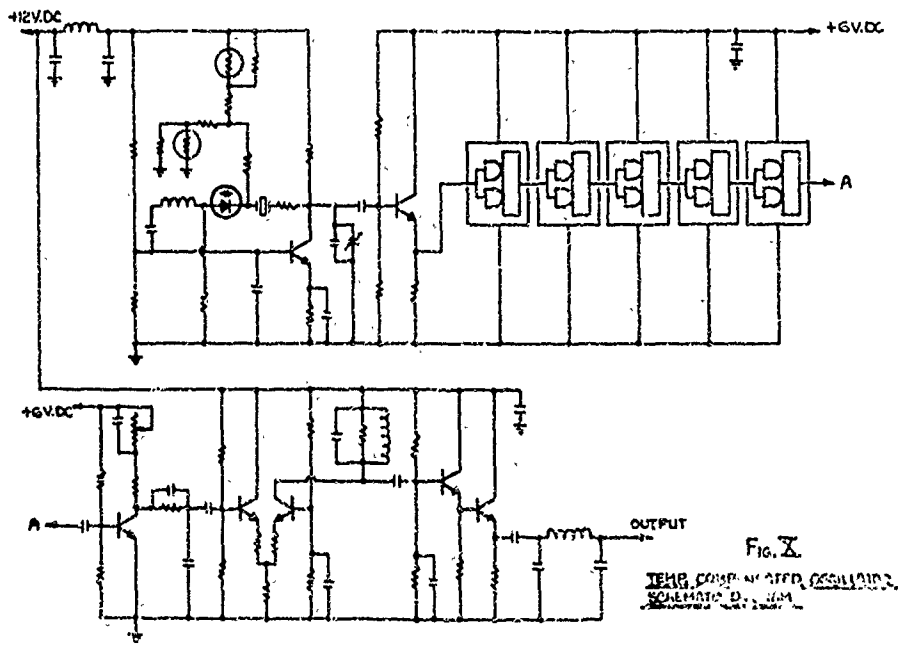


Fig. X

TEMP. COMPENSATED OSCILLATOR
SCHEMATIC DIAGRAM

TEMPERATURE-COMPENSATED CRYSTAL-CONTROLLED OSCILLATORS OPERATING FROM 800 kHz to 1500 kHz

by

H. A. Batdorf
Bell Telephone Laboratories, Inc.
Allentown, Pennsylvania

Abstract

A method of compensating crystal-controlled oscillators operating between 800 kHz and 1500 kHz using CT-cut resonators is presented. These resonators have a parabolic temperature characteristic haveing a parabolic coefficient of approximately $4ppl0^8/°C^2$. Compensation is realized by providing a temperature-dependent bias voltage to a varactor diode placed in series with the crystal unit. The bias network provides a parabolic voltage-temperature characteristic having its turning-point coincident with that of the crystal unit. In order that the temperature compensation be independent of frequency adjustments of the oscillator an inductor is placed in parallel with the varactor and a d-c amplifier is employed. The frequency stability of the oscillator over a temperature range of 10°C to 30°C is within ±1ppm while the oscillator is adjusted ±20ppm from the nominal frequency.

Introduction

Several reports have previously been presented on temperature compensation showing various methods and their results. These methods, however, were used primarily with AT-cut resonators both on their fundamental and overtone modes of operation. There exists, nevertheless, a need for stable frequency sources in a frequency range where the characteristics of the AT-cut resonator are undesirable and where the rectangular DT-cut resonator becomes difficult to manufacture due to its small size.

The frequency range to be considered in this paper is that between 800 kHz and 1500 kHz. Temperature-compensated oscillators in this frequency range, utilizing rectangular CT-cut resonators, have been developed to meet a required frequency stability of ±1ppm. They are being used as time-base generators in data set receivers and frequency generators for carrier systems. The temperature range, the size limitations imposed on the oscillators and the undesirable characteristics of AT and DT resonators in this frequency range led to the choice of CT resonators for these applications.

Crystal Units

Although small AT-cut resonators having a low ratio of capacitances and a low temperature coefficient of frequency can be made at higher frequencies, particularly above 6 MHz, such is not the case in the 800 to 1500 kHz range. The quartz plate must be contoured to keep it as small as possible, and this raises the ratio of capacitances as much as a factor of 2. But contouring is not sufficient. It appears to be very difficult, if not impossible, to find designs at all frequencies between 800 and 1500 kHz which consistently yield units having good frequency temperature characteristics. Also, unwanted flexure modes can unpredictably couple to the main resonance and cause abrupt frequency changes, sometimes exceeding one part per million.

The rectangular DT-cut resonator also would seem to be a reasonable choice in this frequency range. However, the quartz plate becomes extremely small at the higher frequencies. Because of its small size the effective ratio of capacitances increases since the capacitance of the holder becomes as large as, or larger than, the static capacitance of the plate. Also, the effect of the support structure on the plate tends to increase the temperature coefficient of the unit and causes inconsistancy in the turnover temperature among units having the same physical dimensions. Manufacturing would be difficult due to handling problems associated with the small size and the yield would tend to be low since small variations in either plate or electrode dimensions would cause relatively large fluctuations in the electrical characteristics of the crystal unit.

It appears that the rectangular CT-cut resonator is a reasonable compromise for this frequency range. Even though it has a higher temperature coefficient than the DT-cut resonator, it does have a lower ratio of capacitances and the plate size is larger.

Method of Compensation

The frequency-temperature characteristic for a typical rectangular CT-cut resonator is shown in Figure 1. The characteristic is parabolic in the vicinity of turnover, having a parabolic coefficient of approximately 4pp $10^8/°C^2$. For this cut, the ratio of capacitances varies from approximately 300 to 400 over the frequency range. The

magnitude of the ratio of capacitances indicates that the crystal should exhibit reasonable frequency pulling characteristics. On the other hand, temperature compensation can be expected to be difficult due to the large parabolic coefficient of the frequency-temperature characteristic.

Temperature compensation is realized by providing a varactor diode in series with the crystal unit whose bias is determined as a function of temperature. Since the frequency-temperature characteristics of the crystals to be compensated are parabolic, a network consisting of resistors and thermistors is used which has a parabolic voltage versus temperature characteristic. The network is shown in Figure 2.

The turnover temperature of the network is determined by the choice of component values. If it is assumed that thermistors having the same temperature coefficients are used, then the ratio of thermistor resistance at the desired turning point to thermistor resistance at 25°C will be fixed and can be expressed as

$$a(T_o) = \frac{r(T_o)}{r(25°C)} \qquad (1)$$

which will depend only on the turnover temperature T_o. It can be shown that the temperature coefficient of the output voltage of the network will be zero at the temperature T_o, provided

$$a^2(T_o) = \frac{R_s R_p}{r_s(25°C)\, r_p(25°C)} \qquad (2)$$

It is clear from an examination of the network of Figure 2 that the magnitude of the voltage deviation will depend upon the values chosen for R_s and R_p. The curves plotted in Figure 2 show the deviation as a function of temperature for several sets of values for R_s and R_p. It is seen that variations of the order of 1% of the applied network voltage can be obtained for temperature changes of ±40°C from the turning point.

Inspection of the curves of Figure 2 reveal that the network output voltage is not truly parabolic; rather, the curves are steeper for temperatures below T_o than they are at corresponding points above T_o. However, for relatively narrow temperature ranges, the approximation is satisfactory. If a wider temperature range is required, some additional components must be used.

The voltage characteristic required for compensation can be obtained by several methods. It can be measured directly by making a temperature run on a crystal-controlled oscillator and measuring the varactor bias voltage needed to maintain constant frequency as a function of temperature. Another method of determining the required voltage characteristic is to calculate the required bias voltage using the compensating characteristic necessary for compensation and the following equations:

$$\frac{\Delta f}{f} = \frac{1}{2r\left(1 + \frac{C_s}{C_o}\right)} \qquad (3)$$

$$C_s = C_{4V}\left(\frac{4.7}{V + .7}\right)^n \qquad (4)$$

where C_{4V} is the varactor diode capacitance with a four volt back bias, V is the required bias voltage, and n is a variable which can vary between .3 and .6. A third method of determining the voltage characteristic required for compensation is to obtain a graphical solution, which is outlined in Figure 3. In this Figure the required compensating characteristic is plotted in the first quadrant, equation (3) is plotted in the second quadrant and equation (4) is plotted in the third quadrant. The required bias voltage versus temperature characteristic is obtained by sampling the compensating characteristic in quadrant one over the temperature range of interest, moving from quadrant to quadrant normal to the axis at the value determined in the previous quadrant, and arriving at the required solution in quadrant four.

Design Example

For a specific illustration of what has been presented, let's consider the design of an oscillator at 1130 kHz operating over a temperature range of 13° to 33°C. The crystal is a CT-cut designed to be adjusted to frequency with 30 pf series capacitance having its turnover temperature located at 23±5°C.

Since very few crystals exhibit an ideal characteristic, as was illustrated in Figure 3, problems arise due to the tolerances required in manufacture of the crystal. Shown in Figure 4 is a marginal compensating characteristic, i.e., one which deviates from the nominal turnover temperature by five degrees and whose adjustment tolerance is 8ppm from the nominal adjustment point. It is seen that the varactor bias voltage, to initially adjust the oscillator to frequency at 23°C, has changed from 3.3 volts in Figure 3 to 4.6 volts. Also it is seen that the deviation of bias voltage over the temperature range increases under these conditions.

Due to the range over which the nominal varactor bias point can vary, the slope of the varactor characteristic curve over which the varactor operates also varies as is illustrated in Figure 5. Thus each adjustment of the crystal frequency, via the varactor, requires an additional adjustment of the compensation network. A method of minimizing this effect is to effectively linearize the frequency-bias characteristics. This is accomplished by selecting a varactor whose C_{4V} is much greater than that required to just take up the adjustment tolerance and by placing an inductor in parallel with the varactor and operating the

varactor in the region of low bias voltages. The voltage required for compensation is reduced as shown in Figure 6, thus further reducing the change in slope of the varactor characteristic curve over which the varactor operates.

Although it was shown previously that various magnitudes of voltage deviation could be attained for a given turnover temperature by varying R_s and R_p, in accordance with equation (2), it is difficult to attain the correct voltage deviation while maintaining the correct turnover temperature due to the discrete values of components commercially available. In order to increase the flexibility of the compensation network, a d-c amplifier is employed. This allows one to independently adjust the turning point and the magnitude of the varactor bias voltage.

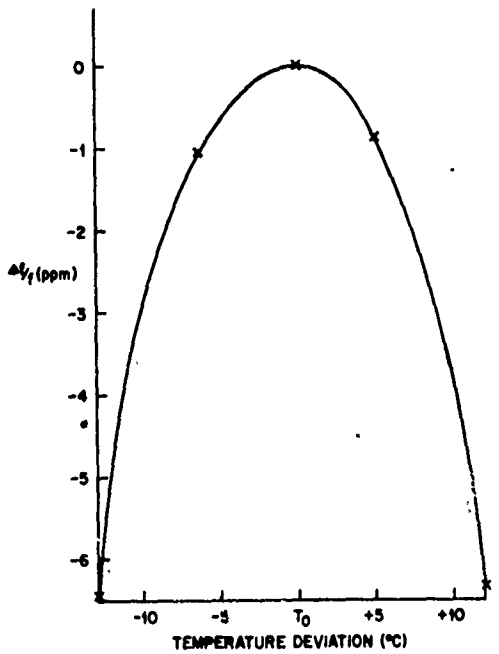
The compensation network employed is shown in Figure 7 and is seen to consist of the following: the bridge circuit which provides a voltage characteristic having the correct turning point and allows adjustment for the adjustment tolerance of the crystal; a d-c amplifier to provide the proper magnitude of the voltage deviation; and the varactor with a parallel inductor which reduces the required bias voltage deviation and effectively linearizes the operating region of the varactor.

Due to the impedance of the varactor and the frequency stability versus drive level characteristic of crystal resonators, it was decided to maintain the crystal drive level at approximately 50 micro-amperes. To achieve this an oscillator as shown in Figure 8 was used. The utilization of Q_1 and Q_2 satisfies the gain requirement while the tank circuit in the collector of Q_2 provides the required phase. Phase adjustment is initially accomplished by replacing the crystal unit and the varactor with a resistor of the equivalent series resonant impedance and adjusting L_1 until the nominal oscillator frequency is obtained. The limiting diodes (CR_2 and CR_3) and the capacitive voltage divider provide the desired drive level.

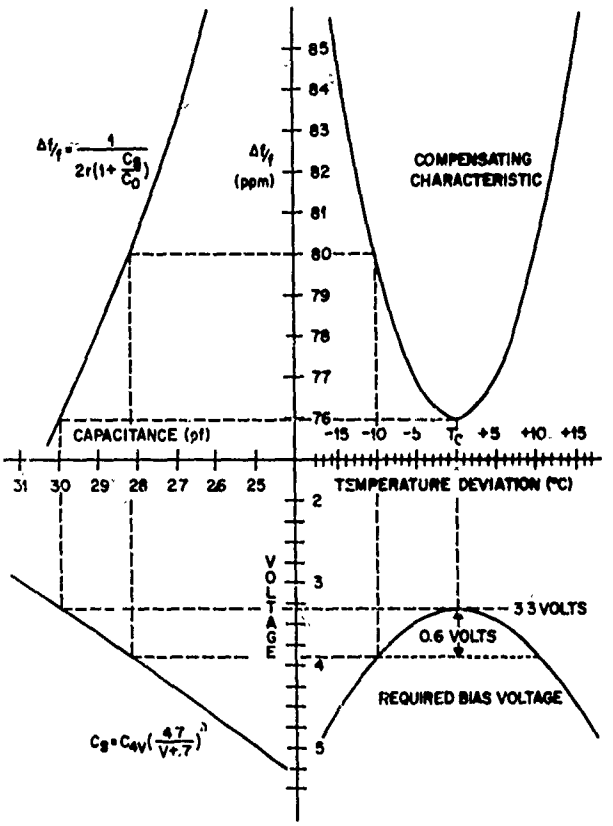
Figures 9 and 10 show the results obtained with one oscillator which is typical of the performance of the other oscillators spanning the frequency range. In Figure 9 is shown the crystal frequency versus temperature characteristic before and after compensation. It is seen that an improvement factor of 7 is readily achieved. In Figure 10 the frequency versus temperature characteristics are shown for which the oscillator is adjusted ±20ppm. It is seen that frequency deviations of less than ±1ppm can be maintained over the temperature range for both extreme operating points.

Summary

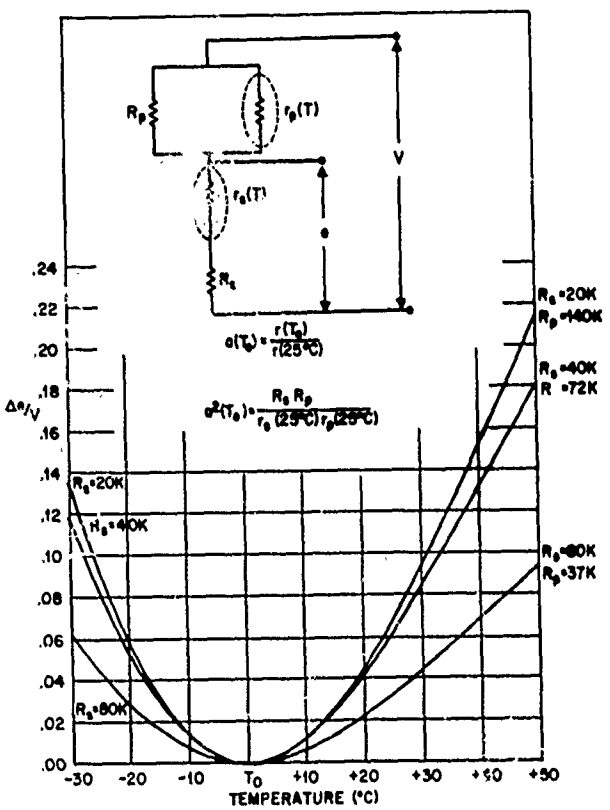
In summary, temperature-compensated, crystal-controlled oscillators have been developed for a frequency range where compensation heretofore has not been discussed. In this frequency range, from 800 kHz to 1500 kHz, it appeared that the rectangular CT-cut resonator would be a reasonable compromise due to poor frequency-temperature characteristics of the AT-resonator and the high ratio of capacitance and handling problems encountered with the rectangular DT-cut resonator. While the CT-cut resonator exhibits a high parabolic coefficient of its frequency temperature characteristics, it was shown that compensation was possible through utilization of a DC-amplifier and the use of varactors whose capacitance at four volts is high compared to the design load capacity. The results of compensating these crystals show that stabilities in the order of ±1ppm over a ±10°C interval is readily accomplished and that frequency adjustment of ±20ppm can be achieved while maintaining frequency stabilities within ±1ppm.



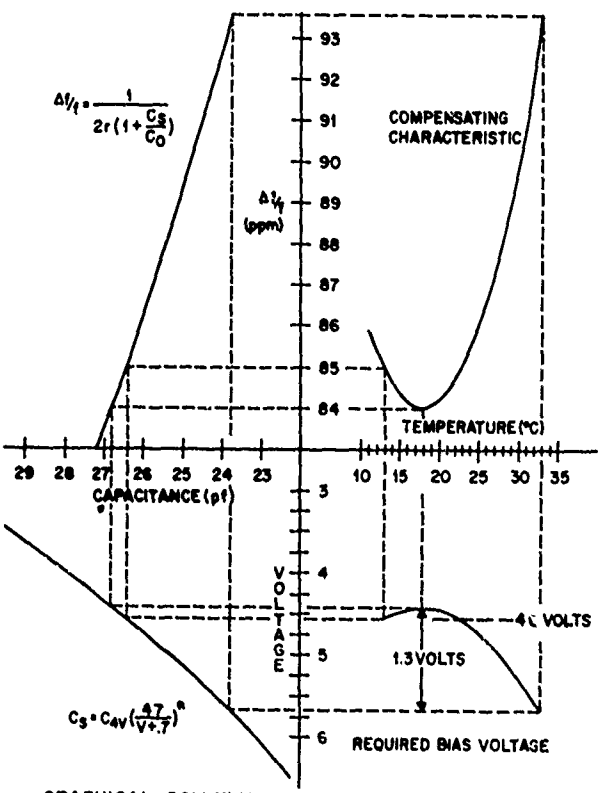
"CT" FREQUENCY-TEMPERATURE CHARACTERISTIC



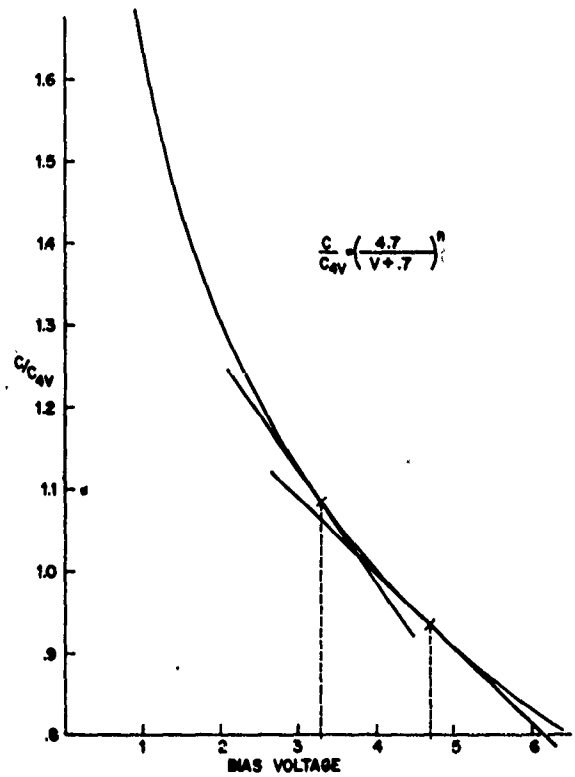
GRAPHICAL SOLUTION



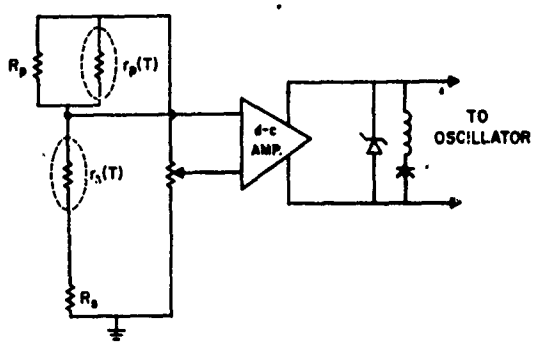
COMPENSATION NETWORK AND CHARACTERISTIC CURVES



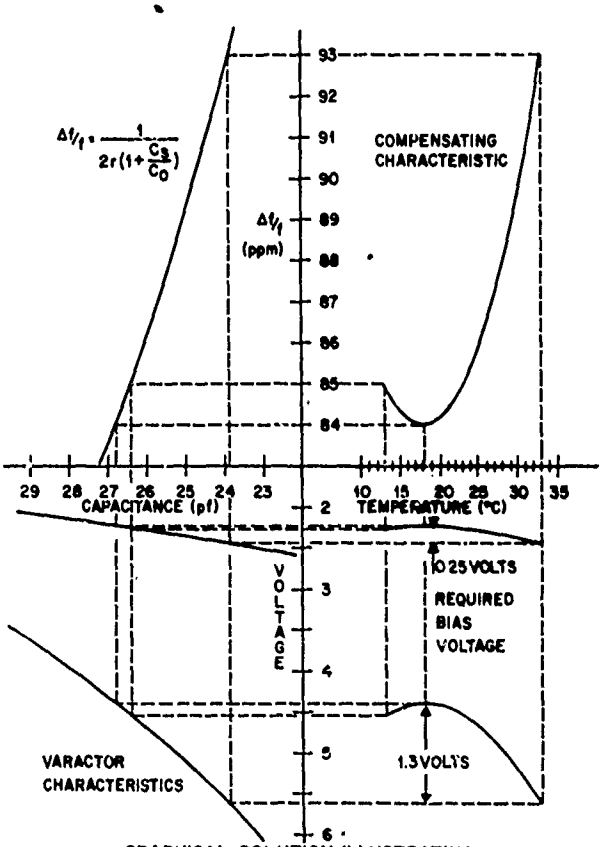
GRAPHICAL SOLUTION ILLUSTRATING BIAS VOLTAGE AT 25°C AND MAGNITUDE OF VOLTAGE DEVIATION



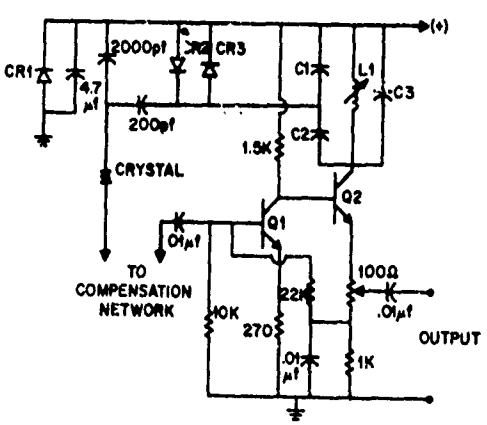
NORMALIZED VARACTOR CHARACTERISTICS CURVE



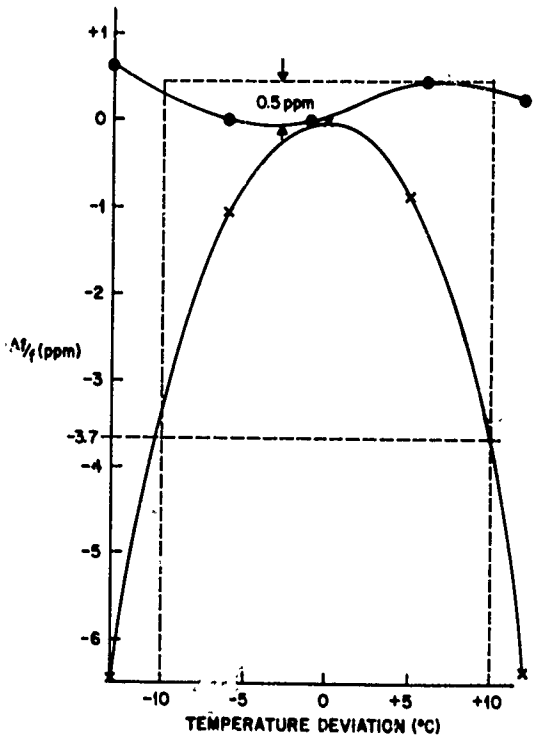
COMPENSATION NETWORK



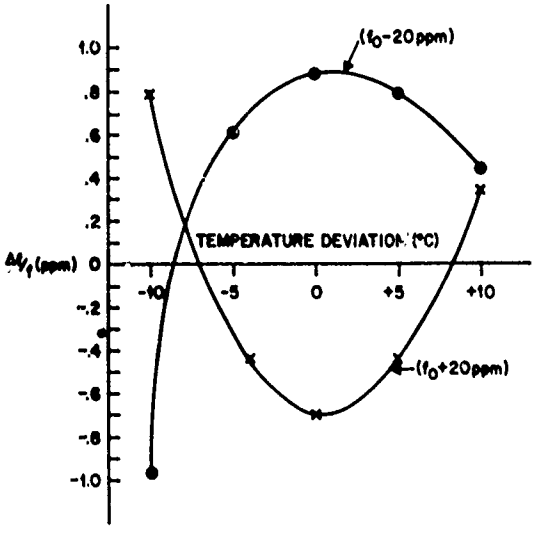
GRAPHICAL SOLUTION ILLUSTRATING CHANGE IN MAGNITUDE OF BIAS VOLTAGE



OSCILLATOR SCHEMATIC



COMPENSATED VERSUS
NON-COMPENSATED CHARACTERISTICS



COMPENSATION CHARACTERISTICS
FOR ±20ppm ADJUSTMENT

FREQUENCY STANDARDS FOR HF COMMUNICATIONS

Eric N. Le Fevre

Royal Navy Scientific Service

Summary

Despite the excellent stability, performance, and m.t.b.f. figures for high grade Crystal frequency standards, recent innovations require better noise purity close to the standard output frequency. This improvement in purity will aid the flexibility of a new transmission system allowing contiguous transmission and reception frequencies. A revised specification suitable for the users of 90% of standards in the U.K. is detailed.

Introduction

The Royal Navy is the major user of high grade frequency standards in the United Kingdom, having installed more than 500 standards (made by Racal Instruments) into operational craft. These are based on a 5 MHz, 5th overtone quartz crystal, and have proved extremely reliable with a mean time between failure greater than 10,000 hours. They have elaborate ovens to produce good stability against environmental fluctuations and a low drive level to maintain long term stability. The stability over periods of seconds to days is for phase modulated and single sideband communication systems two to four orders higher respectively than is required for operational use, although the accuracy over periods of months to years is only barely adequate and relies on transmitted time and frequency information. This is unreliable in war time and introduces further complication and expense.

Frequency Stability and Noise Purity

It is therefore appropriate, in the light of possible future innovations, to restate the actual requirement for HF communications. A stability of 1 in 10^8 measured with a sampling time of 10 ms provides an adequate short term stability provided that a long term drift rate of one or two parts in 10^{11} per month is possible.

This inversion of the specification follows on from the introduction into the Royal Navy of a number of General Technology Rubidium standards. It is well known that these are not Primary standards, but if each Rubidium cell is pre-calibrated and documented, such that any standard can be repaired at sea to a known frequency within 1 in 10^9 then this state is primarily satisfactory. Any sophisticated shipborne communication system must be capable of correcting offset errors in excess of 1 in 10^7 which is the Doppler equivalent to two ships with a relative speed of 60 knots. A ship, anchored fore and aft, in the most sheltered bay, still could not meaningfully use stabilities available in the laboratory due to ship movements. Apart from the frequency stability, for both transmission and reception, the purity at frequencies close to the frequency standard output is of extreme importance. Two problems emerge. For transmission to take place simultaneously with reception at other frequencies own ship transmissions must not extend across the frequency band to the reception frequency and secondly the Receiver Local Oscillator must not allow inverse mixing to take place. This is shown in Fig. 1. A wanted 1 μV signal mixes with the Receiver Local Oscillator to produce the i.f. frequency signal, at which stage information bandwidth filtering may occur. An unwanted adjacent channel signal of large amplitude also acts as the local oscillator to signals spaced at the i.f. frequency from it. One of these is obviously the noise sidebands of the Receiver Local Oscillator. An additional intermediate frequency signal is thus produced with possible masking of the wanted signal. In order to assess the degree to which any particular oscillator is prone to this effect, the experimental rig also shown in Fig. 1 is set up and it corresponds closely to the practical situation. A first grade HF receiver, made by the British G.E.C. is used with cascaded narrowband i.f. filters to determine the level of an unwanted signal marked 'B' at frequencies progressively removed from the wanted signal 'A'. The source under test is the receiver local oscillator, but for comparison is replaced by standard frequency sources. The resulting plot shown in Fig. 2 indicates the degree of filtering necessary to achieve an acceptable level. It is seen from Fig. 2 that the Crystal standard has a low signal-to-noise ratio because the crystal is just oscillating at about 1 μW in order to prevent excessive aging; there is no such limitation with the Rubidium standard whose hard driven crystal is corrected by the Rubidium Servo Loop and has an excellent signal-to-noise ratio.

For existing triple mix synthesisers, harmonics are derived from the frequency standard and selected by filters before combining. If the 20th harmonic of the frequency standard 5 Mc/s output is considered at 100 Mc/s, then the purity at that point the synthesiser is 26 dB worse than is shown in Fig. 2. This worsening of purity is offset in some receivers by the synthesiser frequency being phase locked to the normal free running receiver local oscillator, this acting as a high Q filter to provide an additional 40 dB of filtering.

Usage of Frequency Standards

Existing R.N. HF Transmission Systems use a Marconi wide-band power amplifier to produce 1 kW signals. In order to couple many 1 kW signals to a single wide-band mast-structure aerial, it is necessary to use common aerial working tuned filters. These filters prevent power at other frequencies from feeding back into the transmitter with which the filter is associated. However, the tuned frequencies of these filters have to be separated by 10% to provide the necessary attenuation between transmitters adjacent in the frequency spectrum. At 5% off tune the transmitted noise sidebands are at best 140 dB down on the transmitter peak envelope power, and receivers can detect low amplitude signals only if they are interlaced between local 10% spaced transmissions. However, two innovations alter this situation to an extent that, once again, the purity of the frequency standard becomes predominant. Firstly, receiving aerials are being equalised as shown in Fig. 3. The 30 foot whip aerial was evolved to provide enough signal at 2 Mc/s where nearly 20 dB of mismatch loss occurs when feeding a 50 ohm receiver. Consequently, at the quarter wave resonance much more atmospheric noise is received than is necessary to swamp receiver noise.

Secondly, a new transmission system (Ref. [1]) is being considered which allows transmission without frequency separation restrictions. As shown in Fig. 4, automatic routing of individual modulated signals is possible using the phase switched hybrid transformers as power switches. This untuned system is possible now that Marconi have produced an improved linear power amplifier, able to withstand a 3:1 mismatch whilst maintaining excellent linearity at full power.

Fig. 5 shows a series of modulated transmissions available with any frequency separation; the inband intermodulation products between these transmissions will appear in this format for the relative spacing shown whether this is 1 or 100 kc/s per division. Reception at the particular frequencies at which intermodulation products occur, irrespective of the transmission system used, is difficult and improved linearity of say 10 dB will not alter the situation. However, improved linearity will cause the higher order intermodulation products, some of which appear inband, to fall very rapidly. Reception will, therefore, be possible in the regions between these products providing that the close-in noise of the frequency standard, and hence the transmitter drive unit and receiver, allow >120 dB purity in a 3 kc/s bandwidth, 10 kc/s away from the transmission or reception frequencies. This improvement in close-in purity from 5% off tune to 10 kc/s off tune is possible only if we depart from the triple mix synthesis process with its inherent noise due to cascaded mixers optimised between noise and linearity.

Phase Modulated Communication System

Fig. 6 shows the effect of oscillator phase perturbation on the integration of a 'bit' when compared coherently with the phase of a preceding bit. The stability within the sampling period considered, without regard to the purity, can be seen to be adequate for the triple mix system whilst that of a prototype simple digital synthesiser is shown to be inadequate. However, more advanced digital synthesiser systems now being developed by Plessey, provide the necessary stability for phase modulated communication systems together with the advantage of good purity derived from the phase locked oscillator. This performance can be extended to give fast access tuning systems, where this, too, is an essential requirement.

Conclusion

As a result of the innovations discussed above a new specification is required for frequency standards to provide:-

(i) Atomic Timescale (A1).

(ii) Capability of repair at sea to an accuracy of 1 in 10^9.

(iii) A stability for all sampling times above 10ms of 1 in 10^8.

(iv) A noise content more than 140 dB below the standard frequency output at 5 Mc/s measured in a 200 c/s bandwidth at 10 kc/s from the standard output.

Such a specification would appear to be realisable with a simplified Rubidium standard. Such a standard, with pre-calibrated replacement lamps, can be classed as a primary standard for communication requirements at sea.

Acknowledgement

The author wishes to thank the Director, Admiralty Surface Weapons Establishment for permission to publish this paper.

Reference

The author was assisted by B. Sosin of Marconi Co. in formulating the new transmission system; also by E. Searle and G. Capell of A.S.W.E.

Figure 1 Receiver Test Rig for Measuring L.O. Close-in Noise

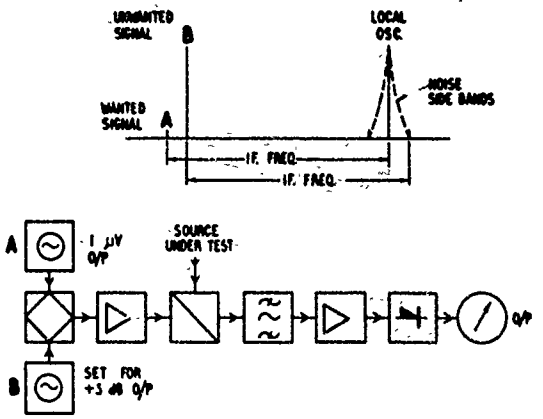


Figure 2 Close-in Noise Purity for Various Frequency Sources

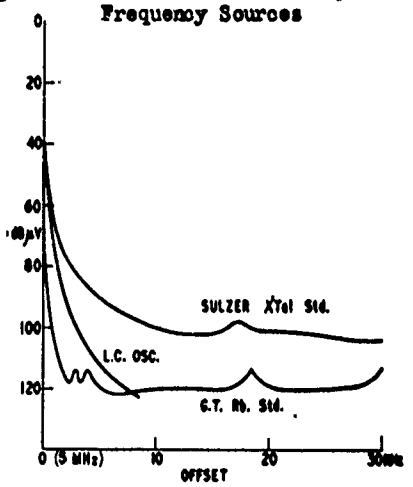


Figure 3 Received Signal Levels from a 500 Watt Two Tone Transmission

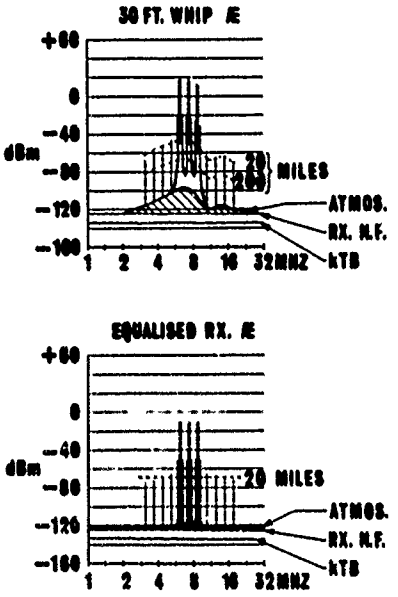


Figure 4 - A Wide-band transmission System

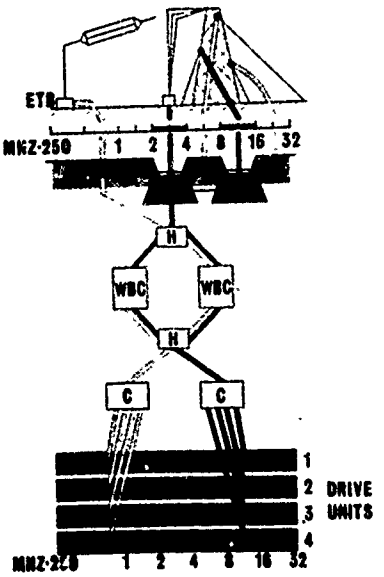


Figure 5 Eight Signal Channels and the Inband Intermodulation Products $2f_1-f_2$; $3f_1-2f_2$ etc.

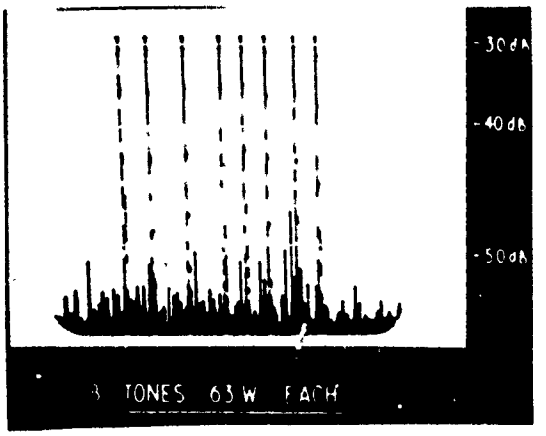
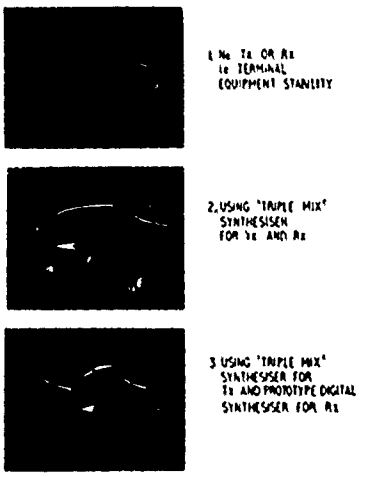


Figure 6 Phase Stability During Integration

LOCAL OSCILLATOR STABILITY
EFFECT ON PHASE MODULATION



DIGIPHASE SYNTHESIZER

Garry C. Gillette
Dana Laboratories, Inc.
Irvine, California

Summary

Indirect or digital synthesizers can have many desirable characteristics, such as integrated circuit construction, few L-C circuits, fast switching, low spurious and phase noise, and small size and cost. Their limitations have been in simultaneously achieving good short term stability and having small frequency increments. The "Digiphase" technique described here is an example of a new form of computing synthesizer which removes these limitations. As a result of digital control, phase and frequency may be simultaneously programmed. Frequencyrampcan also be very accurately generated with a stepped frequency program.

Digiphase Technique

The example treated here is that of a D.C. to 11 MHz synthesizer. The output is obtained by the construction of a digital synthesizer with frequency range 40 to 51 MHz, in 1 Hz steps. The basis of the Digiphase concept is that of computing the desired numerical value of the phase of an ideal signal (in units of cycles and fractional cycles), and then forcing the phase of a measured signal to exactly agree with this computed phase by means of a control loop.

To mechanize this one must note first that the measurement of integer cycles of phase can be performed digitally by simply counting axis crossings of an assumed sinusoidal incoming waveform. The interval remaining between the last axis crossing and a convenient fixed frequency clock yields a time delay. This delay can be converted into an analog quantity proportional to the fractional cycle remainder of the measured phase. In the synthesizer described here we measure, program, and control this interval in units of 200 femptoseconds. By being able to digitally measure integer cycles and analogly measure fractional cycles, and being able to numerically compute their ideal values on a rapid basis, one can compare the two values and obtain a sampled data error signal to be supplied to a correction loop.

Since it is generally not desired to simply have a fixed phase output, but one of a constant phase-rate or frequency, the phase computer portion must digitally integrate the programmed frequency over a sufficiently long interval, and store the result of this computation in a phase control register. This computation must be updated at a rate equal to the sample rate in the control loop. In the example described here the rate is 100 kHz. The front panel frequency is a seven digit decimal number with 1 Hz resolution, and when integrated the numerical value of the phase has a minimum increment of 10^{-5} cycles, for the 10 μ sec. clock interval. Computed phase value is increased or incremented from between 400 and 510 integer cycles every 10 μ sec. to span the range of 40 to 51 MHz. The remaining fractional cycle portion of the computed phase contains a numerical value which should correspond to the analog output obtained from the phase detector if the input frequency and phase are correct as generated by the loop.

In Figure 1 is shown the comparison technique between counted and computed cycles. If one assumes a frequency of 40 MHz + 1 Hz and zero initial phase to be programmed, the integer cycle portion of the phase register increases in increments of 400 cycles every 10 μ sec. The fractional cycle portion increases in increments of 10^{-5} cycles, and only until 1 sec. later does an overflow occur between the two regions of the register. At this time the integer cycle portion has a large multiple of the 400 cycle increment (10^{+5}) plus one cycle. This process continues on, and at the end of two seconds there is an integer multiple of 400 cycles (2×10^{5}) plus two cycles in the integer portion of the register. By simply doing no more than insuring that the average output value of the comparator is zero one can lock the loop input frequency to the programmed frequency, since at no time will the two differ by as much as 1 cycle of phase.

Figure 2 contains a schematic plot of the numerical value of both the fractional and integer components in the phase register for some arbitrary program frequency. When the fractional cycle portion builds up to a value of 1 cycle of phase it overflows into the integer value portion and is itself reset to zero again. The control loop responds to the integer cycles line of the graph in this simplified case, but it is important to note that the difference between this and the ideal performance is exactly the fractional cycles line on the graph. (This must be so by the equation for phase at the top of the graph). One can then construct a supplementary phase modulation signal from the fractional cycle portion of the phase register . Here it is of course assumed the phase detector is linear, as in the case of a flip-flop.

From Figure 1, if one programs into the loop

input frequency 40 MHz and 1 Hz, and the same into the phase computer, there will exist a 1 Hz "beat note" on the output of the flip-flop, due to the manner in which the input frequency and computer are advancing away from the 100 kHz clock. This beat note does not represent a frequency error in the system, since the two are identically at the same frequency and phase. It is simply a linear saw tooth of undesired phase noise. The value of the beat note or sawtooth is exactly proportional to the fractional cycle portion in the phase register, as plotted in Figure 2. The fractional cycle portion of the phase register is then converted into an analog quantity scaled exactly to oppose the value in the F/F phase detector, and summed together with the output of the phase detector. Under these circumstances the output of this particular phase detector is nulled and the loop sees no phase error. The end result is that the loop settles to exactly the phase value programmed by the computer, including fractional cycle increments. Since control of the phase has been achieved it may be programmed along with phase-rate or frequency, as computed by an internal arithmetic unit.

The synthesizer block diagram is shown in Figure 3. The reference timing generator contains a 40 MHz T.C.X.O. with circuitry to lock to an external standard at .1, 1, 5, and 10 MHz. By this technique the good short term stability of the internal T.C.X.O. can be optimally combined with the long term stability of a low power reference oscillator. The 40 MHz signal is then buffered and divided down to generate the various 100 kHz. timing signals used by the phase computer. External outputs are supplied at 10, 5, 1, and .1 MHz from the dividers.

The edge transition output of the comparator and the 100 kHz edge from the timing generator are used by the linear phase detector to generate an output current pulse whose duration is equal to the differential time delay of the input edges. This pulse is summed with the output of the B.C.D. to analog converter at the amplifier current summing node input. Any difference current at this point passes through the series R-C feedback of the loop amplifier to generate a correction signal to the V.C.O. The ripple created by the current pulses is filtered inside the first minor loop G(s), and then again by a 7 pole low pass on the amplifier output. These two combined yield a total rejection greater than 100 dB at the 100 kHz sample rate.

The 40-51 MHz V.C.O. is heavily isolated by multiple use of several stages of dual-gate M.O.S. field effect transistor buffers driven at very high level. This technique reduces the possibility of interaction between logic and output by greater than 100 dB. All inputs and outputs are clamped by hot carrier diodes to minimize any amplitude modulation.

The mixer output is filtered by a 7 pole low pass filter and amplified by a wide-band direct coupled amplifier to the required 2 V behind 50 Ω level. Front panel level adjustment is accomplished by voltage controlling the level of the input 40 MHz over a 10 dB range. A 10 dB per step attenuator is then used to enable continuous level control over a 90 dB range.

Figure 4 is a detailed block diagram of the phase computer and comparator. The computational technique used is that of digitally integrating the front panel frequency. This can be expressed by:

$$\phi_n = \sum_{i=0}^{n} \Delta\phi_i$$

$$\Delta\phi_i = f_i \Delta t \qquad \text{where } \Delta t = 10^{-5} \text{ sec.}$$

which is a sampled version of the continuous expression

$$\phi_t = \int^{t} f \, dt$$

This operation is done by sequencing both the frequency and previous phase value least significant bit first into the B.C.D. adder and entering the result into the phase register. Timing for this process is provided by multiphase 100 kHz signals generated in the sequence generator.

Integer cycles of phase are accumulated first in a high speed bi-quinary divider circuit using MECL II logic. The 4 to 5.1 MHz output rate is then divided by standard T^2L decade counters. The comparison signal is generated by feeding the complement of the desired count into a T^2L hex adder, and then using the carry output as a transition. In the high speed decade the phase B.C. D. information is first converted to decimal so that each of ten possible values is treated independently with minimum delay. The output transition of the comparator is then synchronized to an exact axis crossing of the input V.C.O. frequency to remove any jitter introduced by variable delays in the logic.

Figure 5 is a block diagram of the means used to create a sampled linear ramp whose amplitude is programmed by the fractional cycle value in the phase register. In this process a T.A.C. is used, for time to analog converter. When the count gate is opened the gate flip flops are all simultaneously reset. After the prescribed number of pulses have been counted by the decade counter the hex adder generates a carry, which sets the individual decade flip flop. In this manner the number of counts contained in each gate interval is determined by the number in the phase register decade. By decade weighting the divider the contribution of each gate is scaled to the appropriate phase register decade.

Figure 6 shows the loop block diagram, which is similar to numerous others covered in the literature. By inserting part of the low pass filter

inside the first minor loop its phase shift is cancelled by the loop gain in the minor loop. As seen from the Bode plot, loop bandwidth is 10kHz which assures locking over the entire synthesizer range. The phase detector transfer is described by transconductance since it gives a linear current output for a linear input time delay.

Phase Noise

Particular attention was paid to sources of random phase noise. In Figure 7 the single sideband noise of the V.C.O. is plotted. The V.C.O. determines the output noise for offset frequencies greater than the loop bandwidth of 10 kHz. For offset frequencies less than 10 kHz thermal noise sources in the amplifier and phase detector create a white noise floor for voltage applied to the varactor. This then is transformed into a $\frac{1}{f}$ spectrum of phase by the V.C.O. for offset frequencies close to the carrier. Figure 8 shows the closed loop phase noise performance of the synthesizer.

V.C.O.

Figure 7 contains a schematic of the F.E.T. V.C.O. In order to minimize series resistance to the varactor and still maximize Q the grounded gate F.E.T. was direct coupled to the tank. The dual gate F.E.T. supplies a convenient gain control and excellent isolation to the output buffers. Hot carrier diodes are used as AGC elements by limiting the peak current delivered to the tank. It is interesting to observe that the phase noise performance indicates a 1 Hz offset frequency noise of -112 dB, or L (1) = -112 dB in the notation of Dr. D. Halford as presented at the 22nd Annual Symposium on Frequency Control. This indicates close agreement with his measured values for field effect transistors.

VLF Output

A 400-510 kHz V.C.O. $\div$ 100 output is obtained from the B.C.D. accumulator on the logic. This is mixed with a fixed frequency 400 kHz signal from the reference timing generator to achieve a D.C. to 110 kHz V.L.F. output. Phase noise and spurious are reduced by as much as 40 dB, yielding a high performance output with .01 Hz resolution.

Search

Search provides a decade range of continuously variable offset frequency which is voltage controlled. A block diagram of this section is shown in Figure 9. The input control voltage establishes a rate on the output ramp of the integrator. When either the upper or lower limit is reached the reset control logic supplies the correct reset polarity to the integrator and commands the logic to add the appropriate $\pm 10^{+N}$ cycles of phase to its instantaneous value, where N = 1, 0, -1, -2, -3, -4, -5. The full scale reset rate in the integrator is adjusted to be 100 kHz, which when multiplied by the 10^{+N} phase scale factor yields a search range of ±1 Hz to ±1 MHz around the program frequency. The fractional cycle phase modulation signal required by the Digiphase technique is supplied by a decade weighted attenuator connected to the integrator output. This smoothly advances the loop 10^{+N} cycles before a step of 10^{+N} is added to the logic at the asynchronous reset rate of the integrator.

Performance

Figure 10 lists the important performance specifications of both outputs. Measured performance indicates a 30 kHz signal to phase noise of -60 dB for the 11 MHz output and -90 dB for the 110 kHz output. Typical near-in sideband spurious are less than -70 dB for the 11 MHz output, and -100 dB for offset frequencies greater than 100 kHz from the carrier. Fractional frequency deviation was measured by amplifying the D.C. phase fluctuations on the output of a mixer, and then sampling this output at 1 second intervals with a fast digital voltmeter connected to a printer. The first difference of the readings were RMS averaged on a computer for 100 samples. Settling time was measured to be less than 2 m sec for a 10 MHz step in frequency. Figure 11 shows the output response for a 10 kHz to 30 kHz frequency change. Since only one V.C.O. is used output phase is always continuous and amplitude is constant during switching. Phase stability with temperature is excellent, due to digital division and no tuned circuits in phase sensitive areas.

Construction

As shown in Figure 12 construction of the logic, power supply, and search is on two-sided P.C. cards which are joined to the synthesizer via a common mother board. Front panel switches mount to a common P.C. board which then connects to the logic mother board . Four modules made from a common extrusion mount the remaining 8 P.C. boards. Each module is removeable for servicing with power and signal connectors mounted on the bottom of the module. In the logic section 7400 Series T^2L medium scale integration circuits are used for computing and MECL II series is used for high speed applications.

Conclusion

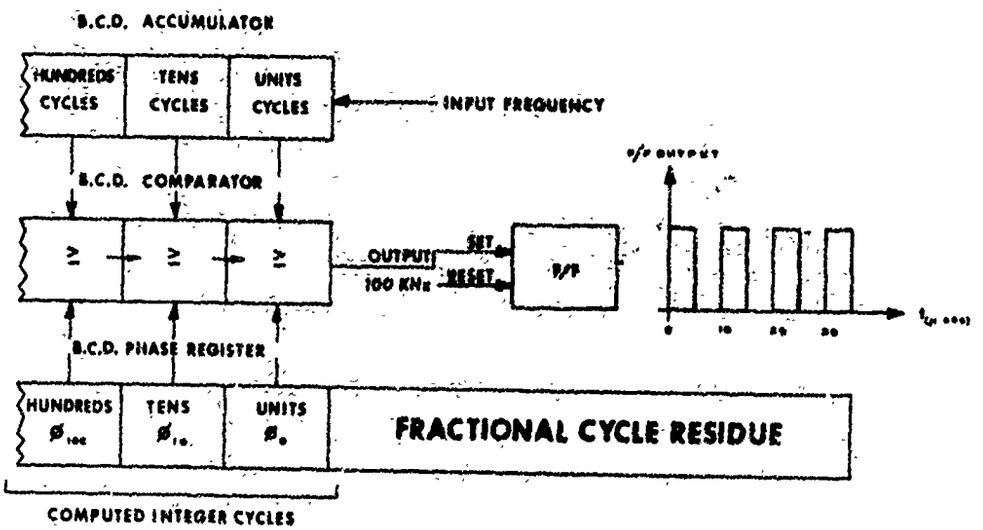
The Digiphase technique described here has many unique properties, such as coherent phase and frequency programmability, search operation, small frequency increments with high sample data rate, and fast switching. As a result of the high sample rate and careful design phase noise per-

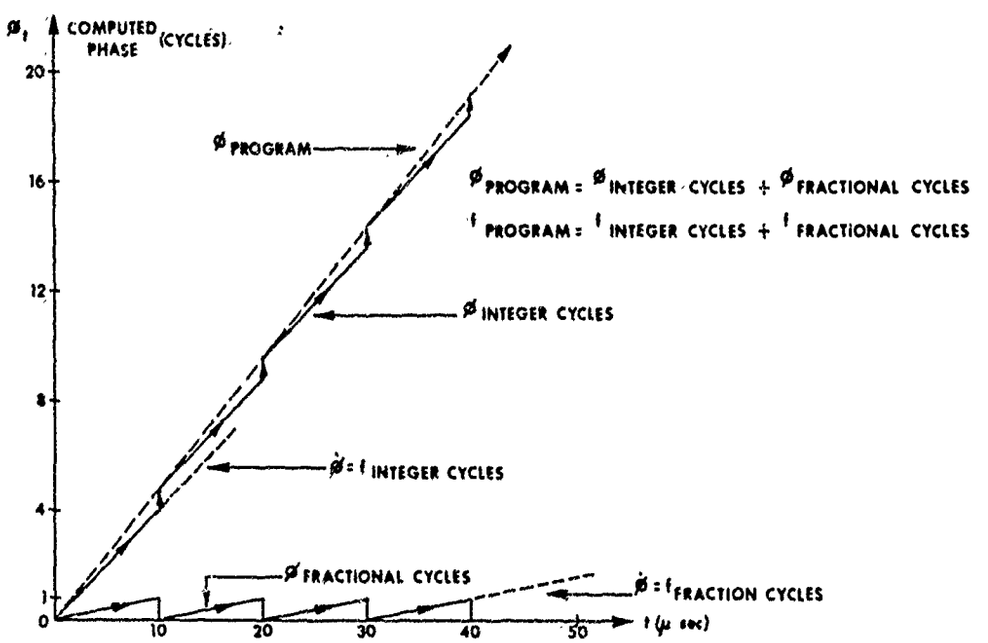
formance is greatly improved. All of the advantages of the digital approach are retained without any implicit performance restrictions.

Acknowledgement

The author wishes to acknowledge the contribution of Mr. Noel B. Braymer at Dana Laboratories whose support and invention (patent pending) of the Digiphase technique made this project possible.

