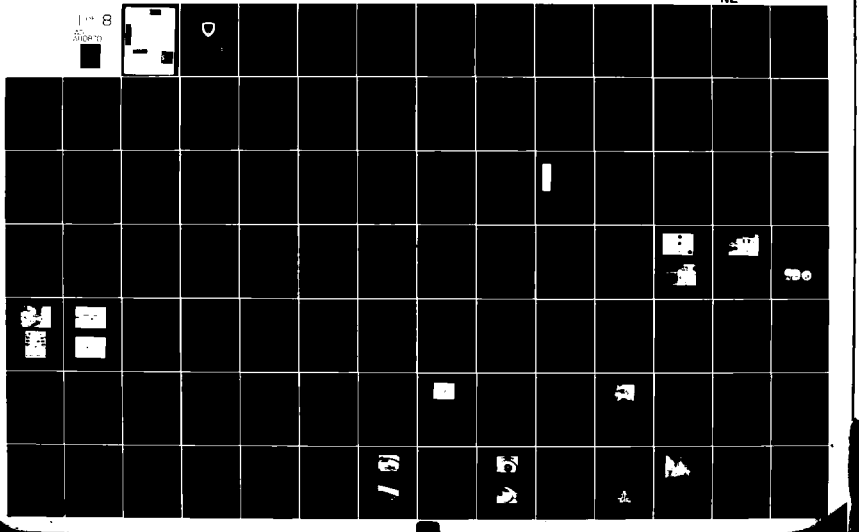
AD-A110 870 ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND AD--ETC F/G 9/5
PROCEEDINGS OF THE 35TH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL, --ETC(U)
1981

UNCLASSIFIED

NL



LEVEL III

AD A110870

DISTRIBUTION STATEMENT
Approved for public release;
distribution unlimited.

DTIC
ELECTE
FEB 12 1982
A

PROCEEDINGS

of the

THIRTY FIFTH ANNUAL FREQUENCY CONTROL SYMPOSIUM

1981

Sponsored By



U.S. ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND
Major General Emmett Paige, Jr.
Commanding

ELECTRONICS TECHNOLOGY AND DEVICES LABORATORY
Dr. C. G. Thornton
Director

© Electronic Industries Association 1981
All rights reserved
Printed in U.S.A.

DTIC
SELECTE
FEB 1 2 1982
A

Copies of the Proceedings are available from:

Electronic Industries Association
2001 Eye Street, N.W.
Washington, D.C. 20006

PRICE PER COPY: $20.00

None of the papers contained in the Proceedings may be reproduced in whole or in part, except for the customary brief abstract, without permission of the author, and without due credit to the Symposium.

27-29 May 1981

Marriott Hotel
Philadelphia, Pennsylvania

Sold by EIA
$20.00
A 21
DTIC COPY INSPECTED

THIRTY-FIFTH ANNUAL FREQUENCY CONTROL SYMPOSIUM

Sponsored By

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

Fort Monmouth, New Jersey

27-29 MAY 1981
Marriott Hotel
Philadelphia, Pennsylvania

SYMPOSIUM EXECUTIVE COMMITTEE

Chairman	Dr. Erich Hafner
Executive Secretary	Mrs. Lee Hildebrandt

TECHNICAL PROGRAM COMMITTEE

Dr. Erich Hafner—Chairman
Dr. Arthur Ballato—Co-Chairman
U.S. Army Electronics Technology and Devices Laboratory

Mr. E. J. Alexander
Bell Laboratories

Dr. J. A. Barnes
National Bureau of Standards

Mr. C. A. Bartholomew
Naval Research Laboratory

Mr. M. Bloch
Frequency Electronics, Inc.

Mr. A. Chi
NASA-Goddard Space Flight Center

Dr. F. Cho
Motorola

Dr. L. Claiborne
Texas Instruments

Dr. L. Cutler
Hewlett-Packard

Mr. M. Frerking
Rockwell International

Dr. H. Hellwig
Frequency & Time Systems

Dr. W. Horton
Piezo Technology, Inc.

Mr. D. Kemper
Tyco Crystal Products

Mr. E. Kentley
Roditi, Inc.

Mr. J. Kusters
Hewlett-Packard

Mr. T. Lukaszek
U.S. Army, ERADCOM

Dr. T. R. Meeker
Bell Laboratories

Mr. D. Reifel
Motorola

Dr. R. Rosenfeld
Sawtek, Inc.