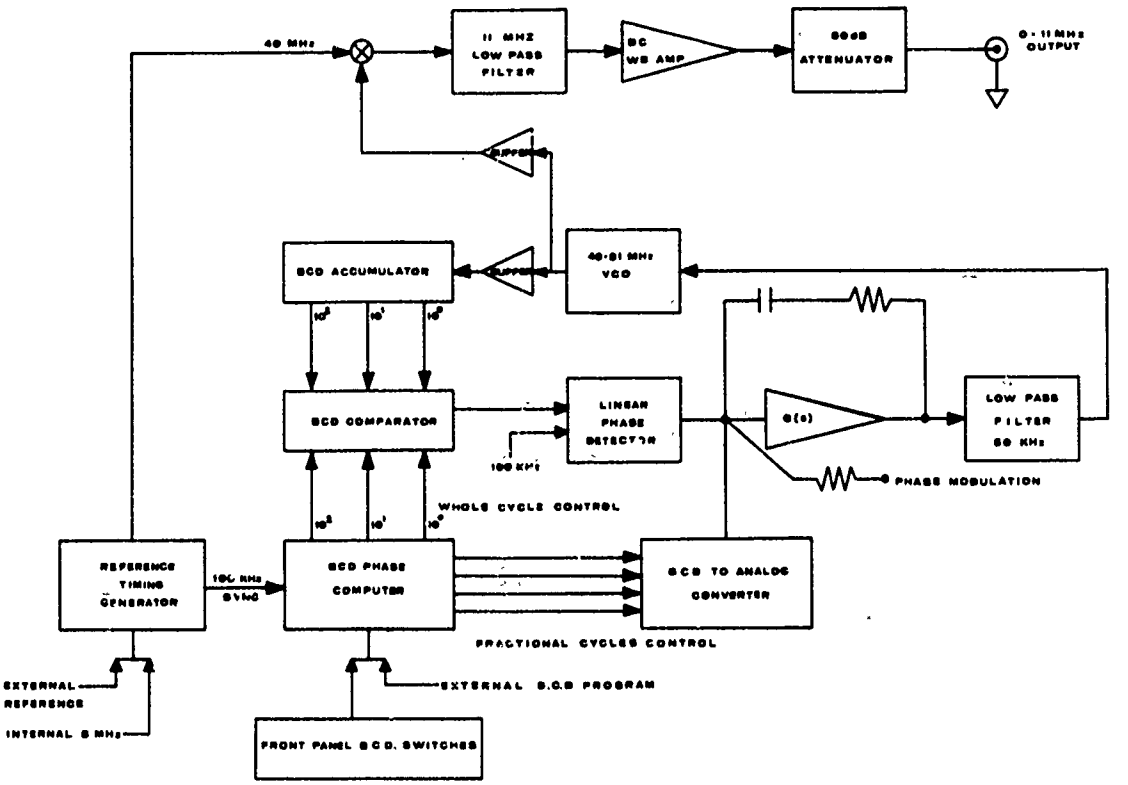
INTEGER CYCLE PHASE COMPARATOR



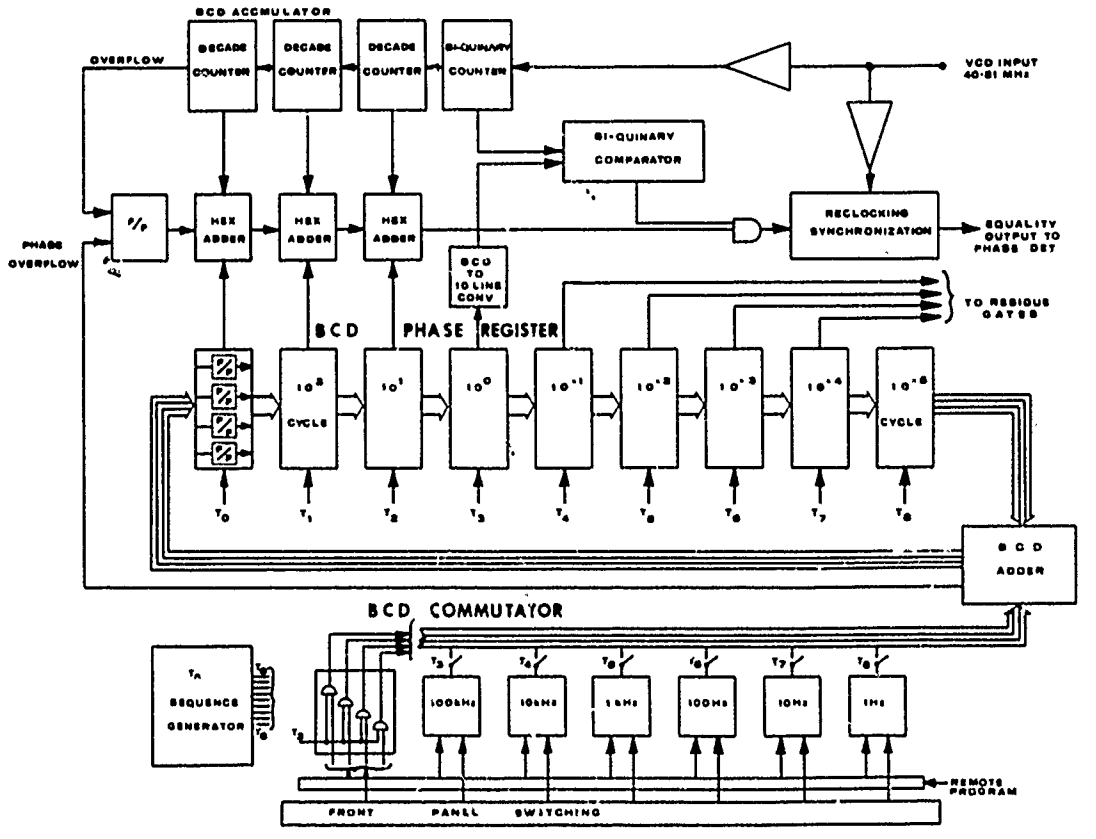


DIGIPHASE SYNTHESIZER

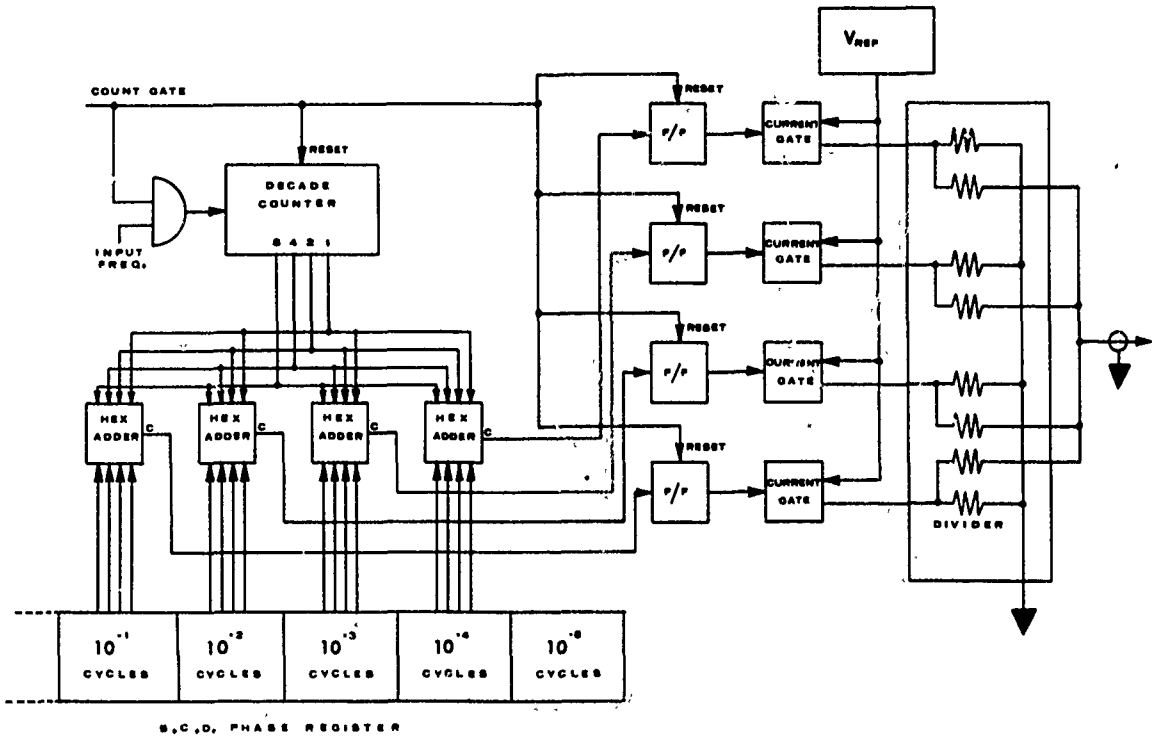
DIGIPHASE SYNTHESIZER



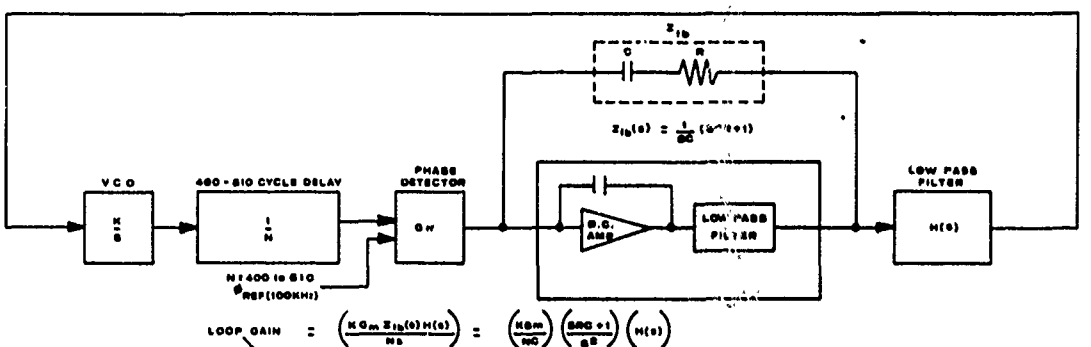
LOGIC DIAGRAM



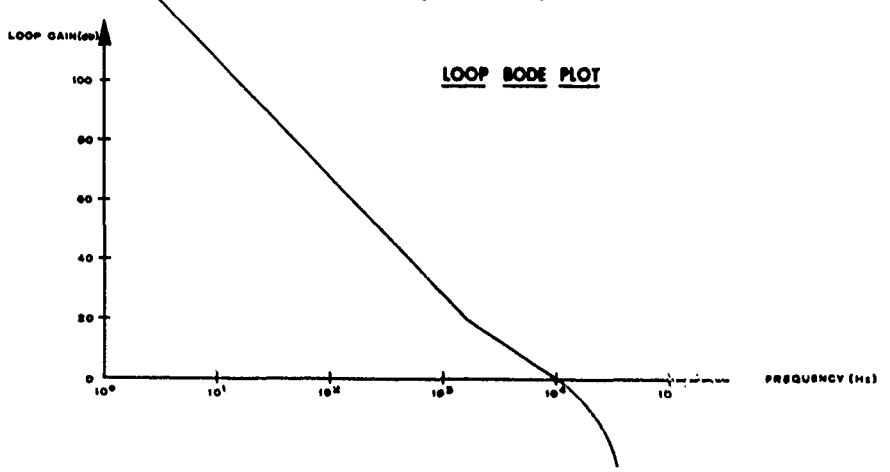
RESIDUE MODULATION GENERATOR



LOOP BLOCK DIAGRAM

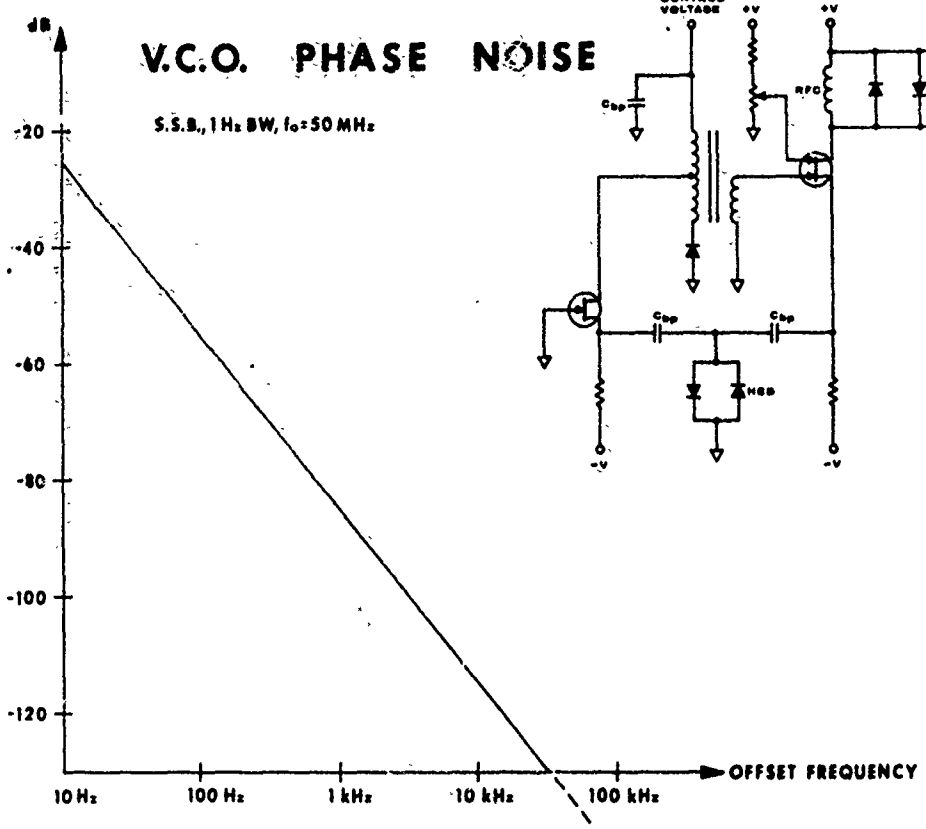


$$\text{LOOP GAIN} = \left(\frac{K G_m Z_{1b}(s) H(s)}{N_b}\right) = \left(\frac{K G_m}{sC}\right)\left(\frac{sRC+1}{s^2}\right)\left(H(s)\right)$$

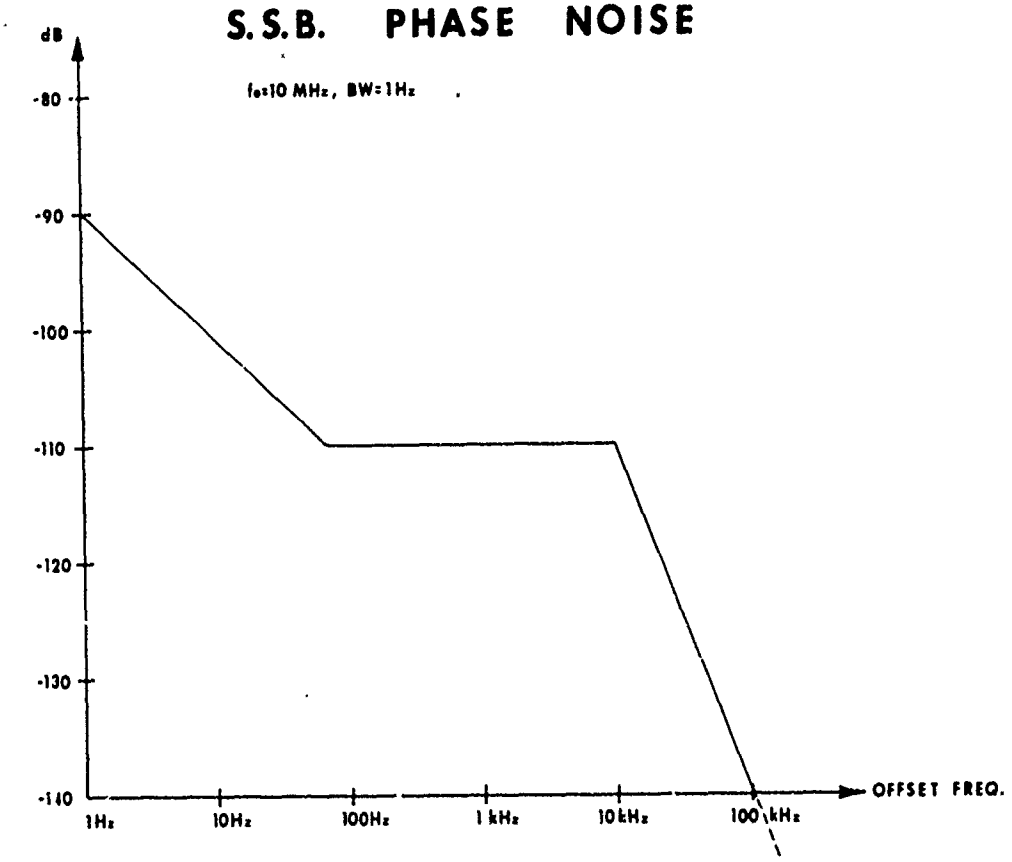


LOOP BODE PLOT

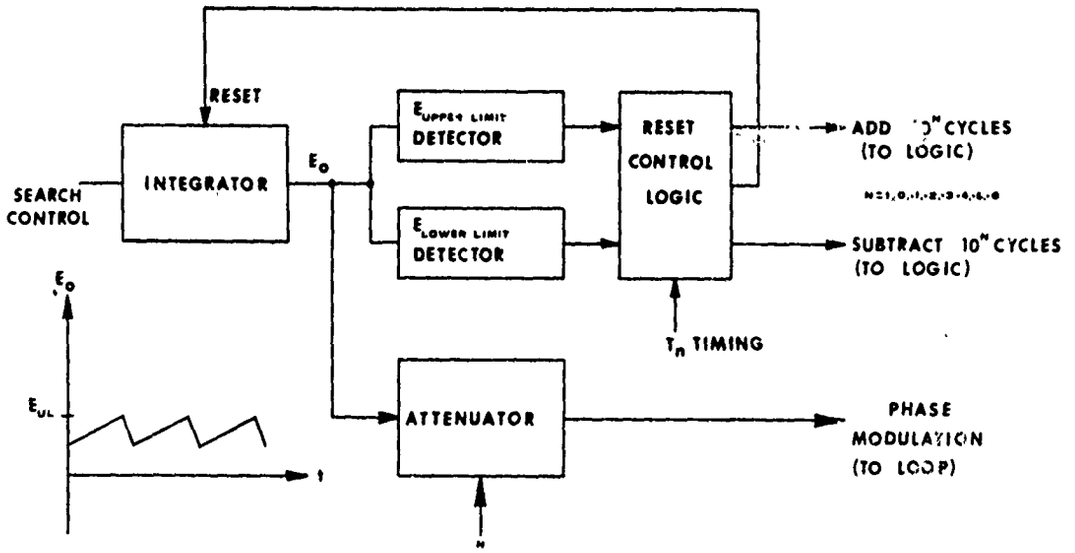
V.C.O. PHASE NOISE



S.S.B. PHASE NOISE

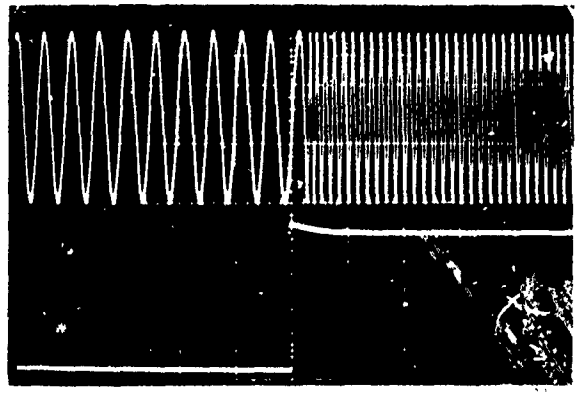


SEARCH

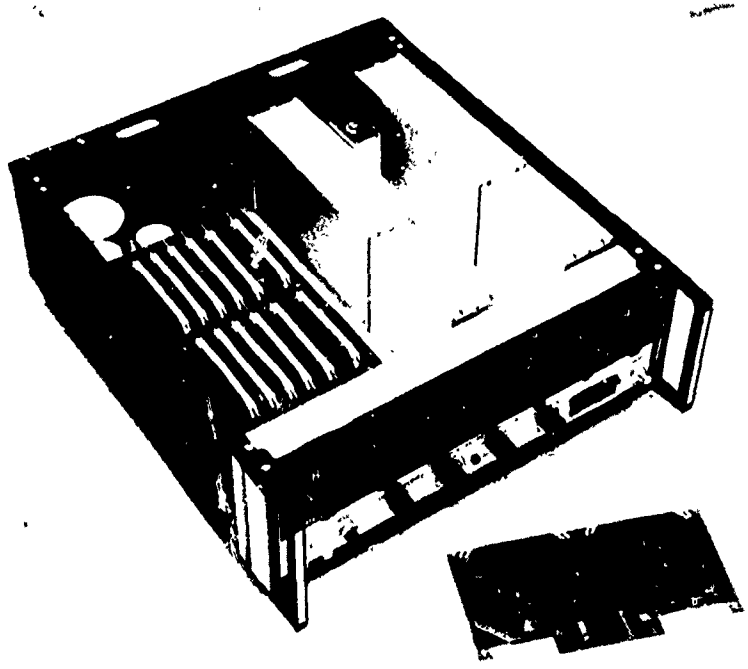


SYNTHESIZER SPECIFICATIONS

* **Frequency Range**		DC to 10,999,999 Hz		DC to 109,999.99 Hz
Least Step		1 Hz		.01 Hz
* **Output Voltage**		1 volt RMS, into 50 ohms from 50 ohm source		1 volt RMS, into 50 ohms from 50 ohm source
* Vernier		0 to −10 dB		0 to −10 dB
Frequency Response		±1 dB, DC to 10 MHz		±1 dB, DC to 100 kHz
Spurious Outputs (Non-harmonic)		−90 dB		−90 dB
(Sidebands within f_o ± 100 kHz)		−60 dB		−90 dB
(Harmonic)		−30 dB		−50 dB
Signal to Phase Noise				
Within f_o ± 15 kHz (excluding ± 10 Hz from f_o)		90 dB/Hz		130 dB/Hz
Beyond f_o ± 15 kHz (excluding ± 10 Hz from f_o)		110 dB/Hz		130 dB/Hz
Total in Band f_o ± 15 kHz (excluding ± 1 Hz from f_o)		54 dB		85 dB
Signal to AM Noise f_o ± 15 kHz (excluding ± 1 Hz from f_o)		84 dB		84 dB
Frequency Switching Time		10μs synchronous, 20μs worst case		10μs synchronous, 20μs worst case
Frequency Settling Time		<1 MHz change, 200μs >2MHz change, 5ms worst case		<10 kHz change, 200μs >10kHz change, 5ms worst case
Frequency Selection				
Manual		Front Panel Switches		
Remote		BCD 1-2-4-8		
		Logic "1" +4V ± 1V	Logic "0" 0V + .5V	
Frequency Stability	Standard	± $2x10^{-9}$ per day, ± $1x10^{-8}$ per °C, 0 to 50°C		
	Option 2	± $2x10^{-10}$ per day, ± $2x10^{-9}$ per °C, 0 to 50°C		
RMS Fractional Frequency Deviation		AVG Time	10 MHz	AVG Time 100 kHz
		10 ms	$3x10^{-9}$	10 ms $3x10^{-8}$
		1 s	$3x10^{-11}$	1 s $3x10^{-11}$
External Reference		1 MHz or 5 MHz, 0.1 V into 100 ohms		
Phase Modulation (3 kHz Bandwidth)		Approximately 36 Volt, linearity ±0.1%		



10 kHz to 30 kHz Settling Time
(200 μsec. per cm.)



Reproduced from best available copy.

A

MINIATURE PRECISION DIGITAL FREQUENCY SYNTHESIZER

Richard J. Hughes

and

Robert J. Sacha
Sylvania Electronic Systems
Williamsville, New York 14221

Summary

This paper describes an indirect precision digital frequency synthesizer developed by Sylvania for experimental use by the U.S. Army Electronics Command, Fort Monmouth, New Jersey, under Contract No. DA28-043-AMC-02408(E). The synthesizer has a tuning range of 70 to 98 MHz in 100 Hz frequency increments. The synthesizer was designed to have high signal purity and stability, small size and low power consumption while still maintaining a large tuning range.

Data is presented which demonstrates that the indirect precision digital frequency synthesizer is capable of meeting the spurious signal level and tuning time requirements. The concept utilized is a two loop phase-locked system, where the output frequency is obtained from a VCO in one of the loops.

The two loop synthesizer uses a narrow band HF loop (minor loop) and a wide band VHF loop (major loop). The minor loop operates over the frequency range 23.0 to 23.999 MHz in 1 kHz increments. In the minor loop an IF signal generated from a single Voltage Controlled Oscillator (VCO) is used as the input signal for an electronically controlled variable digital frequency divider. A 1 kHz reference is phase compared with the variable divider output to generate a VCO error correcting signal. The output of the minor loop passes through a fixed decade divider to produce the band of frequencies 2.3 to 2.3999 MHz in 100 Hz increments. Three thumbwheel switches control the tuning of the minor loop VCO to provide the 100 Hz, 1 kHz, and 10 kHz tuning increments of the synthesizer. The decade divider output is translated up in frequency by a signal derived from the master oscillator, filtered and injected into the major loop. The major loop operates over the frequency range 70 to 97.9999 MHz utilizing two VCO's to cover the range. VCO switching, which is controlled by the 10 MHz switch, is required to tune the required range using a low voltage control signal. An IF signal generated by combining the VCO and minor loop signals drives a variable digital frequency divider. The divider output is frequency and phase compared with a .5 kHz referenced signal to provide error signals which after filtering provide the coarse and fine tuning control for the active VCO. Three thumbwheel switches and a diode decode matrix are used to tune the major loop in 100 kHz, 1 MHz and 10 MHz increments.

Implementation of this concept is presented with emphasis on the advantages of a two loop synthesizer in preference to a single loop synthesizer to achieve signal purity, frequency resolution and tuning speed. The package design is presented to show the modular concept utilized to achieve small size and signal isolation.

Extensive testing was performed to evaluate synthesizer performance and much of the resultant data is presented in this paper. Test results presented in this report include spectrum analysis to determine the level of spurious signals, wave analyzer analysis to determine signal to noise ratio, oscillographic recordings to show the mode and time required to achieve frequency lock.

Design Requirements

The synthesizer provides 100 Hz resolution in the frequency range of 70.0000 to 97.9999 MHz while the tuning dial covered the range of 2.0000 to 29.9999 MHz. Table 1 lists briefly the synthesizer specifications. Figure 1 shows the packaging configuration of the delivered unit. The overall dimensions are 5 x 4.3 x 1.7 inches.

Design Approach

In order to fully understand the problems associated with a 100 Hz resolution synthesizer utilizing indirect synthesis techniques, two basic system designs were studied. The first was a single loop synthesizer having a 100 Hz comparison frequency. A design typical of a single loop system is shown in Figure 2. Using this design and the representative gains as shown, a loop analysis was performed and the following points were noted:

1. Loop acquisition time would exceed the required 0.5 second specification. Assuming a minimum of 60 samples required of the phase detector to bring the VCO into lock, the tuning time

would be 600 milliseconds.

2. The loop bandwidth would be quite low. This would be the resultant effects of both providing enough loop filtering to attenuate the 100 Hz sampling transients sufficiently to meet the -65 dB spurious requirement at the VCO and the low open loop gain of the system due mainly to the high division ratio of the divide-by-N. Attenuation on the order of 80 dB would be required of the loop filter at 100 Hz to maintain a maximum of 4 micro-volt of 100 Hz phase detector noise appearing at the VCO control input. Also, the large loop division ratio required to divide to 100 Hz reduces the open loop gain considerably. The resultant bandwidth would be on the order of 0.1 Hz.

3. The VCO noise specification would not be met since the loop will not provide any filtering of noise due to VCO incidental FM above a 0.1 Hz deviation rate. Tests performed on a variety of oscillator designs utilizing both bipolar and field effect transistors indicated that a -60 dB noise floor could not be achieved by the VCO itself even when frequency locked.

It thus becomes obvious that a single loop synthesizer having a 100 Hz resolution will not meet all the design specifications. The concept was thus proved to be inadequate and a two loop concept established.

The two loop synthesizer uses (1) an HF loop (minor loop) and (2) a VHF loop (major loop). A block diagram of the synthesizer is shown in Figure 3.

The minor loop is completely independent of the main loop and is slaved to the master oscillator through a fixed 25 MHz signal and a 1 kHz reference. As shown in the block diagram of Figure 3, the minor loop VCO (23-24 MHz) is down converted to a 1-2 MHz IF, scaled in the divide-by-M and phase compared with the 1 kHz reference. The phase detector output is filtered in an RC network to attenuate the phase detector sampling noise and provide a means of adjusting the loop response. The filtered phase detector output returns to the VCO as an error correcting signal with a control capability of 100 kHz/volt.

The output of the minor loop passes through a fixed decade divider to produce a band of frequencies 2.3 to 2.3999 MHz in 100 Hz increments. The minor loop controls the 100 Hz, 1 kHz and 10 kHz tuning increments of the synthesizer. The decade divider output is translated up in frequency by a signal generated in the fixed frequency section and injected into the major loop.

The major loop uses a 25 kHz reference signal to control the 100 kHz, 1 MHz, and 10 MHz increments of the synthesizer. The 10 MHz switch also selects one of two VCO's having individual ranges of 70 to 78 MHz and 78 to 98 MHz. This division of the tuning range maintains a maximum tuning ratio of 1.26:1, allowing the VCO's to be controlled with a voltage range of two to eight volts. Frequency error is measured in a digital discriminator which controls a digital-to-analog converter to coarse tune the VCO. Phase error information is generated in a sample and hold phase detector. Noise from the phase detector is filtered out by an RF filter before the phase detector voltage reaches the VCO.

An analysis of the two loop synthesizer system described yielded the following points:

1. 100 Hz resolution could be achieved without the attendant 100 Hz spurious signals at the synthesizer output as would be expected in a single loop system. Further, spurious signals generated in the minor loop could be allowed to be higher than the output spurious specifications by the amount of fixed division on the minor loop signal, in this case 20 dB.

2. Synthesizer lockup time would be the sum of the lockup times of the two loops. For the minor loop operating with a 1 kHz sampling rate, lockup time could be expected to be no greater than 60 milliseconds. The main loop could be expected to lock in 2.4 milliseconds with an additional 0.64 milliseconds taken during frequency search. Overall lockup time would always be less than 64 milliseconds.

3. Close-in noise resulting from VCO incidental FM can be greatly attenuated by increasing the bandwidth of the main loop beyond the lower bound of the noise specification, 300 Hz. The wide bandwidth can be achieved because of the high open loop gains as noted on the block diagram in Figure 3, and because it is possible to use a minimum of attenuation in the loop filter to maintain a fifteen microvolt maximum required 25 kHz signal on the control line for -65 dB spurious at the VCO output.

Packaging

The package developed for this synthesizer divided the unit into functional circuit groups and where possible to have each group on a separate printed circuit card. The synthesizer is comprised of 10 plug-in PC cards, two motherboards, and a switch assembly. Each card has an overall dimension of 1.5 x 3.5 inches.

A photograph of the synthesizer with the cover removed and all components in place is shown in Figure 4. Figure 5 shows a photograph of the component side of each PC card. To ensure good signal isolation, each PC card must be completely enclosed; therefore, a shield was placed between each PC card, and the shields were soldered to the motherboard and the case walls. A second motherboard, containing the diode matrices for control of the variable dividers, is removable for troubleshooting, if necessary, and located below the six (6) section switch. The width of the package was determined primarily by the six (6) section switch, the height of each PC card by the TCXO, and the length was adjusted so the synthesizer could be contained on the 10 PC cards. ELCO pins are used for connection between the PC card and the motherboard. This method is superior to using an edge connector because (1) it requires less volume and (2) the signals enter and leave the PC card at the most desired location. Signal flow is such that there are only two signal crossovers on the motherboard. Connections to the diode matrix card are by wire lead. The design of the variable dividers is such that these control leads carry only a dc signal rather than RF or pulse signals.

Switching Regulator

Two voltages are required to operate the synthesizer. These are: +9.5 VDC, supplied from an external source, and +4.3 VDC generated from a pair of switching regulators operating from the primary supply. Two regulators are supplied to generate separate voltages for digital and analog circuitry to provide a high degree of signal isolation to the digital sections of the synthesizer. A significant amount of power saving was realized by operating much of the analog circuitry at +4.3 VDC rather than from the primary supply. Each regulator switches at a rate of approximately 500 kHz. Because of the large switching transients, the switching regulators were designed using thick film techniques for small size and were placed in a sealed can to prevent radiation of the switching signal. Measurements of the synthesizer output indicated that no switching transients were present at a level of 70 dB below the carrier. Each regulator operates with an efficiency of about 75% and each delivers about 0.5 watt at +4.3 VDC.

The selection of integrated circuits for the counters was made to minimize the power consumption where possible. Higher speed, higher power integrated circuits were used only where speed was a factor. Low speed, low power decade counters were assembled using dual flip-flops having three milliwatt dissipation per binary division.

Test Results

The performance tests most indicative of the capabilities of the two loop synthesizer were found to be: (1) Observation of acquisition time for both the main and minor loops as well as simultaneous switching of both loops; (2) Close-in spurious, particularly in the band 100 Hz to 25 kHz from the carrier; (3) Far out spurious or nonharmonic signals greater than 25 kHz from the carrier; (4) Synthesizer output noise 300 Hz to 3 kHz from the carrier; (5) Loop stability under conditions of worst case loop gains.

The photographs of Figure 6 show the transient VCO control voltage for both loops individually under conditions of maximum frequency change. Minor loop acquisition is seen to occur in 16 milliseconds while the main loop acquires in about one millisecond. Both loops acquire in 17 milliseconds. This demonstrates that the loop acquisition time is governed primarily by the minor loop.

Figure 7 shows a spectrum analyzer display of signal content up to 50 kHz from the carrier. No 25 kHz spurious signals are noted. However, as shown in Figure 8, close-in spurious signals at one and two kilohertz were found. Comparing the amplitude of these spurs with those generated in the minor loop (Figure 9), the assumption that spurious signals generated in the minor loop are attenuated by the fixed division and main loop filtering are confirmed. An investigation into the cause of the one and two kHz spurs at the main output disclosed a deficiency in the isolation between the 1 kHz signal in the divide-by-M and the 25 kHz reference signal. These two signals are near each other on the same card and consequently cross-couple.

Far out spurious are displayed in Figure 10. All nonharmonic signals are seen to be greater than 40 dB below the carrier. The high degree of unwanted signal attenuation is indicative of the adequacy of the packaging concept in the analog portion of the synthesizer.

Figure 11 shows a chart recording of the close-in noise content of the synthesizer output signal. Noise has been plotted in a bandwidth of 10 Hz and 100 Hz per division. The noise pedestal up to the carrier exhibits the effects of a high bandwidth loop wherein the lower frequency incidental FM of the loop has been attenuated by the action of the loop. Noise beyond the pedestal falls off at a 1/f rate, typical of the VCO noise characteristic. At 1 kHz from the carrier, the noise is well below the required -65 dB level.

Loop stability has been verified in two ways. First, a step was inserted on the VCO control voltage and the transient response observed. The damping factor was determined to

be approximately 0.5 under worst case conditions, indicative of adequate open loop gain and phase margins. Secondly, a computer analysis of the loop was performed, to determine the loop bandwidth as well as loop attenuation capability within the loop bandwidth. Figure 12 shows the resultant computer plots. The loop response is shown to peak no more than 2.5 dB at the loop resonance. Also obtained from the computer analysis is the gain and phase margins for incremental steps of N. The margins are tabulated in the table of Figure 12. Worst case gain and phase margins are noted to be -22.24 dB and 45.60 degrees respectively.

Conclusions

The results of the performance tests on the exploratory development model proved that a two loop synthesizer having a high degree of resolution will be far superior to a single loop synthesizer in terms of output spurious, acquisition time, and noise due to incidental FM. In addition, two loop indirect synthesis can be used to generate signals in the 100 MHz region with low power consumption and size.

Requirement	Specification
Frequency Range	Variable: 70.0 to 97.9999 MHz with a corresponding dial readout from 1.0 to 30.9999 MHz
Tuning Increment	100 Hz
Signal/Noise Ratio	60 dB in a 300 Hz to 3 kHz band above the desired signal
Output Voltage	100 millivolts across 50 ohms
Capture Time	Within 15 Hz of output frequency in less than 0.5 second
Spurious Signals	Nonharmonic: 65 dB below desired Harmonic: 40 dB below desired (design goal)
Input Voltage	9.5V ± 0.5%
Input Power	3.0 watts (1.5 watts design goal)
Volume	35 cubic inches
Frequency Stability	$\pm 4 \times 10^{-7}$ from -40°C to +75°C $\pm 1 \times 10^{-6}$ from -55°C to +71°C

Table I Synthesizer Specifications

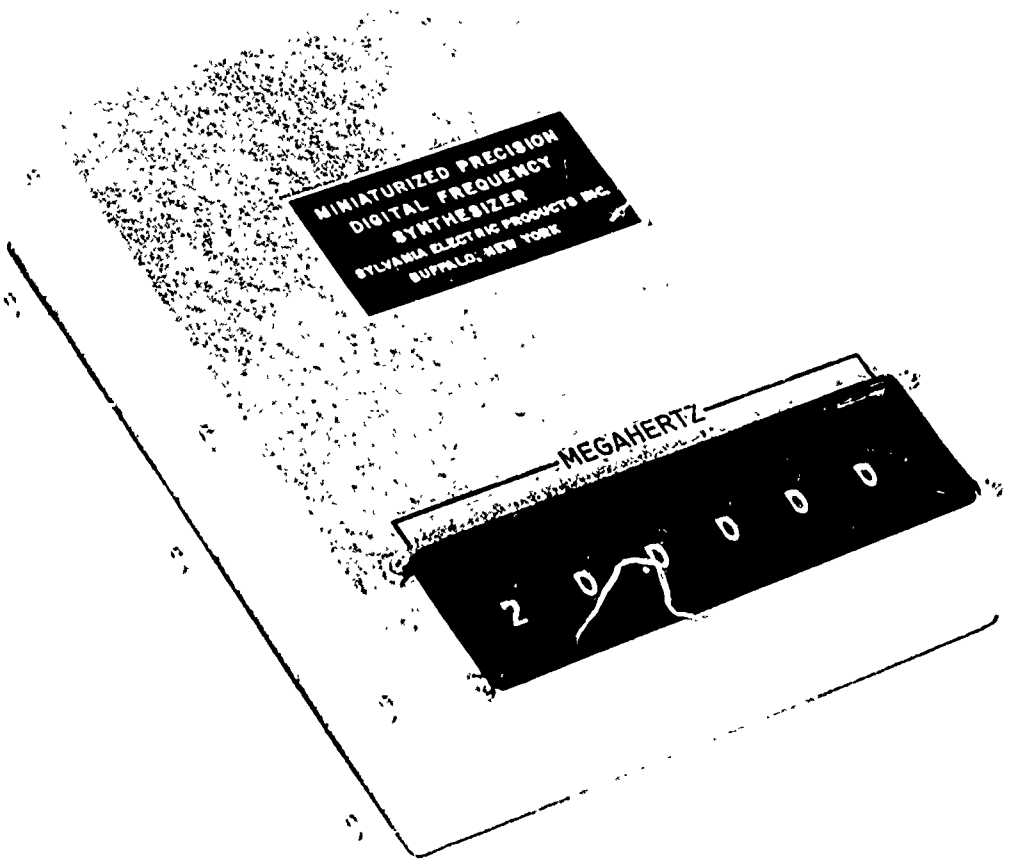


Figure 1 Miniaturized Precision Digital Frequency Synthesizer Unit

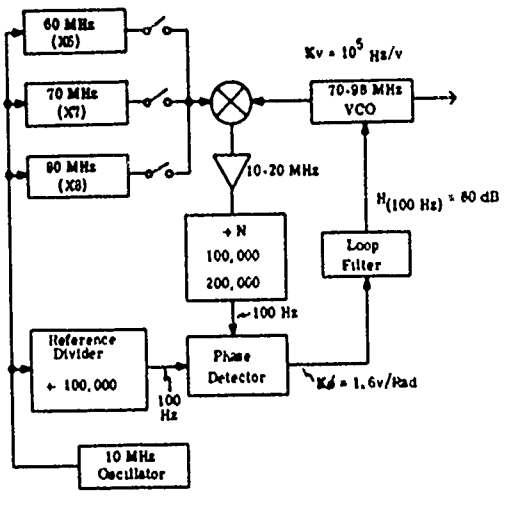
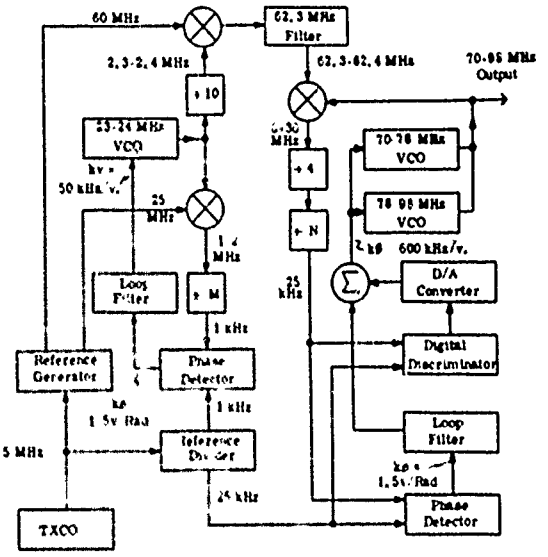
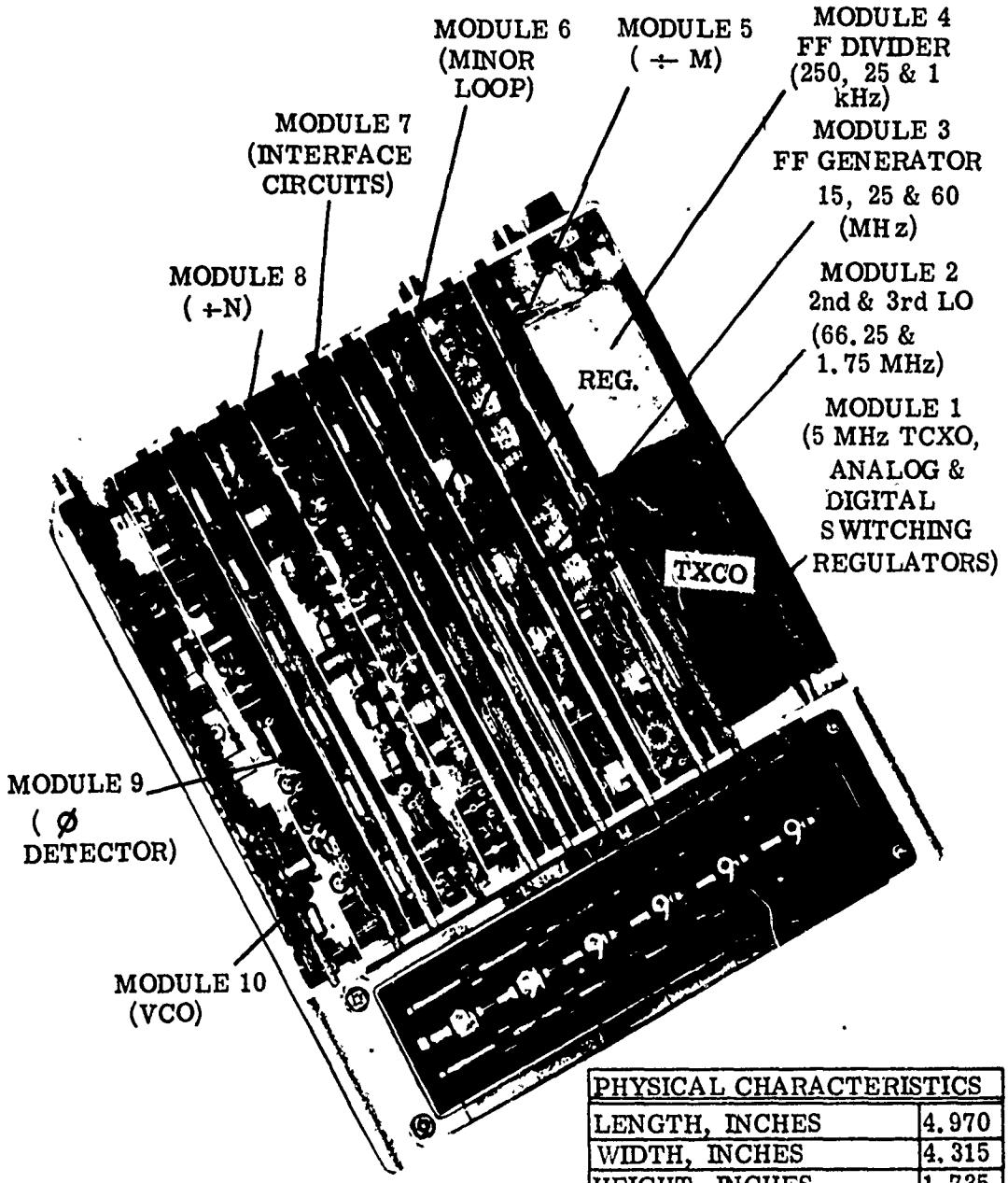


Figure 2 100 Hz Synthesizer, Single Loop



Two Loop Synthesizer Block Diagram



PHYSICAL CHARACTERISTICS	
LENGTH, INCHES	4.970
WIDTH, INCHES	4.315
HEIGHT, INCHES	1.735
VOLUME, CUBIC INCHES	37.2
WEIGHT, OUNCES	25.

Figure 4 Miniature Precision Digital
Frequency Syntheiszer (Top Cover Removed)

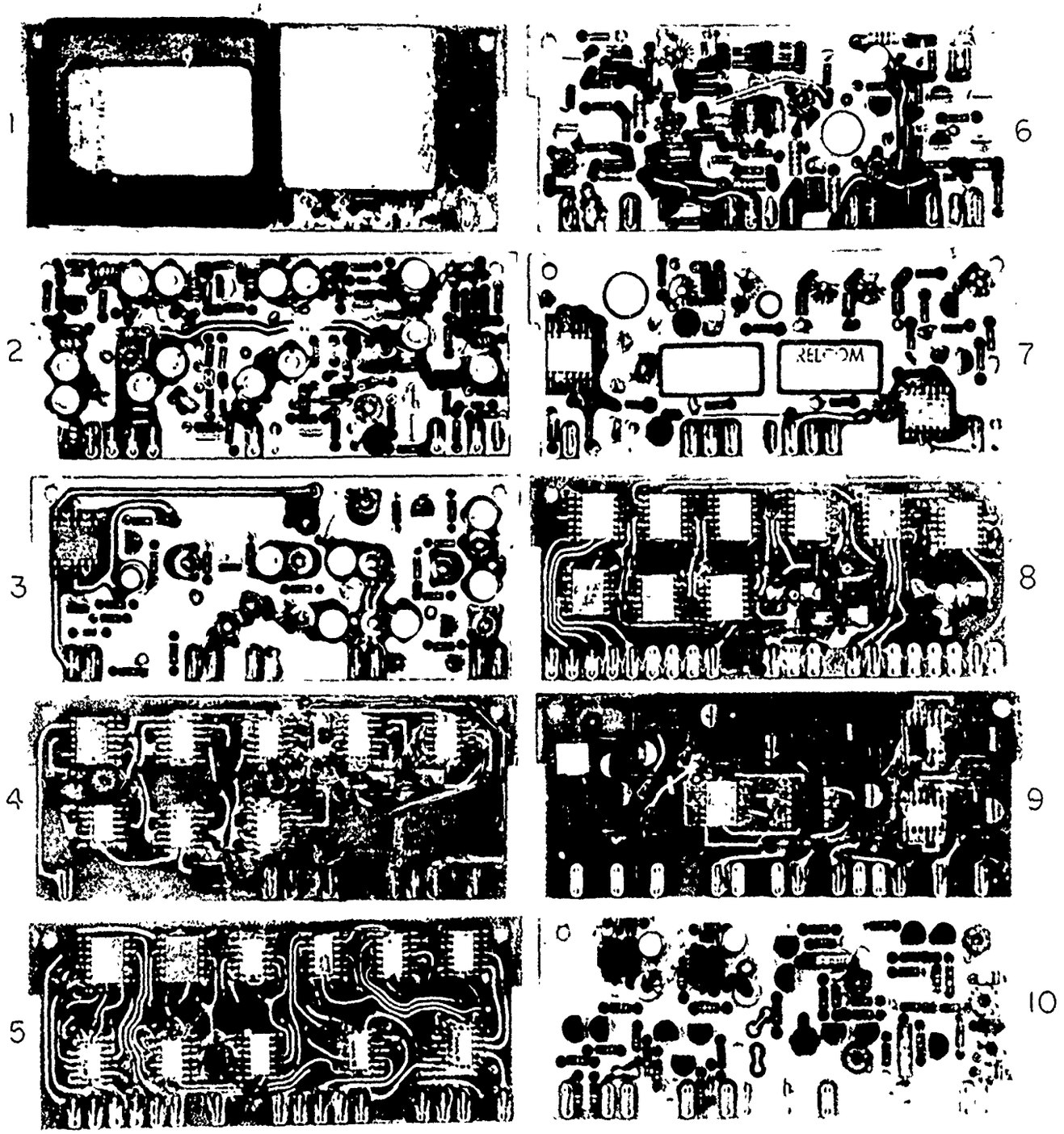


Figure 5. Photograph of Component Side of Modules

Reproduced from best available copy.

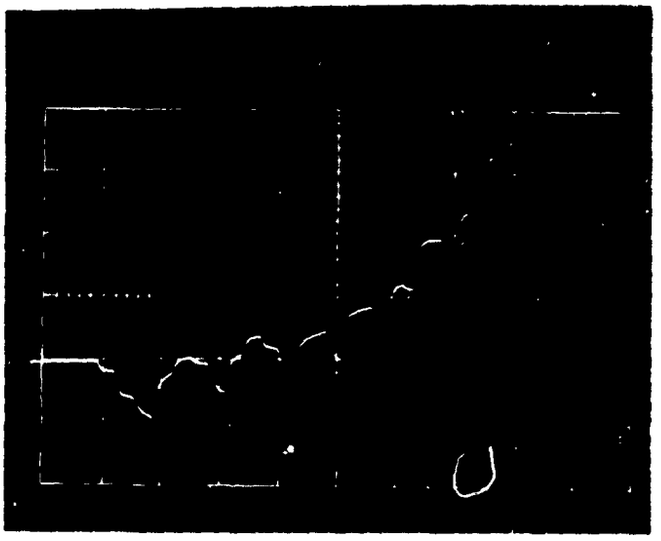


Figure 6a

Main Loop Lock-up Time

0.1 Milliseconds/Division

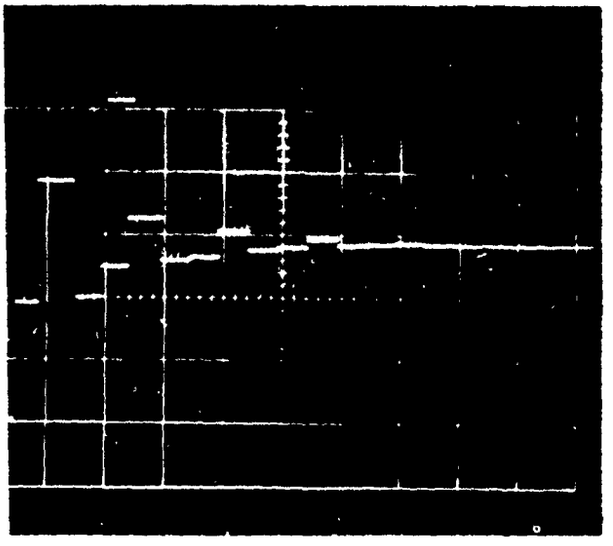


Figure 6b

Minor Loop Lock-up Time

2.0 Milliseconds Division

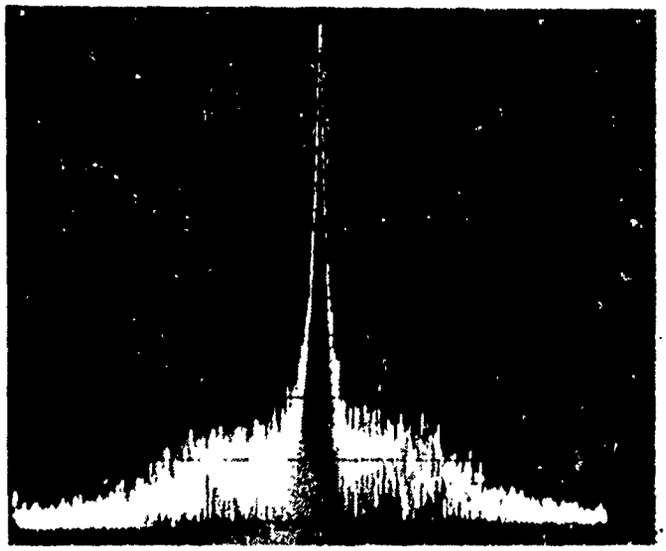


Figure 7

Synthesizer Output Signal

10 kHz/Division

10db/Division

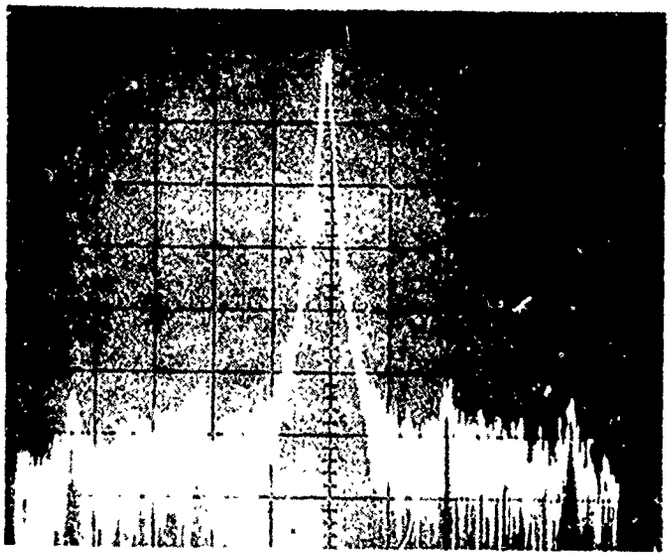


Figure 8

Synthesizer Output Signal

500 Hz/Division

10 db/Division

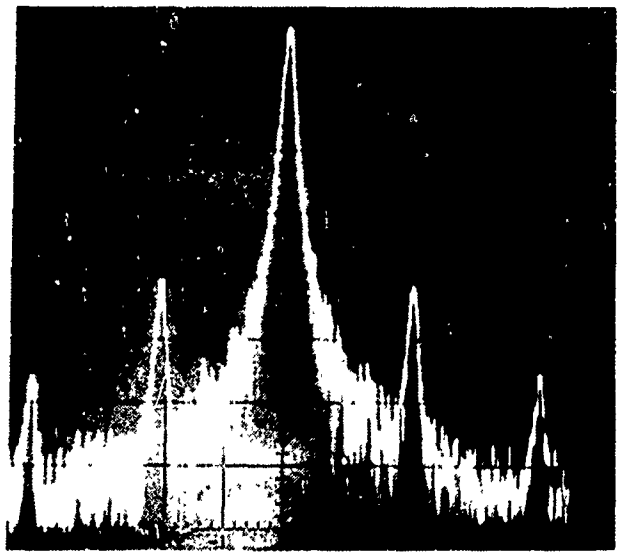


Figure 9
Minor Loop Output Signal
500 Hz/Division
10 db/Division

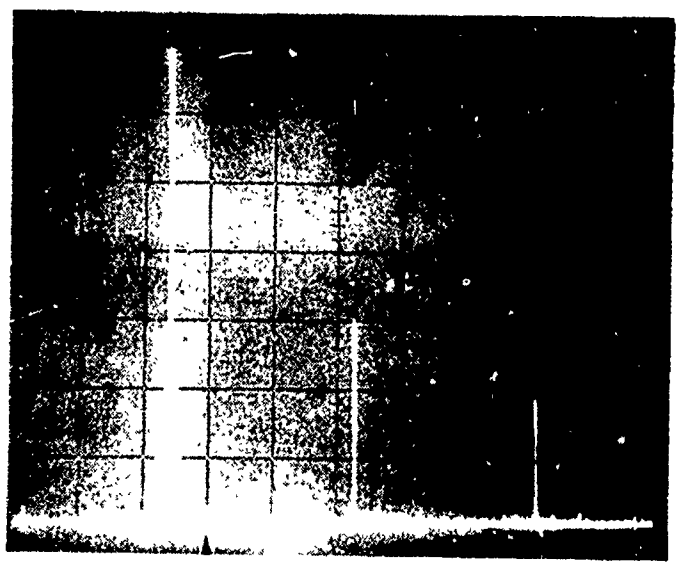


Figure 10
Synthesizer Output Signal
Wideband
30 MHz/Division
10 db/Division

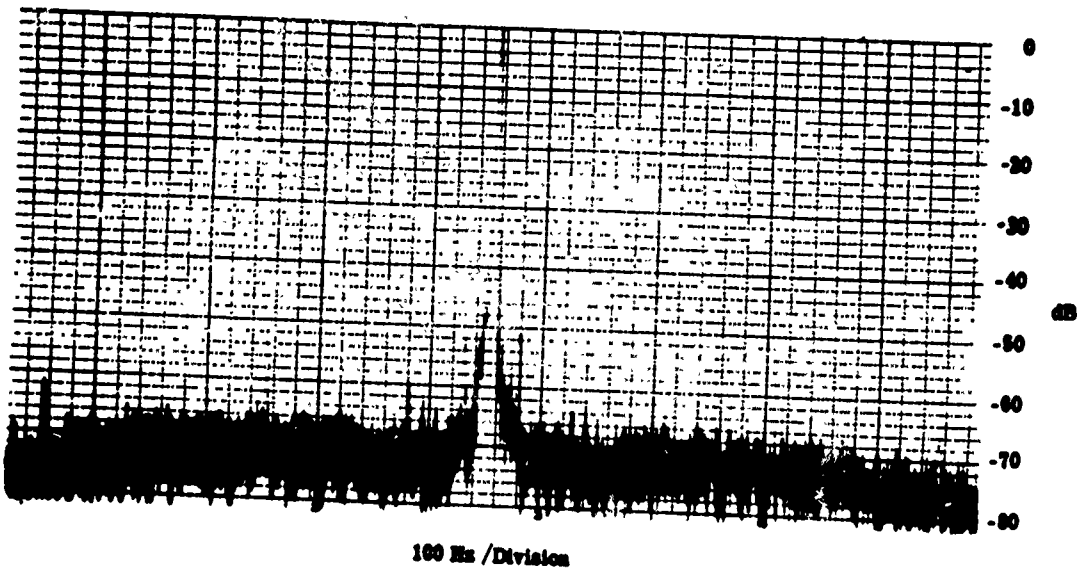


Figure 11 Synthesizer Output Noise Characteristic

GAIN AND PHASE MARGIN

MINIATURE PRECISION 100 HZ SYNTHESIZER ANALYSIS

	GAIN MARGIN (DB)	PHASE MARGIN (DEG)	FREQUENCY (RAD/SEC)
308.00	-22.24	45.60	1396.65
348.00	-23.30	46.19	1308.02
388.00	-24.24	46.77	1219.39
428.00	-25.10	47.35	1130.76
468.00	-25.87	47.93	1042.13
508.00	-26.58	48.52	978.80
548.00	-27.24	49.10	938.40
588.00	-27.85	49.68	897.99
628.00	-28.43	50.27	857.58
668.00	-28.96	50.85	817.18
708.00	-29.47	51.40	787.26
748.00	-29.94	51.93	765.09
788.00	-30.40	52.46	742.92
828.00	-30.83	52.99	720.75
868.00	-31.24	53.52	698.57
908.00	-31.63	54.05	676.40
948.00	-32.00	54.58	654.23
988.00	-32.36	55.11	632.06
1028.00	-32.71	55.64	609.89
1068.00	-33.04	56.10	594.20
1108.00	-33.36	56.57	583.72
1148.00	-33.67	56.91	573.24
1188.00	-33.96	57.32	562.76
1228.00	-34.25	57.72	552.28
1268.00	-34.53	58.13	541.81
1308.00	-34.80	58.53	531.33
1348.00	-35.06	58.94	520.85
1388.00	-35.31	59.34	510.37

FIGURE 12

GAIN MARGIN ATTAINED AT 9.62.9141 RAD/SEC

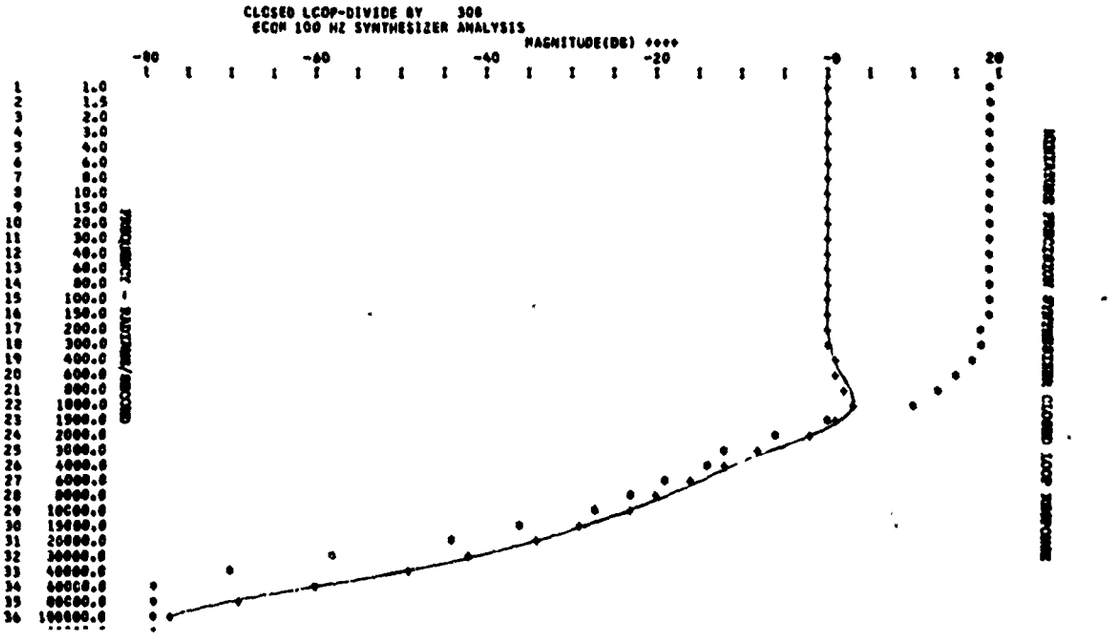


FIGURE 12-A

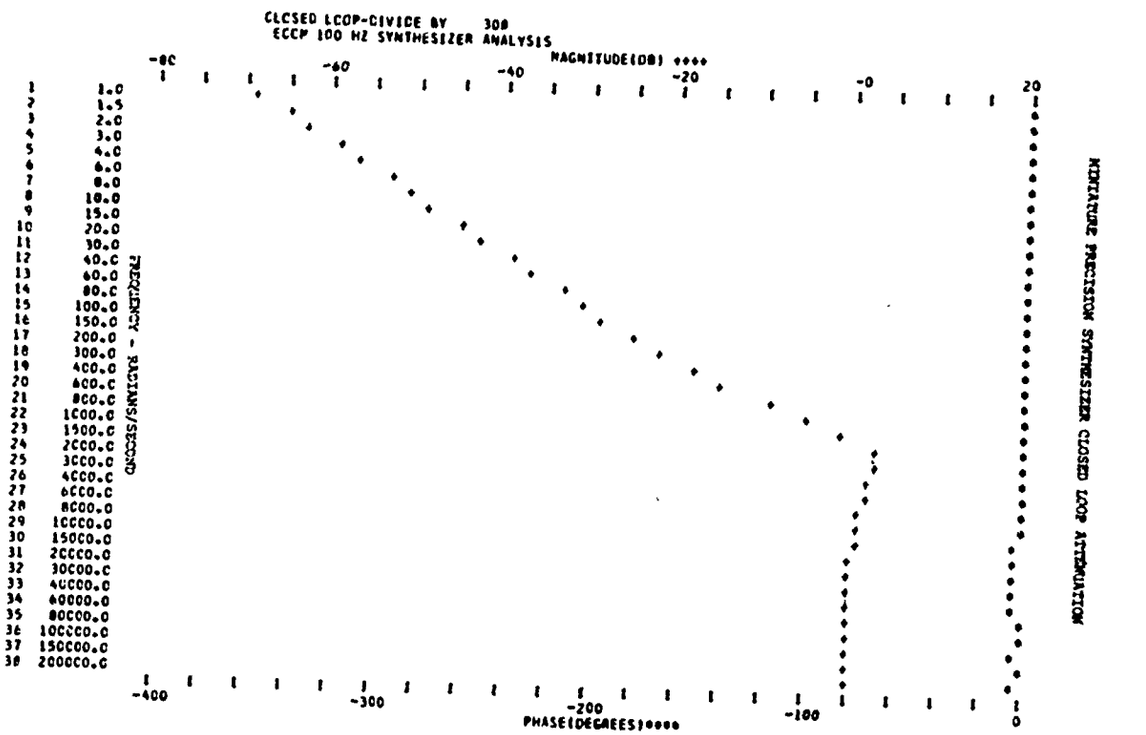


FIGURE 12-B

A CARRIER SUPPRESSION TECHNIQUE FOR MEASURING S/N AND CARRIER/SIDEBAND RATIOS GREATER THAN 120 DB

Charles H. Horn
Collins Radio Company
Cedar Rapids, Iowa

Summary

Requirements for oscillators, synthesizers, and other frequency sources with signal/noise and carrier/sideband ratios in excess of 100 dB have created a need for a method of extending the dynamic range of existing instrumentation. With these problems at hand, a method has been developed for extending the dynamic range of spectrum analyzers, voltmeters, etc., 60 to 70 dB through use of a carrier suppression and reamplification technique. This paper develops the technique in detail, describes a piece of equipment built around this technique, and presents data obtained on various frequency sources using the equipment.

The carrier of the signal to be examined is suppressed by summing it with a generated reference signal equal in amplitude, but opposite in phase. Phase balance and amplitude balance adjustments are provided by manual controls, with the proper phase relationship being maintained by phase locking the reference signal to the incoming carrier. In effect, this system creates a very narrow band rejection filter, with the filter bandwidth being determined by the phase-locked loop characteristic. An interesting innovation to normal phase locking is that the reference signal is locked to the suppressed carrier at the output of the summing point, and not to the incoming signal directly, making the phase balance self-adjusting once a gross phase balance has been achieved with the manual control. Hold range of the reference oscillator is maintained by adjusting its voltage-variable capacitor operating point as the carrier is nulled. Carrier and reference signal summing is done in a hybrid transformer, and the remaining signal then passes through a bandpass filter to an amplification stage. The bandpass filter attenuates any high level, low frequency, or harmonic components which would saturate the reamplification stage or input stages of the measuring instrument. Other refinements, added to the basic technique for operational ease, are also described.

Pertinent operational characteristics for the instrument are described as follows:

Carrier frequency range:	100 kHz to 40 MHz
Carrier suppression range:	dB
Suppression bandwidth (loop bandwidth):	Approximately ±50 Hz (changeable)
System bandwidth (B/P filter bandwidth):	≧±35 percent of carrier frequency
Input signal level:	0.1 to 3.0 V, 50 ohms
System signal/noise ratio:	≧140 dB, (100 Hz bandwidth, 100 kHz from the carrier, 0.25 V input)

Data obtained using this instrument is presented, including the following:

1. HP-5100B synthesizer noise and spurious levels.
2. HP-606A signal generator noise level.
3. Crystal oscillator noise levels with varied drive levels.
4. HP-5061A cesium beam frequency standard 5 MHz output noise level.

These results are of interest in themselves; but moreover, they will demonstrate a measurement capability not otherwise readily obtainable.

Introduction

This paper will develop the carrier suppression and reamplification technique as follows: first, the basic carrier cancellation technique will be discussed; second, a phase-locked oscillator will be added as a cancellation signal source; and third, the fully developed system as is currently being used will be described and an instrument using this fully developed system will be shown. Data obtained with this instrument, in conjunction with a spectrum analyzer, will also be presented.

Basic Carrier Suppression Technique

The basic theory of the carrier suppression technique is, that if a signal with noise and sideband components is properly summed in a linear summing device with a relatively noiseless signal of the same frequency and amplitude, but opposite phase, the output of the summing point will have a substantially reduced carrier component with noise and sidebands relatively unaffected. These signals can be represented by e_1, e_2, and e_3 respectively as shown in figure 1. After carrier suppression the summing point output, e_3, is amplified, effectively increasing the noise and sidebands with respect to the carrier amplitude.

A bandpass filter following the summing point output is required to remove both large low-frequency components, such as those which might exist in a mixer output, and those resulting from the harmonic content of the carrier waveform. If these are allowed to remain, they would saturate the output amplifier or any following instrumentation when the signal is reamplified. In some applications, this filter could be com-

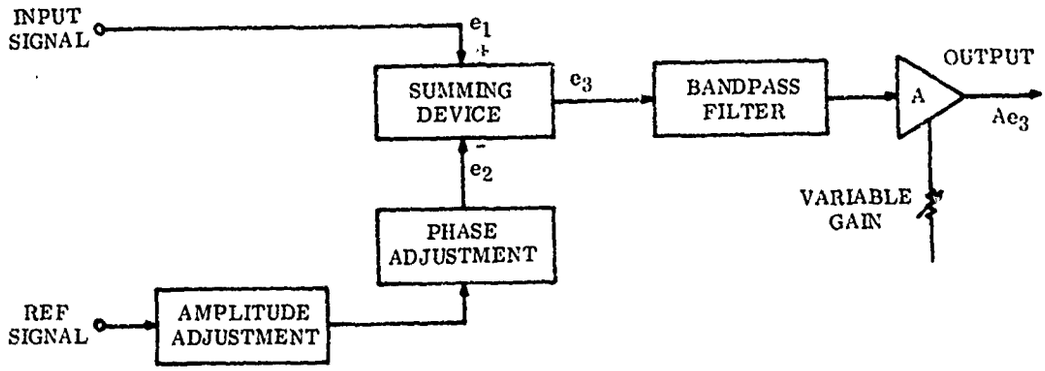


Figure 1. Basic Carrier Suppression Technique.

bined with the output amplifier as a tuned amplifier.

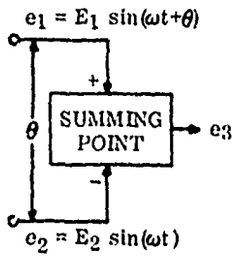
Contrary to the above, there might be instances when one would wish to look at the low frequency or harmonic content; in which case, the filtering would be undesirable.

Summing Point Analysis

An analysis of the summing point can best be performed by differentiating between two conditions which can exist; 1, when phase deviations are present on the carrier to be suppressed and 2, when amplitude variations are present on the carrier. Without a doubt, these two situations could appear simultaneously, and will; but, to simplify the analysis, they will be treated separately as follows:

Carrier Phase Deviations

A summing point with assigned connections and signals as shown in figure 2 is assumed.



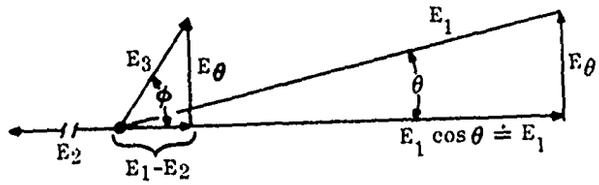
$\theta \ll$ 1 RADIAN

Figure 2. Summing Point Analysis, Phase Variations.

The signal, e_1, is the test signal to be examined and contains incremental phase disturbances, θ, ($\theta \ll 1$ rad). In this analysis, θ can be either a recurring function of time, a random occurrence, a fixed angle or a combination of all of these. In all cases, it can be assumed small because only low level sidebands need be examined using this technique and adequate carrier suppression requires that any static phase error be made small.

The reference signal, e_2, for balancing out the carrier component of e_1, is assumed to have substantially less noise and spurious content than e_1 and is the reference from which θ is compared. Also, nearly exact amplitude balance between e_1 and e_2 is assumed established.

A vector diagram of the signal summation is as shown in figure 3.



$\theta \ll$ 1 RADIAN
$E_\theta \doteq \theta E_1$

Figure 3. Vector Diagram, Phase Variations.

From the vectors, E_3 is the output signal at an angle of ϕ. The output expression for e_3 can be written as follows (see appendix A):

$$e_3 \doteq E_1 (\theta^2 + R^2)^{1/2} \sin(\omega t + \phi) \quad (1)$$

where

$$R = \frac{F_1 - E_2}{E_1} = \frac{\text{suppressed carrier component}}{\text{original carrier component}} \quad (2)$$

$$\phi \doteq \text{Tan}^{-1}\ \theta/R \quad (3)$$

As will be discussed later, a most significant feature of this analysis is that the angle θ is multiplied by the amount of carrier suppression, as shown in equation (3).

The fact that this occurs does not mean that the phase sideband components have been altered or any new ones created because only the carrier component has been suppressed.

Carrier Amplitude Deviations

The same summing point as used for analyzing the phase deviation case will be assumed as shown in figure 4.

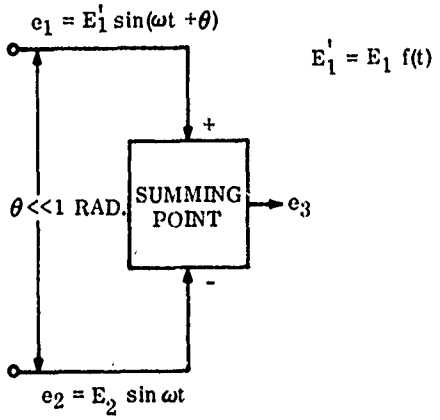


Figure 4. Summing Point Analysis, Amplitude Variations.

In this case, the peak value of the input expression, E_1', varies as some function of time. The reference input, e_2, is assumed constant and is the plane of reference for θ. Variations of E_1' are assumed very small ($\lesssim$ -40 dB). Conditions of approximately 60 dB of carrier suppression are also assumed established. This summing situation can be depicted vectorially as in figure 5.

The vector, E_m, is the amplitude variance portion of E_1' and E_θ and is the phase component of E_1' resulting from the phase difference between E_1 and E_2.

As can be seen from figure 5, the phase of output vector E_3 will vary from ϕ' to ϕ'' as the amplitude and phase components vary. This means that ϕ is a function of E_m and the angle θ, as expressed as follows (see appendix D):

$$\frac{\partial \phi}{\partial E_1'} \doteq \frac{-\theta}{E_1'(R^2 + \theta^2)} \qquad (4)$$

$$\frac{\partial \phi}{\partial \theta} = \frac{R}{(R^2 + \theta^2)} \qquad (5)$$

Even though the summing point output, E_3, is a function of θ, E_1' and R, there are no alternations in the original modulation components nor are there any new ones created. This is because, as stated in the phase deviation case, only the carrier component is suppressed.

Characteristics of the Summing Point Output

From the analysis of the phase deviation case and the amplitude deviation case, the summing point output components are as follows:

1. The original phase components of the input signals.
2. The original amplitude components of the input signals.
3. The remainder of the suppressed carrier.

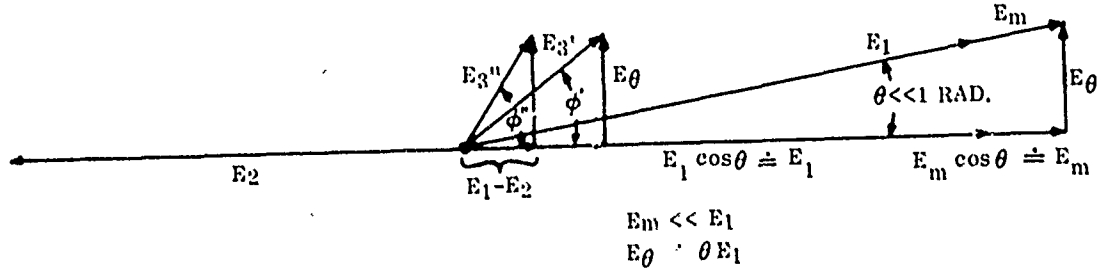
Also, two effects of interest were found to have happened in the carrier suppression process:

1. Input signal phase variance is multiplied by the carrier suppression ratio.
2. Phase of the output, e_3, is varied by amplitude variations of the input signals at the amplitude variation rate.

These two effects are of significance in the phase-locked reference system described in the next section.

Phase-Locked Reference Source

As shown in figure 6, a voltage controlled oscillator, with a phase detector and loop filter, has been added to generate the necessary reference signal. Also, an isolation amplifier has been added to block signals from going through the phase detector into the summation point output. These additions add to the complexity of the system in that now the characteristics of a locked loop must also be considered. Note that the loop is locked through the summing point and not prior to it, as might be expected. This configuration allows the locked loop to maintain proper phase balance. With the signal input to the phase detector being the summing point output, e_3, the characteristics of this signal must be considered in a loop analysis. This is done in the following paragraph.



$E_m << E_1$

$E_\theta \doteq \theta E_1$

Figure 5. Vector Diagram, Amplitude Variations.

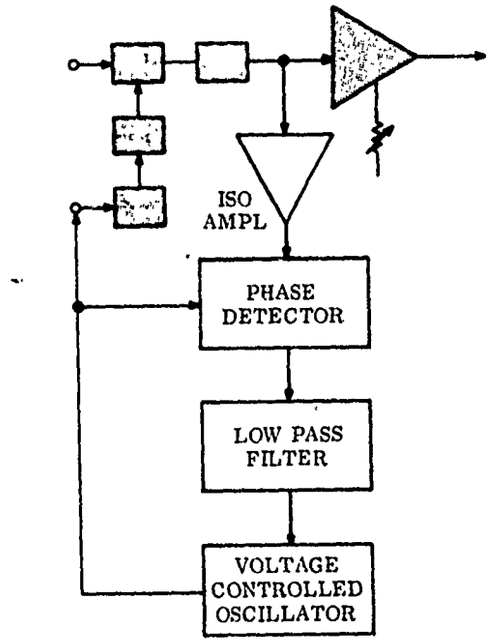


Figure 6. Carrier Suppression System With Locked Reference Source.

Locked Reference Signal Analysis

From the previous section on summing point analysis, it was shown that the summing point output, e_3, has the following characteristics:

1. All original input signal phase and amplitude components are present.
2. The remainder of the carrier component is present.
3. The phase of the output is effected by the carrier suppression ratio times the input signal phase variations.
4. The phase of the output is also a function of the amplitude variations of the input signals.

With e_3 being the signal input to the phase detector, the characterisites as listed will have effects on its output. These are discussed in the following paragraph.

Phase Detector Analysis

Depicting the basic phase detector as shown in figure 7:

$e_3 < < e_4$

$e_4 = E_4 \sin(\omega t + \alpha)$ = reference input from the phase shifter

where

α phase shift contributed by the phase shifter

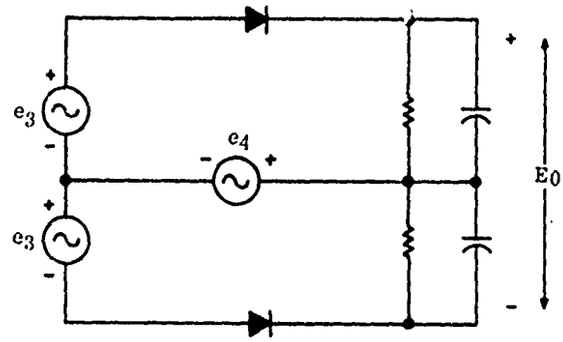


Figure 7. Basic Phase Detector.

The output can be shown to be as follows (see appendix B):

$$E_o = -2\left[(E_1 \cos\theta - E_2)^2 + (E_1 \sin\theta)^2\right]^{1/2} \sin(\phi - \alpha - 90°) \quad (6)$$

Also, the following can be shown:

$$\frac{\partial E_o}{\partial \theta} = 2E_1 \sin\alpha \quad (7)$$

$$\frac{\partial E_o}{\partial E_2} = 2\cos(2\phi - \alpha) \quad (8)$$

Since the phase shifter is initially adjusted for approximately 90 degrees (α), and the locked-loop holds this condition, the phase sensitivity and amplitude sensitivity can be rewritten as follows:

$$\frac{\partial E_o}{\partial \theta} = 2E_1 \quad (9)$$

$$\frac{\partial E_o}{\partial E_2} = -2\sin 2\phi \quad (10)$$

Therefore, the phase detector sensitivity to phase variations is $2E_1$ volts/rad and is not a function of the carrier suppression ratio. Defining a phase component to carrier component ratio N, the peak phase detector output is $4E_1N$.

The amplitude sensitivity of the phase detector varies from -2 to +2 volts/volt with the peak output varying from 0 to $4E_1N$ (N being the ratio of the amplitude component to carrier component).

Therefore, for locked-loop considerations, the phase detector sensitivity can be assumed as $2E_1$ volts/radian, disregarding the carrier suppression and phase variations of the summing point output caused by the amplitude sidebands. This is valid because the phase detector output is not a function of the suppression ratio and the effect of the amplitude variations can be treated using the same sensitivity, $2E_1$ volts/radian.

This means that the loop can be analyzed in a conventional manner as described in the following paragraph.

Loop Analysis

Figure 8 illustrates a loop analysis block diagram.

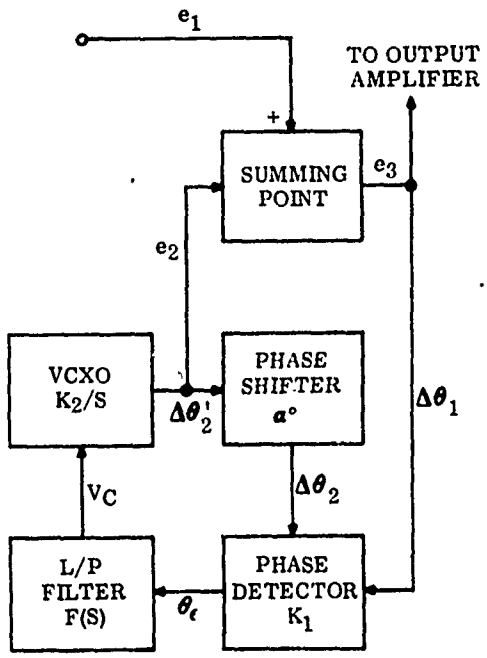


Figure 8. Loop Response, Block Diagram.

As shown in appendix C, the ratio of $\Delta\theta_2$ to $\Delta\theta_1$ is as follows:

$$\frac{\Delta\theta_2}{\Delta\theta_1} = \frac{K_1 K_2 F(s)}{S + K_1 K_2 F(s)} \quad \text{(closed loop gain)} \qquad (11)$$

$$\frac{\Delta\theta_2}{\Delta\theta_1} = \frac{K_1 K_2 F(s)}{S} \quad \text{(open loop gain)} \qquad (12)$$

$\Delta\theta_2/\Delta\theta_1$ must be made small enough so that components generated by the loop modulation of the VCXO do not effect those of interest on e_1 in the summing point, ($\Delta\theta_2/\Delta\theta_1 \leq -20$ dB).

Resume of Locked Reference Source System Characteristics

The locked reference source system characteristics are such that outside of the loop bandwidth, the noise and spurious components will be unaffected. Inside the loop bandwidth a variety of things can happen; sideband components can be suppressed, reinforced, or unsymmetrically effected around the carrier. The effects in this area are interesting, but of no consequence in this application.

A disadvantage which does occur with this configuration is that the phase detector output range is also reduced with the reduction in signal as the carrier is suppressed. This can be compensated for by moving the reference oscillator's voltage-variable capacitors operating point with another slow response loop which senses the phase detector's average output.

Fully Developed Carrier Suppression System

The block diagram of figure 9 shows the carrier suppression system as it is currently being used.

Two circuits have been left off this diagram because of added pictorial confusion. They are, a lock indicator/frequency discriminator circuit which aids in tuning the reference source; and, circuits for indicating when the inputs to either of the two age amplifiers are too low in amplitude for proper age action. Also, several of the blocks in this system are band switched together. Again, this was not shown because of added confusion.

Broadband age amplifiers have been added for convenience in gross amplitude adjustment, with the reference amplifier having a manual gain control for critical amplitude balancing. These amplifiers have been designed to accept 0.1 to 3.0 Vrms and hold a constant output into 50 ohms throughout the frequency range of the instrument.

A hybrid transformer is used as the summing device. It both isolates the two age amplifiers from each other and exhibits a constant impedance to them during carrier suppression.

The hybrid transformer must be terminated by a proper impedance, which is exhibited by the bandpass filters, having been designed to exhibit a nominal 50 ohms throughout their passbands. Twenty-one filters comprise a complete set which covers the frequency range of the instrument in overlapping bands. The center frequency of each filter being 35 percent higher than the previous center frequency, and the bandpass being plus and minus 35 percent. This creates a series of filters which overlap 50 percent, allowing complete carrier frequency coverage through the instrument's frequency range of 100 kHz to 40 MHz.

As will be noted, the phase shifter has been moved out of the reference signal path and is driving the phase detector directly. This permits proper phase shifting without introducing noise on the reference signal. The phase shifter is of the transistor amplifier RC type where the R and C are connected in series from the collector to emitter, the C being band-switched and the R being adjustable for the phase shifting.

The variable gain output amplifier is now a fixed 40 dB broadband amplifier, with a 10-dB/step attenuator preceding it. This allows reamplification of the suppressed signal in a discrete 10-dB incremental manner. The input impedance to the amplifier is 50 ohms to match the attenuator, and the output is 50 ohms to drive the attached instrumentation.

The phase detector is a simple balanced type, as was previously analyzed, using two untuned wideband transformers for the two input signals. A long-time constant operational amplifier integrator has been added across its output to move the operating point of the voltage variable capacitors in the reference oscillator (see figure 10). This is necessary because of the previously mentioned loss in phase detector output range with decreasing input signal. Manual offset on one of the operational amplifier inputs is included, which is used as a fine-balance adjustment in the carrier balancing procedure.

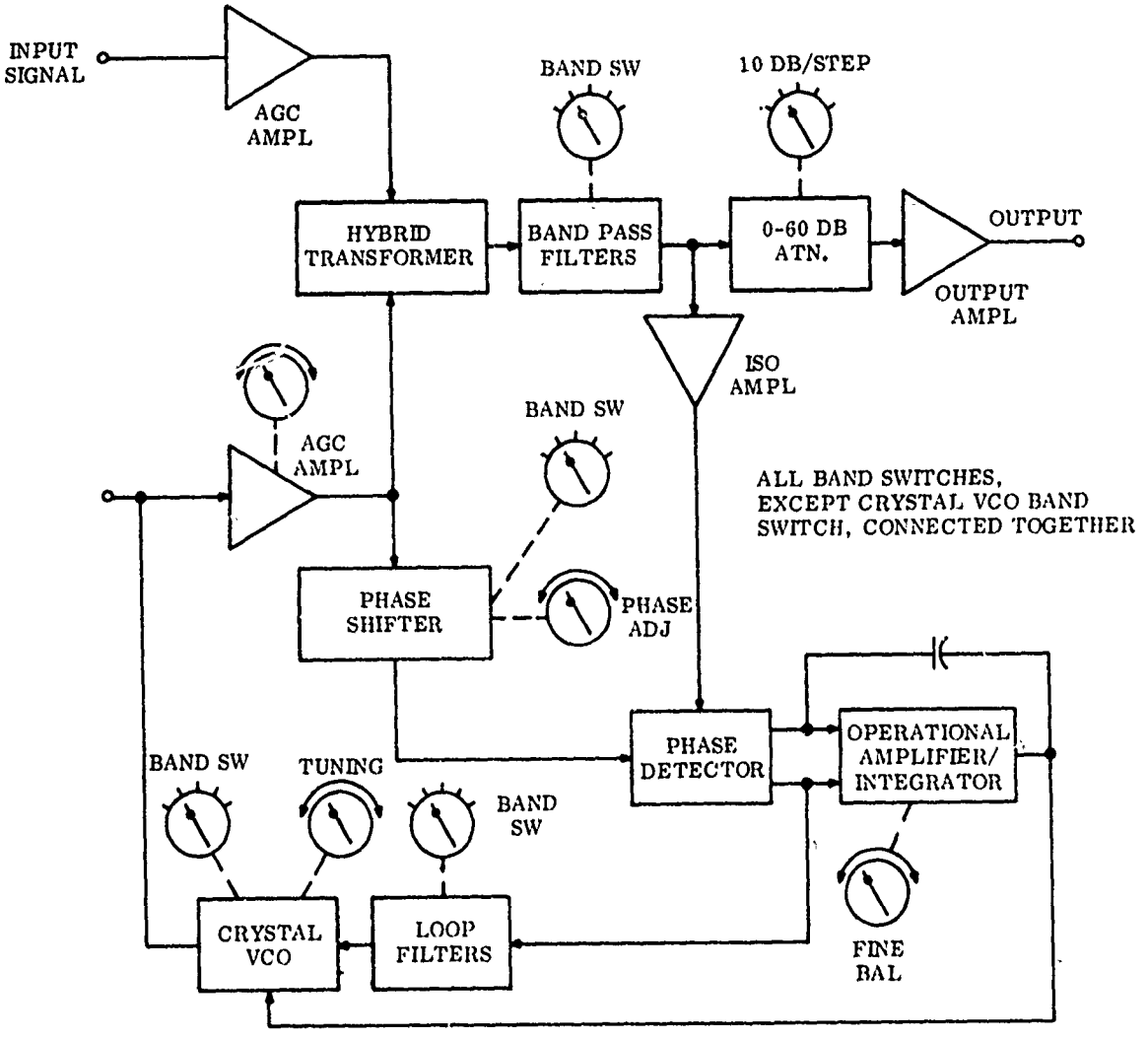


Figure 9. Fully Developed Carrier Suppression System.

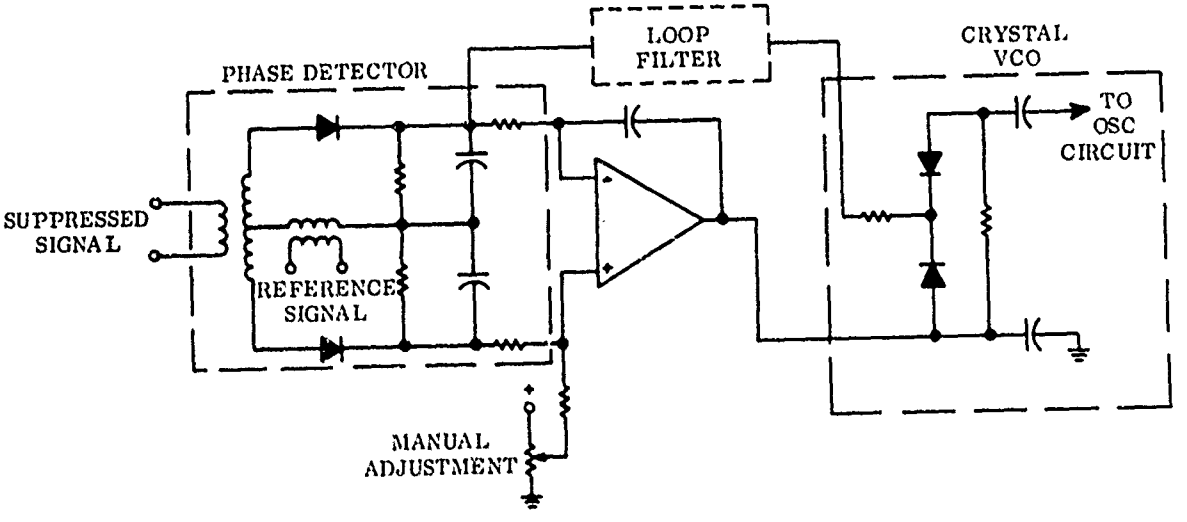


Figure 10. Reference Oscillator, Operating Point, Loop.

The reference oscillator is a voltage-controlled, seven-band, crystal oscillator which is manually tuned with variable inductors for pulling the crystal into the lock range. Different oscillator circuits are employed on the various bands for operation over the required frequency range of the instrument. A crystal oscillator was used because of its lower noise characteristics.

System Noise Measurements

Needless to say, careful design was required to keep the noise level low throughout the system. Low noise devices were used extensively as well as local feedback wherever possible. Figure 11 shows the signal/noise as it was measured in an instrument built using the system just discussed. The measurement method used was the one involving a sensitive communications receiver with a crystal filter keeping the carrier from blocking the receiver.

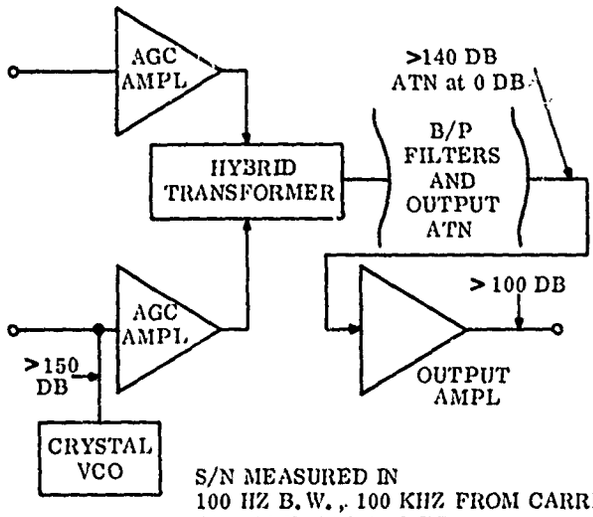


Figure 11. System Signal/Noise Measurements.

Suppressed Carrier Analyzer

Figure 12 is a picture of the instrument which was built utilizing the fully developed carrier suppression scheme. (The name "Suppressed Carrier Analyzer" is somewhat of a misnomer, but it persists. A more fitting name might be "Carrier Suppressor.")

Description

The connections and controls are self-explanatory: 50-ohm INPUT/OUTPUT connections; FUNCTION SWITCH for selecting the operational mode; FILTER SELECTOR for choosing the proper bandpass filter; OUTPUT ATTENUATOR for adjusting the amount of reamplification; PHASE BALANCE for initial phase adjustments; FINE BALANCE, which actually effects the vco operating point; and the AMPLITUDE BALANCE control, a concentric control for a finer degree of adjustment.

The area on the right of the panel contains the LOCK INDICATOR/FREQUENCY discriminator meter, and FAULT LIGHTS for indicating when either of the input amplitudes to the two summing point input age amplifiers are too low.

Directly below the meter and lamps are the BAND SWITCH and TUNING for the reference oscillator. Below these knobs is the CRYSTAL SOCKET. And at the bottom of the panel is a drawer for crystal storage.

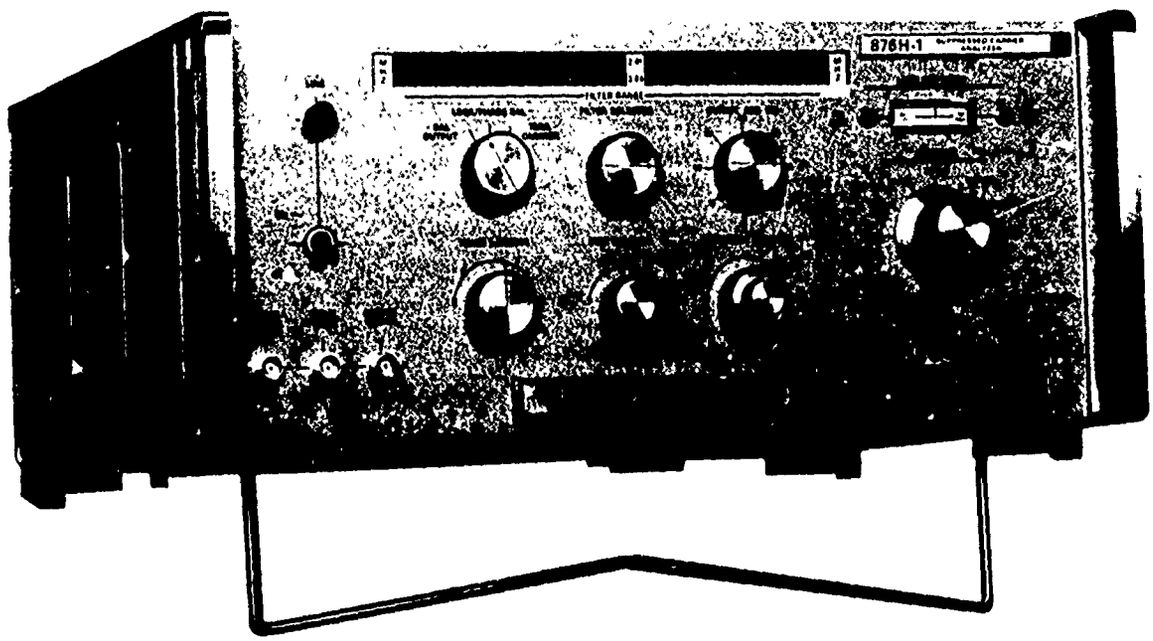


Figure 12. Suppressed Carrier Analyzer, Front View.

On the rear panel, as shown in figure 13, are connections for using external loop filters of the operator's choice. Also, there are proper connections for using lockable frequency sources other than the internal crystal reference oscillator.

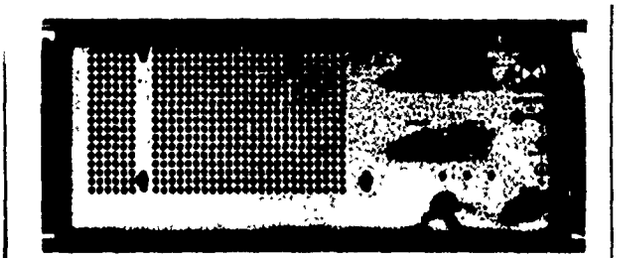


Figure 13. Suppressed Carrier Analyzer, Rear View.

Specifications

The specifications for this instrument are as follows:

Input frequency:	100 kHz to 40 MHz.
Suppression range:	≥60 dB.
System bandwidth (Bandpass filter bandwidth):	±35 percent of input frequency.
Suppression bandwidth (Loop bandwidth):	±50 Hz (can be changed).
*Input signal level:	0.1 Vrms to 3.0 Vrms.
*Input impedance:	50 ohms nominal.
Output impedance:	50 ohms nominal.
Output level:	0.25 V typical (function of suppression and output amplification).
*System signal/noise (without suppression or reamplification):	≥140 dB (100 Hz bw; ±100 kHz from input frequency; input level ≥ 0.25 Vrms).
System power:	117 Vac, single phase.

Test Data

To demonstrate the capabilities of the carrier suppression technique, a test signal from a nondescript crystal oscillator was modulated with a linearly-mixed sum of 1 kHz and 2.5 kHz at levels of approximately -60 dB and -100 dB, respectively. This signal was connected to the instrument just described and the output displayed on a spectrum analyzer and plotted as shown in the upper portion of figure 14. (Figure 14 is a fabricated composite of the original and suppressed/reamplified test signal.) The -60 dB, 1 kHz sidebands are evident, while the -100 dB, 2.5 kHz sidebands are below the range of the analyzer. The signal was then suppressed and reamplified 60 dB and plotted as shown in the lower portion of figure 14. The 1 kHz sidebands are now major components of the data with the 2.5 kHz sidebands being very evident. Also, the basic noise level of the source can be clearly seen at about -120 dB.

Further data taken on some familiar frequency sources follows in figures 15 through 18.

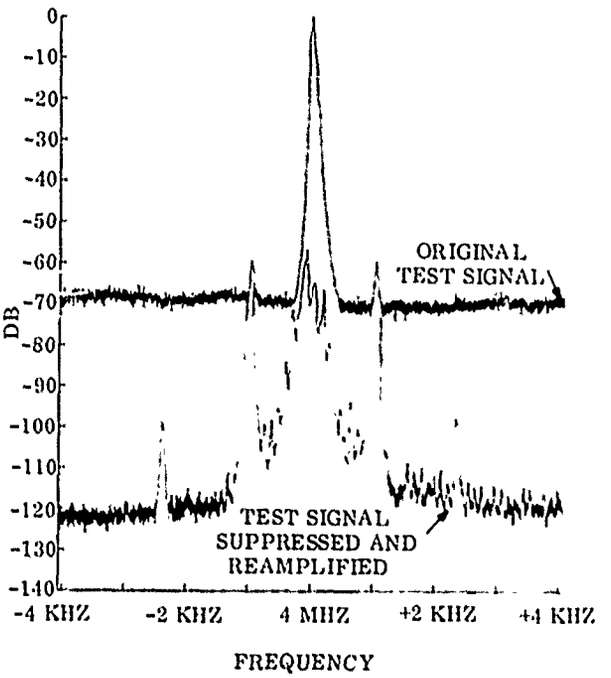


Figure 14. Demonstration Test Data.

*With proposed high impedance probe, input signal level: 0.25 to 2.5 Vrms; input impedance: ≥100 kilohms, system signal noise: ≥130 dB (same requirement as above).

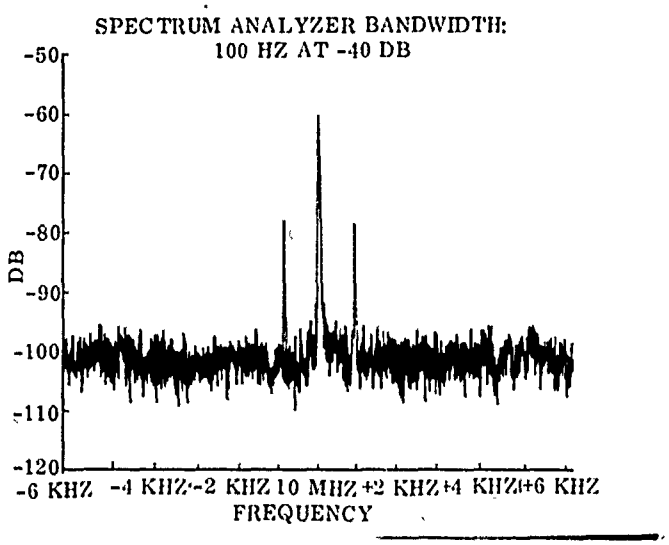


Figure 15. 20 MHz TCXO Buffer Drive Test.

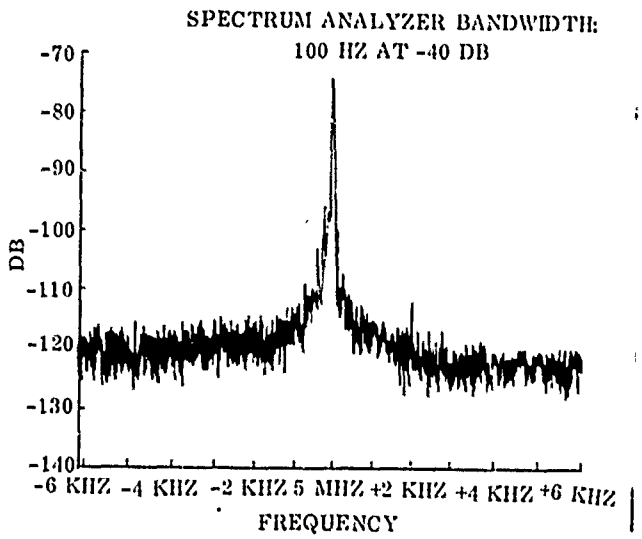


Figure 16. HP-5061A 5 MHz Output.

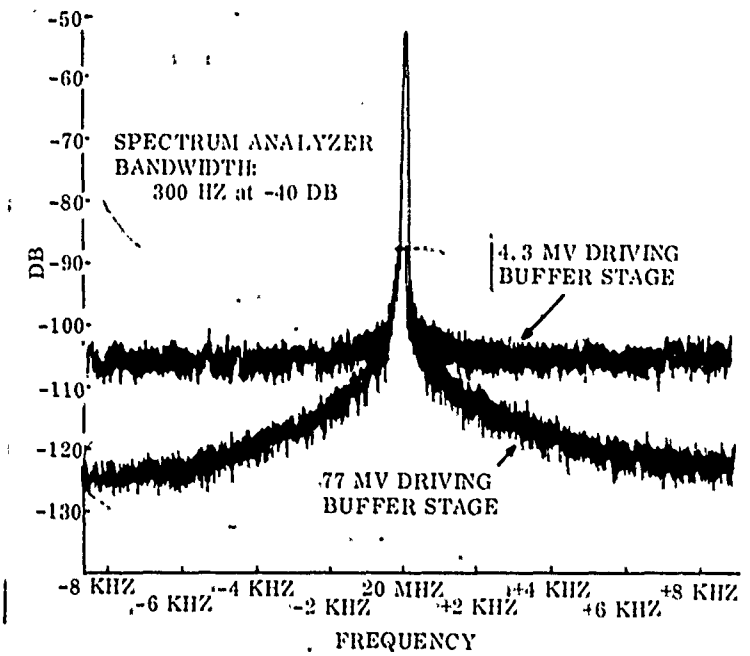


Figure 17. HP-5100B at 10 MHz + 100 Hz.

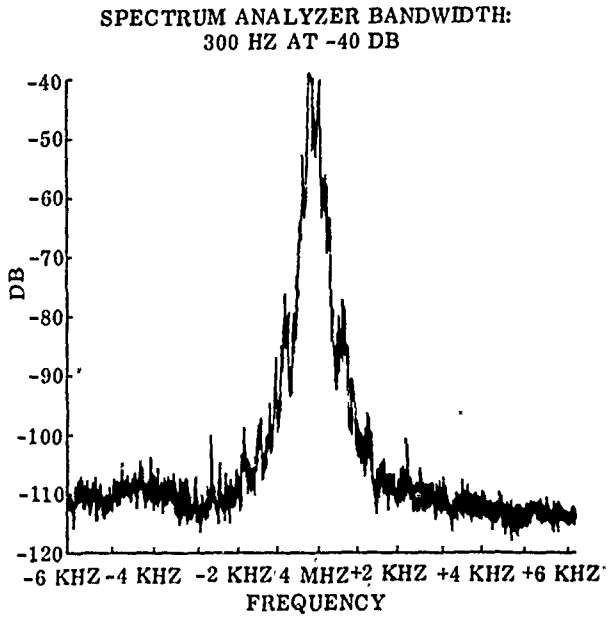


Figure 18. HP-606A at 4 MHz.

Acknowledgment

The author wishes to give credit to Mr. Marvin E. Frerking of Collins Radio Company, for the original conception of the carrier suppression technique, and to gratefully acknowledge the help and suggestions received from him throughout the subsequent development.

Also, I wish to acknowledge the enthusiastic effort extended by Mr. James Doppelhammer to the construction of the Suppressed Carrier Analyzer.

APPENDIX A

Summing Point Analysis

Assuming: Small phase variations on input signal-reference signal assumed constant.

Defining a summing point with + and - inputs (180 degrees apart) as shown in figure A-1:

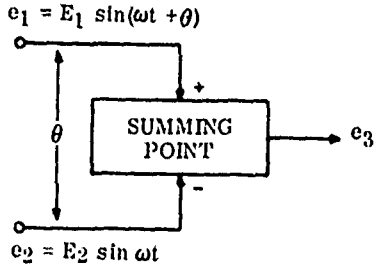


Figure A-1. Summing Point.

The input signal e_1 will be assumed to have small phase variations:

$\theta << 1$ radian

Also, carrier suppression of $\geqq 60$ dB is assumed established.

After summation:

$$e_3 = e_1 - e_2 = E_1 \sin(\omega t + \theta) - E_2 \sin \omega t \quad \text{(A-1)}$$

Using the following identity:

$$\sin(A + B) = \sin A \cos B + \cos A \sin B$$

Then:

$$e_3 = E_1 \sin \omega t \cos \theta + E_1 \cos \omega t \sin \theta - E_2 \sin \omega t \quad \text{(A-2)}$$

$$= (E_1 \cos \theta - E_2) \sin \omega t + (E_1 \sin \theta) \cos \omega t \quad \text{(A-3)}$$

Using the identity:

$$A \cos x + B \sin x = (A^2 + B^2)^{1/2} \sin(\omega t + \phi)$$

Where:

$$\phi = \tan^{-1} A/B$$

Equating:

$$A = E_1 \sin \theta$$

$$B = E_1 \cos \theta - E_2$$

Then:

$$e_3 \quad \left[(E_1 \sin \theta)^2 + (E_1 \cos \theta - E_2)^2\right]^{1/2} \sin(\omega t + \phi) \quad \text{(A-4)}$$

When:

$$\phi = \tan^{-1} \frac{E_1 \sin \theta}{E_1 \cos \theta - E_2} \quad \text{(A-5)}$$

θ was assumed to be very small, allowing the following approximations:

$$\cos \theta \doteq 1 \qquad \sin \theta \doteq \theta$$

Rewriting equations 4 and 5 with the approximation:

$$e_3 \doteq \left[(E_1 \theta)^2 + (E_1 - E_2)^2\right]^{1/2} \sin(\omega t + \phi) \quad \text{(A-6)}$$

Where:

$$\phi \doteq \tan^{-1} \frac{E_1 \theta}{E_1 - E_2} \quad \text{(A-7)}$$

Defining a carrier suppression ratio R:

$$R = \frac{E_1 - E_2}{E_1} \quad \text{(A-8)}$$

Then:

$$\phi \doteq \tan^{-1} \quad \theta/R \tag{A-9}$$

Substituting in equation (A-6):

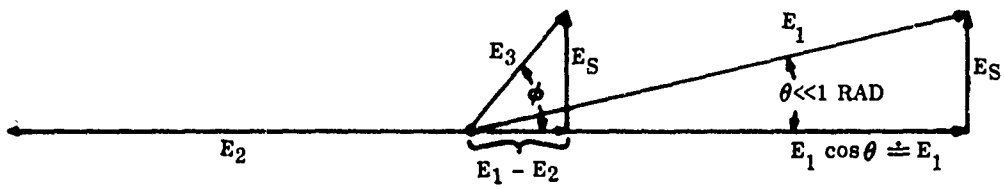
$$e_3 \doteq E_1 \left[\theta^2 + R^2\right]^{1/2} \sin(\omega t + \phi) \tag{A-10}$$

Drawing e_3 in vector representation:

Also it can be seen:

$$E_3 = \left[E_s^{\,2} + (E_1 - E_2)^2\right]^{1/2} \tag{A-11}$$

$$E_3 = E_1 \left[\left(\frac{E_s}{E_1}\right)^2 + R^2\right]^{1/2} \tag{A-12}$$



E_S IS A PHASE MODULATION VECTOR WHICH VARIES IN LENGTH.