Mr. S. Schodowski
U.S. Army, ERADCOM

Dr. J. Vig
U.S. Army, ERADCOM

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

Dr. N. Yannoni
RADC/Air Force

TECHNICAL SESSION CHAIRMEN

PLENARY SESSION

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

RESONATOR PROCESSING & ACCELERATION RESISTANT CRYSTAL UNITS

Mr. D. Reifel, Motorola

MINIATURE RESONATORS & RESONATOR THEORY

Dr. E. R. EerNisse, Quartex, Inc.

FILTER CRYSTALS & CRYSTAL MEASUREMENTS

Dr. T. R. Meeker, Bell Laboratories

CRYSTAL SYNTHESIS & FUNDAMENTAL PROPERTIES OF QUARTZ

Dr. R. Kinloch, Sawyer Research Products

SURFACE ACOUSTIC WAVE DEVICES

Dr. L. Claiborne, Texas Instruments

FREQUENCY GENERATION & FREQUENCY STABILITY

Dr. S. Stein, National Bureau of Standards

FUTURE FREQUENCY CONTROL NEEDS & FREQUENCY & TIME DISTRIBUTION

Mr. M. Bloch

Panel Discussion: The Future of the Crystal Industry—World Views.
Moderator: Dr. D. Hammond, Hewlett-Packard

ATOMIC & MOLECULAR FREQUENCY CONTROL DEVICES

Dr. L. Cutler, Hewlett-Packard

SEMINAR—Properties of Signal Sources and Measurement Methods
Drs. D. Howe, D. Allan and J. Barnes, National Bureau of Standards

TABLE OF CONTENTS

Opening Remarks
- Major General Emmett Paige, Jr., U.S. Army Electronics Research and Development Command Electronics Technology and Devices Laboratory. 1

Plenary Session

A History of the Quartz Crystal Industry in the USA
- V. E. Bottom 3

The Role of Analog Devices in a Digital Age
- R. C. Williamson, MIT Lincoln Laboratory 13

Nonlinear Properties of Quartz Crystal and Quartz Resonators: A Review
- J. J. Gagnepain, CNRS 14

The Effect of Vibration on Frequency Standards and Clocks
- R. L. Filler, USAERADCOM 31

Resonator Processing

Vacuum Processing System for Quartz Crystal Resonators
- J. M. Frank, General Electric Company 40

Metallization of Quartz Oscillators
- A. T. Lowe, Motorola 48

X-Ray Goniometry of the Modified Doubly Rotated Cuts
- E. Knolmayer, Quarzkeramik GmbH 56

An Instrument for Automated Measurement of the Angles of Cut of Doubly Rotated Quartz Crystals
- J. L. Chambers, M.A. Pugh and S. T. Workman, Advanced Research and Applications Corporation; R. W. Birrell and R. J. Valihura, P. R. Hoffman Company 60

Force-and Acceleration- Frequency Effects in Grooved and Ring- Supported Resonators
- M. Nakazawa, Princeton University, T. Lukaszek and A. Ballato, USAERADCOM 71

Acceleration Resistant Crystal Units

Update of SC Cut Crystal Resonator Technology
- B. Goldfrank, J. Ho and A. Warner, Frequency Electronics, Inc. 92

Design of High Performance SC Resonators
- R. W. Ward, Colorado Crystal Corporation. 99

Adjusting the Frequency VS. Temperature Characteristics of SC-Cut Resonators by Contouring
- J. R. Vig, W. Washington and R. L. Filler, USAERADCOM. 104

The Acceleration and Warmup Characteristics of Four-Point-Mounted SC and At-Cut Resonators
- R. L. Filler and J.R. Vig, USAERADCOM 110

Reduction of the Effects of Vibration on SC-Cut Quartz Crystal Oscillators
- V. R. Rosati and R. L. Filler, USAERADCOM. 117

Indirect Amplitude Frequency Effect in Resonators Working on Two Frequencies
- J. P. Valentin, C. P. Guerin, R. J. Besson, ENSMM 122

Miniature Resonators

Quartz Tuning Fork Crystal Using Overtone Flexure Modes
- S. S. Chuang, ASU R&D Corporation. 130

A Miniature Quartz Resonator Vibrating at 1 MHz
- R. J. Dinger, ASULAB S.A. 144

Investigation of Spurious Modes of Convex DT-Cut Quartz Crystal Resonators
- T. Adachi, Y. Tsuzuki, Yokohama National University and C. Takeuchi, Fujitsu Ltd.. . 149

The Edge Mode Resonator
- D. C. L. Vangheluwe and E. D. Fletcher, N. V. Philips Gloeilampenfabrieken, Eindhoven, The Netherlands, Philips Research Laboratories 157