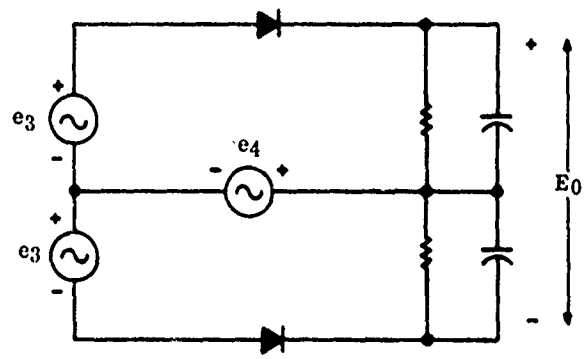
$E_S \doteq \theta E_1$

Figure A-2. Vector Diagram, Phase Variations.

APPENDIX B

Phase Detector Response

Assuming a basic phase detector as shown in figure B-1:



(THIS IS A SIMPLIFIED FORM OF THE PHASE DETECTOR AND DOES NOT INCLUDE THE VOLTAGE RATIOS OF THE INPUT TRANSFORMERS)

Figure B-1. Basic Phase Detector.

And assuming a reference signal e_4:

$$e_4 = E_4 \sin(\omega t + \alpha) \tag{B-1}$$

e_4 is the signal coming from the phase shifter and α is the phase contributed by the phase shifter.

Knowing that:

$$E_0 = \left|e_3 - e_4\right| - \left|e_3 + e_4\right| \tag{B-2}$$

From appendix A, equation (A-4):

$$e_3 = \left[(E_1 \cos\theta - E_2)^2 + (E_1 \sin\theta)^2\right]^{1/2} \sin(\omega t + \phi) \tag{B-3}$$

Defining

$$K = \left[(E_1 \cos\theta - E_2)^2 + (E_1 \sin\theta)^2\right]^{1/2} \tag{B-4}$$

Then the phase detector output (equation B-2) can be rewritten:

$$E_0 = (K^2 + E_4^{\,2})^{1/2} \left[\left(1 - \frac{2 E_4 \cos(\phi - \alpha)}{K^2 + E_4^{\,2}}\right)^{1/2} - \left(1 + \frac{2 E_4 \cos(\phi - \alpha)}{K + E_4^{\,2}}\right)^{1/2}\right] \tag{B-5}$$

But knowing that at carrier null, $E_4 > > K$, E_0 can be simplified:

$$E_0 \doteq -2 K \cos(\phi - \alpha) \tag{B-6}$$

Defining an angle β as $\beta = \alpha + 90°$:

Then

$$E_0 = -2 K \cos(\phi - \beta - 90°) \tag{B-7}$$

But:

$$\cos(x - 90°) = \sin x$$

Therefore:

$$E_0 = -2 K \sin(\phi - \beta) \tag{B-8}$$

Where:

$$K = \left[(E_1 \cos \theta - E_2)^2 + (E_1 \sin \theta)^2\right]^{1/2} \quad \text{(B-9)}$$

And (from appendix A, equation (A-5)):

$$\phi = \tan^{-1} \frac{E_1 \sin \theta}{E_1 \cos \theta - E_2} \quad \text{(B-10)}$$

The expressions for phase detector sensitivity to changes in θ and E_2 can be written as follows:

$$\frac{\partial E_0}{\partial \theta} = -2K \left[\cos (\phi - \beta)\right] \frac{\partial \phi}{\partial \theta} - 2 \left[\sin (\phi - \beta)\right] \frac{\partial K}{\partial \theta} \quad \text{(B-11)}$$

$$\frac{\partial E_0}{\partial E_2} = -2K \left[\cos (\phi - \beta)\right] \frac{\partial \phi}{\partial E_2} - 2 \left[\sin (\phi - \beta)\right] \frac{\partial K}{\partial E_2} \quad \text{(B-12)}$$

Which when simplified and rearranged results in the following:

$$\frac{\partial E_0}{\partial \theta} = 2E_1 \sin \alpha \quad \text{(B-13)}$$

$$\frac{\partial E_0}{\partial E_2} = 2 \cos (2\phi - \alpha) \quad \text{(B-14)}$$

Remembering that α is the phase contribution by the phase shifter and is adjusted initially to approximately 90 degrees and left in that position, equations (B-13) and (B-14) can be rewritten:

$$\frac{\partial E_0}{\partial \theta} = 2E_1 \quad \text{(B-15)}$$

$$\frac{\partial E_0}{\partial E_2} = -2 \sin 2\phi \quad \text{(B-16)}$$

Therefore, the sensitivity of the phase detector to phase variations of the input signal is a constant $2E_1$ volts/radian. The sensitivity to amplitude variations ranges from -2 to +2 volts/volt.

And, for a sideband-to-carrier ratio N_1 the peak phase detector output for phase variations is $4E_1N_1$ and for amplitude variations 0 to $4E_1N_1$.

APPENDIX C

Locked-Loop Analysis

Assuming steady state conditions and arranging the block diagram as shown:

The closed loop response can be found as follows:

$$\theta_\epsilon = (\Delta\theta_1 - \Delta\theta_2) K_1 \quad \text{(C-1)}$$

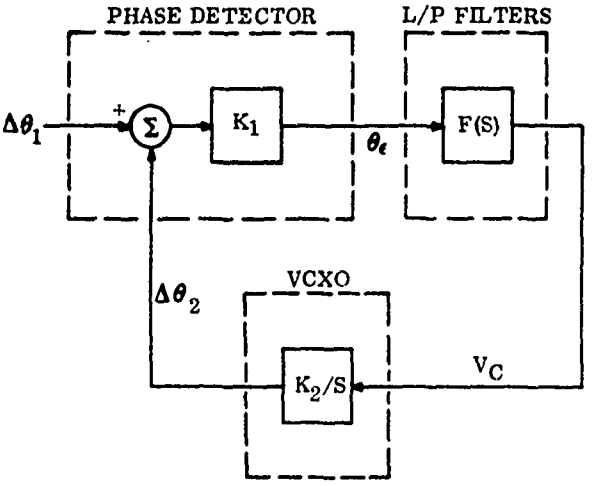


Figure C-1. Basic Locked Loop, Block Diagram.

$$V_c = \theta_\epsilon F(s) = F(s) (\Delta\theta_1 - \Delta\theta_2) K_1 \quad \text{(C-2)}$$

$$\Delta\theta_2 = \frac{V_c K_2}{S} = \frac{K_1 K_2 F(s) (\Delta\theta_1 - \Delta\theta_2)}{S} \quad \text{(C-3)}$$

$$\frac{\Delta\theta 2}{\Delta\theta 1} = \frac{K_1 K_2 F(s)}{S + K_1 K_2 F(s)} = \text{closed loop response} \quad \text{(C-4)}$$

The open loop response is found in a similar fashion:

$$\theta_\epsilon = \Delta\theta_1 K_1 \quad \text{(C-5)}$$

$$V_c = \Delta\theta_1 K_1 F(s) \quad \text{(C-6)}$$

$$\Delta\theta_2 = \frac{\Delta\theta_1 K_1 K_2 F(s)}{S} \quad \text{(C-7)}$$

$$\frac{\Delta\theta_2}{\Delta\theta_1} = \frac{K_1 K_2 F(s)}{S} = \text{open loop response} \quad \text{(C-8)}$$

Using a lag network for the low-pass filter, the response becomes the following:

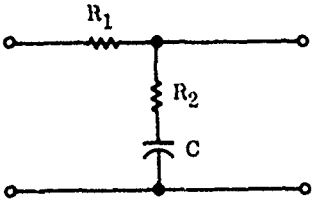


Figure C-2. Lag Network.

$$F(s) = \frac{1 + T_2 S}{1 + T_1 S} \quad \text{(C-9)}$$

Where:

$$T_1 = (R_1 + R_2)\, c \qquad \text{(C-10)}$$

$$T_2 = R_2\, c \qquad \text{(C-11)}$$

Then:

$$\frac{\Delta\theta_2}{\Delta\theta_1} = \frac{K_1K_2\,(1 + T_2\,S)}{S(1 + T_1\,S) + K_1K_2\,(1 + T_2\,S)} \text{(closed loop)} \qquad \text{(C-12)}$$

Also:

$$\frac{\Delta\theta_2}{\Delta\theta_1} = \frac{K_1K_2\,(1 + T_2\,S)}{(1 + T_1\,S)\,S} \text{ (open loop)} \qquad \text{(C-13)}$$

APPENDIX D

Variations in ϕ With Changes in E_1 and θ

Using the expression for ϕ from appendix A (equation (A-7)):

$$\phi \doteq \tan^{-1} \frac{E_1\,\theta}{E_1 - E_2} \qquad \text{(D-1)}$$

Finding $\frac{\partial\phi}{\partial E_1}$:

$$\frac{\partial\phi}{\partial E_1} = \frac{(E_1 - E_2)\;\theta - E_1}{(E_1 - E_2)^2 + (E_1\,\theta)^2} \qquad \text{(D-2)}$$

Using:

$$R = \frac{E_1 - E_2}{E_1}$$

$$\frac{\partial\phi}{2E_1} = \frac{-\theta}{E_1\,(R^2 + \theta^2)} \qquad \text{(D-3)}$$

Finding $\frac{\partial\phi}{\partial\theta}$

$$\frac{\partial\phi}{\partial\theta} = \frac{(E_1 - E_2)\,E_1}{(E_1 - E_2)^2 + (E_1\,\theta)^2} \qquad \text{(D-4)}$$

Again using R:

$$R = \frac{E_1 - E_2}{E_1}$$

$$\frac{\partial\phi}{\partial\theta} = \frac{R}{R^2 + \theta^2} \qquad \text{(D-5)}$$

USE OF THE LORAN-C SYSTEM FOR TIME AND FREQUENCY DISSEMINATION

Paul E. Pakos
Lieutenant Commander, United States Coast Guard
U. S. Coast Guard Headquarters
Washington, D. C.

Summary

The Loran-C system is an excellent medium for time and frequency dissemination. If one is willing to accept certain operating restrictions on the use of the system, sub-microsecond timing accuracies are possible. A frequency reference with a long-term stability equivalent to that attainable with a cesium beam oscillator is available to those located within the groundwave range of a timed Loran-C chain. User costs are not excessive and depend upon the degree of accuracy required to accomplish the intended task.

Introduction

In the early years of this decade, the timing potential of the pulsed low frequency navigation system known as Loran-C was recognized. It was found that the use of a stable oscillator and a relatively minor modification to the existing equipment would allow transmission of a timing pulse once per second in addition to the Loran-C pulses normally transmitted. Actual experiments were conducted in 1961 by the U. S. Naval Observatory and U. S. Coast Guard utilizing the Loran-C chain on the East Coast of the United States. The success of those tests and the growing requirements for precision timing have resulted in a gradual expansion of Loran-C timing services, while the growth of technology has resulted in a parallel path of continuous refinement to the accuracies achievable.

The intent of this paper is to provide a complete description of the Loran-C timing system and to provide an insight into the sources of error which exist in the system. Basic instrumentation methods will be discussed, but full details will be avoided because others have covered that subject adequately.[1] Likewise, a thorough description of Loran-C as a navigational system is considered beyond the scope of this paper, except for those aspects of the navigational system which directly affect Loran-C timing.

Methods of Timing Loran-C Chains

The problem of transmitting time from an entire Loran-C chain reduces to one of timing the master station only, since all slave stations are synchronized to the master in order to satisfy the navigational requirements of the system. The synchronization status of slave stations is monitored continuously by external monitor stations which are located in the prime navigational service area of the chain.

The timing of a master station is accomplished in two ways. First, an additional pulse once per second is transmitted, and second, the Loran-C navigational transmissions are timed. The latter consists of eight pulses per group being transmitted at what is commonly termed the group repetition rate (GRR).[2] The U. S. Naval Observatory (USNO) has established as a datum the arbitrary fact that at 00h 00M 00S Universal Time Coordinated (UTC) 1 January 1958, the start of the first pulse of all Loran-C master groups was coincident with the Universal Time Second (UTS). Thereafter, for each GRR, periodic coincidences with UTS have occurred and will continue to occur. The time interval between coincidences depends upon the GRR and can vary from a minimum coincidence interval of once each second for Loran-C rates SO and SSO to a maximum coincidence interval every 16 minutes 39 seconds for Loran-C rate SS1. The Naval Observatory computes the predicted coincidences with UTS for each year and tabulates them in a Null Ephemeris Table for each Loran-C rate.[3] Using the datum established by USNO, it is therefore possible to synchronize any Loran-C chain to the appropriate Ephemeris Table. The distinct advantage of timing the Loran-C groups is that they offer a greatly increased information rate over that available with the one pulse/second system. This will become clearer when instrumentation methods are discussed.

The equipment necessary to accomplish the timing functions at a master station consists essentially of one or more cesium beam oscillators, three Group Repetition Rate generators, three one pulse/second clocks and majority logic circuitry with automatic reset capability. This equipment enables the master station to remain synchronized to UTC once it has been initially set.

Slave stations do not presently transmit one pulse/second but their normal function of synchronization with the master allows the use of the slave Loran groups for timing. It should be noted that the Ephemeris Tables refer only to master coincidences with UTS; the emission delay of each slave with respect to its master is listed in Table 1.

When a chain is timed, the Naval Observatory normally establishes a time monitor site somewhere within groundwave range of the master station. The type of equipment located at the time monitor site will be described in the in-

strumentation section of this paper, but at this point it is sufficient to say that at least two cesium clocks are present. By predicting the propagation time from the master station to the time monitor, the monitor is able to direct the positioning of the master's transmissions until precise agreement with the Ephemeris Table is achieved. Thereafter, the ability of the time monitor to keep the master on time depends primarily on the offset of the monitor's cesium oscillators with respect to UTC. The Naval Observatory makes regular portable clock visits to the time monitor sites to determine these offsets. With this information and with daily message reports from the monitor on the measured arrival time of the master's Loran and one pulse/second transmissions, the Observatory determines where the transmissions are with respect to the DOD master clock in Washington, D. C. Special users receive daily status reports from USNO while others receive weekly status reports on all the timed Loran-C chains.[4] An example of a weekly report is shown in Figure 1. In February, 1969, it was decided that the long-term stability of the cesium oscillators at the master stations made it perfectly feasible to hold the Loran-C transmissions of timed chains to within ± 5 microseconds of UTC, and this is presently being accomplished. The precise value can be obtained by using the USNO status reports.

It should be noted that USNO's determination of a master's time of transmission with respect to UTC is subject to a constant error which is caused by, among other things, an incorrect prediction of propagation time from master to time monitor. Such system errors could be entirely eliminated only by making measurements of the absolute time of radiation of the Loran pulses from the transmitting antenna, a measurement which is complicated by near field effects. Such measurements are planned, but these system errors are probably less than 5 microseconds and their importance for most categories of timing users is minimized because they are constant. This aspect of Loran-C timing will be discussed further in the section on accuracies.

Coverage

The Loran-C chains which are presently timed are the U. S. East Coast, Northwest Pacific, Hawaiian, and Norwegian Sea chains. Figure 2 shows the area in which a groundwave and skywave from at least one of the stations should be available. Figure 3 shows the groundwave and skywave coverage areas which would exist if all the Loran-C chains presently in existence were timed. The boundaries of groundwave and skywave coverage areas shown are approximate. Exact boundaries are a function of propagation conditions, local noise and interference, and the quality of the receiver employed.

Timing Accuracies Available with Loran-C Groundwaves

It is important to approach a discussion of accuracy cautiously, because the Loran-C timing accuracy available is dependent upon whether one refers to:

1. the ability to synchronize to UTC or to another Loran-C user.

2. "real-time" or "after-the-fact" time synchronization.

3. synchronization accuracies available if two users utilize:

   a. the same Loran-C station
   b. two stations within the same Loran-C Chain
   c. one station from one chain and one from another chain.

These seem to represent a formidable number of conditions, but the significance of some of them can easily be explained by a brief description of Loran-C as a precise navigational system.

When a Loran-C chain is initially established, a calibration is conducted in the navigational service area by monitoring the observed time differences in as many different surveyed locations as is feasible. These measurements are then compared to predicted time difference readings which were based upon assumptions of the conductivities along the various propagation paths. The conductivity assumptions are then continuously refined to minimize the standard deviation of the difference between the observed and the predicted time differences at all the monitor points. When this spatial standard deviation is reduced to 0.1 microsecond or less, the chain is considered calibrated and the resultant conductivities are used for all ensuing time difference predictions.

Using the calibration data, a time difference assignment is made to a navigational service area monitor whose function thereafter is to ensure that slave to master synchronization is maintained to a particular tolerance, typically ± 0.2 microseconds. If this tolerance is exceeded, the system is considered unusable and "blink" (the periodic blanking of certain Loran pulses) is initiated. The probability distribution of slave synchronization error about zero at the navigational monitor is not a Gaussian distribution because the monitor normally will issue a correction to a slave if a synchronization error exceeding ± 0.1 microsecond exists for more than 15 minutes, and will normally not issue any corrections if slave synchronization is within ± 0.1 microsecond. For those wishing to assign a standard deviation to slave synchronization, the standard deviation of the 15-minute averages over a month long period is appropriate, in which case 0.05 microsecond is a typical figure. It is important that such a figure be used correctly. For any given short period of time, ten minutes for example, the slave transmissions are not "noisy" in the sense that an RMS jitter of 0.05 microsecond exists; rather, the Loran

equipment at the transmitting stations is such that any RMS jitter in the transmitted signal is perhaps ten nanoseconds or better, when measured over an integration time of several seconds. This should be considered by users wishing to precisely synchronize clocks using a common slave station. If both were to perform the synchronization during the same 10 minute period, for example, the quality of the synchronization of the two clocks would probably be equivalent to that attained if both used a common master station, which never receives corrections from the navigational monitor. If, on the other hand, the two users performed their individual synchronizations at different times during the day, they would have to consider the 0.05 microsecond figure in their error budget.

It is appropriate that we now look in detail at the error budgets which exist for certain categories of users. To gain some appreciation of the magnitude of the errors involved, one sigma error estimates of the different error sources are assigned as follows:

UTC tolerance: σ_{ut} = 2.0 usec

System error: σ_{se} = 3.0 usec

User prediction error: σ_{pe} = 0.1 usec

Groundwave propagation anomaly (over land): σ_{pa} = 0.2 usec

Slave synchronization error: σ_{ss} = 0.05 usec

User measurement error: σ_{me} = 0.1 usec

The assignment of error estimates inevitably provokes heated discussions; the above figures represent the Coast Guard's experience with the Loran-C system. The UTC tolerance and system error have been discussed previously in the section on timing methods. The user prediction error assigned reflects the results of the chain calibration methods discussed earlier. Such a figure is therefore realistic for those in the navigational coverage area if they use the conductivities which were determined by the calibration. It is not realistic for a user in an uncalibrated area, particularly if he is on land. In this latter case, perhaps a more realistic figure would be 0.4 microsecond, because of the prediction uncertainties inherent in non homogeneous propagation paths and because the user may be experiencing some local terrain effects.

The figure assigned to groundwave propagation anomaly is based upon our experience with the one-way path existing between the East Coast Chain master station at Carolina Beach, N. C., and the slave at Dana, Indiana. The anomalies in the Southeast Asia Chain are significantly lower, perhaps indicating a climatological dependence. The user measurement error presumes a high quality receiving system.

To gain an appreciation of what these error sources mean to a user, first consider one who wishes to synchronize a clock to UTC with the slave of a Loran-C chain. His RMS error would be predicted as follows:

$$\sigma = (\sigma_{ut}^2 + \sigma_{se}^2 + \sigma_{pe}^2 + \sigma_{pa}^2 + \sigma_{ss}^2 + \sigma_{me}^2)^{\frac{1}{2}} \quad (1)$$

Using the previous figures, σ = 3.6 microseconds.

Now suppose he was only interested in "after-the-fact" synchronization, i.e. the next day he removed the UTC error by applying the correction which USNO published in their status report for that particular chain. Clearly, σ_{ut}^2 could be omitted from the calculation. Further, presume that he was once visited by a portable cesium clock and his Loran-C timing system was thereby calibrated, i.e. the system error was removed, and his prediction error was removed. It is quite clear that he is now concerned only with the repeatability of the system, and his errors are now limited to propagation anomalies, measurement errors, and slave synchronization errors, if indeed he is using a slave and not a master. What is the best he can do if he uses a master, has been visited by a portable clock, and can wait a day for the answer?

$$\sigma = (\sigma_{pa}^2 + \sigma_{me}^2)^{\frac{1}{2}} \quad (2)$$

σ = 0.23 microsecond

Now consider someone who wants to synchronize with another, and who is not concerned with UTC synchronization. Suppose both were within groundwave range of the same master station, had not been visited by a portable clock, and wanted "real-time" synchronization. The UTC error and system error would be the same for each and would therefore not affect their ability to synchronize to each other. User #1 would have an RMS error as follows:

$$\sigma_1 = (\sigma_{1\,pe}^2 + \sigma_{1\,pa}^2 + \sigma_{1\,me}^2)^{\frac{1}{2}} \quad (3)$$

σ_1 = 0.25 microsecond

User #2 would have an identical error budget, and, presuming identical errors, the resultant clock synchronization accuracy would be given by:

$$\sigma = (\sigma_1^2 + \sigma_2^2)^{\frac{1}{2}} \quad (4)$$

σ = 0.35 microsecond

Since the UTC error does not come into play, all synchronizations between users of the same chain are "real-time" synchronizations. The use of a master by one user and a slave by the other merely introduces the slave synchronization term in the error budget of the latter. The use of a common slave may eliminate the need for the slave synchronization term if the users accomplish their individual synchronizations at approximately the same time, as discussed earlier. A real advantage can still be gained by a portable clock visit to each, thereby

eliminating the prediction error term which would probably be the largest factor if overland propagation paths were involved.

The justific tion for dropping the system error in these examples is apparent; it is a constant for any particular Loran-C chain. The justification for dropping the UTC error from the error budget is a bit more subtle and goes back to the methods used to time the chain. Because the chain frequency has a typical UTC offset of better than 5 X 10^{-12} (Approximately 0.5 us drift in 24 hours) and because the master station has equipment to keep it from "jumping" in time, the UTC error can assumed to be constant for any given day, and can, therefore, be treated like the system error. This is borne out by Figure 1. These reasons also provide some insight into why not all Loran-C chains can be use for synchronization between two users unless certain operating procedures are observed. Not all chains have the ultrastable cesium oscillators at present, and not all master stations have the equipment necessary to return their transmissions to the same point in time in the event failures occur. This method of operation is acceptable navigationally, because the slave stations simply resynchronize if the master moves in time, but one can easily see that it is disastrous to those interested in precise timing.

Despite the foregoing, if two users utilized an untimed Loran-C chain in such a manner that they both synchronized to the transmissions at approximately the same time, the oscillator drift would not cause significant errors and the probability of a master jump in the time interval between synchronizations would be minimal. Thus, in our parlance, the UTC error for that short interval of time would be a constant, the condition necessary to remove it from the error budget. Because the chain is untimed, the UTC error is also probably very large and is unknown; nevertheless, we have seen that these factors do not affect synchronization between two users of the same chain. The methods to accomplish the mutual synchronization will be discussed in the instrumentation section.

Finally, what about synchronization of users located in different Loran-C chains? The chains themselves are not synchronized to each other except through their individual relationships to UTC. Hence, synchronization accuracies in these cases are identical to those which would be experienced in synchronizing to UTC, except that the resultant synchronization accuracy between the two clocks would be the root sum square of the individual UTC synchronizations.

Use of Loran-C Skywaves

Loran-C skywaves can also be utilized for timing. They offer the advantage of tremendous range at the expense of degraded prediction accuracies and degraded stability. Both the prediction accuracy and the stability are functions of the range to the transmitting station. For the one-hop case, as the range increases, the prediction accuracy becomes less dependent on the estimate of the effective height of the ionosphere. Similarly, as the range increases, the effective length of the propagation path becomes less dependent upon the fluctuations of the ionospheric height. Although quantitative data of the fluctuations of the time arrival of a Loran-C skywave along a single path are not available, our experience with Loran-C navigational systems leads us to believe that a conservative one sigma estimate for the one-hop case would be about 4 microseconds when working beyond 1000 miles from a transmitting site.

The skywave coverage areas shown in Figure 2 and 3 were based upon approximately a 2000 mile radius from the transmitting stations, because at this range the peaks of the skywave are normally above the noise and can be readily seen on an oscilloscope. The user therefore need not utilize any special techniques to handle skywaves vis-a-vis groundwaves. However, a recent experiment by the Austron Company has shown that Loran-C skywaves can in fact be tracked at distances in excess of 4000 miles if special techniques are utilized.[5] Presuming such results are typical of what can be achieved in other areas, timing from Loran-C skywaves should be possible over virtually the entire world.

As in the case of Loran-C groundwaves, the prediction error associated with skywaves can be removed by the single visit of a portable clock. Thereafter, propagation anomalies become the determining factor. More data is needed to determine the statistics of single-hop and multi-hop arrival times, but one way to minimize such anomalies would be to make synchronization checks at the same time each day when propagation conditions are approximately the same.

Instrumentation Methods

Shapiro[1] has gone into full detail in the description of instrumentation methods and no attempt will be made to duplicate those efforts. This section is intended primarily to introduce the instrumentation concepts to the unacquainted reader and to show the basic techniques of recovering time and frequency from the Loran-C system.

Users of the Loran-C system for time synchronization can be put into one of two categories: those who utilize the one pulse/second transmission from master stations, and those who utilize the Loran-C pulse groups themselves from either slave or master stations.

Figure 4 shows a typical instrumentation scheme for time synchronization with the transmitted one pulse/second. This is not the only

way to extract the available information, but most practical methods would employ some variation of this basic scheme.

The Loran signals are received through the bandpass filter and displayed on the oscilloscope which is being triggered by the local clock. Due to the triggering rate of 1 pulse/second, the Loran-C pulse group will appear to drift across the scope, while the received 1 pulse/second will appear stationary. Using a calibrated time base, the time interval from the start of the sweep to the start of the received 1 pps Loran pulse can be measured and will equal propagation time plus receiving system delay plus local clock error, assuming the Loran-C chain is precisely on time.

The accuracy to which this measurement can be made is primarily dependent upon the received signal to noise ratio. The low information rate of one pulse/second and the inherent limitations of visual integration will normally preclude accuracies better than ± 20 microseconds. The use of a storage oscilloscope could be expected to improve upon this, but the basic problem in visual identification of the start of the received pulse is caused by the low signal power in the first few r-f cycles. The Loran-C pulse shape approximates the form $t^2 e^{-ct}$ where "c" is chosen to make the pulse peak at approximately 70 microseconds from the start. Note, however, that it would be incorrect to presume that the received peak of the pulse is 70 microseconds from the start, because of the probable presence of skywaves. The time of arrival of a skywave with respect to the groundwave depends upon, among other things, range from the transmitting station. However, it will not arrive any sooner than 30 microseconds after the groundwave due to the earth's curvature.[6] Observations of the received pulse beyond the 30 microsecond point will most likely be a groundwave-skywave combination and should therefore be avoided. Automatic Loran-C receivers solve this problem by automatically finding a point on the pulse prior to the skywave arrival time.

Figure 5 depicts a typical instrumentation scheme utilized by those who wish to achieve sub-microsecond accuracy with the Loran-C timing system, i.e. those who utilize the Loran-C pulse groups for time synchronization. The Naval Observatory uses similar methods at their time monitor stations, and obtain the required degree of reliability through two independent systems.

The Loran-C rate generator shown provides output triggers at the Group Repetition Rate corresponding to the tracked Loran-C rate and has the special capability of being synchronized by the operator to precise coincidence with any of the input 1 pps triggers. By reference to the Naval Observatory's Ephemeris Tables, the operator synchronizes the rate generator to the appropriate second, thereby providing a "Loran clock" which is running in agreement with both the Ephemeris Tables and the tracked Loran-C chain. An important condition on the operator's clock is that it initially must be accurate to at least one-half the period of the particular Loran-C rate chosen. This means that it must be within ± 25 milliseconds of UTC for Loran-C rate SO, and within ± 50 milliseconds for Loran-C rate SSO, with other rate requirements falling somewhere between those two extremes.

The Loran-C timing receiver shown is simply a single channel Loran-C receiver which automatically detects and tracks a point on the received Loran-C pulse prior to the skywave arrival. This receiver, by utilizing all of the Loran-C navigational transmissions and by using suitable integration times, should be capable of measuring the phase of the received Loran-C signal to 0.1 microsecond. For the sake of example in this discussion, presume that, having accurately found the start of the fourth r-f cycle (30 microsecond point), the receiver provides an output trigger at the particular Loran-C Group Repetition Rate and that this trigger precisely corresponds to the start of the first pulse of the Loran-C group of eight, i.e. 30 microseconds in front of the tracking gate on the first pulse.

All that now remains to be done is to start a time interval counter with the local rate generator and stop it with the output of the Loran-C receiver. The displayed time interval should then equal propagation time plus receiving system delay plus clock error, assuming the Loran-C chain is precisely on time. The time interval displayed is essentially continuous, in that it is updated at a minimum rate of ten times per second.

If the chain being utilized is an "untimed" Loran-C chain, its transmissions are not in agreement with the Ephemeris Table for that rate. Two users can still utilize the chain for synchronization to each other if they accomplish their individual synchronizations at about the same time, as was discussed in the section relating to accuracies available. All they need do is to simply presume that the chain is in fact on time; each would therefore be making an error of the same magnitude and the result would be synchronization to each other.

If the Loran-C receiver utilized is an automatic receiver in the sense that it provides a continuous phase lock on the received signals, it often provides a phase locked 100 kHz or 1 MHz output frequency. Since the frequency of the timed Loran-C chain is controlled by a cesium beam oscillator, the receiver output frequency will exhibit the identical long term stability of the cesium and can therefore be used as an excellent frequency reference. The short term stability will of course be degraded by the receivers measurement fluctuations due to noise and interference.

To implement a precision time and frequency system, it is important that the receiving system delay be accurately known and that it remain stable. Accurate delay measurements are best made with Loran-C simulators; timing equipment manufacturers normally furnish this delay information. Changes to the system delay might be caused by temperature changes or by adjustments to the tuned circuits preceding the actual measurement process. Excluding notch filters, normally the only tuned circuit involved is the bandpass filter which typically has a 3 db bandwidth of 20 to 30 kHz centered on 100 kHz. For high resolution timing systems, the selection of front end bandwidth and number of poles in the filter must not only consider the best compromise for reception of the Loran-C transmissions, but must also consider the effect of delay changes caused by temperature induced changes to filter tuning.

The retuning of narrow band notch filters may cause serious system delay changes. The user is faced with two alternatives. He can use a loop antenna system to null out the most harmful interference or he can calibrate his receiving system with the notch filter locked to the most advantageous position. Most likely he will do both. The latter alternative will not protect him from temperature induced changes but a reasonably well controlled environment should eliminate this as a serious problem.

Care should be taken when tuning notch filters because of the danger of distorting the received pulse shape to a degree where the automatic Loran-C receiver will select an r-f cycle other than the one it was designed to select. Manufacturers should be required to delineate acceptable tuning ranges for any installed notch filters.

Conclusion

Sub-microsecond accuracy is available with the Loran-C timing system if one is willing to accept the restricted operating conditions previously discussed. Without those restrictions, the accuracy to which one user in one part of the world can synchronize to a user in another part of the world is predicted to be five microseconds (one sigma). Probably the most significant advantage of the Loran-C system is that it is fully operational and such accuracies can be achieved without much complexity for the user. The system is also adaptive to the user's needs in the sense that coarse synchronization accuracies are available at minimum user costs while the finer accuracies require a greater investment. All of this is accomplished while still performing the navigational functions, resulting in an efficient use of the electromagnetic spectrum. It is reasonable to postulate that the user's position can be determined by the navigational system, thereby obviating the need for a survey. This, of course, would depend heavily on the degree of accuracy required and whether or not he was located in the prime navigational coverage area of the chain.

The future of the Loran-C system for time and frequency dissemination will undoubtedly see the expansion of the service to all existing chains as the operational requirements for these services arise. On the technical side, investigations are planned to remove the system error from the chains and to investigate the operational feasibility of reducing UTC tolerances to ± 1 microsecond. In that skywaves appear to offer great potential for global timing coverage, they offer a very fertile field for increased studies of the stability of their time of arrival.

References

1. Shapiro, L. D., "Time Synchronization from Loran-C," IEEE Spectrum, August, 1968.

2. "The Loran-C System of Navigation," Jansky and Bailey, Inc., Washington, D. C., February, 1962.

3. "Time Service Announcement, Series 9," No. 13, U. S. Naval Observatory, 18 December, 1968.

4. "Daily Relative Phase Values, Series 4," U. S. Naval Observatory, any date.

5. "A Loran-C Skywave Timing Experiment," Austron, Inc. Technical Note I68-1, July 1968.

6. "Loran," Vol. 1, M.I.T. Radiation Lab Series, 1946.

Chain	Master	Slave Emission Delays (usec)
East Coast	Carolina Beach, N. C.	W - Jupiter, Fla. - 13,695.48 X - Cape Race, Nfld. - 36,389.56 Y - Nantucket, Mass. - 52,541.27 Z - Dana, Indiana - 68,560.68
Central Pacific	Johnston Island	X - Upolo Point, Hawaii - 15,972.44 Y - Kure Island - 35,253.02
Norwegian Sea	Ejde, Faroe Islands	W - Sylt, Germany, - 30,065.22 X - Bo, Norway - 15,048.11 Y - Sandur, Iceland - 48,944.69 Z - Jan Mayen, Norway - 63,216.99
Northwest Pacific	Iwo Jima	W - Marcus Island - 15,283.94 X - Hokkaido, Japan - 36,684.70 Y - Gesashi, Okinawa - 59,463.34 Z - Yap Island - 80,746.78
*Mediterranean	Simeri Crichi, Italy	X - Matratin, Libya - 14,107.58 Y - Targabarun, Turkey - 32,273.28 Z - Estartit, Spain - 50,999.68
*North Atlantic	Angissoq, Greenland	W - Sandur, Iceland - 15,068.10 X - Ejde, Faroe Is. - 27,803.80 Z - Cape Race Nfld. - 48,212.80
*Alaskan	St. Paul Is.	X - Sitkinak, Alaska - 14,284.39 Y - Attu, Aleutian Is. - 31,875.30 Z - Port Clarence, Alaska - 53,069.07
*Southeast Asia	Sattahip, Thailand	X - Lampang, Thailand - 13,182.87 Y - Con Son Is., S. Viet Nam - 29,522.12

*These chains are not presently timed.

Table 1. Slave Emission Delays

U. S. NAVAL OBSERVATORY
WASHINGTON, D. C. 20390

DAILY RELATIVE PHASE VALUES, SERIES 4

Reference: (a) Time Service Letter of 30 September 1968, Series 4

The table gives: (USNO - Station)
Unit = one microsecond

Frequency (kc/s-UTC)	LORAN-C* Iwo Jima 100	LORAN-C* Johnston I. 100	LORAN-C Cape Fear 100	LORAN-C** Ejde 100
1969 Jan. 16	4.1	-	- 0.5	- 6.7
17	4.0	-	- 0.6	- 6.7
18	4.1	-	- 0.5	- 6.6
19	4.0	-	- 0.5	- 6.5
20	4.0	-	- 0.7	- 6.7
21	3.9	-	- 0.6	- 7.0
22	3.9	-	- 0.6	- 7.0

*Measured by USNO Time Reference Station within ground wave range but corrected to refer to USNO Master Clock.
**Computed from differential phase data provided by the U. S. Coast Guard Stations operating on the North Atlantic repetition rate.

Reprinted with permission of the U. S. Naval Observatory.

Figure 1. Example of Weekly Status Report of Timed Loran-C Chains.

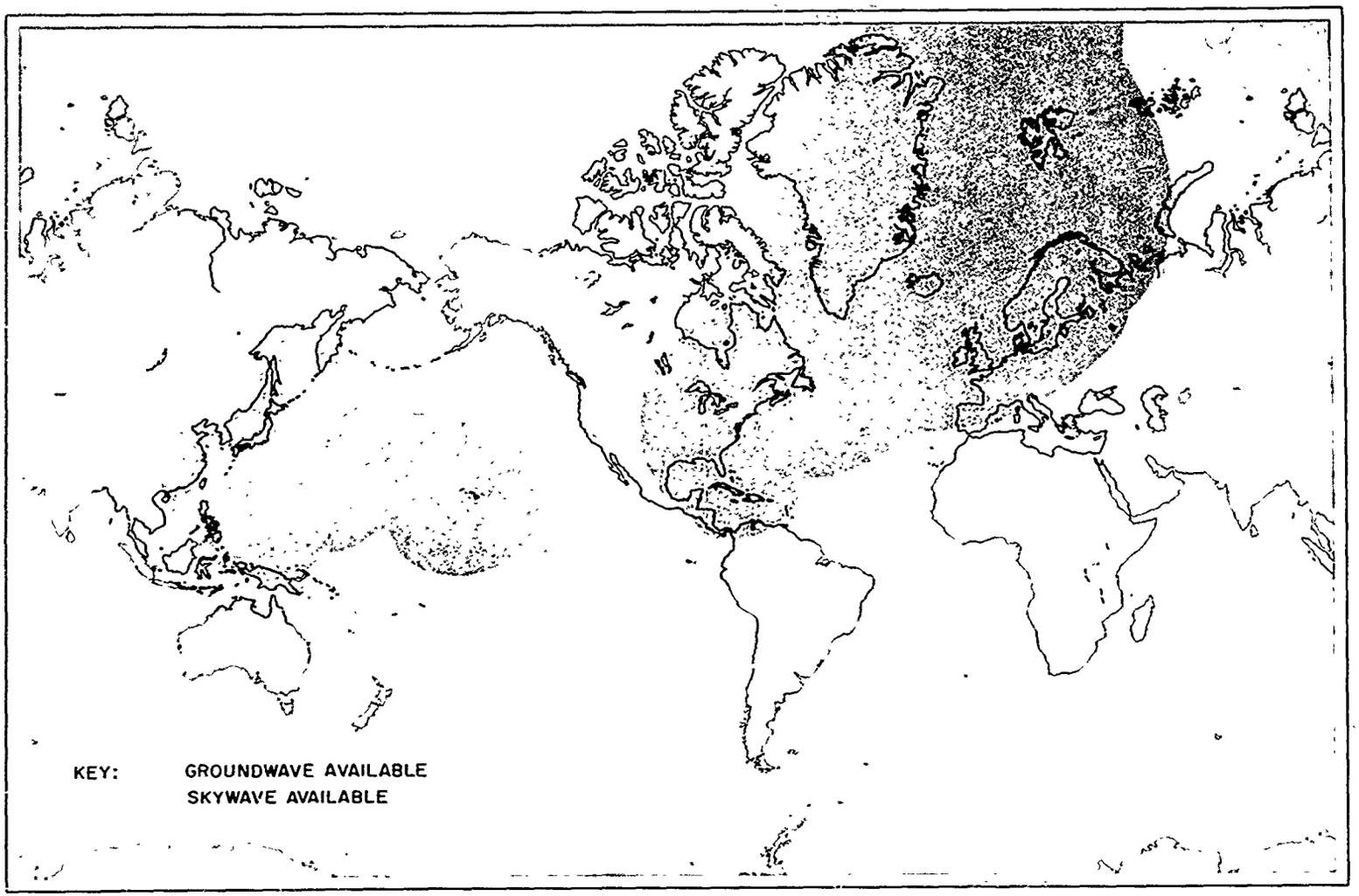


FIGURE 2. EXISTING GROUNDWAVE AND SKYWAVE COVERAGE

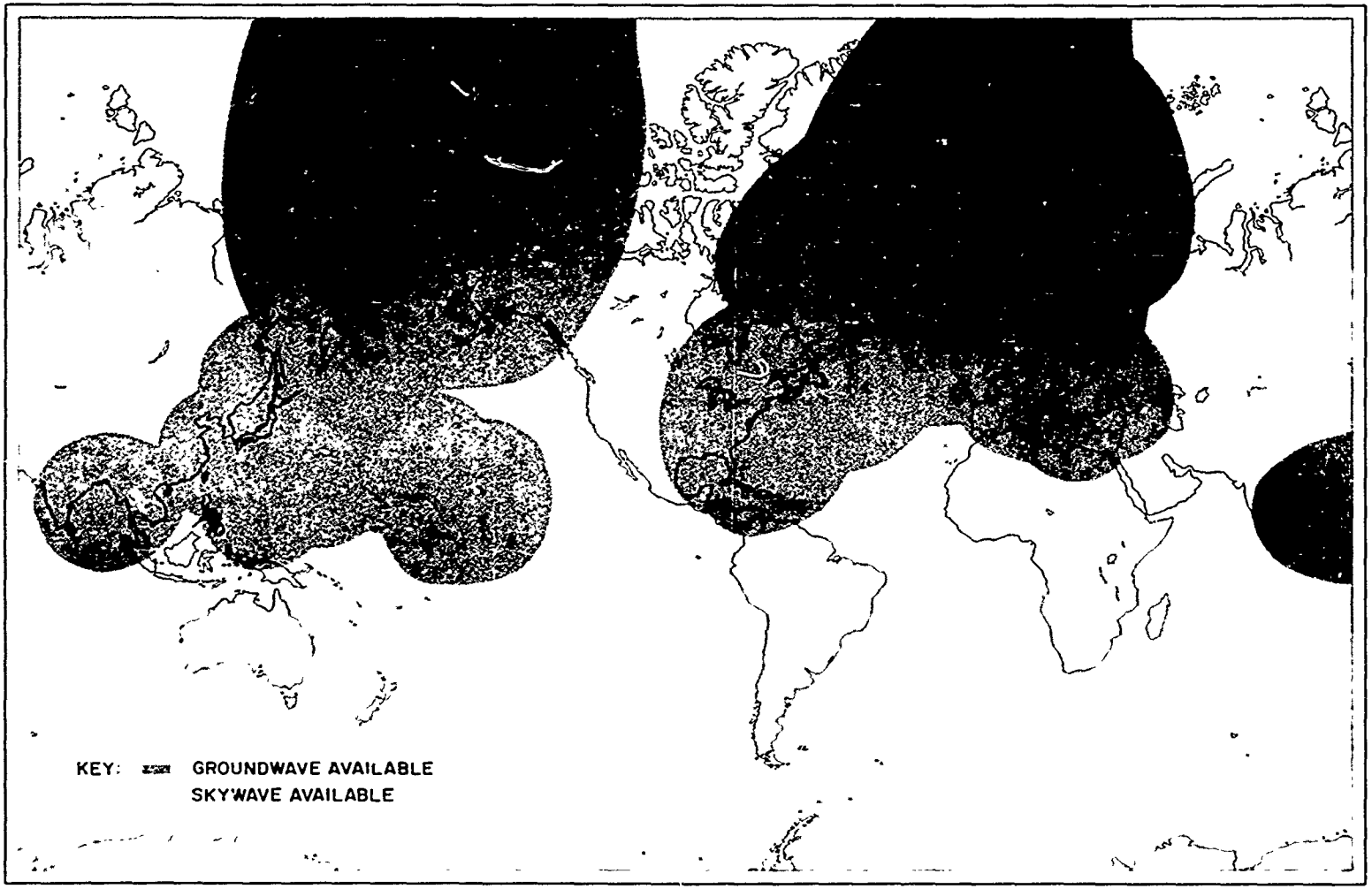


FIGURE 3 POTENTIAL GROUNDWAVE AND SKYWAVE COVERAGE

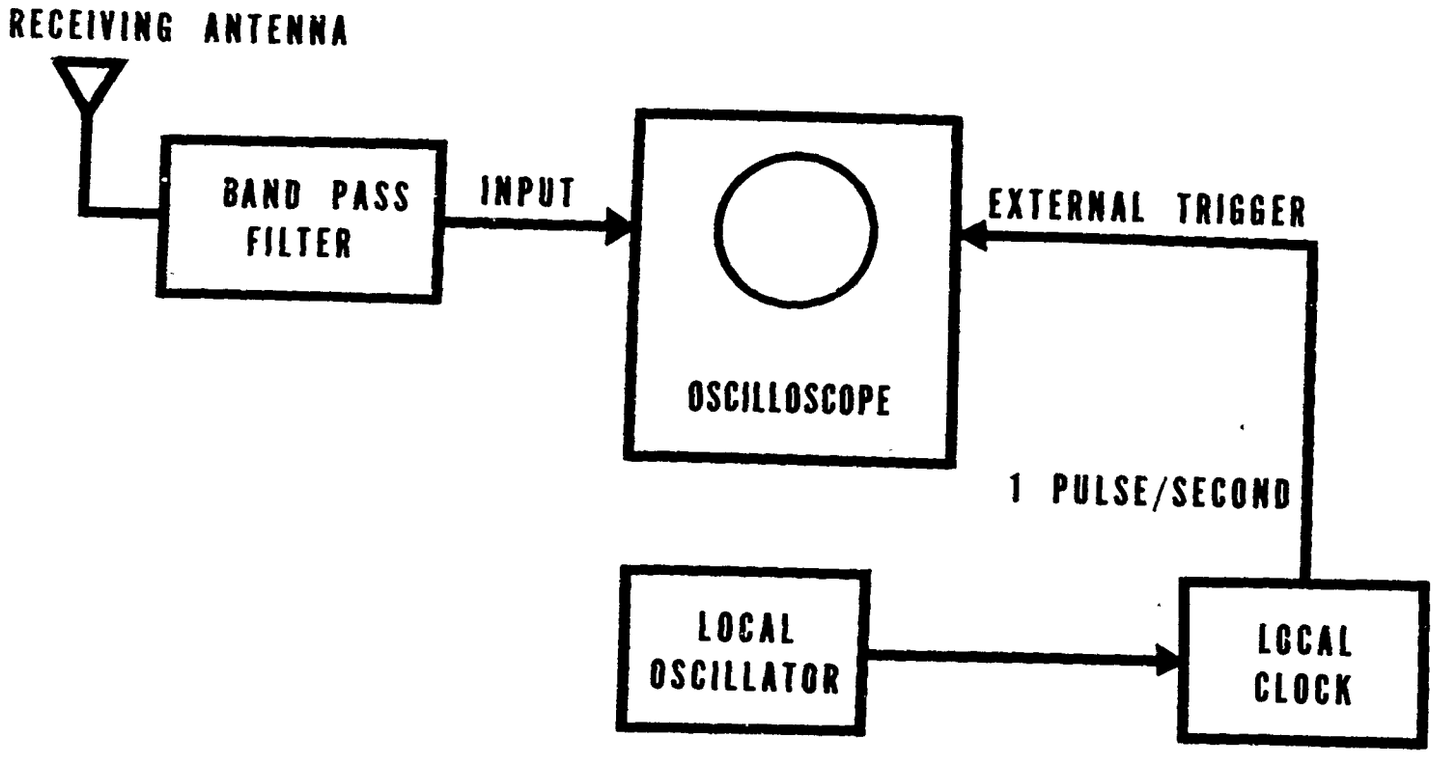


Fig 4 INSTRUMENTATION FOR ONE PULSE/SECOND SYSTEM

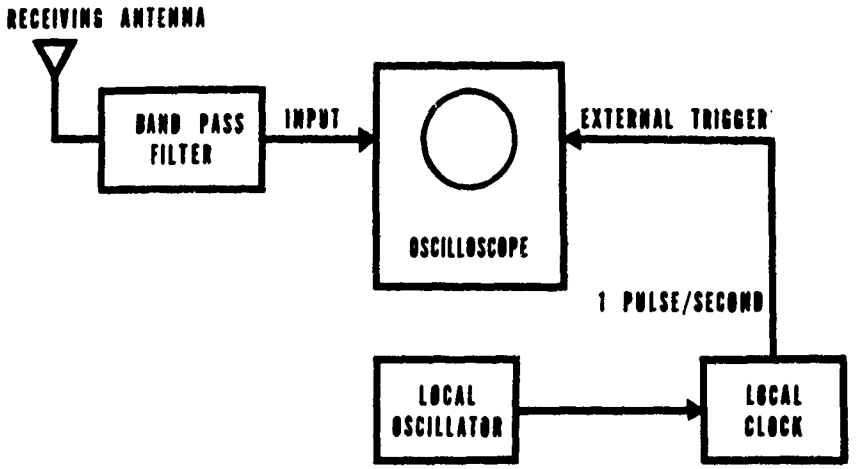


Fig 4 INSTRUMENTATION FOR ONE PULSE/SECOND SYSTEM

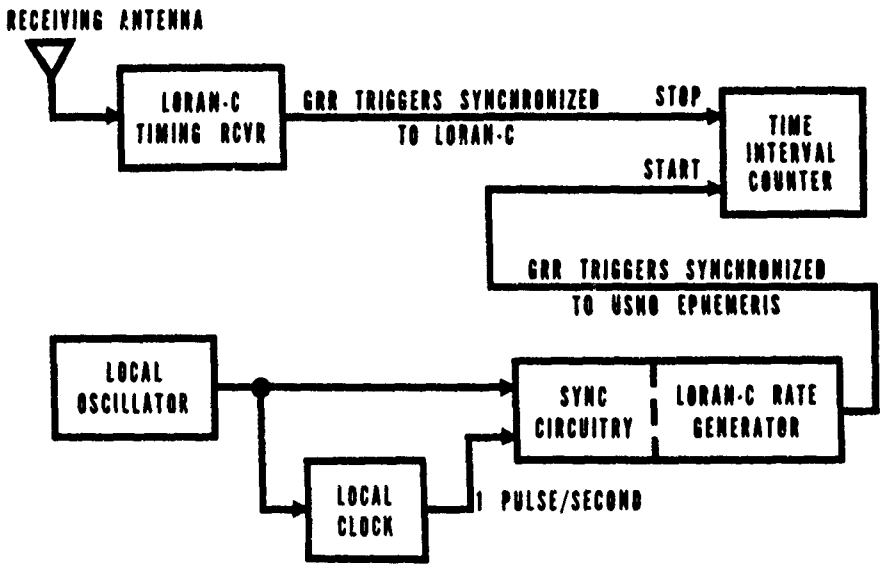


Fig 5 INSTRUMENTATION FOR UTILIZATION OF LORAN-C PULSE GROUPS

AN APPLICATION OF STATISTICAL SMOOTHING TECHNIQUES ON VLF SIGNALS FOR COMPARISON OF TIME BETWEEN USNO AND NBS

Alain Guétrot
Bureau International de l'Heure
Paris, France

Lynne S. Higbie
National Bureau of Standards
Boulder, Colorado 80302

Jean Lavanceau
U. S. Naval Observatory
Washington, D. C. 20390

and

David W. Allan
National Bureau of Standards
Boulder, Colorado 80302

Summary

Recent developments have provided a method for obtaining submicrosecond time comparisons over continental distances. The method was applied to a time comparison between the master clocks at the United States Naval Observatory (USNO) and at the National Bureau of Standards (NBS) in Boulder, Colorado.

There were the following developments. First, if two signals show a reasonable degree of correlation in their fluctuations, then one may derive an optimum linear combination of the two with a mean square error less than for either signal individually. The two signals studied were the transmissions on 21.4 kHz from NSS in Annapolis, Maryland, and on 20.0 kHz from WWVL in Fort Collins, Colorado. It is necessary that receivers be located for both signals at the locations of the controlling clocks. Existence of positive correlation was shown. The positive cross correlation probably was due to the near reciprocal path and the very close transmission frequencies.

Second, the phase fluctuations due to the propagation medium were consistent with a spectral density of the random phase noise proportional to $|f|^{-1}$, commonly called flicker of phase noise. This persisted for Fourier frequencies from one cycle per day down to one cycle per several weeks. The fluctuations on the linear combination of the two signals still behaved as flicker of phase noise but at a lower level.

The phase or time fluctuations of the master clocks however followed a spectral density law proportional to $|f|^{-3}$, flicker of frequency noise, for frequencies lower than one cycle per day.

Third, an optimum linear filter (Wiener filter) giving the minimum mean square error estimate (MMSEE) of the signal has been determined for a random walk of phase signal (spectral density proportional to $|f|^{-2}$) imbedded in white noise (spectral density proportional to $|f|^{0}$). The same filter was shown to be still optimum for spectral densities proportional to $|f|^{-3}$ for the signal and $|f|^{-1}$ for the noise.

Application of the above filter to the appropriate linear combination, defined through correlation properties, of NSS and WWVL signals showed an improvement of 15 dB in the rms day-to-day phase fluctuations. The day-to-day rms time deviations were about 70 ns on the final results. The output estimate of the filter, compared with portable clock measurements, gave a disparity of the order of the final output noise.

The experiment provided an opportunity to determine if there is an effect of mass on frequency and within the uncertainties of the experiment a null result was obtained.

* * *

A COORDINATE FREQUENCY AND TIME SYSTEM

G. E. Hudson, D. W. Allan, J. A. Barnes
National Bureau of Standards, Boulder, Colo.

and

R. Glenn Hall, J. D. Lavanceau, G. M. R. Winkler
U. S. Naval Observatory, Washington, D. C.

16 May 1969

A COORDINATE FREQUENCY AND TIME SYSTEM

G. E. Hudson, D. W. Allan, J. A. Barnes
National Bureau of Standards, Boulder, Colo.

and

R. Glenn Hall, J. D. Lavanceau, G. M. R. Winkler
U. S. Naval Observatory, Washington, D. C.

SUMMARY

A coordinate frequency and time system, suitable for extension to worldwide coverage, is described in relation to the form evolving in the United States. It consists of a network of component primary and associated stations, at fixed locations and altitudes, and a reference coordinating component at a reference location and altitude.

Each primary station is a source of coordinate frequency and time signal emissions steered directly in rate and epoch by a steering element; for coordination purposes this element is offset slightly in rate from an associated independently running "proper" local atomic standard. The local independent atomic standards are each evaluated for internal accuracy of realization of the standard Cesium frequency value assigned in the International System (SI) of Units, and for reliability and stability of operation.

A unified local atomic standard for the system serves as the coordinating component. It is defined in terms of a weighted average of the independent local atomic standard frequencies whose weights are chosen on the basis of reliability, stability, internal accuracy, and independence. It is a "proper paper standard" regarded as normally located at the system origin, at a standard reference altitude. In its definition, the frequency of the unified atomic standard is taken equal to the value f(Cs).

When completely coordinated, the carrier frequencies and signal epochs received at the origin from a participating station have average values equal to the assigned coordinate values when measured by the unified atomic standard. But if the measurement were made by the unified standard at the emission site, this would not be the case, in general, because of the small Pound-Rebka carrier frequency shift. Moreover, referred to emissions from a coordinated station A, there is no fractional difference in average rate and frequency of signals received at A, if they are emitted from another coordinated station, B.

An analysis of this network leads to the necessary conditions on the average frequencies emitted from the participating stations when coordinated and to the frequency values of the independent local atomic standards referred to the unified one.