4.19 MHz Cylindrical AT-Cut Miniature Resonator
- S. Okano, T. Kudama, K. Yamazaki and H. Kotake, Toyo Communication. 166

Resonator Theory

Three-Dimensional Mode-Matching Theory of Rectangular Bar Resonators Using Complex Wave-Numbers
- R. F. Milsom, D. T. Elliott, M. Redwood, Philips Research, University of London. . . 174

Simple Model For an AT Cut Rectangular Quartz Plate
- J. H. Balbi and M. I. Dulmet, C.N.R.S. 187

Stresses in Rectangular Cantilever Crystal Plates Under Transverse Loading
- P. C. Y. Lee, and C. S. Lam, Princeton University 193

An Analysis of SC-Cut Quartz Trapped Energy Resonators with Rectangular Electrodes
- D. S. Stevens and H. F. Tiersten, Rensselaer Polytechnic Institute. 205

Stress Compensated Orientations for Thickness-Shear Quartz Resonators
- B. K. Sinha, Schlumberger-Doll Research. 213

Extensional Vibrations of Rectangular Crystal Plates
- P. C. Y. Lee, M. Nakazawa and J. P. Hou, Princeton University. 222

Filter Crystals

Coupled Thickness-Shear and Thickness-Twist Resonances in Unelectroded Rectangular and Circular AT-Cut Quartz Plates
- H. F. Tiersten, Rensselaer Polythechnic Institute and R. C. Smythe, Piezo Technology Incorporated. 230

Air-Gap Probe Evaluation of Thin Quartz Plates
- L. Dworsky and G. Kennedy, Motorola 237

Laser Processed VHF Monolithic Crystal Filters With On Plate Integrated Matching Impedances
- R. Lefevre, C.N.E.T. 244

Unwanted Modes in 5° X-Cut Crystal Units
- J. J. Boyer, Bell Telephone Laboratories 250

Equivalent Circuit Modeling of Stacked Crystal Filters
- K. M. Lakin, Iowa State University 257

Crystal/Oscillator Measurements

The Quartz Resonator Automatic Aging Measurement Facility
- D. E. Beetley, B. R. Blitch and T. M. Snowden, General Electric Company 263

Comparison of Methods for Measurement of Quartz Crystal Resonators With Load Capacitance
- W. H. Horton, T. S. Payne, R. C. Smythe and D. A. Symonds, Piezo Technology Inc.. . 271

An Automated Resonator Measurement System Using a Reflection Coefficient Bridge
- R. C. Smythe, Piezo Technology. 280

Implementation of an Automatic Microcircuit Measuring System for Quartz Crystals
- G. J. Malinowski, USAERADCOM; G. L. Snider and C. Nyholm, Hughes Aircraft. 286

Crystal Synthesis

Recent Progress on Aluminum Phosphate Crystal Growth
- E. D. Kolb and R. A. Laudise, Bell Laboratories 291

Initial Results With The Air Force Hydrothermal Facility
- A. F. Armington, J.J. Larkin, J. J. O'Connor, J. A. Horrigan, RADC 297

Synthetic Quartz Crystals Grown in NaCl, KCl Solutions and Pure Water, and Their Low Temperature Infrared Absorption
- M. Hosaka and S. Taki, Yamanashi University and K. Nagai and J. Asahara, Toyo Communication . 304

The Influence of the Quality Factor of Quartz on Some Device Properties
- J. C. Brice, E.D. Fletcher, Philips Research Laboratories J. Dowsett Cathodean Crystals . . 312

Fundamental Properties of Quartz

Point Defects in Cultured Quartz: Recent Acoustic Loss, Infrared, and Magnetic Resonance Results
- J. J. Martin, L. E. Halliburton, and R. B. Bossoli, Oklahoma State University 317

Radiation-Induced Conductivity and High Temperature Q Changes in Quartz Resonators
- D. R. Koehler, Sandia Laboratories 322

High Temperature Resonance Loss and Infrared Characterization of Quartz
- H. G. Lipson, A. Kahan, R. N. Brown and F. K. Euler, RADC. 329

Quartz Crystal Oscillator at Cryogenic Temperature
- B. Komiyama, Radio Research Laboratories. 335

Quartz Resonator Thermal Transient Due To Crystal Support
- D. Janiaud and M. Valdois, ONERA; R. Besson, ENSMM and J. J. Gagnepain, CNRS. 340

Surface Acoustic Wave Devices

Commercial Satellite Navigation Using Saw Oscillator
- B. Y. Lao, N. J. Schneier, D. A. Rowe, R. E. Dietterle, Magnavox; J. S. Schoenwald, E. J. Staples, J. Wise, Rockwell International Corporation 345

UHF Voltage-Controlled Narrow-Bandwidth Saw Filters
- J. Henaff, M. Feldmann and M. Carel, CNET 349

Saw Filter Technology and Applications
- B. R. Potter and D. B. MacDonald, Texas Instruments 352

The Status of Magnetostatic Wave Devices
- J. M. Owens, C. V. Smith, Jr. and R. L. Carter, University of Texas 358

Direct Frequency Crystal Oscillators
- L. Bidart and J. Chauvin, CEPE. 365

The Propagation Characteristics of Surface Acoustic Waves on Singly and Doubly Rotated Cuts of Quartz
- D. F. Williams and F. Y. Cho, Motorola; A. Ballato and T. Lukaszek, USAERADCOM. 376

Absolute and Differential Aging of SAW Resonator Pairs
- J. S. Schoenwald, J. Wise and E. J. Staples, Rockwell International 383

Elements of SAW Resonator Fabrication and Performance
- W. J. Tanski, Sperry Research Center 388

GaAs SAW Resonator Oscillators with Electronic Tuning
- M. Gilden and T. W. Grudkowski, United Technologies Research 395

Shallow Bulk Acoustic Waves in Berlinite
- K. F. Lau, K. H. Yen, R. B. Stokes, R. S. Kagiwada, TRW and B. H. T. Chai, Allied Chemical Corporation 401

Frequency Generation

Direct Digital Frequency Synthesis
- A. L. Bramble, USAERADCOM 406

A Low Noise 500 MHZ Frequency Source
- A. Vulcan and M. Bloch, Frequency Electronics and W. Tanski, Sperry Research 415

Spurious Suppression in Direct Digital Synthesizers
- C. E. Wheatley, III and D. E. Phillips, Rockwell International. 428

Advanced SAW-LSI Frequency Synthesizer
- D. J. Dodson, M. Y. Huang, M. D. Brunsman, TRW 436

SC-Cut Quartz Crystal Units in Low-Noise Oscillator Application at VHF
- D. J. Healey, III, S. Y. Kwan, Westinghouse Electric Corporation 440

Digital Temperature Compensation of Crystal Oscillators
- A. M. Renard and K. Barnhill, Raytheon Company 455

Frequency Stability

Characterization of Frequency Fluctuations by Crosscorrelations and By Using Three or More Oscillators
- J. Groslambert, D. Fest, M. Olivier, J. J. Gagnepain, C.N.R.S.. 458

An Ultra-High Resolution Frequency Meter
- J. J. Snyder, NBS. 464

A Modified "Allan Variance" With Increased Oscillator Characterization Ability
- D. Allan and J. Barnes, NBS 470

Relation Between 1/f Noise and Q-Factor in Quartz Resonators at Room and Low Temperatures, First Theoretical Interpretation
- J. J. Gagnepain, J. Uebersfeld, G. Goujon, CNRS and P. Handel, Univ. of Missouri. 476

1/f Frequency Fluctuation of a Quartz Crystal Oscillator and Temperature Fluctuation
- Y. Noguchi, Y. Teramachi and T. Musha, Tokyo Institute of Technology. 484

Frequency Retrace of Quartz Oscillators
- F. Euler and N. Yannoni, RADC 492

Future Frequency Control Needs

Frequency Control Requirements for 800 MHZ Land Mobile Communication
- R. Kinsman and D. Gunn, Motorola. 501

Frequency Stabilization Requirements for Modern Millimeter Wave Systems
- A. Tirkel, TRW Defense and Space Systems Group 511

Frequency Stability Requirements for a 95 GHz Instrumentation Radar System
- D. N. McQuiddy, Jr., Texas Instruments 516

Optical Frequency Control for Wavelength Multiplexed Systems
- F. Welsh and T. Stakelon, Bell Labs. 525

Frequency and Time Distribution

NAVSTAR Global Positioning System (GPS) Clock Program: Present and Future
- D. M. Tennant, Air Force Systems Command Space Division/YEZ 532

Time Dissemination Using NAVSTAR Global Positioning System (GPS) Phase IIB User Equipment
- M. D. Yakos and E. H. Hirt, Rockwell International/Collins Division 537

Construction and Performance Characteristics of a Prototype NBS/GPS Receiver
- D. D. Davis, M. Weiss, A. Clements and D. Allan, NBS 546