The system could be extended internationally, by regarding the national unified standards as components of an international one whose assigned frequency would equal f(Cs). Again, a physical link must be established between the components in real time. This leads to the recognition of small individual frequency offsets of emissions and of national unified standards needed to achieve a well-defined international coordinate frequency and time system.

I. INTRODUCTION

A. General Consideration

There are three aspects of a time and frequency information dissemination system which require three quite distinct types of activity by associated personnel.

1) The system must be administered. This means that its effects on the users served by it must be continually evaluated to see if their needs are met. Policy decisions must be taken with an eye toward future developments, as well as toward safeguarding of commonly accepted practice, and technical feasibility.

2) It must be operated and maintained consistently and efficiently. By its nature all component portions of the system must be coordinated continually to insure internal consistency. Internal records of the regular operations, adjustments, changes in system, and data collected by monitoring activity must be kept in a uniform way and be available to system administrative, operating, and engineering staff.

3) The technical characteristics of the system, including its limitations, must be investigated and continually reexamined in the light of technological advances and the requirements of the system users and potential users. There must be a clear understanding of the existing system physically and from an information theory point of view. This understanding must be both on a conceptual and operational level and be both deep enough and broad enough to yield a proper technical perspective concerning the relation of the system to others, existing or proposed.

These three requirements account for the sextuple authorship of this short paper. The authors can be identified roughly, but not exclusively, in pairs with these aspects. They are paired because the system proposed herein stems from experience with and observation of the joint effort made between two major institutions of the many government agencies interested in establishing a single coordinate frequency and time system for the United States. In making this coordination effort we have observed that it may be feasible to extend its use with small modifications to wider coverage, an aim explicitly stated in the "terms of reference" of Study Group VII of the CCIR. In passing, it should be noted that the terms "coordinated" and "coordinate" -- since they refer to an agreed upon method for referring time to a common reference, are almost synonymous and may be used in the present text interchangeably (except for syntax).

There are, of course, many other individual contributions, including industrial concerns and government agencies, to this systematic effort in addition to the six authors and their respective two agencies. It perhaps suffices to cite in this respect the members of U. S. Study Group VII.

B. The USNO-NBS Coordination

As a guide in discussing the present coordination effort we list seven prerequisites descriptive of the kind of coordinate time system being set up and maintained (although not yet completely formalized!) in the United States for general use.[1]

1) The system contains component member stations and laboratories.

2) Each member has similar identifiable items of equipment or portions of equipment systems, which we shall call elements of the local system component.

3) The component members of the system are evaluated and given statistical weights on the basis of certain criteria agreed upon by the members.

4) Adjustment procedures are formulated in the sense of designating at what portions of the system coordinating adjustments are to be made, and when they are to be made, and tolerance limits of various kinds are specified.

5) Maintenance procedures are formulated and followed in the sense of collection, recording, and reporting of data to be used in order to maintain the smooth operation of the system.

6) Provisions are made for the participation of associated member stations, whose atomic clocks are used at the station for control and stability, but are not maintained sufficiently independently to be included in the component weighting procedure of the system. Such stations are monitored and emit signals within the prescribed tolerance limits -- i. e.,

they disseminate the coordinate time and frequency information in the way agreed upon, but do not have a commercial or government standards laboratory directly working with them as part of the system.

7) A reference location and initial epoch must be designated; a reference time scale (a "paper" clock) must be defined as an average (using the weights chosen in item 3) of the component independent atomic frequency and time standards. It should be noted that the system origin so specified should be in terms of a physical object--e. g., a laboratory building, and a well-defined physical event. Moreover, the record of differences between the reference scale and other well-defined scales such as UT2 and ET must be kept continuously as a necessary part of the coordinate time system.[2]

Since the independent local atomic standards and clocks are usually averages of several physically distinct but similar atomically controlled or calibrated clocks, the name independent local mean standard will be used for them. (Sometimes we shall use the term: "local atomic standards" as an alternative one.) Similarly, the average time scale at the system spatial origin will be designated the national mean scale --or sometimes the "unified atomic standard."

We can retrace these points in terms of the present USNO-NBS coordinated effort for the United States.

1) Obviously, the two present components of the system are:

(1) the U. S Naval Observatory in Washington, D. C., in close association with the laboratories of NRL and commercial standards laboratories which furnish the USNO with information and assistance in evaluating their atomic Cesium standard oscillators and clocks.

(2) the National Bureau of Standards Laboratory at Boulder, Colorado.

2) The essential basic elements for each component have been identified and are illustrated schematically in Figure 1. They are:

(a) local atomic Cesium frequency standards and independent mean time scales. At the USNO Time Service Division, in Washington there is maintained a set of from ten to sixteen independent atomic clocks whose readings are recorded and averaged almost every day by a statistical weighting procedure. The average reading is the independent mean paper time scale IM (USNO). This scale is also known as A. 1. The rate of running of each clock in the set is controlled by the radiation frequency characteristic of the well-known energy transition of atomic Cesium. The statistical procedure chosen insures a high degree of internal stability, reliability, and independence of IM (USNO) from other physical systems. Although not a direct realization in the sense of standards laboratories of the base units of time interval and frequency adopted in the SI, this "proper" time scale yields a close approximation. It is an example of a local atomic standard for the coordinate time system described herein.

At the NBS Time and Frequency Division, in Boulder there is maintained a local Cesium atomic frequency standard known as NBS-III, used to calibrate the rates of running of a set of crystal-oscillator and atomic-oscillator controlled clocks. A statistically weighted average of those readings is used to compute the "paper" atomic time scale known as AT (NBS). The internal accuracy of the reference frequency standard, NBS-III, is continually evaluated to determine the confidence with which the standard realizes the SI base unit of time interval, the second (or its inverse, the hertz). The statistical procedure is chosen to ensure that the AT (NBS) scale realizes the internal accuracy of the local atomic frequency standard, at this moment, 5×10^{-12} (3σ).

Redundancy and reliability are furnished by the crystal and atomic clock system, while long-term stability, as well as accuracy, is computed from the calibrations relative to the NBS-frequency standard. This portion of the NBS system is an example of an independent local proper atomic standard for the coordinate time system, and may be designated IM (NBS). Its scale is identical with what has been known as NBS-A, but is now known as AT (NBS).

(b) coordinating elements whereby necessary corrections and adjustments for coordination are introduced into the local control equipment or perhaps are kept only as a record and guide for coordination. Some detailed methods for the calculation of these corrections are discussed in a later section. No standard name has yet been given to these internal system records; but each component certainly keeps such records. It is probable that they should be continuously available to both components.

(c) Steering elements can be identified at both component stations. They are the respective clocks and oscillators used to control or steer the carrier frequencies, signal pulse rates, and time epochs. One may designate them as coordinate clocks and frequency standards. They should be set physically to maintain the agreed coordination. At the USNO, the steering element is called the master clock, but might be designated for system purposes TC(USNO). Similarly, the NBS time scale or clock keeping coordinate time may be called TC(NBS). When these clocks and scales become coordinated internationally, they could be designated UTC(USNO) and UTC(NBS) respectively; the U denotes "universal", and the TC denotes "coordinated time".

(d) There are, of course, radio emissions of time signals and carrier frequencies closely associated in spatial location with each component. For NBS, this is WWV and WWVL, and for the USNO, this is NSS. There are other emissions closely associated with them, but, because they are not from nearby radio stations, we prefer to distinguish them as "associated stations".

(e) Finally, at each locale there are monitoring and maintenance facilities whereby data required for maintaining continuous coordination are collected and recorded.

3) The initial coordination between the USNO and NBS took place on 1 October 1968. It was necessary at that time to correct relatively large divergences in rate of the respective emissions amounting to about 8 parts in 10^{13}. Hence, it was decided, as an interim measure, without following the format discussed here, to shift the steering rates sufficiently over a period of time, to eliminate the major portion of this discrepancy.[1]

This has now been accomplished. Accordingly, more refined adjustments for coordination, which will take into account small effects such as the Pound-Rebka gravitational red-shift,[3] and relative random walks of the two local atomic standards, will be made in the future. In order to do this, it has been decided to attach equal weight to the two standards. In our notation, this means $\alpha_1 = \alpha_2 = 0.5$. Briefly, this decision resulted from a consideration of several incommensurate requirements at the two components. It is very important that the coordinate scale have nearly one "second" as the base unit of time at the coordinate time origin. This is the "second", as defined in the International System (SI) of units. It is part of the mission of NBS to attempt to realize, via the frequency standard NBS-III, this unit. At the same time, it is also important to keep the coordinate time scale as stable as possible, and insure its reliability, of prime consideration for the USNO.

4) Every few months, as a result of continuous monitoring via radio observations and portable clock intercomparisons, a new adjustment of the steering elements and the radio emissions must be made, to insure that the frequencies and rates of all system emissions, as observed at the USNO (i. e., near sea level), have the nominal values assigned them, measured relative to the unified mean standard of the system. Because of differences produced by propagation, and differences in rate between the local mean standards and the system's unified one, the rates as emitted from the stations and measured by the respective local standards will not have the nominal values but will be slightly offset. This is discussed more fully in the next section.

5) The USNO maintains a large portable clock service, and is charged by the DOD with assuring uniform standards of practice in frequency and time throughout the DOD. Coupled with the radio monitoring facilities both at NBS and the USNO, data are collected which result in a value for the fractional rate difference, S_{12}, between the USNO and NBS independent mean standards.

Based in part on this measurement, and on the weight value chosen, other quantities of importance for maintaining coordination are inferred, as described in the third section.

6) We mention briefly the stations and emissions which can be considered as associated ones. This means that NBS-station WWVH, in Hawaii, being at a different altitude, should emit signals at a rate very slightly different than the rate of WWV -- even when both rates were measured by ideally identical local standards at the two sites. The Loran-C chain and the forthcoming Omega system, as well as other U. S. Navy standard frequency and time signal stations may be considered as associated stations;[4] monitoring data yields differences between their emitted signal rates and those necessary to be correctly coordinated with this system. Broadcast stations, too, if considered as associated coordinate time and frequency stations need similar information.

7) The USNO continually make observations of star transits, and this information is used by the International Bureau of Time (BIH) in constructing the astronomical time scale known as UT2 (roughly the same as Greenwich Mean Time). The knowledge of UT2 and the coordinate time system emissions is a necessary prerequisite for maintaining the coordinate time scale and UT2 as different aspects of the same coordinate time system. Necessary redundancy and convenience in using it for navigational purposes is thereby built into the system.

We have designated the USNO as the spatial origin location of this national coordinate time system. The initial epoch chosen is, provisionally, set at 1 January 1958 at 0000 UT.

II. THE MEANING OF COORDINATE TIME AND METRIC TIME

Coordinate time systems as spatially extended physical objects have long been envisioned in physics, and employed in astronomy. The importance of giving an explicit operational definition of such a system has recently been stressed by several authors -- notably in a presidential address to the Royal Astronomical Society by D. H. Sadler.[5,6,7]

It is our purpose here to explain the distinction between coordinate time and metric time. Very briefly, and roughly, it is analogous to the distinction between master-slave radio stations, and independently running oscillators.

More exactly, let us imagine many clocks, chosen for their stability and reliability, and because each, when calibrated by an atomic frequency standard at the same location, is determined to run freely at a rate very closely approximating that specified in the International System. The clocks can then be distributed over a large area and at different fixed altitudes, at fixed locations on the earth. Each then continues to run with its proper (Fr: propre ≡ self) rate. It is a local atomic standard, and can be used to measure rates and time intervals in terms of the SI base units for frequency and time. It is therefore called a "metric" instrument. This agrees with the strict mathematical sense of metric since we do not envision the possibility of changes in location or "small" accelerations as affecting its rate. It is a good piece of laboratory equipment for measuring time intervals.

Now, two such real clocks when in juxtaposition will normally diverge very slowly, but randomly, in reading and rate, because of tiny random effects which are not eliminated. A suitable average can produce an average reading and rate, which is presumably more stable --

and the stability properties of such a mean clock may be very desirable. However, one cannot necessarily say that the average clock is a better realization of the SI unit of time than either one of the two. This is certainly a matter for further investigation in each case. More important, however, is the fact that two such clocks at a distance would run quite independently, and at their own (proper) rates.

However, to obtain a time epoch and a frequency which is uniform over a large spatial volume, a uniform physical link with some standard clock must be established. It is a remarkable fact that radio signal pulses or carriers, used to establish such a link, and when measured by two standard metric clocks A and B, at a distance from eath other, show very little difference in frequency or rate relative to these. This is illustrated in Figure 2. However, a small systematic effect[3] has been observed in the frequency of electromagnetic radiation when it is propagated through a difference, gH, in gravitational potential, over an altitude difference H- even though the gravitational acceleration field, $\vec{g}$, is a constant. This difference is sometimes ascribed to a difference in rate of the metric clocks. But the argument from physical grounds shown in Figure 3, indicates that the radio waves, or photons, themselves change energy, and therefore frequency, by having work done on them as they move in a direction with a component parallel to the field vector $\vec{g}$. The computed value of this effect on radio carrier frequencies between Boulder, Colorado, and Washington, D. C. is 1. 8 parts in 10^{13} or an accumulative effect over a year of 6 μs. (phase difference in time units).

This can be described geometrically by attributing different radii of curvature to different portions of the spacetime map on which are plotted events happening to two clocks, A and C, at a distance and at different altitudes. In this way the relative difference between the received pulse rate at C and the metric clock rate at C can be portrayed. This is shown in Figure 4.

So it is easy to see, as also shown on Figure 5, that one coordinates time at different locales and altitudes, by sending radio pulses from an origin(say at sea level) so that the coordinate clocks at various locales are "slaved" to the agreed time scale at the system origin. Naturally other propagation effects and random effects must be taken into account, by a suitable collection of data. But the main point made here is that coordinate times (and frequencies) must be carefully determined by establishing physical links with a suitable origin time scale. Moreover, it is by no means obvious that all such physical links will lead to the same coordination. Hence, it is essential to define the method used to establish this link very carefully, and to follow well-prescribed procedures in measuring (by metric clocks) and recording the "behavior " of the extended system.[8]

In this system, coordinate clocks slaved together by radio means are needed for synchronization and epoch interpolation. Independent, proper, metric clocks are needed for time interval and rate measurement and calibration. Radio dissemination is needed in order to define the time coordinate spacing operationally. A unified system mean clock at the origin forms, together with adequate records, a useful reference and transfer standard, if its rate is nearly that prescribed in the International System.

III. PRINCIPLES AND ANALYSIS OF THE SYSTEM

We have enunciated, at least by implication, four main principles necessary for the quantitative definition of the kind of coordinate time system emerging from the USNO-NBS coordination.

1) A national mean standard or unified atomic clock and frequency standard is defined by the weighting process at the system origin. Its fractional rate difference from the average of the set of local mean atomic standard rates is zero.

2) Natural astronomical events are related to the national mean clock readings via UT2.

3) No radio waves or pulses used to synchronize the system should "disappear" -- ideally there is conservation of phase so that in every closed circuit (including equipment) the net phase change is zero.

4) Disseminated coordinated frequencies and rates measured as received at the origin in terms of the national mean standard must have their assigned (i. e., nominal) coordinate values.

In view of these requirements we may write down certain relationships between measured, inferred, or assigned quantities and quantities needed for making adjustments in emissions in order to attain coordination. First, let us define a few symbols. Let

N = number of component stations and laboratories in the system.

$\alpha_1, \alpha_2 \ldots \alpha_N$: weights assigned to the respective components.

$S_{12}, S_{13}, \ldots S_{1N}$: measured or inferred fractional frequency deviations of the rate of the first local mean atomic standard (its metric element) from the other component local mean time and frequency standards, as if in juxtaposition.

$F_1, F_2 \ldots F_N$: fractional frequency deviations of the rate of the respective local mean standards from the national one.

$E_1, E_2, \ldots E_N$: fractional frequency deviations of steering elements from local mean standards--recorded in the coordinating elements and introduced into the steering elements to compensate for random and systematic differences of the component local standards and emissions from the national mean at the system origin.

$G_1, G_2, \ldots G_N$: fractional frequency deviations introduced by the gravitational Pound-Rebka shift into the radio time and frequency signals used for synchronization and dissemination as they travel to the system origin.

$D_1, D_2, \ldots D_N$: ideal fractional frequency deviations of the respective emitted signals from their respective local mean atomic standards in order to follow the presently defined UTC system.

It should be understood that the frequencies as actually emitted and received may not at any one time agree with the ideal values. The system records should show the ideal values, and the departures from them which would be regarded as errors in the coordinate time system.

Figure 6 is a schematic diagram of a national system with N components. On it, one may trace for each component, the part of the component circuit in which each of the quantities defined above appear.

The national mean standard is denoted by the letters UAS (unified atomic standard).

On the figure, certain quantitative relations are shown which we now derive.

$$\sum_{i=1}^{N} \alpha_i = 1 \ . \qquad (1)$$

By the first principle, we may also write

$$\sum_{i=1}^{N} \alpha_i F_i = 0 \ . \qquad (2)$$

The definition of the pairwise measured (or inferred) fractional frequency deviation of the first local standard from the others yields

$$\left.\begin{aligned} S_{12} &= F_1 - F_2 \\ S_{13} &= F_1 - F_3 \\ &\vdots \\ S_{1N} &= F_1 - F_N \end{aligned}\right\} \ . \qquad (3)$$

It is useful to note that $S_{ij} = F_i - F_j$ and therefore that $S_{1j} = S_{1i} + S_{ij}$, for i and j having values 1, 2, ... N.

One immediately infers that

$$F_1 = \sum_{j=2}^{N} S_{1j}\,\alpha_j \quad , \qquad (4)$$

a relation which may be used in place of Equation (2).

The gravitational radiation shifts are easily evaluated from the differences in altitude of the component emission stations from the system reference altitude. Let these altitude differences be $H_1, H_2, \ldots H_N$, in kilometers.

Then

$$G_j = \frac{gH_j}{c^2} \quad (j=1, \ldots N) \qquad (5)$$

where $g = 9.8\ m/s^2$

$c = 3 \times 10^8\ m/s.$

This effect must be included to help satisfy principle 4).

The present UTC-system necessitates the insertion of a constant frequency offset for all UTC-coordinated emissions from the nominal assigned values in order to partially synchronize the time signals with UT2. Thus we have, at present

$$D_j = E_j - 300 \times 10^{-10} \quad . \qquad (6)$$

There are discussions at present aimed at eliminating this large frequency offset, and attaining the very rational requirements of the navigators by a different, simpler, and improved means. When, and if, this is done, the distinction between D and E will vanish. Until then, all UTC-carrier signals used for calibration purposes ought to be corrected on reception by this amount.

In any event, this frequency offset only furnishes part of the information required for a more accurate immediate knowledge of UT2, important to many users. Such information is, or can be, supplied by voice or simple code on the same emissions.

Principles 3) and 4) lead to the equations

$$E_j + F_j + G_j = 0, \quad (j = 1, 2, \ldots N) \quad . \qquad (7)$$

That is, for j=1, for example, the ideal fractional frequency deviation, E_1, between the first local mean atomic standard and the emitted signal rates, added to the gravitational shift, G_1, in the emitted radiation, ought to exactly compensate for the inferred random difference, F_1, in rates of the first local mean standard and the national unified mean standard. The difference between E_1 and D_1 can be ignored, since it is both added and subtracted in making the first, or any, circuit in the system.

Now, if one is given all the α_j-values, and measures, or infers from observations, all the S_{1j}-values, then all the F_j-values can be determined from Equations 1, 3, and 4. The E_j-values are determined from (7), and the D_j-values from (6). One represents most of these relations easily on a frequency level diagram as in Figure 7.

Associated with each component station there are two levels: the level determined by the fractional frequency difference, S_{1j}, of the first station's local mean standard from the j^{th} local mean standard, and the level determined by the gravitational shift, G_j, determined from the altitude difference of the station from the system origin reference level.

Positive quantities are represented by vectors pointing up, negative ones by vectors pointing down. Hence, a station at a high altitude will have a gravitational level below the reference level, indicated by UAS in Figure 7, and conversely. Figure 8 also depicts the situation for the case N= 2 as it existed for the USNO-NBS coordination. As explained, however, the adjustments made did not quite follow this procedure because of the large value of S_{12}. (The values of G_1 and G_2 were taken to be zero for this initial adjustment.)

IV. CONCLUSIONS

Let us summarize the foregoing discussions, briefly. In order to set up this coordinate time system, there seem to be nine essential procedural steps.

(1) The components must be identified-both the associated and the member stations and laboratories.

(2) The essential system elements belonging to each component member must be identified: the local mean atomic standard, the coordinating element, the steering element, and the local coordinated radio station emitting specified time and frequency signals.

(3) The reference origin in space and time must be specified.

(4) Intercomparison measurements of the component local mean atomic time and frequency standard rates and epochs must be made, leading to the values of S_{1j}, (j= 2, . . . N).

(5) Statistical weights (α) must be chosen for each component.

(6) The national mean standard is then defined by determining the deviations (F) of all the local independent mean atomic standards from it.

(7) Astronomical events must be measured in terms of the unified national mean standard, utilizing UT2.

(8) The desired ideal emission rate offsets (E) from their respective local mean standards must be determined and the present UTC value D ($D = E - 300 \times 10^{-10}$) determined for UTC coordinated emissions.

(9) Continuous monitoring procedures must be set up to continually redetermine and publish -- for the system at least -- the weights and

(a) F - values
(b) E, D - values
(c) Differences between actual and ideal emission rates and epochs.
(d) Differences between actual and ideal reception rates and epochs.

If these procedures are permanently implemented, and the present USNO-NBS coordination appears to assure that this will be the case, the system can become a model for more extensive worldwide application, and for other national coordinate time systems. It already is a specific well-defined portion of what is often called the "National Measurement System".[9]

Some modifications may be needed for worldwide use. For example, the weighting process becomes a truly knotty problem, with political overtones. To insure proper statistical procedures, it is essential that all component members keep truly independent standards, and evaluate them exhaustively for reliability, stability, and accuracy. A certain amount of knowledge of how each of these national standards and their components (observatories, laboratories, and radio stations) are linked by physical means and coordination procedures must be available to the designated international coordinating agency. We are fortunate that there is already in existence a traditional agency, the International Bureau of Time (BIH), which receives all the essential astronomical information to construct the UT2 scale, and acts as coordinator for the present UTC system. In fact,

the BIH maintains an average atomic scale of time which satisfies some of the prerequisites to serve as an international mean time scale and standard (the scale known as A. 3). Questions concerning the weighting procedure, i. e., how to use the data furnished the BIH --(both national mean scale data and individual component data) should be answered.[10,11,12]

Again, the CCIR has established an International Working Party charged for the moment with considering the improvement of the UTC system to meet modern requirements. So it would be natural for this group, IWP VII/1, to continue to act as an advisory body to the FAGS (steering committee for the BIH) and to the BIPM (the international bureau engaged in metric standardization activities). These are clearly interdisciplinary matters which are the concern of the sciences of astronomy, geophysics, radio engineering, physics, and politics. This is a matter for future consideration in the CCIR.[13]

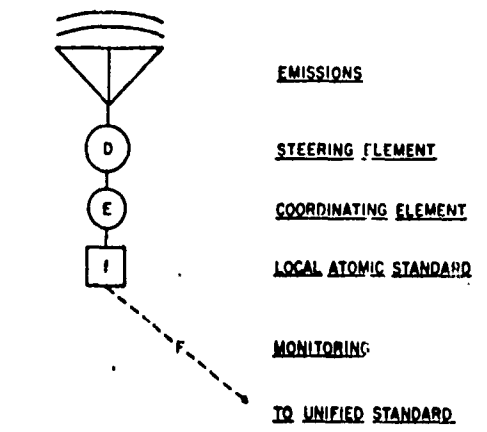


Figure 1: The symbols D, E, F refer to fractional frequency offsets explained in the text. The symbol "f" stands for the rate or frequency of the local mean clock or frequency standard--close to the prescribed SI-value.

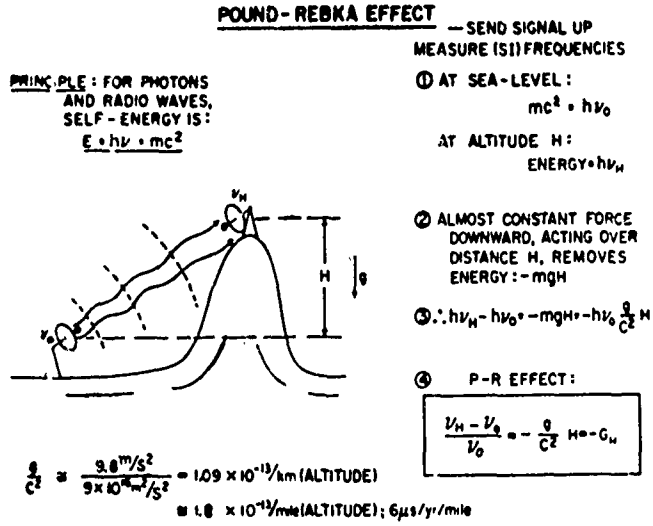


Figure 3: The Pound-Rebka, or gravitational frequency shift has been experimentally verified to one percent for gamma rays using the Mössbauer effect. It is a consequence of the general relativistic theory of gravitation.

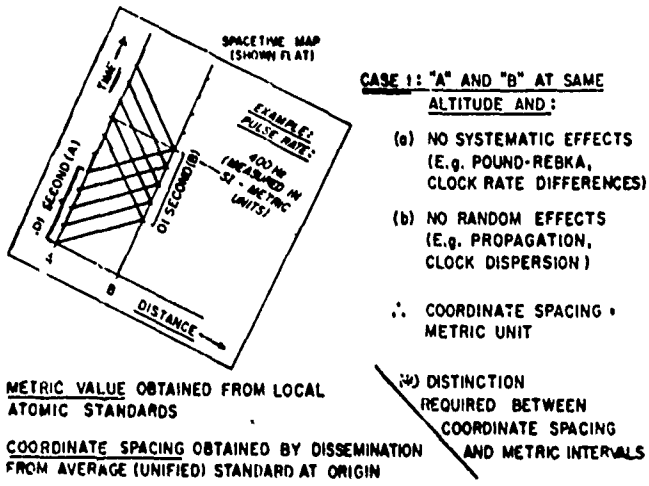


Figure 2: The schematic spacetime map shows radio pulses transmitted to B and returned to A to synchronize their clocks. A and B are separated by about 2.25×10^3 km (~1400 miles).

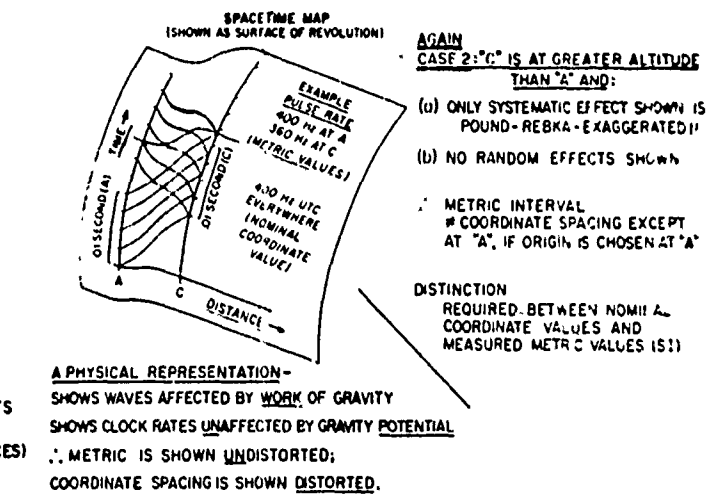


Figure 4: A and C are again separated by about 2.25×10^3 km. Their coordinate clocks are synchronized by radio pulses, and the rates of these are measured by the metric clocks running at their own (proper) SI rates. The curvature of the schematic map is introduced to show that the intervals of the two metric clocks are equal at the two respective locations differing in altitude. Only altitude changes produce the P-R effect, although for pictorial purposes, horizontal distances as well are combined with differences in altitude.

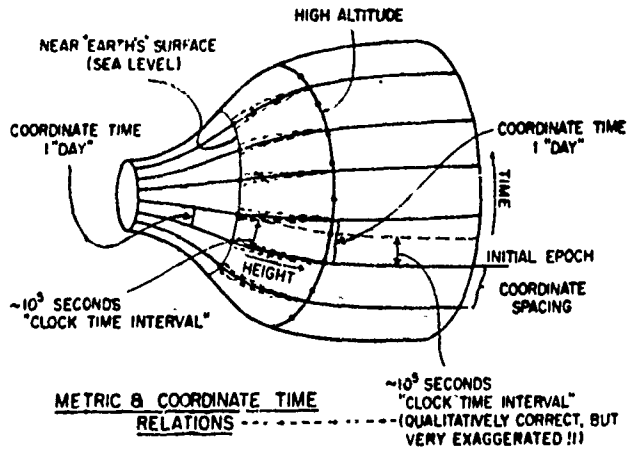


Figure 5: This is an even more exaggerated portrayal of the spacetime curvature needed to show the difference between coordinate intervals and metric intervals. The center of the "earth" is at the smallest cross section of the "bottleneck". The generators of the surface follow roughly the variation in gravitational potential with altitude. Unlike Figure 4, no horizontal distances are involved.

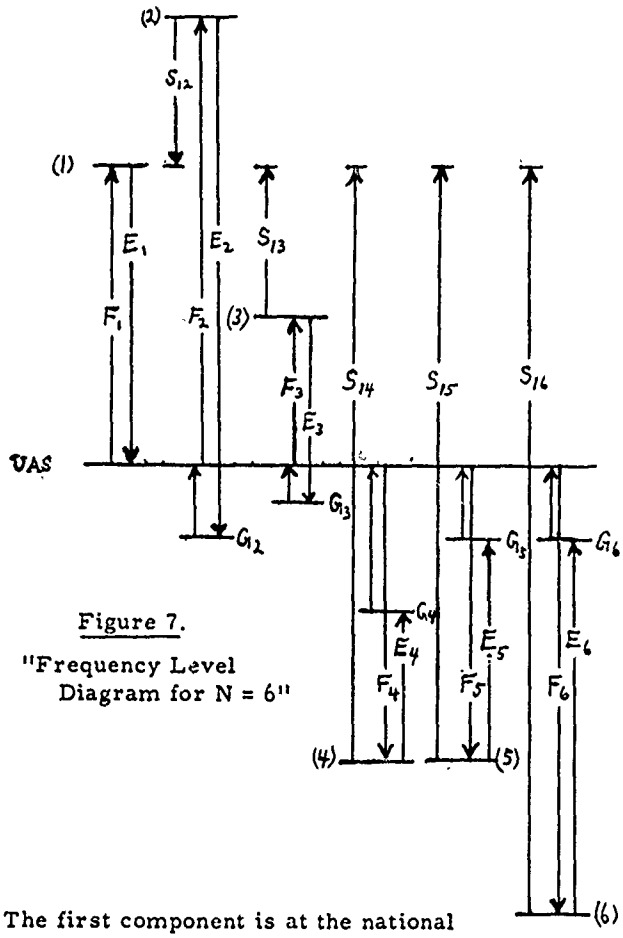


Figure 7.

"Frequency Level Diagram for N = 6"

The first component is at the national mean reference level, (UAS), so $G_1 = 0$. Positive quantities are associated with upward arrows. Weights are not shown.

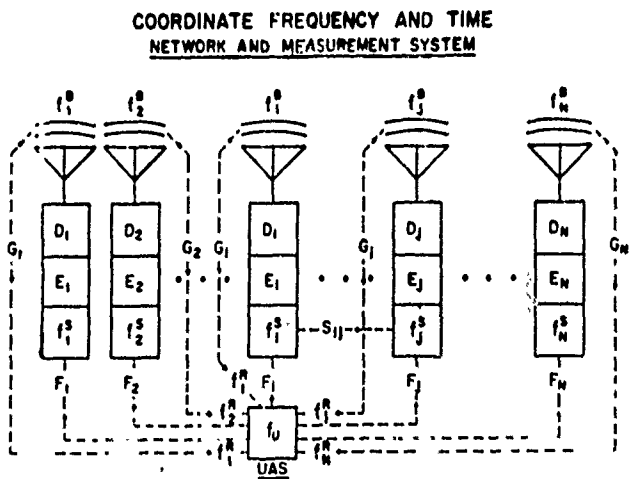


Figure 6: The coordinate time system resembles an electrical network. Ideal values of emitted frequencies f_j^B, are indicated, as are ideal values f_j^R, of received frequencies, and the rates f_j^S of the standards, and f_u of the system unified mean standard. Not shown are the observed emitted frequencies f_j^E, and actually received frequencies f_j^A. The measured or inferred rate intercomparison of the i^{th} and j^{th} local standards is indicated by the fractional frequency difference S_{ij}.

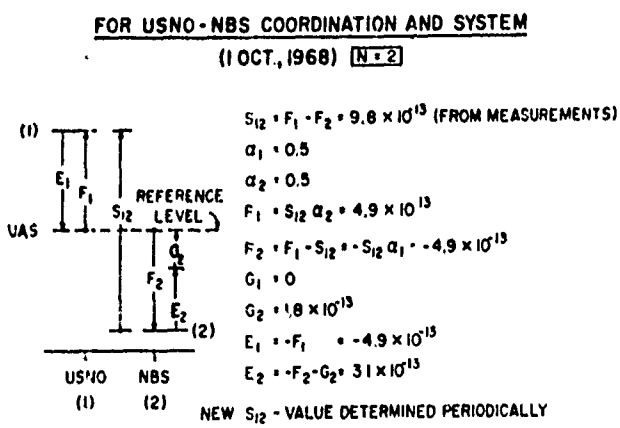


Figure 8: The values shown are of the correct orders of magnitude, but are for illustrative purposes only. Presently, the difference in rates of the local mean standards is considerably less, so that the importance of the gravitational shift is relatively greater.

V. REFERENCES

1) A. Guetrot, D. W. Allan, and L. S. Higbie, National Bureau of Standards and J. Lavanceau, USNO, "An Application of Statistical Smoothing Techniques of VLF Signals for Comparisons of Time between USNO and NBS", Proc 23rd Annual Frequency Control Symposium, USAECOM, (6-8 May 1969).

2) Nautical Almanac Office, USNO and Her Majesty's Nautical Almanac Office, Royal Greenwich Observatory "The American Ephemeris and Nautical Almanac for the Year 1969" --U. S. Government Printing Office, Washington, D. C. (1967). (Also see: "Explanatory Supplement to the Astronomical Ephemeris and the American Ephemeris and Nautical Almanac", Her Majesty's Stationery Office, London (1961)).

3) R. V. Pound and G. A. Rebka, Physics Review Letters 4, p. 337 (1960).

4) G. M. R. Winkler "Recent Improvements in the U. S. Naval Observatory Timekeeping and Time Distribution Operations", Proc. 22nd Annual Symposium on Frequency Control, USAECOM, p. 383, (April, 1968).

5) D. H. Sadler, "The Presidential Address -- Astronomical Measures of Time", Q. Jour. Royal Astronomical Society, 9, pp. 281-293 (1968).

6) G. C. McVittie, "General Relativity and Cosmology", p. 89 & ff, University of Illinois Press, Urbana (1965).

7) G. E. Hudson, "Spacetime Coordinate Systems" Proc. of the International Conference on Chronometry, 1, pp. 197-221, Lausanne (June 1964)

8) J. L. Synge, "Relativity, the Special Theory" (p. 7), J. Wiley and Sons, 2nd ed. (1965).

9) R. D. Huntoon, "Concepts of the National Measurement System", Science, 158, No. 3797, pp. 67-71 (1967).

10) B. Guinot, "Formation of a Mean Atomic Time Scale" (to appear in Astro. Bulletin)(1967).

11) J. A. Barnes, "The Development of an International Atomic Time Scale", Proc. IEEE, 55, No. 6, pp. 822-826, (June 1967).

12) L. Essen, "Time Scales" Metrologia 4, No. 4, pp. 165-168, (Oct. 1968).

13) G. E. Hudson, "Some Characteristics of Commonly Used Time Scales", Proc, IEEE, 55, No. 6, pp. 815-821, (June 1967).

Light Modulation at the Rb^{87} Hyperfine Frequency*

H. Tang and W. Happer

Columbia Radiation Laboratory, Department of Physics,

Columbia University, New York, New York 10027

We wish to report on the observation of microwave light modulation at the ground-state hyperfine structure (hfs) frequency of Rb^{87} since our experiment can be regarded as an outgrowth of the Zeeman light-modulation experiments which were first performed by Bell and Bloom[1] following a technique proposed by Dehmelt,[2] we would like to first give a short review of their experiment. The apparatus for the Bell and Bloom experiment is shown schematically in Fig. 1. Here, two light beams circularly polarized in the same sense are used. The first beam, which propagates along the direction of the static magnetic field in the z-direction, optically pumps the vapor in the resonance cell. The second beam, which is perpendicular to the first beam, is the one that is intensity modulated. The vapor in the resonance cell, upon being pumped, becomes oriented and acquires a static longitudinal macroscopic magnetization $\langle \vec{M}_z \rangle$ along the z-direction. Magnetic resonance may then be induced by applying an rf magnetic field H_1 perpendicular to the z-direction at a frequency ω_f, which is close to the Larmor frequency. Then the total magnetization of the vapor is no longer longitudinal but acquires a transverse component, $\langle \vec{M}_\perp \rangle$, rotating at frequency, ω_f. The absorption of the second beam depends on

$$\langle \vec{M}_x \rangle = \langle \vec{M} \rangle \cdot \vec{s} \ , \qquad (1)$$

where $\vec{s}$ is the average spin of the photons. Since $\langle \vec{M}_\perp \rangle$ rotates about H_0 at ω_f, $\langle \vec{M}_x \rangle$ will oscillate at ω_f and the transmitted light from the second beam will be intensity modulated at frequency ω_f. The modulation frequency ω_f, which is close to the Larmor frequency, is dependent upon the magnitude of H_0; hence, this is called the Zeeman light modulation and is the basis of optically pumped magnetometers.[3] Light modulation at frequencies corresponding to the ground-state hfs splittings was first observed K^{39} by Firester and Carver.[4] The frequency of modulation was 462 MHz, in the UHF region. There was some controversy over the existence of microwave light modulation at the much higher hfs frequencies of Rb and Cs. However, last year we were able to observe light modulation at the 6.835 GHz hfs frequency of Rb^{87}.[5]

To demonstrate the mechanism for hfs light modulation in Rb^{87}, let us first consider the ground-state level structure of the two isotopes of rubidium as shown in Fig. 2. Rb^{87} has two Zeeman multiplets of total angular momentum F=2 and F=1 due to coupling of the atomic spin J = 1/2 to the nuclear spin I = 3/2. The frequency separation of the two hfs levels of Rb^{87} is 6.835 GHz. Rb^{85} has a similar pair of hfs levels corresponding to total angular momentum F=3 and F=2 due to the nuclear spin of I = 5/2. The frequency separation of the two hfs levels of Rb^{85} is 3.035 GHz. Figure 3 shows the D_2 absorption line of Rb^{87} for the $(5^2S_{1/2} \rightarrow 5^2P_{3/2})$ transition and the profile of the 7800-Å D_2 line of the Rb^{85} resonance lamp. Here, the excited-state hyperfine intervals are not well resolved, and the line of each isotope consists of two hyperfine components. It is seen that the F=3 component of Rb^{85} light overlaps the F=2 component of Rb^{87} absorption. Hence, Rb^{85} light will pump only the F=2 component of the Rb^{87} atoms.

Consider a resonance cell containing a vapor of Rb^{87} atoms placed in a microwave cavity tuned to resonate at the hfs frequency ω_m, as shown in Fig. 4. A small static magnetic field defines the z-axis. If a beam of Rb^{85} D_2 light were incident on the vapor, the Rb^{87} atoms would be pumped into the F=1 hfs level of the ground state. The microwave field in the cavity will couple the $|F=1, m_F = 0\rangle$ and $|F=2, m_F = 0\rangle$ states. Figure 5 shows this situation in a simple vector model for an individual atom. Initially the atom is in the $|F=1, m_F = 0\rangle$ state. The total angular momentum vector $\vec{F}$ is perpendicular to the z-axis, as are the nuclear spin vector $\vec{I}$ and the electronic spin vector $\vec{J}$. Since in this state there is no longitudinal component of $\vec{J}$, there is no longitudinal component of magnetic moment, $\langle \mu_\parallel \rangle$. When the microwave field H_1 is applied, the atom will be in a superposition of states; $\vec{I}$ and $\vec{J}$ will tend to separate, and because of the coupling between $\vec{I}$ and $\vec{J}$, they will precess about $\vec{F}$ at the 0-0 hfs frequency. The vector $\vec{J}$ will now have a longitudinal component which, because of the precession of $\vec{J}$, will be oscillating at the hfs frequency; consequently, there will be a longitudinal component of the magnetic moment oscillating at the 0-0 hfs. The microwave field will continue to separate the $\vec{I}$ and $\vec{J}$ vectors until they are once again collinear in the state $|F=2, m_F = 0\rangle$,

at which time the longitudinal magnetic moment $\langle \vec{\mu}_{\parallel} \rangle$ will again be zero. Hence, with the microwave fields on, the atoms in the vapor will go back and forth between the two states $|F=1, m_F = 0\rangle$ and $|F=2, m_F = 0\rangle$. Upon averaging over the magnetic moments of all the atoms, there will be a net macroscopic longitudinal magnetization $\langle \vec{M}_{\parallel} \rangle$ in the vapor which is oscillating at the hfs frequency. There will be no net transverse magnetization because the atoms are randomly oriented in the transverse plane. The absorption of a light beam propagating along the z-direction will be proportional to

$$\langle \vec{M}_{\parallel} \rangle = \langle \vec{M} \rangle \cdot \vec{s} \ , \qquad (2)$$

where $\vec{s}$ is the average photon spin. From (2) we see that the light beam must have some degree of circular polarization in order to be intensity modulated.

To make quantitative calculations of the properties of a vapor with hfs coherence, one must describe the process of modulation in more general terms. It is possible to define[6] for the vapor a susceptibility $\langle \chi \rangle$ which oscillates at the 0-0 hfs frequency. This rapidly varying susceptibility will cause the optically pumped Rb^{87} vapor to behave as a parametric medium. A light wave of frequency ω propagating in this medium can be coupled to waves of frequencies $\omega + \omega_m$ and $\omega - \omega_m$. Similarly, a wave of frequency $\omega + \omega_m$ will couple to waves of frequencies $\omega + 2\omega_m$ and ω. The frequency dependence of the up-coupling and the down-coupling susceptibilities are shown in Fig. 6. The up-coupling susceptibility $\chi_+(\omega)$ attains its maximum value in the neighborhood of the F=2 component of the atomic absorption line. The down-coupling susceptibility $\chi_-(\omega)$ is small for such frequencies but attains a maximum value for frequencies in the neighborhood of the F=1 component of the atomic absorption. The up-coupling polarizability is small for the latter frequencies. Furthermore, it may be shown that for a linearly polarized carrier wave, the sideband waves are polarized at right angles to the carrier wave. Therefore, in order to observe intensity modulation of the carrier wave by photoelectric mixing at the photocathode of the detector, a circularly polarized carrier wave must be used.

A schematic diagram of the apparatus used to detect this modulation is shown in Fig. 7. A glass absorption cell containing Rb^{87} and about 10 Torr of neon buffer gas, was placed in a cylindrical microwave cavity, tuned to resonate at 6.835 GHz in the TE_{011} mode. The cell was about 4 cm long and the cell temperature was maintained at 45°C. A beam of Rb^{85} resonance radiation from a Varian spectral lamp passed through the vapor and pumped the Rb^{87} atoms into the lower hfs level. A small, static magnetic field was placed along the axis of the cavity to establish the z-axis of the coordinates. The F=2, m=0 and the F=1, m=0 sublevels of the Rb^{87} ground state were coupled by the microwave field of the cavity. The microwaves were generated by harmonic multiplication of the 999,225-Hz signal from a crystal-controlled oscillator. By varying the oscillator frequency, we could sweep the microwaves through the hfs resonance. The transmitted light beam from the cavity was passed through a circular analyzer and a D-line filter. The light intensity was detected by a crossed-field photomultiplier tube.[7] The 6.835-GHz signal from the photomultiplier tube was mixed with a 6.805-GHz signal from a local oscillator. The 30-MHz difference frequency was amplified and mixed with a 30-MHz reference signal obtained by mixing the local oscillator with the cavity excitation. The dc output from the 30-MHz mixer could be observed directly on a millivoltmeter. A typical modulation signal of 7800-Å D_2 light is shown in Fig. 8, which is the recorded output of the 30-MHz balanced mixer when the microwaves were swept through the hfs resonance. The top trace is the signal with the circular analyzer in place, while in the lower trace, the circular analyzer is removed. We see that only circularly polarized light is intensity modulated. The bandwidth of the detection system was about 1 Hz, and the modulation index was measured to be between 10^{-3} and 10^{-1}. Similar signals could also be observed with 7947-Å D_1 light and isotope-filtered Rb^{87} light.

The photomultiplier tube was very graciously loaned to us by N. Wittwer of the Bell Telephone Laboratories. Without this tube we could not have observed light modulation of such a high frequency. Figure 9 shows the arrangement of the photocathode and dynodes of the tube. This is a crossed-field device in which the electric field is maintained by a potential difference between the dynodes and a conducting bar which runs nearly parallel to the dynodes. The magnetic field is applied externally and is perpendicular to the electric field. The photoelectrons leaving the photocathode follow a curved path to the first dynode where they excite secondary electrons which in turn follow a curved path to the next dynode. The electrons, therefore, go from the photocathode to the anode in a trochoidal path, and multiplication takes place at each dynode. The trochoidal traversed by an

individual electron in going from one dynode to the next is essentially congruent to the paths of all the other electrons between the same pair of dynodes. This quality eliminates transit-time spread and gives this tube its large bandwidth capabilities. The photocathode has a quantum efficiency of about 10^{-4}, and the active area is about 0.1 cm^2, as it is the image of the slot masking the anode. The anode is just the center conductor of a 50-Ω coaxial cable going to the output connector of the tube, which has a current gain on the order of 50 dB and a bandwidth of about 10 GHz.

Aside from its intrinsic interest, hfs modulation has several potential applications. One of the first suggested uses[8] for hfs modulation is to monitor cavity pulling and light shifts in the Rb^{87} maser. The Rb^{87} maser is optically pumped with Rb^{87} resonance light which has been filtered with a Rb^{85} filter. The filtered Rb^{87} light will have only t the F=1 hfs component and achieves population inversion by pumping Rb^{87} atoms into the F=2 level. The maser field will produce hfs coherence in the Rb^{87} vapor; hence the transmitted pumping light will be intensity modulated. Both cavity pulling and light shifts result in the displacement of the oscillation frequency of the maser from the unperturbed hfs transition frequency.

Cavity pulling is due to the shift of the cavity resonance frequency from the atomic resonance frequency as a result of thermal expansion of the cavity. Light shifts are due to a shift in the atomic resonance frequency as a result of the interaction of the atoms with the pumping light. When the center of the atomic resonance coincides with the maser cavity resonance and when there are no light shifts, the modulation of the pumping light is 90° out of phase with the microwave fields in the cavity. Detuning of the center of the cavity resonance from the center of the atomic resonance results in a deviation from this 90°-phase relationship between the modulation and the cavity field. Although one cannot discern whether a phase deviation in the hfs modulation is due to cavity pulling or light shifts, cavity pulling is ordinarily a larger source of instability. Hence a servo-system to compensate for the thermal drifts in cavity tuning may be based on locking the phase of the hfs modulation of the pumping light at a value differing by 90° from the phase of the microwave fields in the maser cavity.

A more interesting potential use for hfs modulation is perhaps the possible construction of self-oscillating spin generators on the field-independent 0-0 hfs transition. A schematic representation of such a spin generator is shown in Fig. 10. The intensity-modulated light emerging from the cavity is detected by the photomultiplier tube, and the output signal at the hfs frequency ω_m is amplified, phase shifted by 90°, and then fed back to the cavity. When the sum of the phase shifts is zero around the closed loop and the closed loop gain is equal to unity, the system will oscillate spontaneously. Such a system is analogous to a maser[3] in that the atomic spin system provides self-sustained oscillations at its own resonance frequency although the gain is provided by an external amplifier. The stability of such a spin generator should be comparable to that of the maser. A Rb^{87} spin generator would have no obvious advantages over the Rb^{87} maser; however, a Cs^{133} spin generator would perhaps be desirable, since it does not appear possible to construct a Cs^{133} maser at this time. Figure 11 shows the level structure of Cs^{133}, which has a nuclear spin I = 7/2 and a ground-state hfs separation of 9.192 GHz. Optical pumping may be accomplished by using the 3888-Å emission line of He, which overlaps the F=4 component of the third resonance line of Cs^{133}. This He emission line would pump the Cs^{133} atoms into the F=3 ground-state sublevel, but this is the wrong level for population inversion and would not produce maser action. However, it is possible to obtain modulated light at the Cs^{133} hfs frequency by the technique which we have already described.

Another potential use for hfs modulation is one suggested by Carver[9] for the construction of a wideband magnetometers. Such a device would be a spin generator oscillating on a magnetic field-dependent transition. In the case of Rb^{87}, for example, the F=1, m_F = 1 to F=2, m_F = 1 transition could be used. In such an oscillator, the frequency of oscillation would change proportionally to a change in magnetic field. The device would be wideband in a relative sense, i.e., one can measure large field variations and stay within the bandwidth of the microwave cavity.

In order to construct a spin generator or to study the parametric generation of sidebands in a vapor of optically pumped atoms with hfs coherence,

one would need a larger index of modulation than that which we obtained in our previous experiment. Larger indices of modulation may be obtained by passing light through longer lengths of optically pumped vapor. A calculation of the conversion efficiency of power in the carrier wave into power in the sideband waves has been made for Rb^{87} vapor. In this calculation we have made the assumption that only two sidebands of frequencies $\omega \pm \omega_m$ are coupled with the carrier wave of frequency ω_o. The results of this calculation are shown in Fig. 12, where we have plotted the conversion efficiency of power from a monochromatic carrier wave into the strong sideband. We have plotted this for carrier waves of several frequencies in the vicinity of the F=4 component of the D_1 line of Rb^{85}. We see that sizable conversion efficiencies are obtainable for reasonable lengths of vapor. However, light from a resonance lamp is not a monochromatic wave but is of quite broad spectral profile. Based on the results of Fig. 12, Fig. 13 shows now the spectral profile of the D_1 line of Rb^{85} light is modified in passing through a 40 cm length of pumped Rb^{87} vapor with 0-0 hfs coherence. We have neglected the weak sideband since it is small compared with the strong sideband. We can see that near the center of the strong sideband, the power conversion efficiency is 5-6%. This suggests that the sideband may be observed directly with a high-resolution optical spectrometer such as a Fabrey-Perot interferometer.

It must be mentioned here that all the calculations on the conversion efficiencies of the sidebands were made assuming good phase matching between the carrier wave and the oscillating susceptibility of the vapor. The most efficient coupling of a carrier wave to the strong sideband light wave occurs when the difference between the propagation constants of the two light waves in the vapor is equal to the propagation constant of the microwaves in the cavity. This latter condition may be stated as

$$k' = K + k ,$$

where K is the microwave propagation constant in the cavity, k' is the propagation constant of the optical sideband in the vapor, and k is the propagation constant of the carrier in the vapor.

In order to obtain modulated light with a large index of modulation, we have constructed a long microwave cavity to hold a long ($\approx$40 cm) resonance cell. The cavity, cell, and resonance lamp are shown in Fig. 14. The cavity is made from a stack of metal plates with holes cut in the middle. The spacing between the plates is made small enough to prevent the transverse electric waves from propagating out of the cavity. Since this cavity permits only azimuthal currents, only TE_{omn} modes will be contained in the cavity. The reason for this peculiar cavity design is to allow the pumping light to enter the cavity from the sides to provide uniform pumping along the length of the cell. This cavity has an unloaded Q on the order of 10^4. In order to achieve phase-matching conditions, the wavelength of the standing wave in the cavity must be shortened to make it nearly equal to the free-space wavelength of the microwaves. This is accomplished by dielectric loading of the cavity with the heavy quartz walls of the resonance cell. By carefully choosing the wall thickness and the diameter of the quartz tube used for the cell, the proper phase-matching conditions may be achieved.

In summary, large indices of modulation of light at hfs transition frequencies may be obtained by using long lengths of optically pumped atomic vapor provided that phase-matching conditions are rigorously met. Efficient generation of sidebands makes it possible to study parametric frequency conversion at optical frequencies spectroscopically and to facilitate the construction of microwave spin generators.

REFERENCES

*This research was supported wholly by the Joint Services Electronics Program under Contract No. DA-28-043 AMC-00099(E).

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

(2) H. G. Dehmelt, Phys. Rev. 105, 1924 (1957).

(3) A. L. Bloom, J. Appl. Opt. 1, 61 (1962).

(4) A. H. Firester and T. R. Carver, Phys. Rev. Letters 17, 947 (1966).

(5) B. S. Mathur, H. Y. S. Tang, R. Bulos, and W. Happer, Phys. Rev. Letters 12, 1035 (1968).

(6) W. Happer and B. S. Mathur, Phys. Rev. 163, (1967).

(7) R. C. Miller and N. Wittwer, IEEE J. Quantum Electron. QE-1, 49 (1965).

(8) R. Novick, P. Davidovits, W. Happer, and W. A. Stern, Physics of Quantum Electronics: Conference Proceedings, edited by P. L. Kelley, B. Lax, and P. E. Tannenwald (McGraw-Hill Book Co., Inc., New York, 1966) pp. 267-279.

(9) A. H. Firester and T. R. Carver, Phys. Rev. 164, 76 (1967).

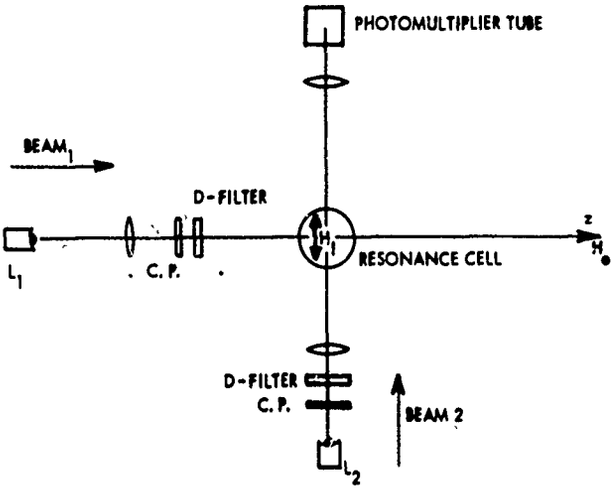


Fig. 1 Crossed beam apparatus for the detection of Zeeman light modulation.

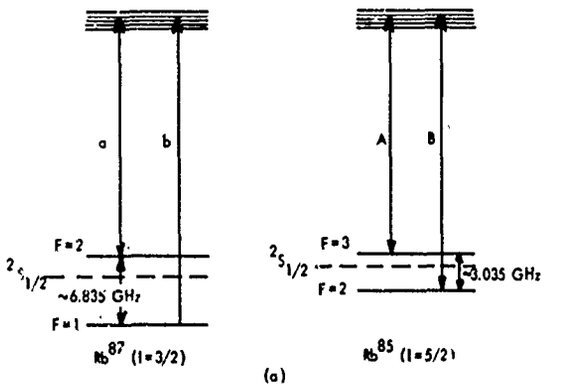


Fig. 2 Ground state level structure of Rb^{87} and Rb^{85}.

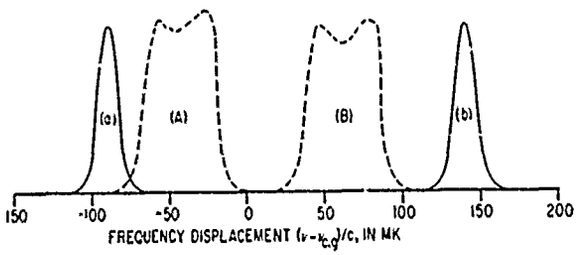


Fig. 3 D_2 absorption line of Rb^{87} (a) and (b) and the D_2 emission line of Rb^{85} (A) and (B).

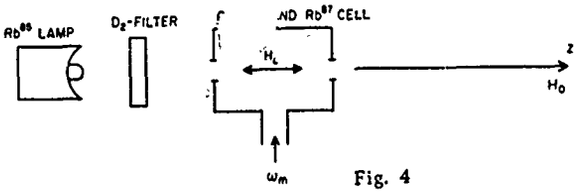


Fig. 4

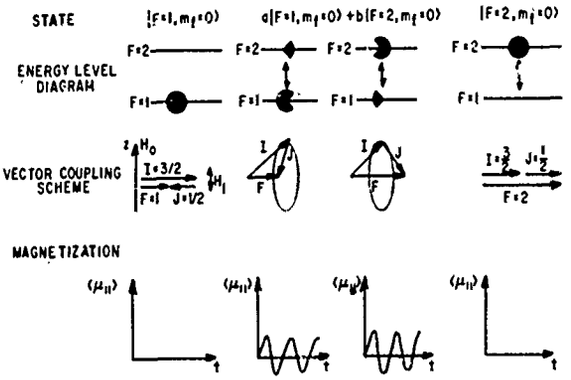


Fig. 5 Vector model of a Rb^{87} atom undergoing the 0-0 hfs transition.

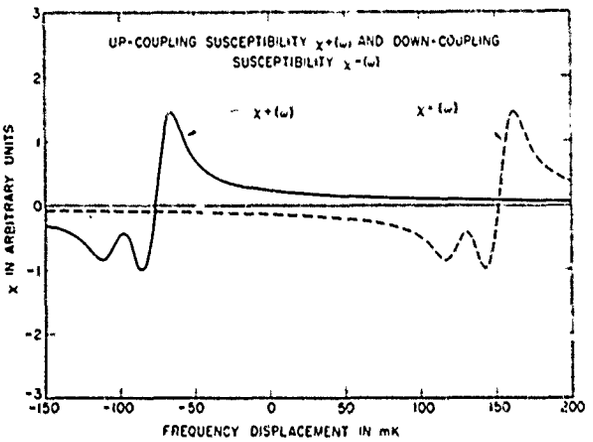


Fig. 6 Frequency dependence of the up-coupling and the down-coupling susceptibilities.

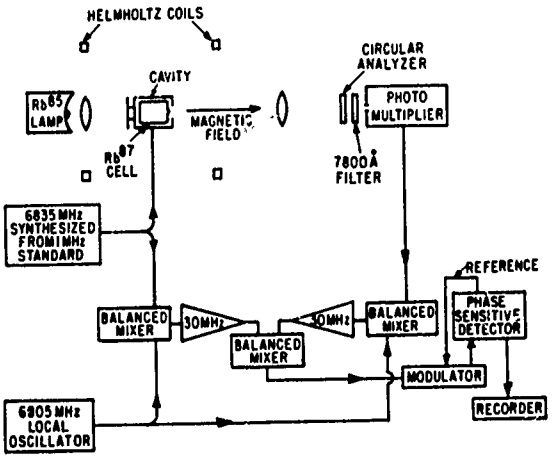


Fig. 7 Experimental apparatus for observing hfs modulation in Rb^{87}.

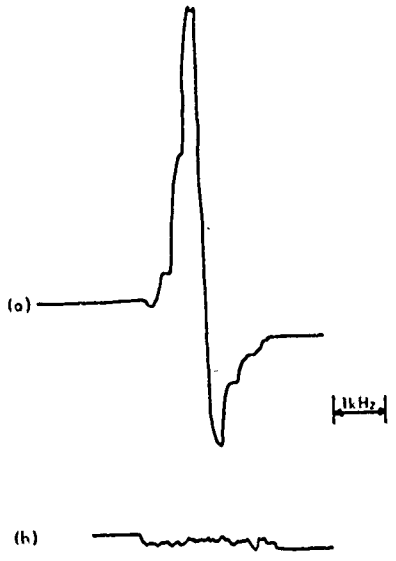


Fig. 8 Modulation signal for the D_2 line of Rb^{85} light,a) with circular analyzer in place,b) with circular analyzer removed.

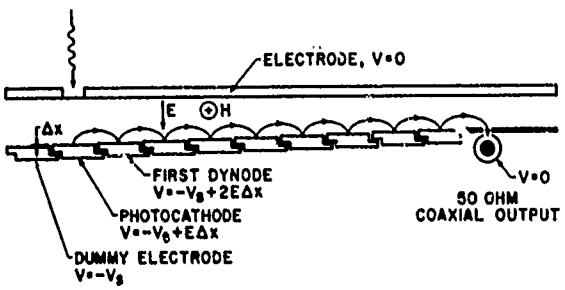


Fig. 9 Electrode configuration of the crossed-field photomultiplier tube (after Miller and Wittwer).

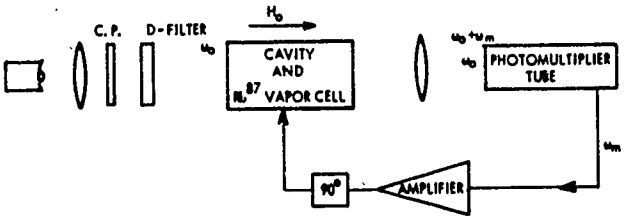


Fig. 10 Schematic representation of a Rb^{87} spin generator.

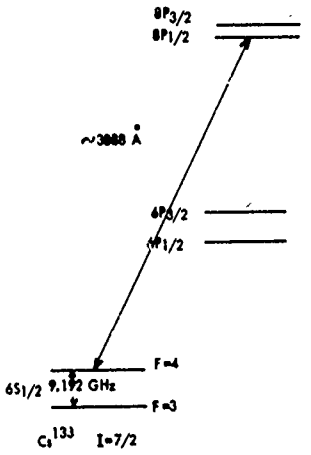


Fig. 11 Level structure of Cs^{133}.

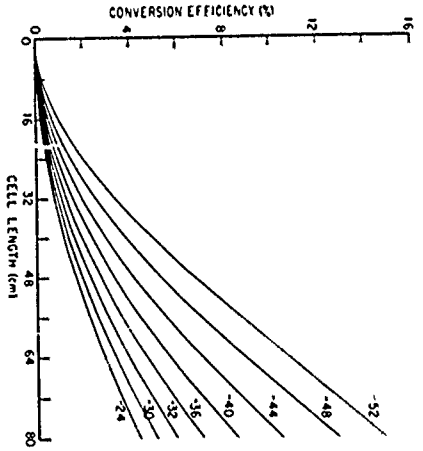


Fig. 12 Plot of the conversion efficiency of power from the carrier wave into power in the strong side-band vs. cell length for an initially monochromatic carrier wave. The numbers beside the curves are the frequency displacements of the carrier waves from the center of gravity of the Rb^{87} D_1 absorption line.

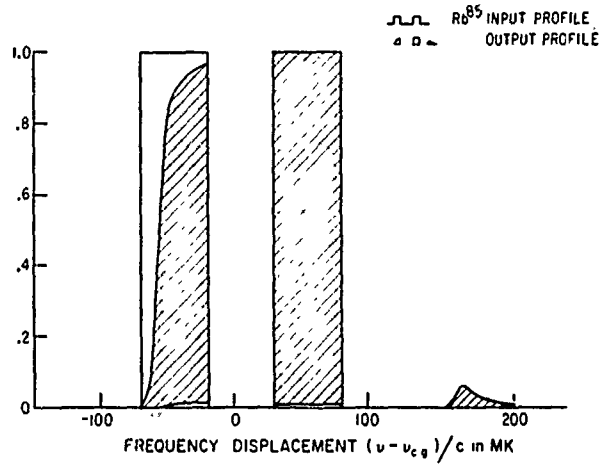


Fig. 13 Representation of the spectral profile of the Rb^{85} D_1 line before and after passing through a 40cm. length of Rb^{87} vapor.

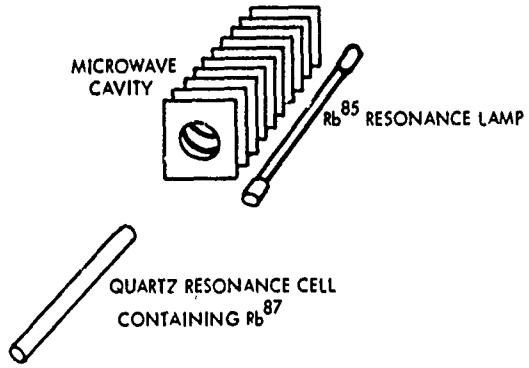


Fig. 14 Drawing of the long cavity, resonance cell, and resonance lamp to be used to obtain enhanced modulation indices.

An Optically Pumped Rb^{85} Maser Frequency Standard*

W. A. Stern and R. Novick

Columbia Radiation Laboratory, Department of Physics,

Columbia University, New York, New York 10027

An optically pumped Rb^{85} maser frequency standard with good short-term characteristics is under construction. This frequency standard will operate in the ground state of Rb^{85} at 3035.7345 Mhz. The maser will be small in size, have both a high power output and a high short-term frequency stability. The Rb^{85} maser oscillator frequency standard will operate between the $5^2S_{1/2}$,F=3 and $5^2S_{1/2}$,F=2 ground-state hyperfine transition. In the Rb^{85} maser, intensity pumping is possible because of the accidental isotope coincidence in the spectra of Rb^{85} and Rb^{87}. The Rb^{85} resonance radiation, containing both the $5^2S_{1/2}$,F=3 – 5^2P and $5^2S_{1/2}$,F=2 – 5^2P resonance lines, is passed through a filter cell filled with Rb^{87} and 70 Torr of helium. With sufficient buffer-gas pressure broadening and shifting, the $5^2S_{1/2}$,F=2 Rb^{87} absorption line can be made to absorb the unwanted Rb^{85} $5^2S_{1/2}$,F=3 emission line normally present in the resonance lamp almost completely. Since the filtered resonance radiation contains only the $5^2S_{1/2}$,F=2 – 5^2P resonance line, it causes an overpopulation in the $5^2S_{1/2}$,F=3 level in the ensemble of Rb^{85} atoms. In the steady state, assuming that the $5^2S_{1/2}$,F=2 Rb^{85} level is completely depopulated, each $5^2S_{1/2}$,F=3 magnetic substate has a population density of 1/7. When such an ensemble of atoms is placed in a cavity tuned to the hyperfine transition frequency, coherent microwave emission takes place. If this microwave emission exceeds cavity losses, oscillation occurs. It is expected that a power output of 10^{-10} watts should be obtained. For averaging times ranging between 1 and 10^{-2} seconds the maser should be stable to a few parts in 10^{12}. The Rb^{85} maser will be limited in its long term stability by cavity pulling, frequency shift due to buffer gas, chemical changes in the cavity environment due to outgassing and changes in the pumping lamp profile. However, several schemes have been proposed to solve these problems.[1]

Detailed studies of the filtering of Rb^{85} light by Rb^{87} vapor buffered with helium were made first on a 3.4 meter Ebert scanning spectrometer and later on a high finesse, 2-inch piezoelectrically scanned Fabry-Perot interferometer. The Rb^{87} – Rb^{85} hyperfine lines at 7800Å as well as the pumping scheme are shown schematically in Figure 1. Figure 2 shows the ratio of the wanted line to the unwanted line as a function of temperature for low Helium buffer gas pressures. It is evident that at these low-pressures the pumping scheme is relatively inefficient although the filter cell temperature is quite low. Figure 3 is a similar plot of data for a filter cell filled with Helium at 76 Torr. Figure 4 shows the actual profile of the filtered and unfiltered light at 77°C with 66 Torr of Helium. Here the unwanted F=3 Rb^{85} hyperfine line disappears while the desired pumping F=2 hyperfine line is attenuated by 65%. The Rb^{85} and nitrogen buffer gas are contained in a TE_{021}-mode copper-plated stainless-steel cavity. The unloaded cavity Q is about 85,000. The nitrogen buffer gas reduced the collision frequency of the aligned Rb^{85} atom with the cavity walls, and quenches reradiation from the upper P states.

Salient features of the cavity are shown in Figure 5. It is coupled to the outside world by a magnetic loop which excites the field pattern shown in Figure 6. The loop is rotated before evacuation to provide the desired coupling. Tuning is accomplished by a copper stub which is moved in and out by a stainless steel bellows. Pumping light is admitted through two windows; each having a stainless steel housekeeper seal.

An evacuated glass cylinder is placed between each perforated end plate and window to prevent absorption of the pumping light by Rb vapor, which would be there in the absence of the cell. The cavity is pumped out through a side arm on one of the windows. A conventional

vacuum system is used to achieve pressures of the order of 10^{-6} Torr in the cavity. When the cavity is closed to the vacuum station, the Rb metal reacts with residual gases in the cavity and reduces the pressure to less than 10^{-9} Torr.

The Rb^{85} maser oscillator frequency standard is useful as a tool to study light modulation in the ground state; frequency shifts due to pumping light, buffer gas, and cavity pulling; and finally, the quenching mechanism in the Rb^{85} system. The Rb^{85} maser frequency standard can also be used to study the processes causing frequency instability in the Rb^{87} maser oscillator frequency standard.

Reference

*This work was supported wholly by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S Air Force) under Contract DA-28-043 AMC-00099(E).

1. P. Davidovits and R. Novick, Proc. of IEEE 54, 155 (1966).

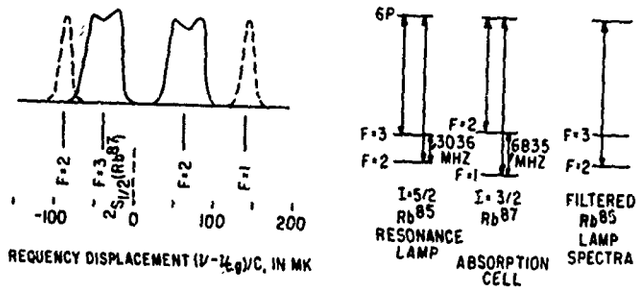
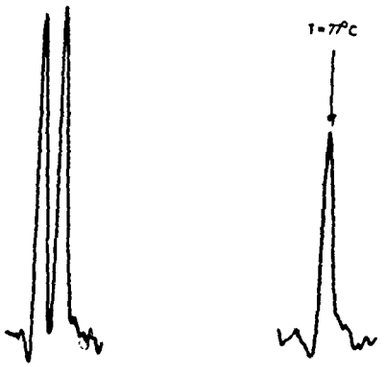


Fig. 1 Pumping Scheme in Rb^{85}



Scan of the filtered Rb-85 7949-Å line at a Rb-87 filter-cell temperature of 40.5°C.

Scan of the filtered Rb-85 7949-Å line at a Rb-87 filter-cell temperature of 77°C.

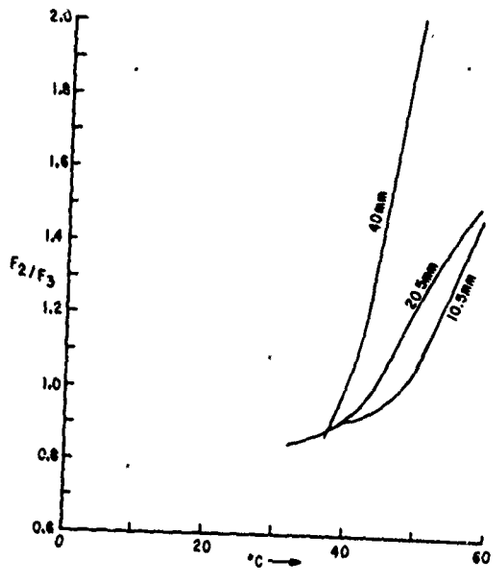


Fig. 2. Efficiency of Pumping Lamp for Low Helium Pressures

Fig. 4.

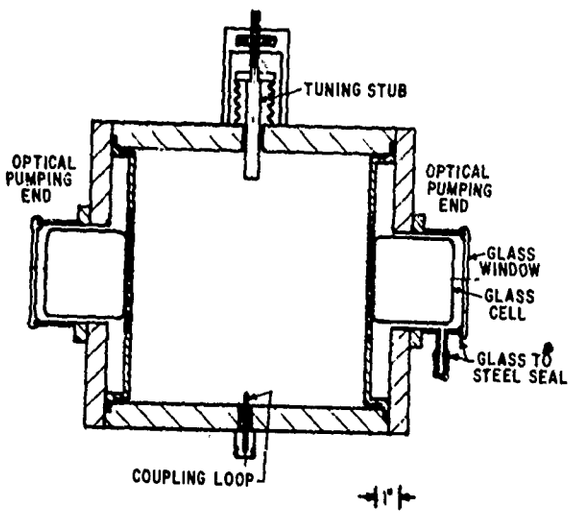


Fig. 5. TE_{021} Mode Maser Cavity

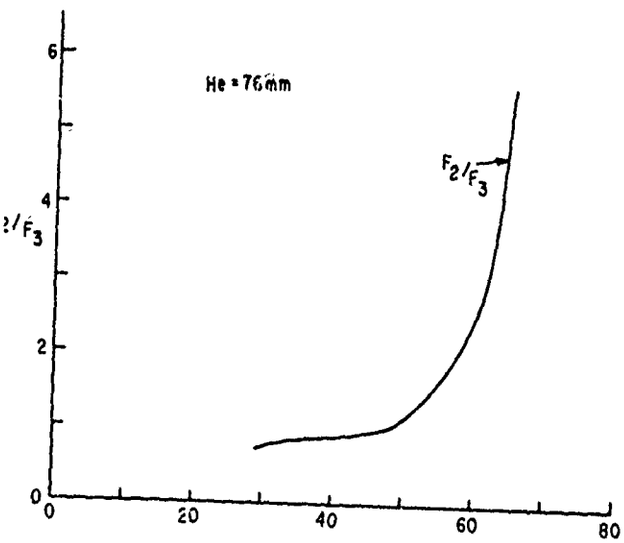


Fig. 3. Efficiency of Pumping for 76mm Helium

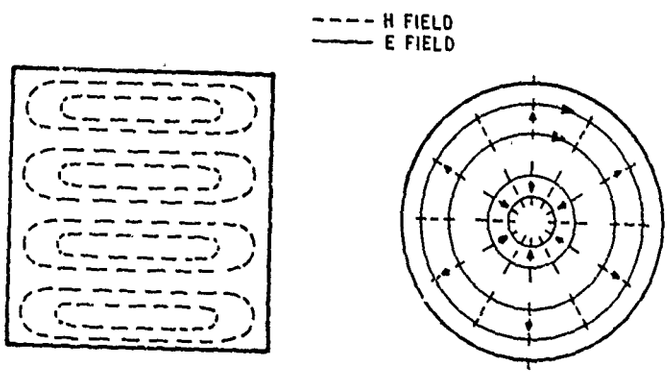


Fig. 6

A REPORT OF THE PERFORMANCE CHARACTERISTICS OF A NEW RUBIDIUM VAPOR FREQUENCY STANDARD

Darwin H. Throne

Hewlett-Packard Company
Palo Alto, California

Summary

The Model 5065A Rubidium Frequency Standard was introduced by the Hewlett-Packard Company about one year ago. Data now exists to characterize the performance of the standard in regard to long-term stability (>1 week), short-term stability (<1 hour) and environmental stability.

The long-term stability of rubidium standards has been the subject of several studies in recent years.[1,2] The long-term specification for the 5065A is $\pm 2 \times 10^{-11}$/month; but, due to the many processes that can occur in the atomic resonator, different aging rates are observed for each instrument.

The short-term stability of the standard for averaging times long compared to the servo time constant is determined by the signal-to-noise ratio of the atomic resonator and the frequency multiplier chain. The standard achieves a closed loop stability of 3.6×10^{-12} rms for 1 second averaging. Typical standards have a flicker level of $\approx 3.6 \times 10^{-13}$ for averaging times longer than 100 seconds.

To utilize the intrinsic stability of a frequency standard in an applications situation, the standard must be reasonably insensitive to environmental changes. Rubidium standards are perturbed by changes in temperature, magnetic field and vibration. Effort was made to minimize the effect of each of these environmental effects on the standard.

The calibration of the standard and time scale changes are facilitated by a variable frequency synthesizer and a linearized magnetic field control. This feature enables the user to make time scale changes without reference to another standard with an error of less than 3.5×10^{-11}.

Frequency Stability

Long-Term Stability

The long-term stability of the rubidium vapor resonator is mainly determined by the stability of the rubidium spectral lamp and chemical processes within the absorption cell. Since these may be aiding or opposing effects, the net result for the independent processes may produce positive, negative or zero frequency change with time.

Figure 1 shows a long-term stability plot of two 5065A Rubidium Standards referenced to the Hewlett-Packard house cesium standard. The reference was within 1×10^{-12} with respect to the National Bureau of Standards and the U.S. Naval Observatory from November 1967 to April 1969.

5065A #1 exhibited a frequency drift of -6×10^{-12}/month from 3 November 1967 to 20 February 1968. There was a magnetic field correction of $+7 \times 10^{-12}$ on 20 February 1968. A change of slope was observed in early March to a rate of about -3.3×10^{-12}/month. This slope continued to early June 1968. The aging rate changed again in early June to a slope of $<2 \times 10^{-12}$/month. The discontinuity in August 1968 was due to adjustment of the magnetic field to correct the 5065A frequency to agree with the reference standard. Following the frequency adjustment, the slope remained at a value of $<2 \times 10^{-12}$/month through February 1969.

5065A #2 exhibited a frequency drift of -5×10^{-12}/month from 10 July 1968, to 20 September 1968. In September the frequency stabilized with daily variations about a constant offset of -1×10^{-11} with respect to the reference standard. The standard began to exhibit a positive aging rate in late February. The offset at the end of the record is -5×10^{-12} with respect to the reference standard. The slope is less than 2×10^{-12}/month.

The data in Figure 1 is presented to indicate very long-term aging characteristics of the rubidium resonator. The daily variations of the frequency for 5065A #1 are much larger than the variations of the frequency of Instrument #2. This can probably be attributed to the prototype electronics in #1 and need not concern us here. This instrument is in the Hewlett-Packard Standards Laboratory and has been in operation since September 1967.

Chi et. al. have reported a systematic drift rate of 6 to 7×10^{-14}/day for three rubidium standards. The frequency of the standards was measured for four months. A linear extrapolation of the daily drift rate yields a rate of $\approx 2 \times 10^{-12}$/month.

To summarize, we should note that the long-term stability of a rubidium standard is not adequately defined by either a constant drift rate as indicated by the 2×10^{-11}/month reference line in Figure 1 or a statistical measure where the stability is defined as an rms value of the frequency measured over a long period of time. The data

clearly shows that both specifications apply to this type of standard. The drift rate specification gives a worst case performance characteristic. The rms value of the frequency variations over a given time interval is much more difficult to specify without actually making measurements over the time interval.

Short-Term Stability

The stability of frequency standards which are perturbed by various spectral densities of noise has been treated in detail by Cutler and Searle and others.[3,4] Allan has shown that to estimate the variance of the frequency fluctuations in an oscillator one must specify the averaging time, the sample size and the time between adjacent measurements.[4]

Figure 2 shows the expected value of the variance of the frequency fluctuations versus averaging time for the 5065A Rubidium Standard. The data for averaging times from 10^{-4} seconds to 450 seconds was obtained with the beat frequency measurement systems described by Cutler and Searle.[3] The data were analyzed on an N=2 basis and corrected for measurement dead time as discussed by Allan and Barnes.[4,5] The data at 3600 seconds was taken from a continuous phase record using the method described by Allan.[4] Each data point represents 100 or more measurements of sample size 2. The data points with the error bars were calculated with a computer program. The error bars represent the standard deviation of the successive frequency differences with respect to the expected value. The triangular data points were obtained with an HP 5360A Computing Counter. The counter was programmed to automatically calculate the expected value of the frequency fluctuations.

The 5065A has a quartz oscillator which is slaved to the atomic resonator. For averaging times small compared to the servo time constant (<100 msec), the frequency fluctuations are determined by the quartz oscillator. The oscillator output is filtered by a single pole crystal filter. This is indicated by the theoretical line with slope -1/2 at about 3 msec.

The change in slope from -1 (oscillator additive phase noise) to -1/2 (white frequency noise) at about 200 msec indicates the point at which the rubidium resonator begins to determine the frequency fluctuations. The frequency fluctuations vary as $\tau^{-1/2}$ from τ = 200 msec to τ = 100 seconds. The frequency instability does not improve for averaging times from 100 seconds to 3600 seconds indicating the presence of f^{-1} type noise.

In summary, the rubidium standard achieves a short-term stability of $\approx 3.6 \times 10^{-12}$ for 1 second averaging and reaches a flicker level of $\approx 3.6 \times 10^{-13}$ for averaging times between 100 seconds and 3600 seconds.

Environmental Stability

The performance of a frequency standard in an applications situation is affected by the existing environmental conditions.

The effect of temperature on the frequency of the 5065A Rubidium Standard is shown in Figures 3a and 3b. The histograms in Figures 3a and 3b show the frequency change observed at -5°C and +55°C relative to the frequency at +25°C for the first production of thirty-five instruments. The data was obtained from a continuous phase recording using a 5 MHz linear phase comparator and a strip chart recorder. In production tests the frequency standards were placed in a temperature chamber. The temperature was held at -5°C for four hours and +55°C for eight hours with a two to three hour reference at +25°C. The frequency offset at +25°C was subtracted from the offset at the temperature extremes. These residuals are plotted in the histograms. The average frequency offset for the sample at -5°C is -2.9×10^{-12} with a variance of 1×10^{-11}. The average frequency offset at +55°C is $+3.5 \times 10^{-12}$ with a variance of 7×10^{-12}.

The atomic resonator in a rubidium standard is susceptible to changes in the ambient magnetic field. This effect is typically held to less than 6×10^{-12} for a 1 gauss change in ambient field. This limit is achieved by limiting the calibration magnetic field to less than 175 milligauss and by using triple magnetic shields.

The 5065A has passed the MIL-167 vibration test.

Synthesis of the Atomic Resonance Freq.

The control system is similar to the system used in the HP 5060A described by Bagley and Cutler in 1964.[6] The notable differences are: 1) the microwave frequency is 6.8346...GHz; and 2) the synthesizer frequency can be adjusted by four thumbwheel switches.

The microwave frequency is synthesized by multiplying 5 MHz to 6.840 GHz and effectively subtracting the 5.3 MHz synthesizer frequency to produce 6.8346...GHz. The synthesizer block diagram is shown in Figure 4. The 5.3 MHz VCXO is phase locked to a subharmonic of the 5 MHz input signal. The subharmonic frequency is adjusted by thumbwheel settings which control the preset digital divider. Any frequency within the control range of the phase lock loop which satisfies the relation f = 5(M/N) for N <5000 can be synthesized.

The synthesizer has a range of 1×10^{-7} with incremental steps $<2 \times 10^{-9}$. This combined with a magnetic field range of 2×10^{-9} enables the user to easily make precise time scale adjustments.

Frequency Adjustment

A rubidium standard requires initial calibration and may require periodic frequency adjustments to correct for aging. Calibration and time scale adjustments are facilitated by a linearized magnetic field control and the 5.3...MHz synthesizer discussed earlier.

A linearity curve for the magnetic field control is shown in Figure 5. The frequency reference is chosen to correspond to a dial setting of 0.00. The dial was adjusted in ten equal steps from 0.00 to 10.00. The range of frequency offset was adjusted to be 2×10^{-9}. The frequency was measured at each dial setting using the beat frequency described by Cutler and Searle with an HP 106 as a reference oscillator.[3] The data points are the difference between the measured frequency and a calibration factor of 2×10^{-10} per major division. We see that by using the magnetic field as a fine tuning frequency adjustment, the frequency of the standard can be adjusted over a range of 2×10^{-9} with an error of less than 3.5×10^{-11}.

We must emphasize that small adjustments can be made with a precision of better than 10% of the magnitude of the change. For example, a frequency change of 2×10^{-11} would be in error by less than 2×10^{-12}.

Conclusion

The performance characteristics of the 5065A Standard have been presented. The long-term stability of the rubidium standard is not a uniquely defined term but rather involves random variations about a linear drift. The drift rates observed were from $<2 \times 10^{-12}$/month to 6×10^{-12}/month.

The short-term stability of the frequency is more uniquely determined and was shown to be $\approx 3.6 \times 10^{-12}$ for averaging times of 1 second. The specification for the standard is 7×10^{-12} for 1 second averages.

The environmental stability and ease of frequency adjustment make the rubidium standard extremely useful in many timekeeping and radar applications.

Acknowledgements

The author wishes to express his gratitude to B. E. Swartz who designed several of the electronic circuits in the 5065A; to R. A. Baugh who did much of the development work on the rubidium vapor resonator and to Bruce Fowler who contributed to the development work on the atomic resonator.

Appreciation is also expressed to Syble Spencer who reduced the long-term stability data and to Ilhan Gozaydin who did the computer program analysis and contributed to the mechanical design of the 5065A.

References

1. A. R. Chi, J. H. Roeder, S. C. Wardrip and B. Kruger, "Long-term Frequency Stability Measurement of Rubidium Gas Cell Frequency Standards," Proc. 22nd Annual Symposium on Frequency Control, p. 592.

2. A. O. McCoubrey, "A Survey of Atomic Frequency Standards," Proc. IEEE 54, 2, 116 (Feb. 1966).

3. L. S. Cutler and C. L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards," Proc. IEEE 54, 2, 221 (Feb. 1966).

4. D. W. Allan, "Statistics of Atomic Frequency Standards," Proc. IEEE 54, 2, 170 (Feb. 1966).

5. J. A. Barnes, "Tables of Bias Functions, B_1 and B_2, for Variances Based on Finite Samples of Processes with Power Law Spectral Densities," NBS Technical Note #375 (Jan. 1969).

6. A. S. Bagley, L. S. Cutler, "A Modern Solid-State Portable Cesium Beam Frequency Standard," Proc. 18th Annual Symposium on Frequency Control, p. 344.

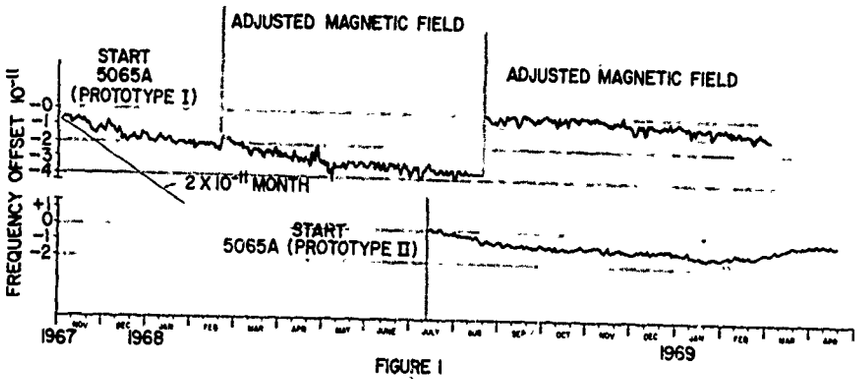


FIGURE I

DAILY AVERAGE FREQUENCY OF
TWO 5065A RUBIDIUM STANDARDS
REFERENCED TO HP HOUSE STANDARD

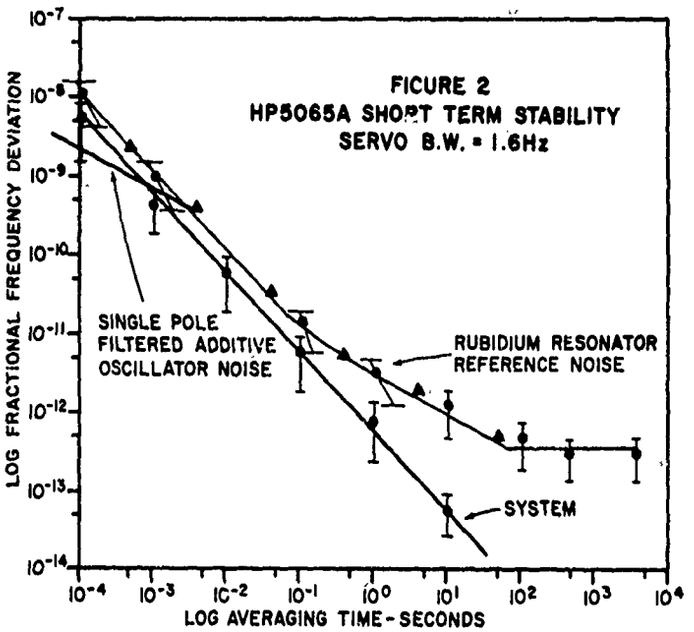
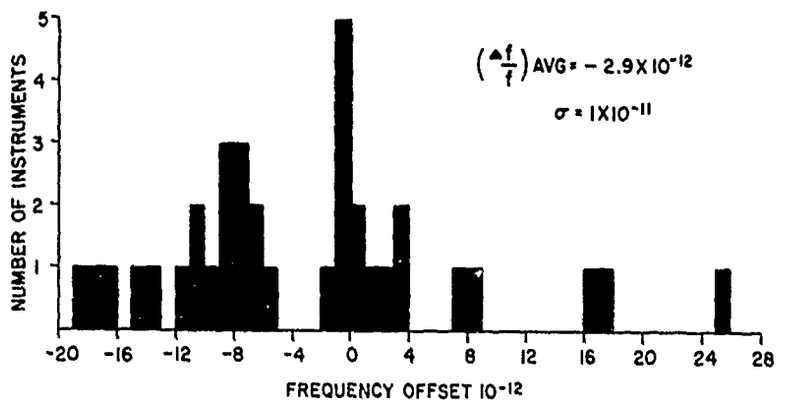


FIGURE 3A

NUMBER OF INSTRUMENTS VS
FREQUENCY OFFSET AT -5° C



IMPROVED STATE SELECTION FOR HYDROGEN MASERS*

Richard F. Lacey and Robert F. C. Vessot
Hewlett-Packard Company
Beverly, Massachusetts

Summary

The performance of the atomic hydrogen maser can be markedly improved by increasing the fraction of hydrogen atoms in the (F=1, m=0) state that enter the maser's storage bulb. In the conventional maser design, the beam of atoms entering the storage bulb consists principally of a mixture of atoms in the desired (F=1, m=0) state and the (F=1, m=1) state. The latter do not contribute to the maser's oscillation but degrade it by line-broadening spin-exchange collisions with the radiating atoms that provide the maser's output signal. Atoms in the unwanted state can be eliminated by the use of an improved beam optics design that uses two state selection magnets in succession, with a space between them. The pair of magnets is arranged so that hydrogen atoms effused from the source (within a limited velocity range and solid angle) are focussed into the entrance aperture of the storage bulb. By inducing a transition in the region between the two magnets that reverses the orientation of each hydrogen atom with respect to the static magnetic field (i.e., $m \rightarrow -m$), atoms originally in the (F=1,m=1) state end up in the (F=1, m= -1) state and are deflected out of the beam by the second magnet.

Calculations of the performance of a beam optics system show that it is possible to have 90 per cent of the atoms focussed into the storage bulb in the desired (F=1, m=0) state and that the fraction of hydrogen atoms in this state effused from the source that reach the storage bulb can be comparable to conventional designs. The best design we found was to have the first hexapole magnet long enough to make the trajectories of atoms with a typical velocity parallel to the axis, and the second long enough to bring these trajectories together inside the magnet, to make a virtual source, with a sufficient length of magnet remaining to focus atoms from this virtual source to the aperture of the storage bulb. It is necessary to have the second magnet very long in order to eliminate atoms that have been transferred into the (F=1, m= -1) state by the transition between the state selection magnets, and desirable to have the first magnet as short as possible to maximize the useful hydrogen atom flux.

A hydrogen maser with an eight-inch storage bulb was modified to use the beam optics system described above. With no transition induced between the magnets, the maser oscillated better than it did before the modification. When adiabatic fast passage transitions were induced, the power output typically increased about 2db, the threshold of oscillation was lowered, and the hydrogen resonance linewidth decreased. By measuring the relative power and relative flux as a function of linewidth it was estimated that the ratio of useful flux to total atomic hydrogen flux into the storage bulb was increased 20 per cent when transitions were induced in the adiabatic fast passage transition region between the state selection magnets.

The improvement, though distinct, is smaller than expected. This might be the result either of a partial scrambling of states in the beam by Majorana transitions or of a background of unpolarized atoms that find their way into the storage bulb.

I. Introduction

From the theory of the hydrogen maser it is apparent that spin exchange collisions are a major limitation on performance.[1,2] Spin exchange collisions decrease the length of time an atom interacts with the oscillating magnetic field in the maser's cavity, and thereby broaden the hydrogen resonance line, increase the flux intensity at which self-sustaining oscillation begins, decrease the power output, and limit the extent to which the beam flux can usefully be increased. In the conventional hydrogen maser beam optics, at least half of the hydrogen atoms focussed into the storage bulb of the maser are in the (F=1, m=1) state and do not contribute to the maser's oscillation, but do undergo spin exchange collisions with atoms interacting with the oscillating magnetic field. Elimination of the atoms in this state could in principle reduce spin exchange collision effects by a factor of two.

The unwanted state can be eliminated from the beam focussed into the storage bulb of a maser by the use of a beam optics system consisting of two hexapole state selection magnets with a region in between them in which atoms in the beam that are in the (1, 1) state undergo a transition to the (1, -1) state. Atoms in the (1, 0) state will remain in that state. Atoms in the (1, -1) state are deflected out of the beam by the first magnet, along with those in the (0, 0) state. The second magnet deflects out those atoms transferred into the (1, -1) state by the transition induced between the two magnets.

The transition may be induced by passing the beam through a coil in which there is a magnetic field oscillating at the resonance frequency of the transition. The resonance frequency depends on the strength of the applied magnetic field: f= 1.4 MHz/Gauss. This Rabi resonance method requires that the applied magnetic field be reasonably homogeneous over the coil and constant with time, and that the velocity distribution in the atomic beam be relatively narrow. Such requirements can be considerably relaxed if the method of adiabatic fast passage is used. This method was first used on hydrogen masers by Audoin et al,[3] and consists of applying an oscillating magnetic field in a region where the applied d.c. magnetic field varies linearly

* This work was supported by Goddard Space Flight Center of NASA.

with distance as the beam of atoms passes through. The magnitude of the d.c. field and the frequency of the oscillating field are adjusted so that resonance occurs in the middle of the interaction region. As we will discuss below, the transition probabilities between the states of the F=1 level become relatively independent of the exact power level of the oscillating field, and insensitive to small changes in its frequency.

II. Improvement of Maser Operation

If quenching due to magnetic inhomogeneity over the volume of the storage bulb of a hydrogen maser can be neglected, the mechanisms that contribute to the hydrogen resonance linewidth are spin exchange collisions, escape of atoms from the storage bulb, and loss of atoms due to recombination on the wall of the storage bulb. Experimentally, the last two terms cannot be distinguished, and we lump their effects on the linewidth together. We have

$$\Delta\nu = \Delta\nu_{min} + \Delta\nu_{s.e.}$$

where $\Delta\nu$ is the hydrogen resonance linewidth, $\Delta\nu_{min}$ is the portion of the linewidth due to escape from the storage bulb and recombination; i.e., the linewidth limit as the density of hydrogen atoms approaches zero. $\Delta\nu_{s.e.}$ is the portion of the linewidth due to spin exchange collisions between hydrogen atoms and is proportional to atomic hydrogen density in the storage bulb, or to the total flux of atoms entering the storage bulb:

$$\Delta\nu_{s.e.} = \frac{\sigma\bar{v}I_{tot}}{2\pi V_b \gamma} = \Delta\nu - \Delta\nu_{min}$$

where σ = spin exchange collision cross section,
$\bar{v}$ = mean relative velocity between H atoms,
V_b = volume of the storage bulb,
γ = reciprocal of mean lifetime of hydrogen atoms in storage bulb,
I_{tot} = total flux into the storage bulb.

The total power radiated by the atoms in the storage bulb in the absence of magnetic quenching can be expressed as:

$$P = \frac{1}{2}h\nu\left[I - \frac{hV_c}{8\mu_0^2\eta Q}\Delta\nu(2\Delta\nu - \Delta\nu_{min})\right]$$

where I = net useful flux,
V_c = volume of the cavity,
μ_0 = Bohr magneton,
η = "filling factor" (reference 1),
Q = loaded cavity quality factor,
h = Planck's constant,
ν = radiated frequency.

The width of the hydrogen resonance line is found by measuring the shift in the radiated frequency as the cavity is detuned:

$$\nu - \nu_0 = \Delta\nu Q/\nu_0$$

where ν_0 is the exact resonance frequency, as modified by the wall shift, second-order Doppler effect, and magnetic field correction. $\Delta\nu_{min}$ can be obtained by extrapolation of the linewidth as a function of flux to zero flux.

Reducing the linewidth $\Delta\nu$ while maintaining I will obviously result in a marked increase in power output. In addition, reducing the linewidth lowers the value of flux required for the maser to oscillate at all (P>0). Finally, the theoretical limit to the frequency stability of a maser depends on linewidth and power:

$$\sigma_f/\nu = \Delta\nu/\sqrt{2}\nu)\ (kT/Pt)^{1/2}$$

where σ_f is the rms deviation in output frequency,
k is Boltzmann's constant,
T is the absolute temperature,
t is the averaging time interval.

As a practical matter, however, this limit is never reached because the instability due to additive noise in the external circuitry that receives the maser's signal, and long term drifts due to temperature fluctuations and the like, drown it out.

The power delivered by the maser beam is usually expressed in terms of a parameter q:

$$q = \frac{\sigma\bar{v}_{rel}}{8\pi\mu_0^2 Q}\ \frac{V_c}{\eta V_b}\ \frac{I_{tot}}{I}.$$

q can be expressed in terms of the linewidth also, making use of the expressions given above. It is difficult to measure power or flux absolutely, but the value of I for which the power is zero can be found in terms of the linewidth by extrapolating the curve of power as a function of linewidth to zero power, $\Delta\nu'$. When P=0 we have

$$I = \frac{hV_c}{8\mu_0^2\eta Q}\ \Delta\nu'(2\Delta\nu' - \Delta\nu_{min}).$$

If we assume that the ratio I_{tot}/I is a constant for all flux levels, then

$$q = \frac{\Delta\nu_{min}(\Delta\nu' - \Delta\nu_{min})}{\Delta\nu'(2\Delta\nu' - \Delta\nu_{min})}$$

The smaller the value of q, the better the performance of the maser. Note that the value of q can be obtained experimentally independent of any knowledge of cavity Q, coupling coefficient, filling factor, collision cross section, or cavity and bulb volume.

To measure the effectiveness of using adiabatic fast passage, the ratio of the values of q without and with it can be made. This will give the improvement in the ratio of the useful flux I to the total flux, I_{tot}. Ideally, the ratio of the two values of q should be 2:1.

III. Beam Optics

The design of the proper beam optics for use with an adiabatic fast passage transition to eliminate unwanted states from the beam was the subject of extensive study. The design criteria were to focus (F=1, m=0) state atoms in a range about the most probable velocity into the storage bulb, and to minimize the flux of atoms in other states, in particular due to atoms transferred into the (F=1, m=-1) state by the adiabatic transition. As shown in Figure 1, the layout consists of a source of hydrogen atoms which are emitted through a collimator placed on the axis at one end of a hexapole magnet. Atoms in the (F=1, m=1) and the (F=1, m=0) states are deflected toward the axis, while those in the other two states are deflected away. The beam then passes through a transition region in which the adiabatic fast passage transition may be induced.

For design purposes, the length of this region was fixed at four inches. The length of the first magnet can be set so that the trajectories of atoms originating on the axis of the system that have the most probable velocity ($\sqrt{3kT/m}$) will be parallel to the axis at the exit of the magnet. Another possibility is to make the magnet longer so that atoms with this velocity will be brought to a focus at the entrance of the second magnet. Calculations showed this latter case to be less efficient than the former in an over-all system designed to produce a focus at the entrance to the storage bulb.

The length of the second magnet should be designed to focus atoms with the optimum velocity, given the trajectories with which they enter it, at the entrance to the storage bulb. If the typical entering trajectory is parallel to the axis, then the second magnet need have a smaller length to gap diameter ratio than the first magnet because the amount of deflection the atoms require is relatively slight. However, atoms transferred into the (F=1, m= -1) state by the adiabatic fast passage transition will also receive only a small deflection, and if the distance from magnet to bulb aperture is relatively short, some of these atoms may not be deflected from the beam sufficiently far to miss the bulb aperture.

The best way to overcome this problem seemed to be to lengthen the second magnet so that the typical trajectory has the form shown in the middle illustration of Figure 1. Calculations showed that the overall efficiency of such a system could be comparable to a simple single magnet system. Somewhat lower velocities also contribute, following a trajectory whose typical shape is shown in the bottom illustration of Fig. 1.

Beam optics calculations were made in two different ways. In the first, the assumption was made that all atoms originated exactly on the axis of symmetry of the magnets. The exact force law on the atoms in the hexapole magnets was used to compute the trajectories. The intensity reaching the storage bulb was obtained by integrating over the range of initial slopes of the trajectories and over the range of velocities emitted by the source. In the second, the assumption of a linear force law (valid for m=±1 states) was made, only the entrance and exit planes of the magnets and the entrance to the storage bulb were considered to have limiting apertures, and the intensity reaching the storage bulb was obtained by integrating successively over the two components of initial slope for a source point a given distance from the axis, over the area of the sources, and over the velocity distribution. The agreement between calculations based on the two approximations was fairly close for source radii small compared to the radius of the gap of the first magnet. The disagreement is greatest for long distances between the second magnet and storage bulb, since the effect of off-axis source points increases with the length. On the other hand, the limited number of apertures at which atomic trajectories can be terminated, and the linear force law assumed, makes intensities calculated by the second method greater when source and magnet-to-storage bulb dimensions are small.

IV. Adiabatic Fast Passage

Adiabatic fast passage refers to a system in which the population distribution of a system is reversed (i.e., m→ -m) by sweeping the frequency of a perturbation that can induce transitions through resonance at a rate fast compared to the natural lifetime of the states, but slow compared to $H_{mn}/\hbar$, where H_{mn} is the magnitude of the perturbation connecting two states. In a hydrogen atomic beam this is done by passing through a coil in which an oscillating magnetic field is induced. The coil lies between the poles of a tapered magnet, so that the atom passing through the coil sees a uniformly varying magnetic field. The frequency of oscillating field is adjusted so that resonance for the transition between the magnetic sublevels of the F=1 level occurs at the middle of the coil.

The transition probability for a spin ½ particle can be calculated readily by approximating the continuously varying magnetic field by a stepped one, divided into many regions in which the magnetic field is constant. If the transition probability for a spin ½ particle is known, those for one of any degree of magnetic degeneracy can be calculated by use of formulas due to Majorana that give the transition probability between any two magnetic sublevels in terms of that for a spin ½ particle.[4]

The calculations show that the transition probability becomes essentially constant if the total interaction time is long enough. The calculation assumes that the perturbing oscillating field can be decomposed into two contra-rotating components, one of which is ignored, and that the interaction with the perturbing field begins and ends suddenly.

The adiabatic fast passage transition region was built with a tapered magnet 3.25 inches long in which was a coil 2 inches long to produce the oscillating field. The taper was such that the field varied 8 per cent of its mean value over the length of the coil, and it generally was set to 20 gauss in the middle.

V. Experiment and Results

The two hexapole state selection magnets had effective lengths of 3.734 inches and 7.750 inches, and the peak strength of their fields was 6500 and 7040 gauss, respectively. Their bores were 1/8 inch. The maser had an eight-inch diameter storage bulb, and four sets of magnetic shielding. Its fine tuning was done mechanically with a plunger inserted into the cavity. The distance from the second magnet to the storage bulb aperture was about 20 inches.

The experimental data was taken by measuring the beat between the experimental maser and a reference as a function of maser cavity resonance frequency. For a given beam intensity and composition, the maser frequency is a linear function of cavity resonance frequency, and the slope of the line is proportional to linewidth. The slopes and estimates of the errors to the slopes were obtained by standard techniques of regression analysis. The relative power and source pressure, assumed proportional to hydrogen beam flux, were plotted as a function of this slope for a range of source pressures, and with and without the adiabatic fast passage resonance operating. The slopes for zero power out and for zero beam flux were obtained by extrapolation, as shown in Figures 2 and 3.

By calculating the value of q with and without the enhancement of the adiabatic fast passage transition and taking their ratio, the improvement in the ratio of useful flux to total flux with the enhancement was found to be approximately 20 per cent. The power output was increased, typically, by 2db.

The effect of the adiabatic fast passage transition on linewidth and maser output power was found to be remarkably insensitive to the frequency and power level applied to the coil, as long as the power was above a minimum value at which quenching of the maser resonance occurred, and the applied frequency was close to the hydrogen resonance frequency for $\Delta F = 0$, $\Delta m = \pm 1$ transitions at the middle of the coil.

The improvement in the operation of the maser using the adiabatic fast passage enhancement was marked, and very useful in operating the maser, particularly at low flux levels near the threshold of oscillation.

The degree of improvement is smaller than expected. There are several possible explanations. One is that Majorana transitions are occurring between the state selection magnets and the storage bulb aperture that scramble the state populations. Also, magnetic inhomogeneities in the cavity might be causing similar effects. There may be a flux of unpolarized atoms finding its way into the storage bulb. The beam optics of the adiabatic fast passage resonance may not be as efficient as our calculations show. However, a beam optics system far more complicated than the conventional one did work, and the operation of the maser was greatly improved by its use.

References

1. D. Kleppner, H.M. Goldenberg, and N.F. Ramsey, Phys. Rev. 126, p. 603 (1962).
2. D. Kleppner, H.C. Berg, S.B. Crampton, and N.F. Ramsey; and R.F.C. Vessot, H.E. Peters, and J. Vanier, Phys. Rev. 138, p. A972 (1965).
3. C. Audoin, M. Desaintfuscien, P. Petit, and J.-P. Schermann, IEEE Trans. on Instrumentation and Measurement IM-17, p. 351 (1968).
4. E.P. Wigner, Group Theory, New York:Academic Press (1959).

Acknowledgment

We wish to thank Dr. Fouad Major of NASA for his support of the research on which this paper was based, Gary Seavey for his help in measuring the field strengths of hexapole magnets, Hans Van Heyst for the mechanical design of beam optics system, and Louis Mueller for his aid in assembling the modified maser.

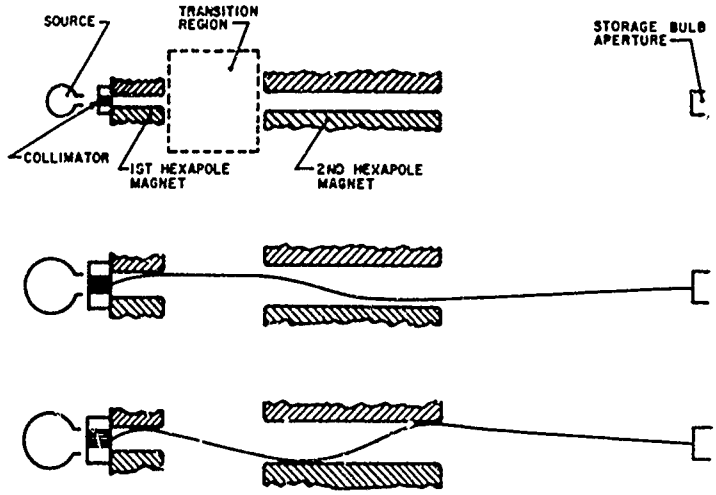


FIGURE 1

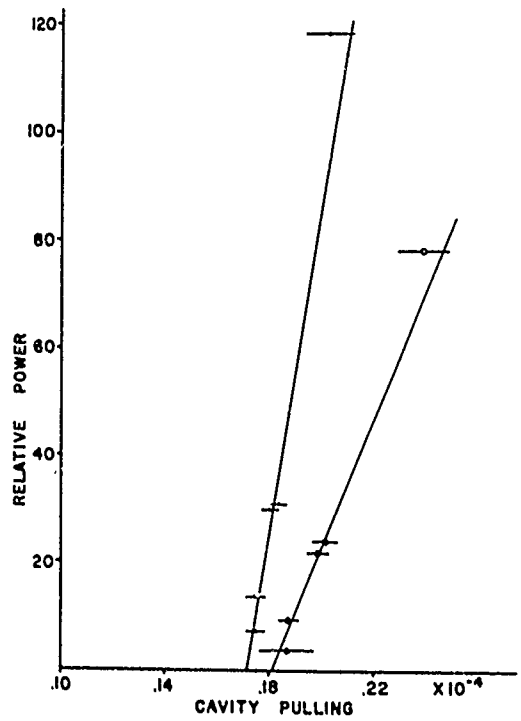
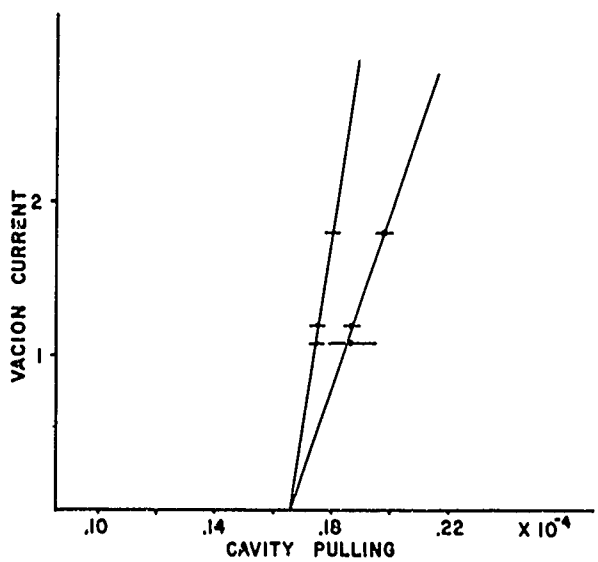


Figure 2



FIGURE

RECENT RESULTS CONCERNING THE HYDROGEN MASER WALL SHIFT PROBLEM*

E. E. Uzgiris and P. W. Zitzewitz†

Harvard University

Lyman Laboratory of Physics

Cambridge, Massachusetts 02138

Summary

A series of measurements to determine the absolute wall shift of FEP Teflon in the hydrogen maser yields a result which is approximately 25% smaller than determined by previous measurements. The absolute effect of the wall on the maser frequency for a 6 inch diameter spherical bulb at 40°C. was found to be -24.7 ± .5 mHz. These measurements allowed the evaluation of the Large Storage Box Maser Teflon surface, and it was found in this case that the absolute wall effect was - 2.4 mHz. Therefore the effect is reduced in magnitude by the expected amount. In addition, the temperature dependence of the wall shift was studied. Confirming earlier measurements, a large decrease was found at higher temperatures (approximately 80°C.). A value for the temperature dependence in the region of measurements is presented. Finally, several experiments to be undertaken in this laboratory which deal with the wall shift problem are described.

Introduction

Several workers have studied the wall effect and have determined the absolute frequency shift for a wall of FEP Teflon.[1,2] Although these experiments were done in 1964, there did not appear any conflicting data[3] until absolute wall measurements were undertaken in this laboratory last fall. The project was begun because knowledge of the wall shift was necessary to evaluate the large hydrogen maser[4] (by E.E.U.) and to study temperature characteristics of various walls in detail (by P.W.Z.). Initial measurements appeared anomalously small compared to existing data, and a full scale effort was undertaken to evaluate the absolute wall shift of FEP Teflon.

The procedure involved measuring the relative frequency offset of a test maser from a reference maser. The reference maser was undisturbed throughout the course of the experiment and the measurements were repeated for different coatings and different bulb sizes. In this way a dependence of the relative wall effect on bulb size was generated from which the absolute wall shift can be inferred. A total of 26 separate Teflon coatings were measured and 19 were used in the final determination. The result, which is smaller than previous determinations[1,2] by approximately 25%, is most likely due to a change in the composition of the FEP Teflon.

Experimental Results

Fig. 1 is a graph of the results with the relative wall shift shown versus the inverse bulb diameter.

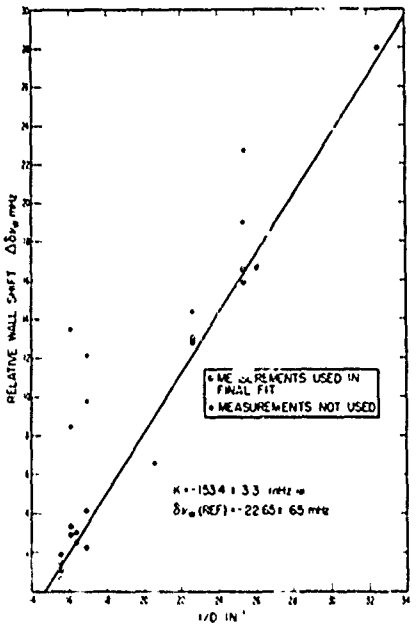


Fig. 1
Relative Wall Shift Versus Inverse Bulb Diameter

* Supported by the National Science Foundation, the Office of Naval Research, and the National Aeronautics and Space Administration.

† Danforth Graduate Fellow.

This dependence can be represented by

$$\Delta\delta\nu_w = K/d - \delta\nu_w \text{ (ref.)}$$

where $\Delta\delta\nu_w$ is the relative measured frequency shift, $\delta\nu_w$ (ref.) is the absolute wall shift of the reference maser, d is the bulb diameter, and K is the slope of the line. All the bulbs are spherical except for the smallest, which is a cylindrical bulb with a mean free path[5] corresponding to that of a 3.07" sphere. Error bars for each measurement are omitted for clarity. For the most part, the errors, the standard deviation of all measurements of a particular coating, range between .3 mHz and 1 mHz, but two poorly measured bulbs were uncertain by 3 mHz. All 28 points were fit at first. Only those which had an overlap of the 95% confidence limits of the line and the point were retained for the final fit. Points were statistically weighted in the least squares fit. The errors quoted for the parameters of the line are standard deviations of the mean.

The scatter in the points about the line cannot be explained fully by magnetic field instability, cavity drifts, electronic noise, and deficiencies in measurement technique, and must be attributed to variations in the characteristics of the Teflon. Of the ten bulbs not used in the final fit, most did not pass the water drop test,[6] at least under the scrutiny we had developed by the end of the experiment. One bulb that did not pass and had an estimated area of .01 cm^2 where water would wet the surface had a measured wall shift 5% higher than the norm. Several successive coatings were usually required before the bulb passed. The coating procedure itself was standard:[6] a 380°C. fusing oven bake for 20 minutes with circulating air which was dried with a dry-ice trap, and a slow cooling rate (several hours).

There was no detectable difference between test masers pumped by mercury vapor pumps and those with sputter-ion pumps. There was no evidence of a drift in the frequency of the reference maser.

We thus conclude that one can coat a bulb in a standard way, test the surface with a water drop, and find scatter in the wall shift of a few parts in 10^{13}.

Besides corrections for magnetic field offset, cavity and spin exchange shifts and the second order doppler shift, corrections were made for the temperature variation of the wall shift and the effects of the bulb stem on the average mean free path.

Fig. 2 shows the temperature variation of the wall shift for two bulbs in the 23-50°C. range.

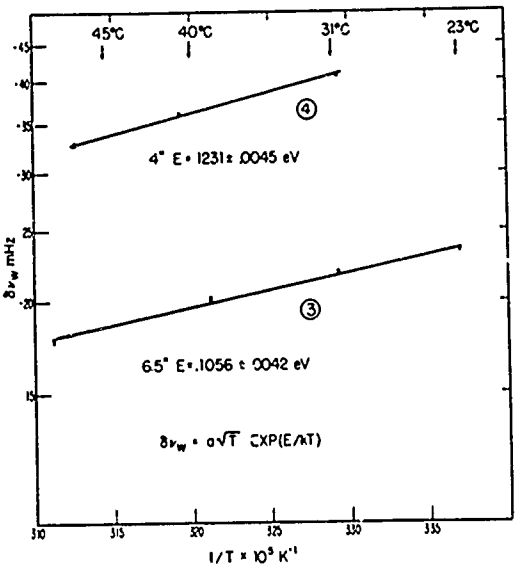


Fig. 2
Wall Shift Versus Inverse Temperature in 23°C to 50°C range

The 6 1/2" bulb has a smaller than average wall shift and has a smaller temperature dependence. The 4" bulb has a wall shift 5% larger than normal and has a larger variation. The implications of this are being investigated. The fit to a theoretical curve given by a simple physical adsorption model is quite good. The linearized shift in the region about 31.5°C. where our experiment was done is

$$\alpha \equiv \frac{\partial\delta\nu_w}{\partial T} / \delta\nu_w = -(12.7 \pm 1.3) \times 10^{-3} \text{ deg}^{-1}.$$

Most runs in the experiment were within 2° of this, but some, especially earlier ones, were 9° below, at room temperature.

The stem effect is the increase of the measured wall shift caused by the larger number of collisions incurred by those atoms which enter the bulb stem and later return to the bulb. We have estimated the effect and have subtracted from the measured values corrections ranging from .03% to 2% depending on the particular bulb used. As the effect is proportional to the ratio of stem to bulb volume, it is larger for the smaller bulbs. The present estimate of the stem

correction, details of which will be published later, could be in error by 25%. Such an error has been allowed. However, this correction and associated uncertainty does not change the line slope beyond the errors of the slope.

The final results are summarized in Table I.

COMPARISON OF WALL SHIFT RESULTS

(1) $\delta\nu_w = \frac{K}{d}[1+\alpha T_i(T-T_i)]$ (3) $\delta\nu_w(T) = \alpha\sqrt{T}\exp\left(\frac{E_{ads}}{kT}\right)$

(2) $\phi = \frac{4\pi}{3v}K$ (4) $\alpha_{T_i} = \frac{\partial\delta\nu_w}{\partial T}\Big/\delta\nu_w\Big|_{T=T_i} = -\frac{E_{ads}}{kT_i^2} + \frac{1}{2T_i}$

ALL REFERENCED TO 40°C

	K(mHz-in)	$\delta\nu_w(6")$(mHz)	ϕ(rad)	E_{ads} (eV)	α (K^{-1})
VANIER-VESSOT 1964	-208±2	-34.7±.3	$-(8.7\pm.1)\times10^{-8}$	.065±.02	$-(5\pm1)\times10^{-3}$
MATHUR 1964	-198±30	-33.0±5	$-(8.3\pm1.3)\times10^{-8}$	—	—
UZGIRIS-ZITZEWITZ 1969	-151.8±3.2	-24.7±.5	$-(6.3\pm.1)\times10^{-8}$	.114±.010	$-(12.0\pm1)\times10^{-3}$

Table I
Summary of Data

Eq. (1) defines the line slope K; Eq. (2) the phase shift per collision ϕ. The interpretation of the results of the temperature variation depends on the model used. Eq. (3) defines the adsorption energy, E_{ads}, for our model, and Eq. (4) shows the method used to determine the linearized shift α. The change in K and α from the results in Fig. 1 and the text is due to the use of the temperature reference point of 40°C. rather than 31.5°C.

It should be noted that the Teflon used in the measurements by Mathur was DuPont FEP 120, Blend 33,53.9% solids, purchased in 1964. The Teflon used in the present work is Blend B-107,54.5% Solids, purchased in 1967. It has been noted that the newer Teflon seems to coat differently from the older variety[7] and this may be due to differences in the physical characteristics. If this is in fact the case -- although it must be emphasized the evidence is by no means certain -- then the difference in wall shift between these measurements and those on the older Teflon in 1964 would not be surprising.

Other Work

Among other work under way at Harvard, Fig. 3 shows the present status of work on temperature dependence.

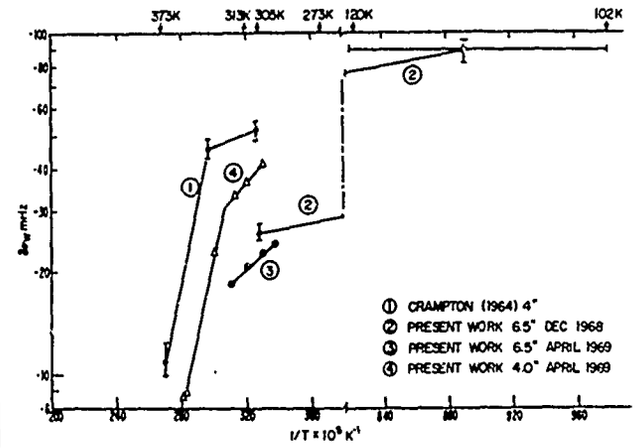


Fig. 3
Wall Shift Versus Inverse Temperature in 100K to 400K range

The 1964 work by Crampton[8] used bulbs coated by Mathur.[1] The two room temperature runs are those described above. The break in the line at higher temperatures is most exciting and measurements are being extended to higher temperatures. Comparison of this data with the low temperature run suggest the existence of a similar break below room temperature, possibly correlated with the observed phase change in crystalline TFE Teflon at 19°C.[9].

We have long been interested in developing a wall surface made of a material that has a high enough vapor pressure so that the surface would be continuously renewed. This would afford a surface independent of the impurities in the vacuum system and would hopefully lead to more reproducible wall shifts. Xenon at 50°K is an excellent candidate. Another is Fluorolube, a perfluoronated lubricating oil with a mean molecular formula of $C_{21}F_{44}$, a product of the Oak Ridge Gaseous Diffusion Plant. Although the vapor pressure at room temperature is too large for practical use, we have measured the wall shifts of two bulbs at two temperatures. The results are as follows:

Bulb Size	Wall Shift (mHz)	Temp.	Vapor Pressure (Torr)
6 1/2"	- 27.5	23°C.	2×10^{-6}
4"	- 43.0		
6 1/2"	- 78.2	32°C.	1×10^{-5}
4"	- 62		

The large increase in wall shift is correlated with the vapor pressure rise and thus it seems

possible that collisions in the vapor phase are responsible for the change. Fractions of this oil with a larger molecular weight and thus lower vapor pressure are being obtained.

A flexible bottle to measure the absolute wall shift without changing bulbs has been proposed and is being developed by Douglas Brenner. This method of varying the mean free path while maintaining the same bulb surface may prove advantageous and yield accuracies higher than heretofore achieved.

The wall shift for the Large Storage Box Hydrogen Maser[4] has been measured with respect to our reference maser. With all numbers referenced to 27°C., the absolute shift is -2.4 ± 0.8 mHz. The larger error comes from the small number of comparisons and the uncertainty in the reference. Since the extrapolation of the wall shift of Teflon used in smaller masers to a bulb with a 58.3" diameter is -2.8 mHz, it can be seen that the purpose of the larger maser, obtaining a wall shift lower by a factor of ten by increasing the bulb size, has been achieved. In order to realize the accuracy potentially obtainable by such a maser, a technique to vary the mean free path in this bulb (similar to the flexible bulb approach above) is being developed. This technique would be unlimited by the small maser wall uncertainties. Full details of this experiment will be published later.

Acknowledgements

The authors wish to thank Professor N. F. Ramsey for support and encouragement, Professor Stuart Crampton for many useful discussions and helpful criticism, and Sunchee Chou for help in coating the bulbs.

References

1. B. S. Mathur, S. B. Crampton, D. Kleppner and N. F. Ramsey, Phys. Rev. 158, 14 (1967).

2. R. Vessot, et al, I.E.E.E. IM-15, 165 (1966).

3. A. O. McCoubrey, Proc. of the I.E.E.E. 55, 805 (1967).

4. E. E. Uzgiris and N. F. Ramsey, I.E.E.E. J. of Quan. Elec. QE-4, 563 (1968), Proc. of the 22nd Freq. Cont. Symp. 452 (1968).

5. D. Kleppner, et al, Phys. Rev. 138, A972 (1965). The formula as given here in the appendix is incorrect. The correct mean free path is $\lambda = 2R/(1+R/L)$.

6. H. C. Berg and D. Kleppner, Rev. Sci. Instr. 33, 248 (1962).

7. R. F. C. Vessot, Private Communication.

8. S. Crampton, Private Communication.

9. C. A. Sperati, H. W. Starkweather, Jr., Fort. Hochpolymeren-Forsch., 2, 465 (1961).

A NEW METHOD FOR MEASUREMENT OF THE POPULATION DIFFERENCE OF HYPERFINE-LEVELS OF STORED ATOMS

C. Audoin*, M. Desaintfuscien*, P. Piéjus, J-P. Schermann*
Institut d'Electronique Fondamentale
Laboratoire associé au C.N.R.S.
Bâtiment 220, Faculté des Sciences
91-Orsay - France

Summary

We describe a double resonance experiment for measuring the atomic beam composition of a hydrogen maser. This method requires only one single static magnetic field.

An audio frequency field whose frequency is close to the frequency spacing between the Zeeman sublevels of the F=1 state is applied to the stored atoms. The hyperfine $\Delta m = 0$ transition line is then split into a triplet (Autler-Townes effect), the component of which are very well separated owing to the high Q factor associated with the stored hydrogen atoms.

We show that the intensities of the triplet lines are related to the differences of population of the three F = 1 sublevels on the one hand and the F = 0 level on the other hand.

The method is used to measure the efficiency of an improved state selector which eliminates 85% of atoms in the undesirable F = 1, m = 1 state.

1. Introduction

A single hexapolar magnet, used as a state selector of hydrogen atoms, focuses into the storage bulb of a hydrogen maser[1] the atoms in both energy levels F = 1, m = 1 and F = 1, m = 0. Only atoms in the F = 1, m = 0 level are used for maser operation with the F = 1, m = 0 $\longrightarrow$ F = 0, m = 0 transition. Atoms in F = 1, m = 1 level, which do not take part in the Δ m = 0 clock transition, increase the atomic density in the bulb and the atomic linewidth by means of the spin exchange process.

It is of primary interest to operate the H-maser with an atomic linewidth a small as possible in order to obtain a very good long term stability and to make easier the automatic tuning of the cavity So, it is important to improve the state selection to eliminate as much as possible the atoms in the undesirable F = 1, m = 1 level.

For this purpose, we have proposed and realized[3] a state selector, for H masers[2], including two hexapolar magnets separated by an adiabatic rapid passage (ARP) region, whose principle has been given by Abragam and Winter [4].

To analyse the operating conditions of this device and to measure its efficiency, we have considered the means of determinating the population difference between each of the three F = 1 sublevels on the one hand and the F = 0 level on the other hand.

The known possibilities are as follows

1. Measurement of the spin exchange contribution, $T_{2\,\text{spin exchange}}$, to the T_2 relaxation time. It is related to the atomic density in the bulb, n, by the formula :

$$\left(\frac{1}{T_2}\right)_{\text{spin exchange}} = \frac{1}{2}\, n \sigma \bar{v}_r \qquad (1)$$

where σ is the spin exchange cross section and $\bar{v}_r$ the relative mean velocity of hydrogen atoms.

This measurement gives the total atomic density in the bulb, and, then, the atomic flux of the beam, but not its detailed composition.

2. By operating the maser as an amplifier, it is, in principle, possible to measure its gain for the F = 1, m = 0 $\longrightarrow$ F = 0, m = 0

---

* Section d'Orsay du Laboratoire de l'Horloge Atomique

transition and for the F = 1, m = ± 1 → F=0, m=0 transitions and to compare the involved population differences.

Observation of the Δm=0 transition requires a static magnetic field parallel to the R.F. field and observation of the Δm = ± 1 transitions a static magnetic field perpendicular to the R.F. field. In a H-maser, the direction of the R.F. field is imposed and, for the comparison, it is necessary to produce two perpendicular static fields. But, for practical reasons, one of them cannot be made sufficiently homogeneous to avoid uncoherent mixing of the F=1 sublevels. So, this method fails.

These difficulties are overcome by a new method, based on the Autler-Townes effect [5]. It requires the single static magnetic field which is used for the Δm=0 clock transition, and an audio frequency field, whose homogeneity is not critical, to produce a coherent mixture of the sublevels of the F=1 state.

2. Energy levels of hydrogen atoms submitted to an A.F. field

At first, we consider the energy levels of hydrogen atoms in presence of an A.F. magnetic field whose frequency is close to the spacing (expressed in frequency unit) of the F = 1 Zeeman sublevels.

The high frequency field inducing the Δm=0 hyperfine transition will interact with the whole system atoms + A.F. field. But due to the great frequency difference between the A.F. field and the H.F. field (a few kHz compared to 1,42 GHz), the F=0 level is not involved in the described state mixing. It is neglected for the moment. It will be reintroduced, without further complication, when we will consider the detection of state mixing and the measurement of the differences of population.

2.1 - Energy levels of the atoms

In the static magnetic field H_0 of about 1 milligauss applied to the hydrogen atoms in the storage bulb of a H maser, the three sublevels of the F=1 state can be considered as equidistant, and they can be described as the three sublevels of a spin 1 (figure 1A). Their respective energies are :

$$+\hbar\omega_0,\ 0,\ -\hbar\omega_0, \text{ with } \omega_0 = \frac{1}{2}\gamma_e H_0$$

where γ_e is the gyromagnetic ratio of the electron.

2.2 - Energy levels of the A.F. field

The energies of the eigen-states $|n\rangle$ of the A.F. field of frequency ω_x are $n\hbar\omega_x$, where n is the number of photons.

2.3 - Energy levels of atoms and A.F. field without interaction

The eigen-states of the whole system atoms + A.F. field, supposed at first without interaction are called $|F,m\ ;\ n\rangle$. Their energies are obtained by adding the energies of the atoms and of the A.F. field.

For example, on the energy diagram shown in figure 1B, the state $|a\rangle = |1,1\ ;\ n-1\rangle$ is the tensorial product of the atomic state F = 1, m = 1 and of the A.F. state $|n-1\rangle$. The states $|b\rangle = |1,\ 0\ ;n\rangle$ and $|c\rangle = |1,\ -1\ ;\ n+1\rangle$ are also states of the whole system, without interaction.

The energies of these states are :

$$\begin{cases} \mathcal{E}_a = \hbar\omega_0 + (m-1)\hbar\omega_x \\ \mathcal{E}_b = n\hbar\omega_x \\ \mathcal{E}_c = -\hbar\omega_0 + (n+1)\hbar\omega_x \end{cases} \qquad (2)$$

We have to consider states, the m quantum number as also the number of photons of which differs by 1. A level crossing of such states is shown in figure 1B.

For the resonance condition $\omega_x = \omega_0$, the states $|a\rangle$, $|b\rangle$, $|c\rangle$ are degenerated (level crossing).

2.4 - Energy levels of interacting atoms and A.F. field

The interaction between the atoms and the A.F. field is now considered. The direction of this field is perpendicular to the static magnetic field H_0, and ΔF=0, Δm = ± 1 transitions are induced. H_x is its amplitude, averaged by the atomic motion if the A.F. field is not uniform.

It can be shown that the Hamiltonian describing the interaction is [6] :

$$\mathcal{H}_{int} = \frac{\gamma_e}{2}\ \frac{H_x}{\sqrt{n}}\ S_x\ (a + a^+) \qquad (3)$$

where a and a^+ are respectively annihilation and creation operators of a photon of energy $\hbar\omega_x$. S_x is the component of the spin 1 parallel to H_x.

A convenient way to calculate the perturbed energy levels and the perturbed states of the system is to use the so called "resolvant method" [6].

As a result of the quantum mechanical calculation [7], the states $|\varphi_+>$, $|\varphi_o>$, $|\varphi_->$ of the perturbed system which are a coherent mixture of the unperturbed states $|a>$, $|b>$ and $|c>$ are given by :

$$\begin{vmatrix} |\varphi_+> \\ |\varphi_o> \\ |\varphi_-> \end{vmatrix} = \begin{vmatrix} \frac{1-u}{2} & \sqrt{\frac{1-u^2}{2}} & \frac{1+u}{2} \\ \sqrt{\frac{1-u^2}{2}} & u & -\sqrt{\frac{1-u^2}{2}} \\ \frac{1+u}{2} & -\sqrt{\frac{1-u^2}{2}} & \frac{1-u}{2} \end{vmatrix} \begin{vmatrix} |a> \\ |b> \\ |c> \end{vmatrix} \quad (4)$$

The dimensionless parameter u equals :

$$u = \frac{\omega_x - \omega_o}{\sqrt{(\omega_x - \omega_o)^2 + \left(\frac{\gamma_e}{2} H_x\right)^2}} \quad (5)$$

The eigen-energies of the system are :

$$\begin{cases} E_+ = E_b + \hbar\sqrt{(\omega_x - \omega_o)^2 + \left(\frac{\gamma_e}{2} H_x\right)^2} \\ E_o = E_b \\ E_- = E_b - \hbar\sqrt{(\omega_x - \omega_o)^2 + \left(\frac{\gamma_e}{2} H_x\right)^2} \end{cases} \quad (6)$$

The perturbed energy levels are shown in figure 1C. The horizontal level is not modified by the interaction and, as it is classical, the degeneracy for $\omega_x = \omega_o$ is removed by the interaction.

3. Population of the states $|\varphi_+>$, $|\varphi_o>$, $|\varphi_->$

The population of the states $|\varphi_+>$, $|\varphi_o>$, $|\varphi->$ are related to those of the states $|a>$, $|b>$, $|c>$.

The relation depends upon the length of time, τ, necessary to apply the perturbation due to the A.F. field.

The perturbation is applied adiabatically or suddenly whether the frequency spacing $\omega_o/2\pi$ between the F = 1 sublevels is very much larger or smaller than τ^{-1}.

In the adiabatic process, the intrinsic properties of each quantum states do not change. Then the population of the states $|\varphi_+>$, $|\varphi_o>$, $|\varphi_->$ are respectively equal to the population of the states $|a>$, $|b>$, $|c>$.

When the atoms of the beam enter very rapidly the A.F. field region, like in a H-maser, the perturbation is suddenly applied. In this case, the dynamical state of the atoms does not change. The state $|a>$, for example, is not affected. But, the systems must be described as a coherent mixture of states $|a>$, $|b>$, $|c>$ when the atoms are submitted to the perturbation. So, the contribution of each state $|a>$, $|b>$, $|c>$ to the population of states $|\varphi_+>$, $|\varphi_o>$, $|\varphi_->$ equals the square of its coefficient in the equation (4). For example, the number of atoms in state $|\varphi_+>$ which were in state $|a>$ before the perturbation is proportional to $\left(\frac{1-u}{2}\right)^2$.

If we call $P_{1,1}$, $P_{1,o}$, $P_{1,-1}$ the respective populations of the atomic states $|1,1>$, $|1,o>$ and $|1,-1>$ when the A.F. field is not present the populations P_+, P_o, P_- of the states $|\varphi_+>$, $|\varphi_o>$, $|\varphi_->$, are given by

$$\begin{vmatrix} P_+ \\ P_o \\ P_- \end{vmatrix} = \begin{vmatrix} \left(\frac{1-u}{2}\right)^2 & \frac{1-u^2}{2} & \left(\frac{1+u}{2}\right)^2 \\ \frac{1-u^2}{2} & u^2 & \frac{1-u^2}{2} \\ \left(\frac{1+u}{2}\right)^2 & \frac{1-u^2}{2} & \left(\frac{1-u}{2}\right)^2 \end{vmatrix} \begin{vmatrix} P_{1,1} \\ P_{1,o} \\ P_{1,-1} \end{vmatrix} \quad (7)$$

4. Detection of coherent level mixing by double resonance

The mixing between the F = 1 sublevels cannot be detected directly, but can be observed owing to its influence on the hyperfine transition $\Delta m = 0$, as indicated in figure 2. The atomic F = 0 level is simply translated by the amount $n\hbar\omega_x$ by applying the A.F. field.

The best way to observe this mixing is to operate the maser as an amplifier because all the parasitic effects connected to the dynamical properties of the oscillator are then avoided.

Below the threshold of oscillation, a high frequency signal induces transitions between the states $|1,0>$ and $|0,0>$. In presence of the A.F. field the $|1,0>$ state is coherently mixed to the $|1,1>$ and $|1,-1>$ states and the $\Delta m=0$ line is splitted into a triplet, the components of which are frequency separated by :

$$\frac{E_+ - E_o}{\hbar} = \frac{E_o - E_-}{\hbar} = \sqrt{(\omega_x - \omega_o)^2 + \left(\frac{\gamma_e}{2} H_x\right)^2} \quad (8)$$

This splitting is known as the Autler-Townes effect [5]. In this case, the main feature is the very good separation between the lines, as shown in figure 3.

This particularity is associated with the very small width of the $\Delta F=1$, $\Delta m=0$ line (about 1 c/s) obtained with stored hydrogen atoms. Such a good separation allowed us to go further in the Autler-Townes effect study to relate the height of the three lines to the population of the F = 1 sublevels [8].

Note that this double resonant experiment is usually performed, when the maser is oscillating, to measure accurately the value of the static magnetic field H_0. A dip in the oscillation level is observed when the resonant condition $\omega_x = \omega_0$ is fulfilled [9]. When the maser is used as an amplifier, the height of the central line is minimum under the same condition and we have checked that the static field is then measured with the same accuracy as before.

5. Relation between the intensities of the triplet lines and the population of the F=1 sublevels

So far, for simplicity purpose, we have only considered the populations of the $|\varphi>$ states and of the unperturbed $|a>$, $|b>$, $|c>$ states. Nothing at all is to be changed if the P factors are differences of population between the states associated with the index of the P factors on one hand and the F=0 state on the other hand.

The intensities I_+, I_0, I_- of the triplet lines depend upon two factors. The first is the difference of populations between the $|\varphi>$ states and the $|0,0\,;\,h>$ state described by P_+, P_0, P_-. The second is the probability that the observed transition occured. It is proportional to the squared coefficient of the $|b>$ state in the development of the $|\varphi>$ states as given by the relation 4.

Thus, we obtain for the relative intensities :

$$\begin{cases} I_+ = \dfrac{1-\mu^2}{2} P_+ \\ I_0 = \mu^2 P_0 \\ I_- = \dfrac{1-\mu^2}{2} P_- \end{cases} \qquad (9)$$

We have specified the influence of the H.F. observation field on the measurement of the intensity of the lines. The height of the linearly detected lines are proportional to I_+, I_0, I_- under the following conditions :

- The maser must be far from the threshold of oscillation for each line. Then, the regenerative properties of the amplifier are negligeable.

- The amplitude of the H.F. field must be very small for avoiding the saturation of the lines (the detected power is about 10^{-16} watts).

6. Check of the method with a classical state separator

We have checked the validity of the sudden application of the perturbation using a classical state separator which focuses into the storage bulb both $|1,0>$ and $|1,1>$ states. In our case it is made of two hexapolar magnets (figure 4). As long as an adiabatic rapid passage is not performed between the two hexapolar magnets, they are optically equivalent to a single hexapolar magnet.

We have calculated the transmitting factor of this magnetic lens and its separation efficiency by a phase plane method developped for this purpose [10]. The theoretical composition of the beam entering the bulb is very nearly 50% of atoms in the $|1,0>$ state and 50% of atoms in the $|1,1>$ state. (In the maser used for this experiment, the length of the drift space between lens and bulb is 90 cm).

The H-maser is voluntarily held well below threshold condition of oscillation by operating its atomic hydrogen source at very low pressure (about 10^{-2} Torr).

A 5 054 825 c/s signal delivered by a 1163 A G.R. frequency synthesizer is frequency multiplied to induce the $\Delta m=0$ transition (figure 5). The synthesizer is driven by a very stable 1115 B G.R. quartz oscillator. The heterodyne receiver gives a 405750 c/s signal whose frequency is held constant by a tracking filter system when the excitation signal is frequency swept. The constant frequency signal is mixed down to 5 750 c/s, linearly detected and filtered (1 s. time constant). A record of the triplet lines is shown in figure 3.

The A.F. field is produced by a pair of rectangular coils wounded around the cavity, under vacuum, approximately in Helmholtz position.

It is applied to the atoms when they enter the bell jar assembly, within a time of about 30 microseconds. The frequency of the field is chosen to be 7 kc/s to

achieve the condition of sudden application. The static magnetic field is then H_o = 5 milligauss.

The parameter u is measured by the following method :

- the A.F. field is produced using a frequency synthesizer. The resonant frequency ω_o is determinated with a great accuracy by observing the disparition of the central line of the triplet for a very weak value of the A.F. signal amplitude. The value of $\omega_\infty - \omega_o$ is then obtained by direct reading on the frequency synthesizer dial.

- At resonance, the separation between the two sideband lines provides the value of the quantity $\frac{\gamma_e}{\omega} H_\infty$.

The figure 6 shows the comparison between the theoretical and experimental intensities of the triplet lines for the following values of the differences of populations :

$$P_{1,1} = 0.5 \quad , \quad P_{1,o} = 0.5 \quad , \quad P_{1,-1} = 0$$

As it must be, there is a very good agreement between the measured points and the theoretical curves calculated under the hypothesis of sudden application of the perturbation.

7. Application to the measurement of the beam composition obtained with an improved state selector

7.1 - State selector with adiabatic rapid passage

A schematic representation of the device is given in figure 4.

The atomic beam is produced by a single tube. Its diameter is 1 mm and its length 20 mm. The gap of the hexapolar magnets are 3 mm for the nearest of the source and 7 mm for the other. The magnetic field at the pole tips of both magnets is 7.000 G. The first magnet is 47 mm long and the second 69 mm. They are separated by a adiabatic rapid passage region where static magnetic field perpendicular to the beam is applied. It varies by 1 G/cm around 8 G. A R.F.coil produces an alternative field parallel to the beam whose frequency is about 12 MHz with an amplitude of a few Gauss. This frequency corresponds to the spacing between the F = 1 sublevels in the coil's center.

It can be shown that the F = 1 sublevels can be considered as equidistant in the A.R.P. conditions and the A.R.P. process can be easily understood with the results of the first part of this report.

In figure 7 we see that, if the atoms are introduced in the R.F. field far from resonance (D.C. magnetic field very different from 8 Gauss), they are then in states $|\varphi\rangle$ very close to the unperturbed states. The perturbation is suddenly applied with $u \# + 1$ and formulas 7 show that

$$P_+ \# P_{1,-1} \; ; \quad P_o \# P_{1,o} \; ; \quad P_- \# P_{1,1} \tag{10}$$

For moving atoms, the magnetic field is adiabatically swept across the resonance value (no transition between the $|\varphi\rangle$ states) and rapidly (interaction time much smaller than relaxation times of atoms in the beam). After crossing the coil, the R.F. field is suddenly suppressed far from resonance.

Thus the effects of the A.R.P. are :

- to keep constant the population of the $|1,o\rangle$ state. This has been checked by using the maser as an amplifier on the $\Delta m=0$ unperturbed line and by switching on and off the A.R.P.

- to exchange the populations of the $|1,1\rangle$ and $|1,-1\rangle$ states.

Thus the atoms in $|1,1\rangle$ and $|1,-1\rangle$ states are repelled from the beam-axis either in the first hexapolar magnet or in the second one.

In conclusion, the state separation acts like :

- a doublet of convergent lenses for atoms in $|1,o\rangle$ state

- a convergent lens and a divergent lens for atoms in states $|1,1\rangle$ and $|1,-1\rangle$.

- a doublet of divergent lenses for atoms in state $|0,0\rangle$.

The atoms in $|1,0\rangle$ states are then very well separated from the others.

7.2 - Measurement of the beam composition obtained with this improved state selector

The differences of population between the F = 1 sublevels and the F = 0 level have been obtained by the double

resonance method. The beam composition, with a probably overestimated uncertainty, is as follows :

$$P_{1,1} = (13 \pm 3)\%$$

$$P_{1,o} = (84 \pm 5)\%$$

$$P_{1,-1} = (3 \pm 3)\%$$

These results have been checked by a direct measurement of the spin exchange contribution to the T_2 relaxation time. The T_2 measurement is classical [11], the absolute atomic flux is obtained as indicated in our last year's report [12]. The results are shown in figure 8.

The comparison with the ideal case where all the F = 1, m = 1 atoms are supposed eliminated gives the ratio of the atomic densities (formula n° 1) obtained when the A.R.P. device is operated or not. Then, the global beam composition is as follows :

$$P_{1,o} = (87 \pm 2)\%$$

$$P_{1,1} + P_{1,-1} = (13 \pm 2)\%$$

8. Conclusion

The validity of this double resonance method for measurement of population differences of hyperfine levels of stored atoms can be extended to more complicated hyperfine structures than that of hydrogen atoms.

It may be useful for studies of population inversion obtained by atomic beam technics (deuterium, cesium, rubidium) or by optical pumping.

References

1. D. Kleppner, H.M. Goldenberg, N.F. Ramsey, Physical Review, 126, 1962, p. 603

2. J-P.Schermann, Comptes Rendus Ac.Sci. 263, 1966, p. 295

3. C.Audoin, M. Desaintfuscien, P. Petit, J.P. Schermann, I.E.E.E. Transactions on Instrumentation and Measurement IM-17, 1968, p. 351

4. A. Abragam, J.M. Winter, Comptes Rendus Ac.Sci., 255, 1962, p. 1099

5. S.H. Autler, C.H. Townes, Physical Review, 78, 1950, p. 340

6. C. Cohen-Tannoudji, S. Haroche, Journal de Physique 30, 1969, p.125 et 153

7. P. Piéjus - Thèse Paris 1968 (unpublished)

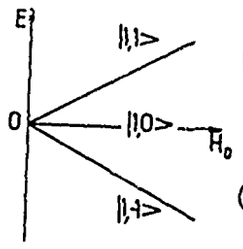
8. C.Audoin, M. Desaintfuscien, P. Piéjus J-P. Schermann, Journal of Quantum Electronics (to be published)

9. H.G. Andresen, Zeits. für Physik, 210, 1968, p. 113

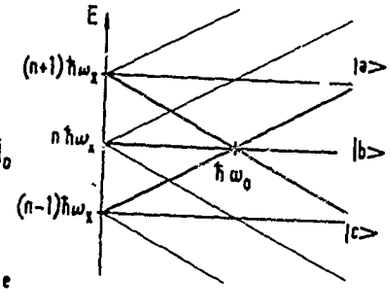
10. C. Audoin, M. Desaintfuscien, J.P. Schermann, Nuclear Instruments and Methods, 69, 1969, p. 1

11. D. Kleppner, H.C. Berg, S.B. Crampton. N.F. Ramsey R.F.C. Vessot, H.E.Peters, J. Vanier Physical Review, 138, 1965, p. A972

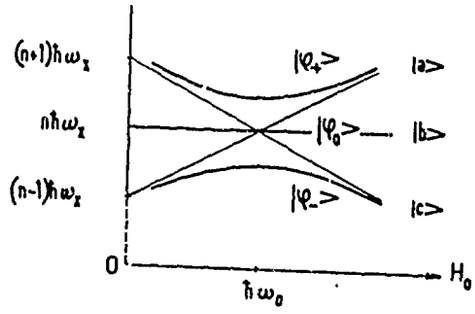
12. C. Audoin, M. Desaintfuscien, J-P. Schermann, Proceedings of the 22nd Annual Symposium on Frequency Control - Atlantic City, 22-24 April 1968, p. 493



1A_Energy levels of the F=1 multiplicity of H-atoms in the static field H_0



1B_Energy levels of H-atoms and A.F. field (without interaction)



1C_Energy levels of interacting H-atoms and A.F. field

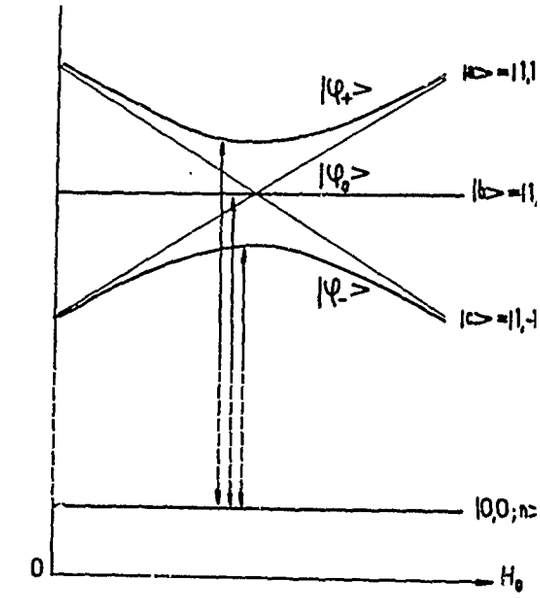


Fig.2 _Triplet splitting of the $\Delta F=1, \Delta m=0$ transition

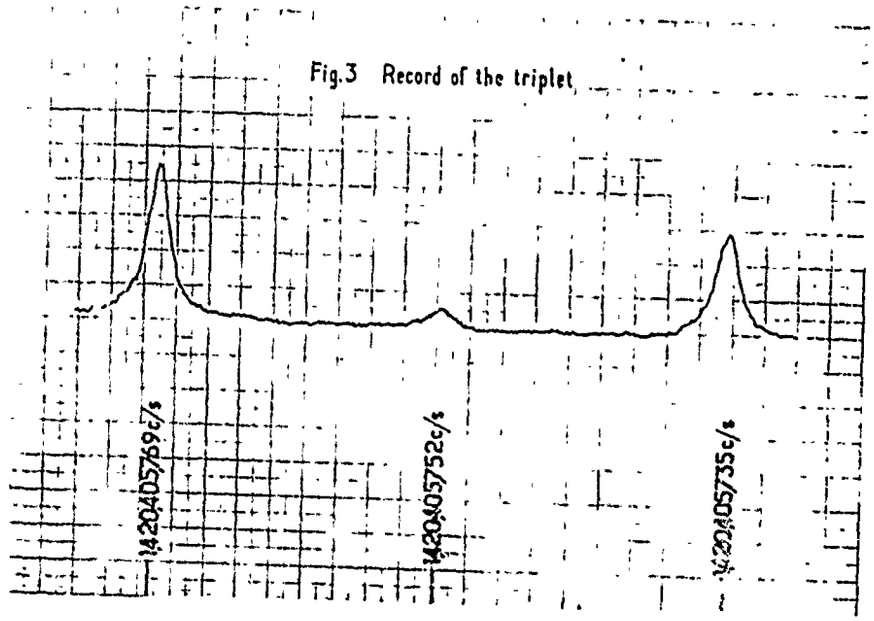


Fig.3 Record of the triplet

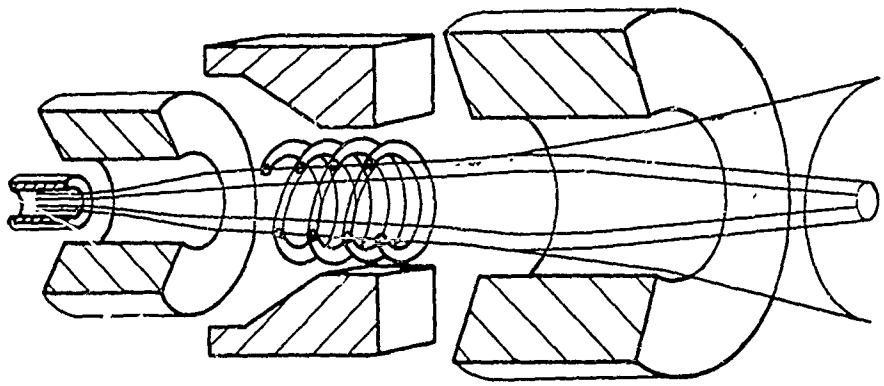
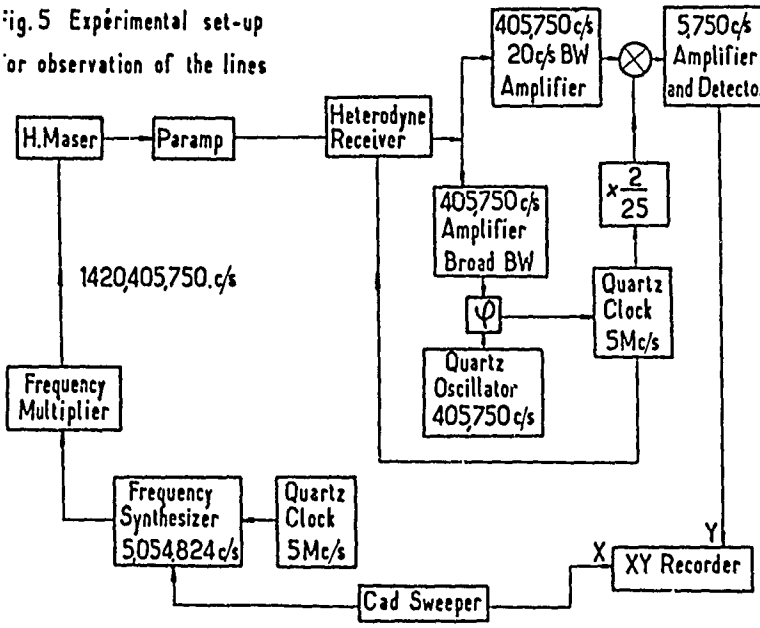


FIG.4.SCHEMATIC REPRESENTATION OF THE STATE SELECTION WITH ADIABATIC RAPID PASSAGE. THE SEPARATION OF F=1, m=0 AND F=1, m=1 STATES IS SHOWN.

Fig. 5 Expérimental set-up for observation of the lines



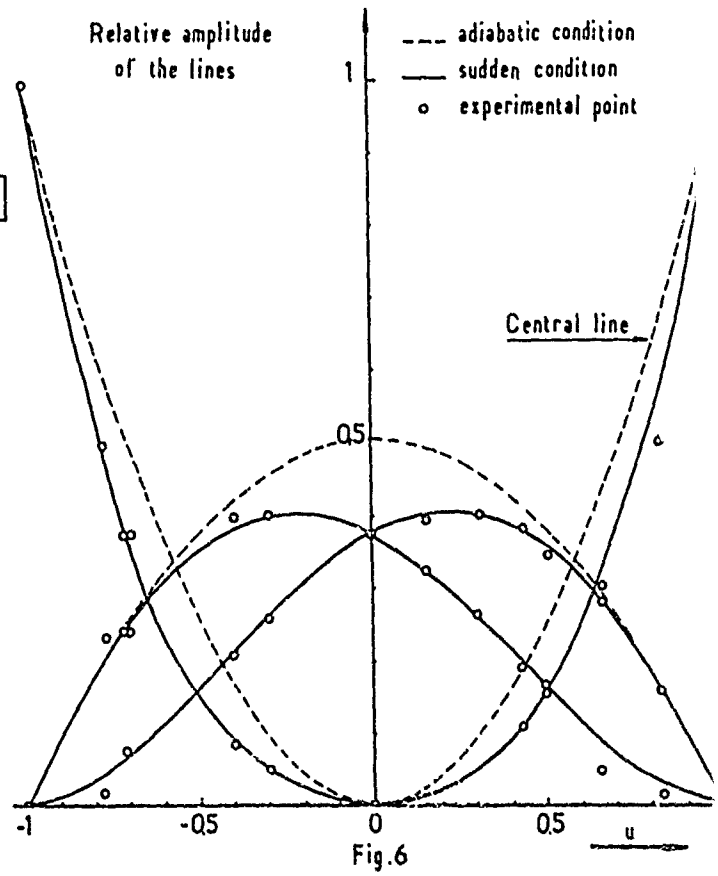


Fig.6

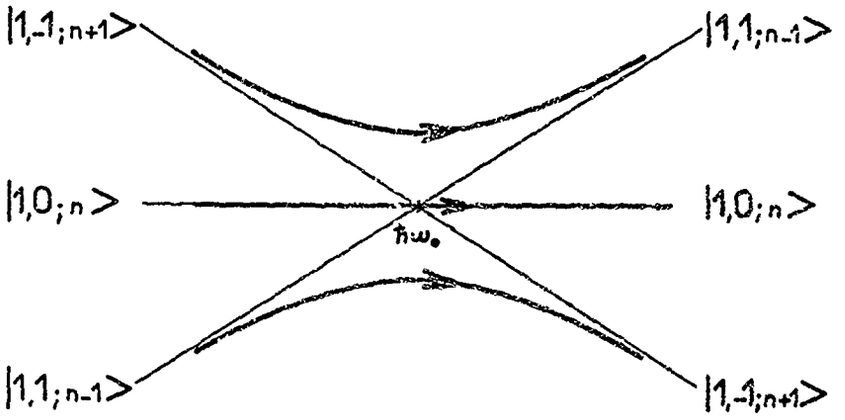


Fig.7 Effect of the adiabatic rapid passage on H atoms

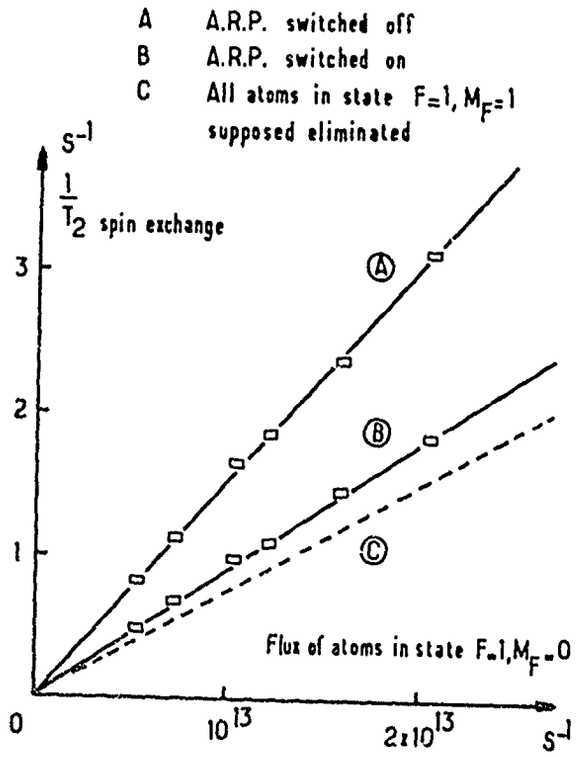


Fig 8 Effect of the A.R.P. on the spin-exchange contribution to the relaxation time T_2

ATOMIC HYDROGEN STANDARDS FOR NASA TRACKING STATIONS

H. E. Peters
E. H. Johnson
and
T. E. McGunigal
Goddard Space Flight Center
Greenbelt, Maryland 20771

Summary

Work at Goddard Space Flight Center to adapt the hydrogen maser for field use at tracking stations has been in progress for three years. Prototype standards based upon this work are now in the test and evaluation phase. Experimental work leading to the design of these prototypes was presented at the 22nd Annual Frequency Control Symposium.[1] In the present paper we will discuss the design of the prototypes and give test results. We will also present further experimental data on an experimental maser representative of the new design which has been in continuous operation for approximately two years.

The prototype standards have been designed as self contained frequency and time standards. A unit contains a hydrogen maser, receiver-synthesizer, automatic cavity tuner, clock, standard output frequencies and time signals, and other system electronics. Designed for a negligible amount of operator attention, the standard contains status monitoring and alarm circuits, and the entire system transfers uninterruptedly to battery operation in the event of power failure.

Introduction

The stability and accuracy of a standard based upon the atomic hydrogen maser arises largely from the fact that atomic hydrogen may be confined within a properly treated enclosure for relatively very long periods of time (over one second) in a manner which perturbs, but very slightly, the magnetic hyperfine quantum states of the atom. Additionally, the frequency of the transition involved has almost negligible dependance upon magnetic field, temperature, cavity pulling, and atom-atom or atom-molecule collisions, under conditions which have been achieved in present day standards.

These conditions, however, have required the attainment of very high temperature stability, of magnetic field homogeniety, and of cavity frequency control, as well as the use of advanced electronic techniques in receivers, multipliers, synthesizers, etc. The large physical structure of the maser, and the several specially designed subsystems required to operate it optimally, are perhaps, in striking contrast with the relative simplicity and beauty of the basic physical principles. The hydrogen hyperfine energy level diagram shown in Figure 1(a), and the hydrogen maser basic elements shown in Figure 1(b) illustrate this point. It is only recently that hydrogen masers which approach the stability and reproducibility now known to be achievable have been constructed in many research laboratories or that the reality of the masers present performance and future promise have begun to be recognized by authorities in the fields of timing and frequency control.

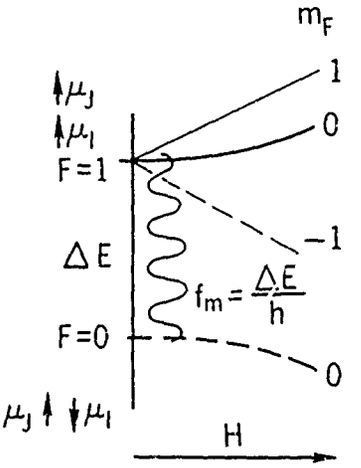


Figure 1(a). Hydrogen Hyperfine Energy Levels

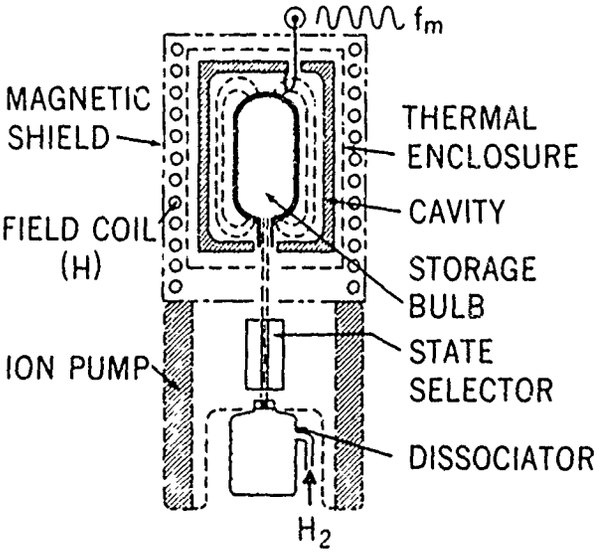


Figure 1(b). Elementary Hydrogen Maser Oscillator

The scientific applications in very long baseline stellar interferometry,[2] in earth navigation, geodesy, basic research, aircraft collision avoidance, etc, as well as use as a laboratory comparison standard are exciting the interest of scientists in these several disciplines. The prospective applications in space tracking and navigation[3] are NASA's primary motivation for undertaking the present work. Figure 2 is a photograph of one of the four prototype hydrogen standards which we have recently constructed at NASA's Goddard Space Flight Center and which are now being tested and evaluated for field application. Many aspects of the design of these units differ substantially from designs conceived elsewhere. Our main effort in this paper will be to describe the design, as well as the functional characteristics of these standards.

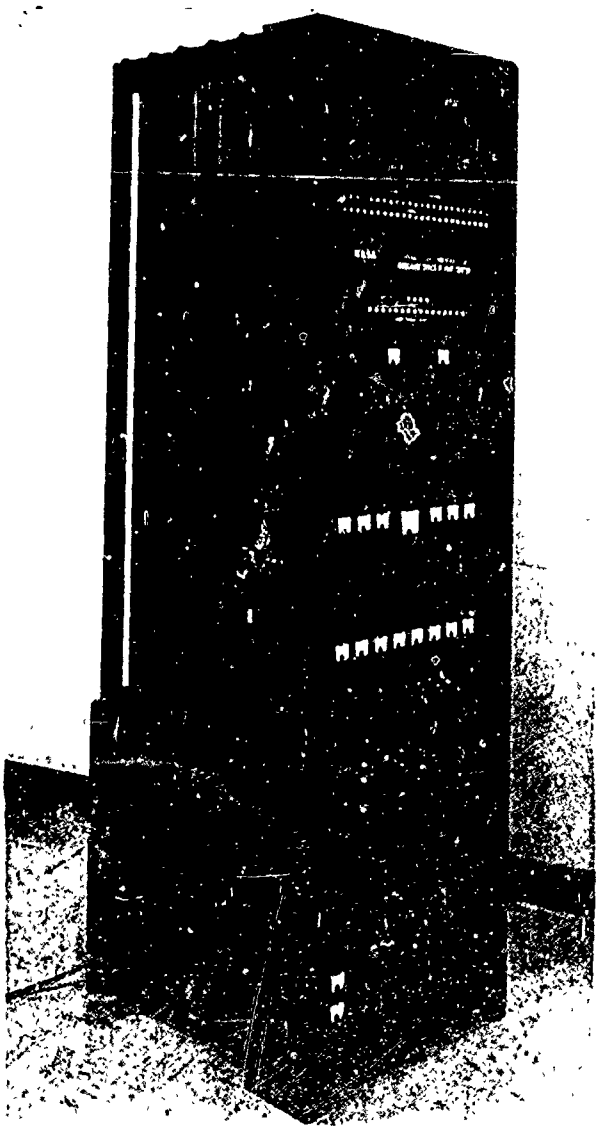


Figure 2. NASA Prototype Hydrogen Standard NP-1

Overall System Arrangement

Figure 3 shows the overall configuration of the hydrogen maser and the associated subsystems. These are located within a framework which is 6 feet high, 22 inches wide, and 27 inches deep. Since these standards have been designed primarily for ground based station operation, much consideration has been given to the spacious arrangement of subsystems in easily accessible, replaceable, and repairable modules. The hydrogen maser assembly itself, 4 feet in height by 18 inches in diameter, is mounted on a rugged aluminum frame and is surrounded by a cage structure which allows free ambient convection cooling while protecting the maser from mishandling in shipment or in use. The electronic modules are mounted on the front of the main frame and fold out as illustrated. Batteries which provide six hours of uninterrupted operation during a-c line failures or during transport, are mounted either in the lower section or externally, as desired. Three hundred and thirty watts at usual a-c line voltages and frequencies is the net normal power required.

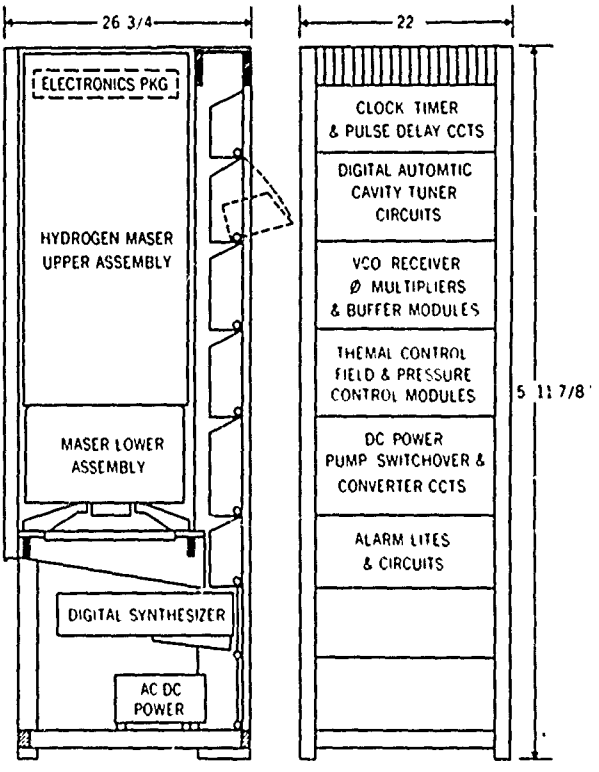


Figure 3. NASA Hydrogen Standard Layout

Hydrogen Maser Design

Lower Assembly

The maser lower assembly is shown in Figure 4. The 1.6 liter bottle of hydrogen, with 1,000 psi initial

pressure, is adequate for twenty years of normal operation; but it may, if required, be refilled without shutdown. The pump assembly consists of a six element, 150 liter per second titanium ion pump. This pump, at measured hydrogen flow rates, and using extrapolated performance data on several other pumps and masers, should provide ten years of continuous service between pump element changes. The pyrex RF source dissociator bulb is 2 inches in diameter and 2.5 inches long. This bulb, as well as a five watt, 120 MHz solid state exciter circuit, is located within an aluminum reentrant enclosure at the pump bottom. This arrangement effectively shields the source and prevents RF interference. A multitube collimator having 97 tubes, each 0.1 inch long and 0.002 inch diameter, confines the hydrogen atoms to a narrow beam as they enter the state selector.

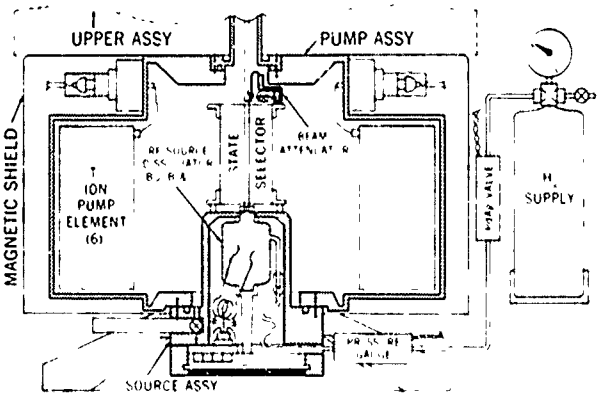


Figure 4. NASA Hydrogen Maser Lower Assembly

The state selector is a hexapole magnet with 0.05 inch diameter aperture. It is four inches long and has a maximum field strength of approximately 10,000 gauss. A beam attenuator, which is required in tuning the maser cavity, is located at the state selector exit. It consists essentially of a single very light spring with a "flag" which has fixed teflon stops. A preset proportion of the beam is intercepted by this "flag" as a soft iron pellet attached to the spring is moved by a miniature magnetic solenoid which has its axis transverse to the beam direction. The pump assembly has a single magnetic shield of 0.063 inch thick soft iron and has a 2-3/4 inch diameter hole at the top for the vacuum connection to the maser upper assembly. A built in vacuum valve in the source assembly provides for initial pump down of the system.

Maser Upper Assembly

Figure 5 shows the maser upper assembly. The FEP teflon lined quartz storage bulb is 5.5 inches in diameter by 16 inches long and has a volume of 5.5 liters. The bulb collimator, which is an integral part of the bulb, is 6 mm inside diameter by 1.0 inch long and provides a calculated exit relaxation time for thermalized hydrogen atoms of 1.56 seconds. The bulb is held securely to the cavity in radial directions by compressed viton o-rings at top and bottom, while a spring at the top holds the bulb down with a force several times its weight.

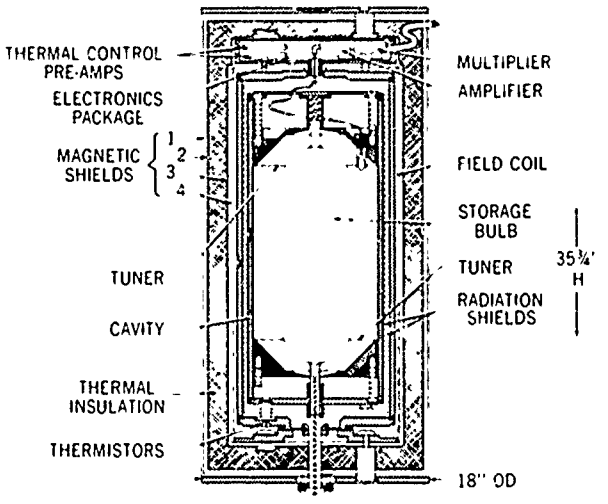


Figure 5. NASA Hydrogen Maser Upper Assembly

The 6061-T6 alloy aluminum cavity is nominally 10 inches in diameter. Shaped cavity tuners at both ends are each held in place by six 1/2 inch diameter adjustment bolts, 3 in tension and 3 in compression. The cavity length may thus be varied from 16 inches to 21 inches for adjustment of the cavity frequency. Aluminum end caps are securely screwed to the cavity cylinder; this makes the entire cavity assembly a thermally massive and isothermal structure. A precision thermistor and bifilar wound heater windings are located on the external cylindrical cavity surface. Low thermal conductivity metal supports at three locations provide a strong mechanical support for the cavity. Two polished aluminum cylindrical radiation shields are located between the cavity and the vacuum enclosure, and these are supported at each end by three small dielectric posts of negligible thermal conductivity.

The vacuum enclosure is of heavy aluminum with primary thermal control on the outer cylindrical surface, and additional thermal control near the cavity supports on the lower end. A large diameter mounting plate of aluminum provides a strong support as well as an isothermal surface adjacent to the vacuum enclosure bottom. This plate is attached to the assembly bottom cover by three widely spaced, low thermal conductivity, metal supports.

Four 0.025 inch thick moly-permalloy magnetic shields are used in this design. Two inside shields, with 0.25 inch separation, surround the vacuum enclosure. A field coil wound on an aluminum cylinder is placed between the enclosure and the inner shield, and extends from shield top to bottom. Two similar shields

are located outside these; they also have a 0.25 inch separation, but the inside diameter is approximately four inches greater than the outside diameter of the inner set. This space is filled with a two inch thickness of expanded polystyrene bead thermal insulation.

Near the top of the maser assembly, between the inner and outer magnetic shields, is located the thermally controlled electronic package. This is of heavy aluminum and is bolted to the vacuum enclosure top at three widely separated locations through one-half inch diameter holes in the inner shield caps. This arrangement provides an isothermal surface facing the vacuum enclosure top, as well as a thermally controlled, electrically, magnetically, and thermally shielded location for critical electronic modules in close physical and electrical proximity to the maser cavity. An incidental advantage is the reduction of possible microphonics by the mass of the vacuum enclosure. The cavity coupling loop is enclosed in a fused silica tube, and teflon insulated coaxial cable with copper inner and outer conductors connects this to the receiver input through a centrally located vacuum feed through. Three top supports attach an additional thermal shield (which is fastened to the electronics package) to the assembly top cover plate. Thus the entire assembly is supported top and bottom.

A very important consideration in hydrogen maser design is the requirement that no net d-c currents, which would destroy the magnetic field homogeniety, flow through any hole in the magnetic shields. Experience has shown that thermoelectric currents, or other current shield "Icops" must be avoided at all costs. Thus, all circuitry and components are electrically insulated from structures within the shields; all currents (and cables) going into the assembly must emerge through the same shield holes they enter; inner thermal and magnetic shields are electrically isolated; and the support bolts use nylon insulating bushings. The cavity coupling loop and cable make no d-c connection to the cavity or vacuum enclosure. Initial shield degaussing is accomplished by attaching a tubular connector to the top of the vacuum enclosure through an annular space between the coupling vacuum feed through and central holes in the magnetic shields, package, and top cover. The lower vacuum connection is not electrically insulated, and thus the a-c degaussing current can pass axially through the shields.

An additional magnetic field coil is wound on the lower vacuum connection tube where it passes through the outer magnetic shields. This has been found desirable in order to bias the magnetic field to eliminate mixing of the Zeeman levels of the F=1 states as the atoms pass through magnetic null regions in the shield holes where the field may have transverse components and undesirable direction changes. The coil is a single layer of wire two inches long with one hundred turns one inch in diameter. A fixed bias current of five to ten milliamperes d-c is usually adequate. (This must be in a particular direction, which is easily found by experiment.)

Electronics Systems

Thermal Control

There are five temperature control points with associated servo systems in each standard; these are: the maser cavity, main vacuum enclosure, vacuum enclosure lower support, electronics package, and the body of the source pressure sensors. A typical thermal control servo is illustrated in Figure 6(a). A most important feature of this servo is that the thermistor sensor is in a bridge circuit with d-c bias, rather than the a-c biased type frequently used in less critical applications. Typical thermistors of 50 K-Ohm rating in a bridge with 10 volts bias provide a sensitivity of about 0.1 volt per degree Celsius. In consideration of the low noise level and negligible drift of solid state differential amplifiers such as the ones used in the present case, the system noise and thermal drift is determined primarily by the thermistor characteristics. Individual heater current meters have been used, since, in the prototypes, concurrent observation is particularly desirable, and switching transients are also avoided.

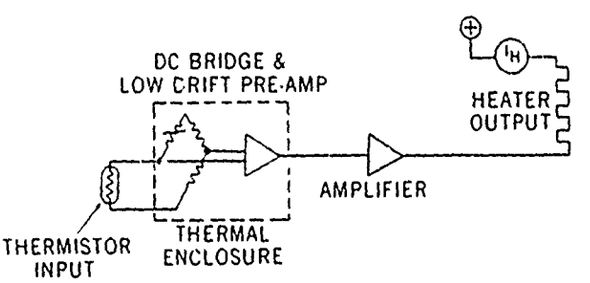


Figure 6(a). Typical Thermal Control Channel

Pressure Control

Figure 6(b) shows the source pressure control system. Separate servos are used for the pressure

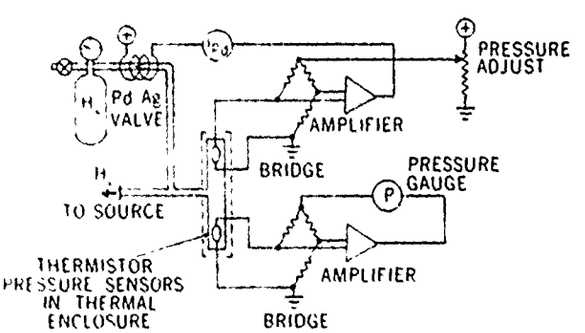


Figure 6(b). Source Pressure Control System

gauge and the pressure adjustment. The palladium-silver hydrogen purifier and control valve uses a very small, low thermal mass flow element, which results in relatively short pressure change times (typically 45 seconds) as well as stable response characteristics.

Receiver-Synthesizer

The basic receiver-synthesizer system is shown in Figure 7. A 5 MHz voltage controlled crystal oscillator with good low noise and low drift characteristics is phase locked through a multiheterodyne receiver-synthesizer system to the maser output signal. The components most sensitive to environmental noise or RF interference are located within the thermally controlled electronics package in the maser assembly. These include a 4 db noise figure solid state preamplifier, the first and second IF amplifiers, local oscillator multipliers and mixers. A digital synthesizer with divider provides 0.0001 Hz incremental adjustment (relative to 1420 MHz) of the output frequency; the maximum synthesis adjustment is limited by the crystal lock range to approximately $\pm 5 \times 10^{-7}$. Not shown in Figure 7 are auxilliary alarm circuits which indicate when VCO course adjustment is necessary, or when malfunctions occur such as loss of signal from the maser, the synthesizer, dividers or loss of lock, etc. Besides the usual 5 MHz, 1 MHz, and 100 KHz output frequencies, a 100 MHz output is provided directly from the first local oscillator multiplier for use in phase locking higher frequency signal sources to the maser.

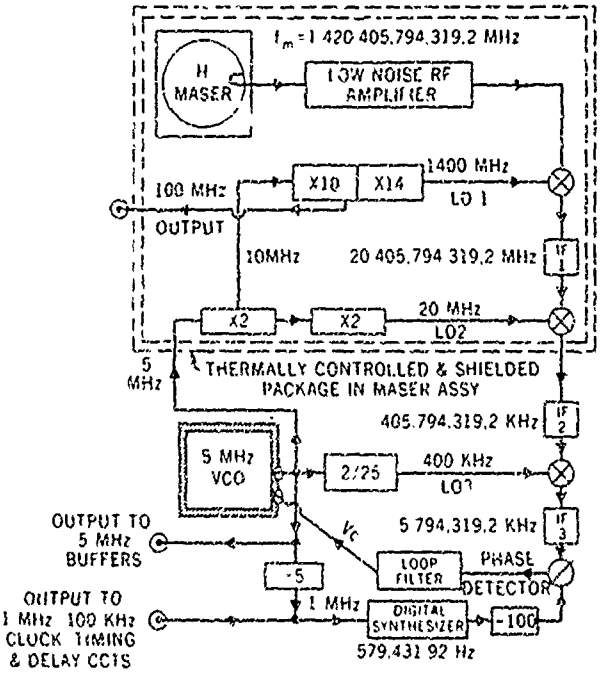


Figure 7. Basic Receiver-Synthesizer

Automatic Cavity Tuner

A most important electronic subsystem in these standards is the automatic cavity tuner, which is shown in Figure 8. This system was described at the 22nd Annual Frequency Control Symposium (Reference 1). The present system is unique in that it does not require a second hydrogen maser as a reference oscillator; by using digital integration and automatically variable period averaging techniques, the maser cavity is maintained at the proper frequency so as to affect the maser output one part in 10^{13} or less with the use of a precision quartz crystal oscillator as reference. This essentially eliminates the cavity pulling and spin exchange effects on the maser frequency. Panel indicator lights are provided to show frequency comparator counts, averager accumulation, and cavity correction register number. The cavity frequency may be manually offset when desired, and observation of the tuner response using these indicators provides confirmation of proper operation. The tuner response (see results below) agrees very well with the theoretical predictions and data given in reference 1.

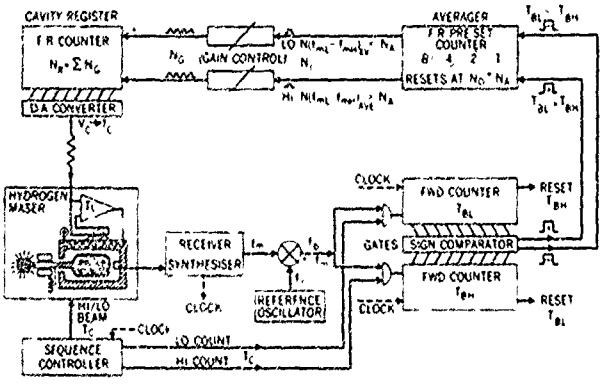


Figure 8. Auto Tuner Block Diagram

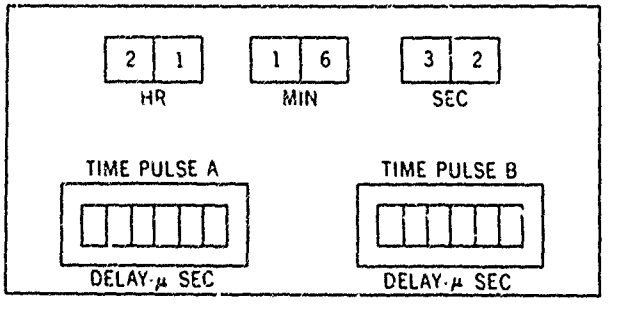


Figure 9(a). Clock and Time Pulse Panel

Clock and Time Pulse Circuits

Figure 9(a) shows the digital clock, and the time signal control panel. Two separate one pulse per second outputs are provided, each having independent delay adjustable in one microsecond increments by panel decade switches.

Audio Alarm and Red Light Monitor Circuits

Twenty-four "red light" circuits and a very audible audio alarm are provided on these standards. Some of the functions presently monitored are listed in Figure 9(b). This feature is particularly useful in rapid diagnosis of malfunctions, and of course, will alert station operators or other users to improper operating conditions. A push button on the front panel also allows one to test the alarm circuits.

XTL PHASE LIMIT (& LOCK)	5 MHz OUTPUT
XTL HEATER	1 MHz OUTPUT
AUTO-TUNER BEAT	100 KHz OUTPUT
CAVITY FREQ. CONTROL LIMIT	400 KHz L O
A-T MANUAL	23V MAIN POWER
A-T HOLD	20V. REGULATED POWER
405 KHz IF AMPL.	5V REGULATED POWER
ION PUMP PRESSURE	-15V REGULATED POWER
SYNTHESIZER OUTPUT	CHARGER VOLTAGE
ON BATTERIES	OTHER (SPARES)

Figure 9(b). Alarm Monitor Points

NASA Prototype Standard Data

Testing and evaluation of the new prototype standards has just begun. One standard (NP-1) has been oscillating for four months, and much of the data presented here are the results of measurements with this one. Tests on the other units (NP-2, -3, and -4) are just beginning at this writing, so intercomparison data with very meaningful statistics are not yet available, although some tentative first results from these are included.

On all of the units, the loaded cavity Q's are 36,000 (±2,000) and the coupling factor has been adjusted to 0.26. The measured line Q at oscillation threshold, from cavity frequency pulling data, is 3.2×10^9. The typical operating line Q at high beam is 2.5×10^9, and at low beam is 2.9×10^9. The beam intensity attenuation ratio has been set at 2:1. This relatively small change is all that is required with the present design to achieve the tuning accuracy desired (10^{-13} or better). Thus there are no seriously large changes in maser power level or in stability due to tuning modulation. The resulting ratio of tuning modulation to average frequency offset is 1:10 so that, for example, a net maser frequency offset of one part in 10^{13} due to a detuned cavity produces a modulation of one part in 10^{14}.

On NP-1, the recorded average cavity correction rate, referred to the output, is presently 6.6×10^{-14} per day. This represents the extent to which the maser would drift in a day if the cavity were not automatically tuned. Thus, for critical applications requiring the ultimate in short term stability, the autotuner may be stopped, and the maser operated at optimum beam strength for periods of a day or more, without significant detuning of the cavity.

The first stability measurements between two of the new standards have just been made. The comparis is between NP-1 and NP-3, with the power level, at high beam for both masers, of 10^{-13} watts (this was no accurately determined in this first run). The data was taken using the 5 MHz outputs from both masers and multiplying each to 200 MHz. One maser synthesizer was offset so as to produce a beat frequency of 100 Hz upon mixing the 200 MHz signals. A beat frequency filter with a bandwidth of approximately 200 Hz was use and the relative frequencies were determined for variou measuring times using a period counter. The data anal ysis was based upon the successive difference method (Reference 4 see articles by Barnes, Allen, and Vesso et al.).

RMS Fractional Frequency Deviations

Measuring Time, T, Secs.	1	10	100	1000
Both Beams High	1.9×10^{-12}	1.6×10^{-13}	1.6×10^{-14}	1.2×10^{-14}
Both Beams Low	2.7×10^{-12}	3.2×10^{-13}	3.2×10^{-14}	8×10^{-15}

The deviations appear to be characteristic of additive noise (1/T) for T up to 100 seconds, and are about 3.5 times the theoretical values for the present power levels and other design parameters, thus they may be improved upon in the future. For T = 1000 seconds and low beam, the data is consistant with the theoretical minimum due to fundamental maser pertubing noise (Reference 4, see article by Cutler and Searles) at this power level.

The hydrogen consumption rate data and the ion pump operating pressures and maser power levels are consistent with the design goals, as well as with calculations based upon the atomic trajectory and efficiency analysis given in reference 1. The net high beam flow rate, calculated from hydrogen bottle pressure drop experimental data on NP-1 is 5×10^{15} atoms per second (7.1×10^{-5} torr-liters per second of molecular hydrogen).

NASA Experimental Maser Data

The NASA experimental maser (NX-1) from which much of the experimental data of reference 1 was obtained, has been oscillating continuously since Sept. 1967 and has been automatically tuned essentially all of this time. Thus some very significant long term experience with a system representative of the new standards is

presently available. The long term stability and reproducibility, the cavity thermal control system performance, and the experience with the NASA design of the RF source dissociater is the most interesting of this experience.

Many frequency comparisons have been made with NX-1 during this period. In 1967 comparisons were made with two NASA modified H-10 hydrogen masers, with two similar hydrogen masers at the Naval Research Laboratory (using a telephone line), and with several cesium standards. Since 1967 nearly continuous comparisons have been made with respect to GSFC operated cesium "house standards," and for over three months in the summer of 1968 NX-1 was the comparison standard in tests of several cesium beam standards which were being evaluated for use at Apollo tracking stations. Just recently, comparisons with NP-1, the first of the new hydrogen standards, have been made. The 1967 measurements indicated agreement within one part in 10^{12} among the several hydrogen masers. The cesium comparisons in 1967, 1968, and present cesium data indicate no change in the NX-1 frequency during this period within the accuracy capabilities of the cesium standards. To look at this another way, the significant cesium data, when related to NX-1, show the various cesium standards to have the same frequency, in general, to well within their accuracy specifications. The recent comparisons with NP-1 show agreement to well within one part in 10^{12}. The significance of all these comparisons is that (a) the intrinsic reproducibility inherent in hydrogen standards is very superior, and (b) if there is any long term variation in the wall shift correction of continuously operating hydrogen standards, it is of the order of 5 parts in 10^{13} per year or less. (Much other data by other authorities also supports these conclusions, see Reference 4, article by McCoubrey, and References 5 and 6.)

A continuous record has been kept, since 1967, of the cavity frequency corrections of NX-1 made by the automatic cavity tuner. This data indicates the rate of thermal drift (or other systematic changes) of the maser cavity. At first, this was -0.0012°C per day, while at present, it is -0.0001°C per day. Initial data on NP-1 indicate a correction rate equivalent to -0.0004°C per day. This encouraging data is consistent with the thermal control experiments reported in reference 1, upon which the rather adventurous hydrogen maser design inovation of using an aluminum cavity, with thermal tuning, was based.

The 5 watt solid state RF source dissociater used in NX-1 has not exhibited any change in characteristics over the entire experimental period. Most significantly, the pyrex glass source bulb has not developed the intense internal coating and discoloration which was characteristic of elderly maser sources of earlier design. The sources on the new standards appear to reproduce the discharge color and intensity of NX-1, so reliable, long lived sources may be realistically anticipated.

Conclusion

From the present results, it appears that the new prototype hydrogen standards will achieve their design goals. However, the next, and most important and interesting phase of their development is just starting. During the coming months, in addition to stability and reproducibility intercomparisons, many environmental and systematic tests will be performed. Possible "bugs" will be hunted, and optimum operating conditions will be determined. With successful completion of the local tests, several field tests are being planned. These will involve experiments in three way doppler tracking[3] with standards at two or more remote sites, timing experiments, and experiments in collaboration with radio astronomers in very long base line (VLB) stellar interferometry.[2] Finally, three of the standards may be deployed at the Apollo tracking stations having 85 foot antennas. These are located at Madrid, Spain; Carnarvon, Australia; and Goldstone, California.

It is appropriate to illustrate the portability of the standards at this point. Figure 10 shows NP-1 mounted on a "maser dolly" and operating on external batteries. One man can quickly attach the dolly, and easily maneuver and position the standard.

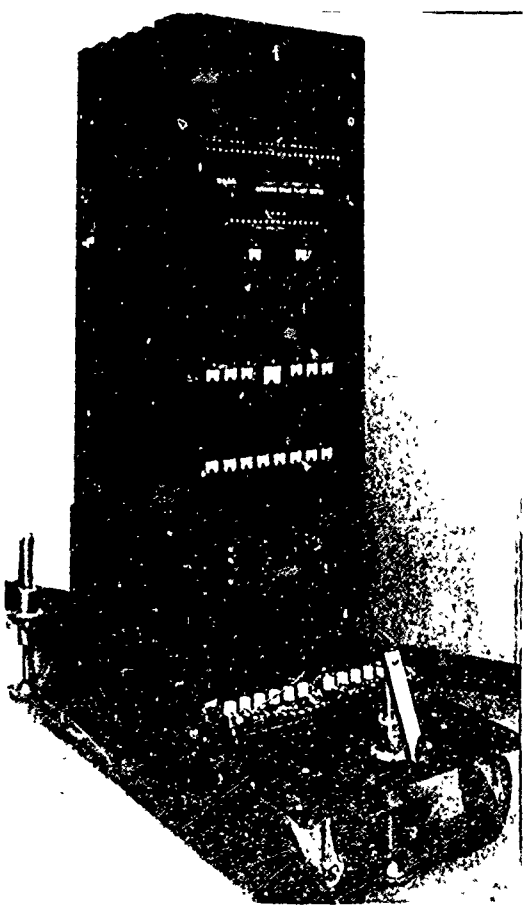


Figure 10. NASA Hydrogen Standard on "Dolly"

References

1. H. E. Peters, T. E. McGunigal, and E. H. Johnson, "Hydrogen Standard Work at Goddard Space Flight Center," Proceedings of the 22nd Annual Frequency Control Symposium, Atlantic City, 1968.

2. M. H. Cohen, et al, "Radio Interferometry at One-Thousandth Second of Arc," Science, Vol. 162, No. 3849, 4 October, 1968.

3. F. O. Vonbun, "Space Trajectories and Errors in Time, Frequency, and Tracking Station Location," GSFC Report X-507-67-163, 1967.

4. Proceedings of the IEEE Special Issue on Frequency Stability, Vol. 54, No. 2, February 1966.

5. H. E. Peters and T. E. McGunigal, "Recent Accuracy Comparisons of Cesium and Hydrogen Atomic Frequency Standards," Goddard Space Flight Center Report X-523-67-529, 1967.

6. R. Vessot, et al., "An Intercomparison of Hydrogen and Cesium Frequency Standards," IEEE Transactions on Instrumentation and Measurement, Vol. 1 M-15, No. 4, December 1966.

FREQUENCY MIXING AND MULTIPLICATION IN THE FAR INFRARED AND INFRARED

A. Javan
Massachusetts Institute of Technology

This talk reviewed the M.I.T. activities in the area of frequency mixing and multiplication in the far infrared and infrared. The characteristics of a frequency multiplier chain using "laser links" were described. In this chain, high order harmonics of a far infrared laser is mixed with the fundamental frequency of another laser and an intermediate frequency in the microwave range. The "phase locking"of a far infrared signal against high order harmonics of a microwave fundamental frequency was also discussed. The talk also included a description of frequency mixers and rectifiers used in these experiments.

USE OF LASER-SATURATED ABSORPTION OF METHANE FOR LASER FREQUENCY STABILIZATION*

R. L. Barger and J. L. Hall
Joint Institute for Laboratory Astrophysics†
Boulder, Colorado

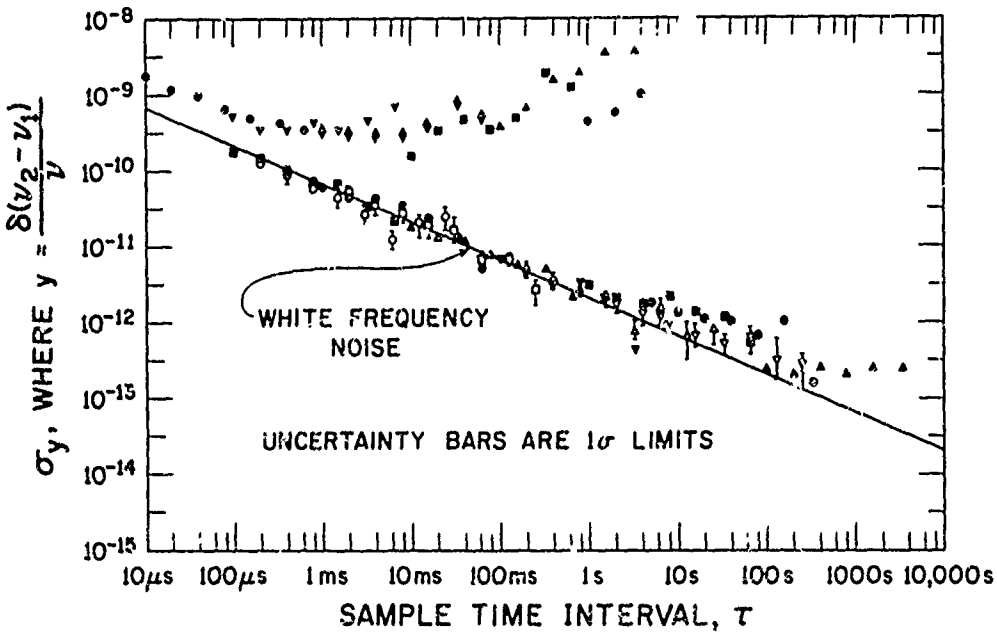
Using laser saturation of molecular absorption, we have obtained at 3.39 μ an emission feature in laser power centered on a methane vibration-rotation line [P(7) line of the ν_3 band].[1,2] We report here the results of using this feature to stabilize laser frequency. The width of this very sharp Lorentzian line is determined by collisional and interaction time effects. Methane collisional self-broadening is measured to be 8.1 ± 0.3 kHz/m Torr, and an upper limit for the exceptionally low self-induced pressure shift is 75 ± 150 Hz/m Torr. We have obtained widths as narrow as 150 kHz halfwidth at half-maximum intensity, i.e. $\Delta\nu_{1/2}/\nu \approx 1\times10^{-9}$. Of this, about 100 kHz is due to time of flight across the laser beam; the remainder is due to power broadening. The asymmetry previously observed[2] in power broadening has been reduced to a negligible level. Beat frequency measurements between two independent, dissimilar methane stabilized lasers show that we have stabilized laser frequency to line center with a reproducibility of better than ± 1×10^{-11}. An analysis of the frequency noise spectrum is given in the figure.

*Supported in part by NASA, Advanced Development Division, Goddard Space Flight Center, Greenbelt, Md. and in part by the National Bureau of Standards, of which the authors are staff members.

†of the University of Colorado and the National Bureau of Standards.

1. J. L. Hall, IEEE J. Quantum Electron. QE-4, 638 (1968).

2. R. L. Barger and J. L. Hall, Phys. Rev. Letters 22, 4 (1969).



Frequency fluctuations versus sample time interval of (a) Beat between two free running He-Ne lasers (upper set of measurement points). (b) Beat between two He-Ne lasers, each independently controlled by a methane absorption cell. (Lower set of measurement points).

Plotted is the Allan variance.* Number of samples $N = 2$; ratio of dead time plus sample time to sample time $r = 1$; bandwidth $B \gtrsim 1$ MHz.

*D. Allan, Proc. IEEE 54, p. 221 (1966).

STABILITY INVESTIGATIONS OF HCN LASER

V. J. Corcoran, R. E. Cupp, W. T. Smith
and J. J. Gallagher

Electromagnetics Department
Martin Marietta Corporation
Orlando, Florida

Summary

Experiments have been performed on the heterodyning of the far infrared hydrogen cyanide laser. Investigations of the beat between the laser and a harmonically generated signal originating from a 5 MHz frequency standard have provided information relevant to the laser stability and phase-locking of the laser to the microwave source. Linewidths of the beat are approximately 5 kHz. Computer programs have been employed to analyze the conditions necessary for locking of the laser. The experimental details of phase-locking investigations will be presented to demonstrate the progress made thus far, the correlation of experiments with computer analysis and the necessary experiments for the achievement of laser phase-locking.

Frequency stabilization of the laser to the frequency standard by an AFC system has been achieved with stability on the order of a part in 10^8. Frequency pulling of the laser over a range of approximately 600 kHz results from tuning the frequency of the microwave signal.

Introduction

During the past five years, laser activity has been extended to the far infrared wavelength region. As a result, the output wavelength of gas lasers has been steadily increased to the point that the longest achievable wavelength is overlapped by the shortest wavelength that is generated harmonically from microwave oscillators. The availability of these lasers now makes possible several investigations which were not previously considered possible.

Because of the overlap of laser transitions with the signals obtained by harmonic generation from conventional oscillators, it has been possible to mix the output of lasers with the output of klystrons, and thereby measure very accurately the laser output frequency. This has been achieved with the HCN laser transitions at 890 GHz and at 960 GHz by mixing with the 12th and 13th harmonics of a 70 GHz klystron[1]. More recently, it has been possible to observe the DCN line at 190 microns[2], and the H_2O line at 118 microns[3] by beating against a harmonically generated signal.

Several applications of the far infrared lasers require greater stability of these sources than is available when operated in a free running mode. Thus, the establishing of standard frequency signals, velocity of light measurements, the extension of existing frequency standard techniques to the optical region and the eventual use of precise tunable signals depend upon the availability of far infrared signals with stabilities on the order of parts in 10^8 or better. Previous work, performed in the Electromagnetics Department of Martin Marietta, with the support of the United States Army Electronics Command, has provided the preliminary basis for the extension of frequency control to the far infrared[4].

The objective of the investigation discussed here has been to develop the means of obtaining from the HCN and/or H_2O lasers now in operation in the Martin Marietta laboratories an IF signal of sufficient spectral purity so as to permit eventual phase-locking of the laser to a frequency standard. It has initially been desired to obtain stabilities on the order of a part in 10^8 by either phase-locking to the multiplied frequency standard signal or by locking with an AFC circuit to the frequency standard signal. With the AFC loop, the stability of one part in 10^8 has been achieved.

A theoretical treatment of phase-locking of the laser and various experimental studies of the phase-locking have been performed, while it has also been possible to concentrate on the stability characteristics of the laser, the beating of two HCN lasers, and the eventual AFC of the laser by the harmonically generated frequency standard signal. The work has been concerned mainly with the HCN laser since it is more accessible by harmonic generation, but all techniques are applicable to the H_2O laser.

Laser Stability Measurements

In order to determine the stability and factors affecting the stability of an HCN laser, a series of experiments were performed in which two free-running lasers were mixed in open-structure, point contact mixer. One laser is mounted on a cement beam for stability purposes. The laser uses the box configurations for the mounting of the mirrors[4] and is approximately 12 feet long. The gas inlet is in the grounded box and the pump outlet is in the high voltage box. The other laser is mounted on a granite slab and essentially has the same configuration as the other laser, but is only 10 feet long.

An open-structure point contact diode is constructed on the same principles as the conventional waveguide mixers, except that the laser signals are coupled to the diode through two 1/2 inch holes. Small mirrors, mounted on gimbal supports, reflect the signals back to the diode structure. It was found, during these investigations, that the optimum part of the diode for the laser signals to strike is the hook in the tungsten whisker rather than the junction of the whisker and crystal.

The output of the mixer is amplified in a video amplifier. The amplified signal is displayed on an oscilloscope and mixed with a local oscillator. The sideband that is generated is displayed in the frequency domain on a spectrum analyzer. Widths in excess of 50 kHz were obtained in all observations.

The first modification to the above experiment was to isolate each laser from the ground by setting it on six inner tubes. Only slight improvement in the stability occurred.

The next modification was to stabilize the discharge in the laser tubes by using an ethyl ether and ammonia mixture instead of natural gas and nitrogen. The line has narrowed considerably, in most cases being less than 20 kHz. In the time domain representation, the drift was found to be at a 120 Hz rate, thus indicating that the ripple on the power supply could be a major source of instability. By the use of filtering and a feedback network, a constant current supply can be generated; with the reduced rippler and drift of the current from the power supply, each laser should be more stable with a resultant reduction in the linewidth of the beat. With such a circuit, the laser employed in the phase-lock experiments has not shown a significant narrowing of the beat.

While further reduction of the beat width should occur, the linewidth 10-20 kHz is more narrow than the beat between the laser and klystron harmonic has been until recently. This indicates a need for improvement in the spectral purity of the multiplying chain, which will be discussed in a later section.

Phase-Locked Laser Techniques

The earlier attempts at phase-locking the laser to the harmonic of the frequency standard have been previously described.[4] The techniques employed were similar to those described below. However, a number of modifications were made in these experiments; some changes were suggested by the phase-lock analysis and the laser stability studies.

The system used to attempt phase-locking the HCN laser signal to the submillimeter frequency standard is the following: A Varian VC705 reflex klystron is phase-locked to the ninth harmonic of an X-band reflex klystron, which itself is phase-locked to the sum of an harmonic of a rubidium frequency standard signal and an interpolation oscillator signal. The eleventh harmonic of the Varian VC705 is generated and mixed with an HCN laser signal at 890,758 GHz in a point contact multiplier. This beat frequency signal is amplified in an LEL 30 MHz IF amplifier and sent to the Model 2650A Dymec Transfer Oscillator Synchronizer. The phase error signal from the synchronizer before filtering is amplified, filtered, and fed to the piezoelectric translator on which one of the laser mirrors is mounted. The amplifier and filter were designed to provide sufficient loop gain and loop bandwidth to capture the beat frequency signal.

The signal generated by beating the eleventh harmonic of the Varian VC705 with the laser output is observed after amplification by the 30 MHz IF amplifier. A time domain representation of the beat frequency signal reproduced from a type 647 Tektronix oscilloscope is obtained. The frequency spectrum of the beat is also obtained. The spectrum is obtained from a Polarad spectrum analyzer with an STU-1A RF tuning unit. The short term drift range or the bandwidth integrated over the short term drift is approximately 50 MHz.

Signal-to-Noise Improvement

A number of variations of the equipment were tried. With the HCN laser operating at 890,758 MHz the following signals were used:

Tube	Harmonic
Varian VC714	6
OKI 90V10	10
OKI 70V11	12
Varian VC715	7

Also, the seventh harmonic of a Varian VC715 tube was mixed with the 960 GHz line from an HCN laser. All of these changes were made with the intention of improving the signal-to-noise ratio out of the IF amplifier. It was found, however, that the primary limitation on the signal-to-noise ratio was due to saturation in the IF amplifier. When that problem was eliminated, then the signal-to-noise ratio was satisfactory.

Mechanical Isolation

In an attempt to narrow the width of the beat between the laser and the submillimeter frequency standard, modifications were made to isolate the laser from mechanical vibrations. The laser was isolated from the floor with six inner tubes. Any coupling of vibrations through hose connections was also eliminated by turning off the sources of vibration and by isolating them. The use of geophones showed that the vibrations were reduced; however, no change occurred in the beat frequency signal.

Discharge Stability

Another suspected source of the beat frequency drift was the instability of the discharge and the current. The change from a mixture of natural gas and nitrogen to a mixture of ethyl ether and ammonia resulted in a significant change in the discharge stability. The instability of the discharge caused it to appear uniform. The appearance of striations indicates the temporal stability of the discharge with the ethyl ether-ammonia mixture. In addition to the visual stability of the discharge, the stability of the current was examined. Neither improvement in the current nor the discharge resulted in a change in the beat frequency signal.

Breakdown

The laser used in these experiments had been operated with the mirrors in the gas mixture atmosphere. Whenever a voltage above approximately 200 V

was applied to the piezoelectric translator, on which the output mirror was mounted, breakdown would occur. That problem was circumvented by making a Brewster window from polypropylene and operating the output mirror in the atmosphere. With the Brewster window system, the beat frequency signal-to-noise ratio was reduced by a factor of two and the breakdown voltage was g greater than 1500 V.

Automatic Frequency Control of the Hydrogen Cyanide Laser

In order to achieve an improvement of the stability of a far infrared laser, it is also possible to consider automatic frequency control of these devices. With such techniques, stabilities on the order of a few parts in 10^8 can be readily attained. The phase-locking of the laser to the harmonic of the microwave source requires further work; however, with equipment similar to that described in the previous section, the laser has been locked by AFC methods to the generated millimeter signal.

The AFC of the HCN laser has been performed on the 890 GHz line by beating the laser with the 12th harmonic of an OKI 70 V10 to produce a 60 MHz IF signal. Figure 1 shows the experimental apparatus. The 60 MHz signal was amplified, and a discriminator response was produced in a Micro-Now Model 201. The output signal is amplified in a DC amplifier and applied to the piezoelectric element controlling the position of one of the laser mirrors. Under AFC, a significant change in the characteristics of the 60 MHz beat is observed on the spectrum analyzer. Vibrational effects, which are evident in the unlocked free-running operation, are no longer present in the AFC mode. By tuning the microwave harmonic, it is possible to tune the laser frequency. A variation of the harmonic of the frequency standard by as much as 600 kHz shows no variation in the frequency of the IF, indicating that the laser is following the microwave signal when AFC is employed. No significant change in the linewidth is observed as it remains on the order of 10 kHz; however, the signal remains well within the 10 kHz for periods of several hours. This corresponds to a stability on the order of a part in 10^8 for this period.

Further improvement of this figure is evident from recent investigations. In order to provide an improvement in linewidth during the unlocked operation, a laser, 9 feet in length, with 4-inch glass tubing has been assembled. Invar rods are employed to maintain the separation of the laser mirrors. The laser mirrors have 3.875 inch diameters, a 24-foot radius and 0.090 inch holes for coupling out the energy. With this laser, linewidths of approximately 5 kHz have been observed on the Marconi spectrum analyzer. With this laser, it should be possible to realize stabilities on the order of 5 parts in 10^9 by AFC techniques.

Several difficulties still remain in the overall system, which have delayed the phase-locking of the laser and make the AFC technique difficult. Improvement of the piezoelectric elements is necessary. Improved sensitivity and lack of resonances is required. With the original mirrors, made of 2.875 inch diameter, 0.500 inch thick quartz, a 500 volt signal from the DC amplifier resulted in a 150 kHz swing up to 700 Hz (frequency input) but produced a resonance at 450 Hz with a 300 kHz swing. In order to move the resonance to a higher frequency, a 0.250 inch thick, 2.938 inch diameter flat aluminum plate was fabricated. This lighter plate moved the resonance to 680 Hz. Modulation was possible to 1500 Hz. It would be desirable to have the resonance at a higher frequency (2 kHz or higher) so that a flat frequency response up to this frequency is available.

Investigations are now being performed to provide an accurate quantitative figure for the stability. Counting techniques and discriminator calibration are being considered as methods for providing this figure. When phase-locking is eventually achieved, the desirable technique will be the beating of two lasers locked to a microwave source.

Conclusions and Recommendations for Future Work

The problem of phase locking a laser signal to a submillimeter frequency standard has been investigated analytically and experimentally. In addition, laser stability studies to augment the laser-microwave phase-locking experiment were performed. Improvements in signal-to-noise ratio and laser and microwave stability have been made. Although phase-locking has not been achieved, it appears to be analytically feasible, and no experimental information has indicated an impasse. It is expected, therefore, thatcontinued work in this area would result in the phase-locking of a laser signal to a submillimeter frequency standard.

The HCN laser has been locked to a harmonic of a frequency standard with a long-term stability on the order of a part in 10^8. With linewidths on the order of 5 kHz now being achieved in a new laser, it should be possible to improve this stability by approximately an order of magnitude.

On the basis of the work performed and the prospects for locking a far infrared laser, work should be continued on the frequency control of infrared lasers. The following investigations warrant further consideration in future programs:

1. Continued improvement of the spectral purity of the microwave oscillator from which the millimeter wave signal is generated. With the linewidth of the beat signal now reduced to 5 kHz, it is possible that furtner reduction of this width can result from an improved spectrum of the originating source. Two new multiplying chains have been received in the electromagnetics Laboratory and the spectral purity of the signals from these devices will be ascertained in the beating experiments.

2. Improvement of the laser power supplies. Instability and noise can be originating from the power supplies. A current regulator, assembled

to provide further improvement in the discharge stability, has shown no change in the spectral characteristics.

3. With the improved linewidth, a reassessment of the effects of vibrations should be performed. The new system, withinvar rods to minimize the effect of physical motion of the two laser mirrors, will provide a means for investigating these effects.

4. Investigations should be continued to improve the discharge configuration by improving the electrode geometry.

5. Improved mirror holders are needed and will contribute to the stability of the system.

6. An improved means of feeding the control signal to the laser is needed. Thus far, the use of the piezoelectric elements has not been successful, but work will continue on improving this technique. Present plans call for feeding an AFC signal through the power supply while the phase-locking signal is applied to the piezoelectric element on the laser mirror.

7. Methods for enhancing the signal-to-noise ratio of the heterodyne signal between the laser and the microwave source should be studied. The possibility of using a higher frequency millimeter wave source can be investigated. Both a 127 GHz Varian klystron and a 300 GHz carcinotron are available for these studies. Phase-locking of the carcinotron will be necessary. Improvement of the mixer structures could result in an enhanced S/N. The initial use of open structures has proven relatively successful for heterodyning two lasers, and with the appropriate focussing mechanism, should be applicable to the mixing of the laser and microwave signal. A new open-structure mixer will accept two laser signals and a microwave signal.

8. The actual phase-locking of the laser to the multiplied signal should be achieved in the near future. Work should continue on this aspect and should eventually be extended to higher frequency laser systems. The locking of the H_2O laser at 118 microns is the next most likely system to be continued.

9. The development of mixing techniques for use with the CO_2 laser at 10.6 microns is required for the application of frequency control techniques for the near infrared wavelength region, and consideration of these techniques will be made in future investigations.

10. In future work in this laboratory spectroscopic investigations will be performed with the HCN laser. Free radicals and other molecules which have energy levels tunable with magnetic fields will be studied. In addition, a continued effort toward obtaining better detectors, non-linear devices and components will be made. An absorption cell, inserted into the laser interferometer, is being constructed for magnetic resonance of paramagnetic molecules. Experimentation is being performed in preparation for the heterodyning of the $^{13}CO_2$ laser with a harmonic of the carcinotron.

Acknowledgement

This work was supported by the Frequency Control and Solid State Division, United States Army Electronics Command, Fort Monmouth, New Jersey, under Contract DAAB07-68-C-0206 and the Electronics Research Center, National Aeronautics and Space Administration under Contract No. NAS 12-2008.

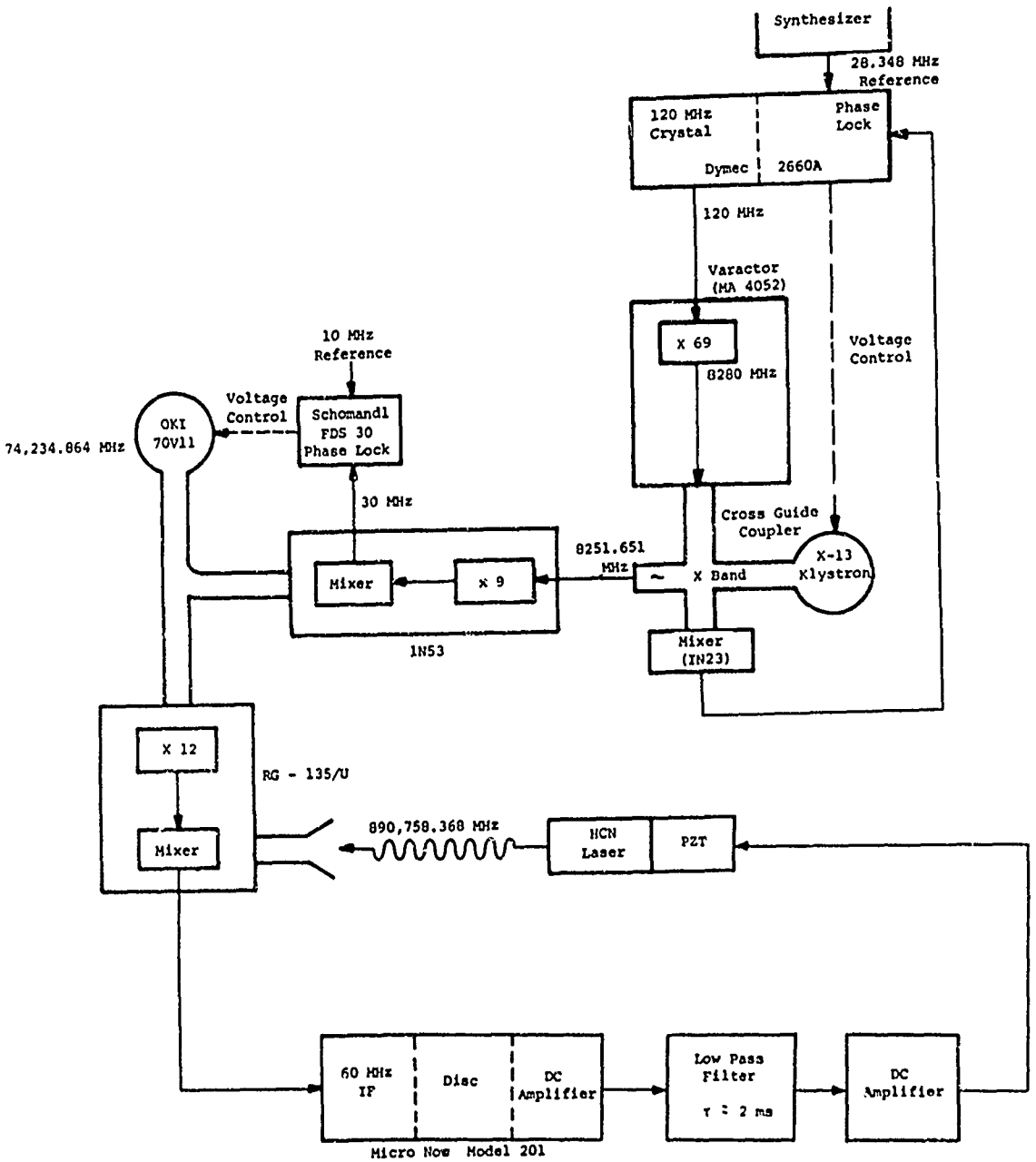
References

1. Hocker, L. O., Javan, A., Rao, D. R., Frenkel, L., and Sullivan, T., "Absolute Frequency Measurement and Spectroscopy of the Laser Transitions in the Far Infrared," Applied Physics Letters 10, 147 (1967).

2. Hocker, L. O. and Javan, A., Physics Letters 26A, 255 (1968).

3. Frenkel, L., Sullivan, T., Pollack, M. A., Bridges, T. J., "Absolute Frequency Measurement of the 118.6 μm Water Vapor Laser Transition," Applied Physics Letters 11, 344 (1967).

4. J. J. Gallagher, R. E. Cupp and R. A. Kempf, "Millimeter and Submillimeter Device Research," Final Report, Contract DA-28-043-AMC-02484(E), January, 1968, OR 9111; J. J. Gallagher, V. J. Corcoran and R. E. Cupp, "Extension of Frequency Control to the Infrared Region," Contract DAAB07-68-C-0206, Semiannual Report, August, 1968, OR 9625.


Synthesizer
28.348 MHz
Reference
120 MHz
Crystal
Phase
Lock
Dymec
2660A
120 MHz
Varactor
(MA 4052)
X 69
8280 MHz
Voltage
Control
10 MHz
Reference
Voltage
Control
OKI
70V11
74,234.864 MHz
Schomandl
FDS 30
Phase Lock
30 MHz
Cross Guide
Coupler
8251.651
MHz
Mixer
x 9
~
X Band
X-13
Klystron
1N53
Mixer
(IN23)
X 12
RG - 135/U
Mixer
890,758.368 MHz
HCN
Laser
PZT
60 MHz
IF
Disc
DC
Amplifier
Low Pass
Filter
τ = 2 ms
DC
Amplifier
Micro Now Model 201

LASER FREQUENCY STABILIZATION USING A PRIMARY FREQUENCY REFERENCE

Shaoul Ezekiel
Massachusetts Institute of
Technology

The use of an atomic or molecular resonance, observed in a beam, as a primary frequency reference has of course been successfully exploited in the long term stabilization of microwave oscillators. At microwave frequencies, the design criterion is the suitability of the molecular beam transition as regards to perturbation by external fields and ease of detection. The local oscillators offer no problems since they have a large tuning bandwidth. To find a suitable molecular beam reference for the stabilization of the laser oscillator is not so simple because of the narrow tuning range of the laser. However, we have reported the observation of resonance fluorescence of iodine in a molecular beam, induced by light from a single longitudinal mode 5145 A argon laser.[1] The nearest I_2 resonance is 600 MHz away from the center of the laser gain curve. The individual width of the iodine resonances has been inferred to be 50 kHz from life time measurement of the excited state of I_2 in the beam. The use of one of the iodine resonances is being exploited as an external frequency reference in the feedback stabilization of the argon laser.[2] The narrow linewidth plus the fact that there is no Stark shift due to collisions or collision broadening in the beam promise a high degree of long term frequency stability and resetability. Calculations based on the signal to noise ratio obtained in the resonance fluorescence data indicate that it would be possible to detect a shift of 200 Hz from the center of the iodine resonance, i.e., 5×10^{-3} of the 50 kHz width, in an integration time of one second. The molecular beam stabilization scheme will be discussed and new data will be presented.

---

[1] S. Ezekiel and R. Weiss, "Laser-Induced Fluorescence in a Molecular Beam of Iodine", Physical Review Letters, January 15, 1968.

[2] S. Ezekiel and R. Weiss, "A Molecular Beam Reference for Laser-Frequency Stabilization", Paper #16P-2, International Quantum Electronics Conference, Miami, Florida, 1968.

PRECISION TIME MEASUREMENTS

OF

OPTICAL PULSARS

P. E. Boynton, R. B. Partridge and D. T. Wilkinson
Princeton University

Recent measurements at Princeton of the optical pulsar in the Crab Nebula (NP 0532) are discussed. Of particular interest is pulse time-of-arrival. Such phase measurements over a long interval lead to estimates of the pulsar period and its time derivatives. Primary concern is with the second derivative whose value would be an important check on recent pulsar models. Timing techniques involve visual cycle identification of the Loran-C one-second tick leading to synchronization (to $\pm$ 20 μs) of a locally generated one-second pulse. The pulsar signal is recovered with the usual signal averaging technique, using a multiscaler scanned at the pulsar frequency.

Preliminary results for the period and first derivative are consistent with numbers derived from radio observation of NP 0532. Further work on ephimeris corrections and additional data reduction is required to extract the second derivative.