The NATO III 5 MHZ Distribution System
- A. Vulcan and M. Bloch, Frequency Electronics, Inc. 553

Low Noise Buffer Amplifiers and Buffered Phase Comparators for Precise Time and Frequency Measurement and Distribution
- R. A. Eichinger, Bendix Field Engineering Corporation 565

Panel Discussion

The Future of the Quartz Crystal Industry - Worldwide. Moderator: D. L. Hammond, Hewlett-Packard

Panelists' Remarks
- Arthur Ballato, Electronics Technology & Devices Laboratory, USAERADCOM 576
- Juergen H. Staudte, Staudte Engineering 583
- W. H. Horton, Piezo Technology Inc. 592
- Toshiaki Takeuchi, Nihon Dempa Kogyo Co., Ltd 593
- Dr. Rudolf Fischer, K V G Germany 595

Atomic & Molecular Frequency Control Devices

Mono-Ion Oscillator as Potential Ultimate Laser Frequency Standard
- H. Dehmelt, University of Washington 596

Proposed Stored $^{201}Hg^{+}$ Ion Frequency Standards
- D. J. Wineland, W. M. Itano, J. C. Bergquist, and F. L. Walls, NBS 602

Optical Pumping by Lasers in Atomic Frequency Standards
- L. Lewis and M. Feldman, NBS 612

Preliminary Investigation of a New Optically Pumped Atomic Rubidium Standard
- M. Feldman, J. C. Bergquist, L. L. Lewis, and F. L. Walls, NBS. 625

Development of a Sapphire Lamp for Use in Satellite-Borne Atomic Rubidium Clocks
- T. C. English and E. Jechart, Efratom 637

A Miniature, High-Performance Rubidium Frequency Standard
- T. Hashi, K. Chiba and C. Takeuchi, Fujitsu Limited 646

Performance of the GPS Cesium Beam Frequency Standard in Orbit
- M. W. Levine, Frequency and Time Systems 651

Long Term Performance of VLG-11 Masers
- J. White and K. McDonald, NRL 657

Feasibility of Extremely Small Hydrogen Masers
- H. E. Peters, Sigma Tau Standards Corporation 662

Appendix---Table of Contents 667

Properties of Signal Sources and Measurement Methods
- D. A. Howe, D. W. Allan, and J. A. Barnes . . . . . A-1(669)

AUTHOR INDEX 717

SPECIFICATIONS AND STANDARDS GERMANE TO FREQUENCY CONTROL. 718

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

OPENING REMARKS

Major General Emmett Paige, Jr.
U.S. Army Electronics Research and Development Command
Adelphi, MD 20783

On behalf of the United States Army Electronic Research and Development Command, parent command of the sponsoring Electronics Technology and Devices Laboratory, I take great pleasure in welcoming you all here, and formally open this Symposium.

This is the Thirty-Fifth Annual Symposium on Frequency Control, and since thirty-five years is a mile-stone, I'd like to use that motif to frame my remarks this morning by literally taking a page out of a book - in this case, from the first published proceedings volume of twenty-five years ago.

In the opening remarks delivered at that time - May, 1956 - Brigadier General Earle F. Cook said, "This being the Tenth Annual Symposium, my feeling is that it is getting to be a permanent affair". His prediction is turning out to be pretty accurate!

Since its inception in 1947 in a conference room at Ft. Monmouth, attended by a handful of contractors, through the 1956 meeting with an attendance of about one hundred, this symposium has continued to grow - in numbers and stature. Today it is the world's largest symposium in its field, with over 800 participants from all over the globe.

One change that has taken place in these years has been the steady internationalization of the symposium. From 1956 until 1966 the number of international authors hovered around 5%. It has increased with the years to an average of 40%, indicating the technical strengths of the programs, and the world-wide interest in their contents.

Another change that has taken place over the years has occurred in the market place - and this brings me to the heart of my anniversary comparisons.

In 1956 the military share of the quartz crystal and frequency control market was very large - perhaps half. Today it is less than 2%. Why has this come about? One reason is that the commercial market has opened up for crystals, first for color TV sets, then for CB radios, and most recently for wristwatches, and for microprocessors; the latter for use in everything from games to automobile electronics to industrial process control. Another reason for the shrinkage in the military market has been the development of sophisticated means of frequency synthesis, whereby one reference crystal is used to derive a multitude of accurate frequencies.

Does the small military share of the current frequency control market indicate a dwindling military interest in the field? My answer is an emphatic "NO". Modern C^3, IFF, navigation, surveillance, and EW systems are dependent upon ultra-precise frequency control. The frequency control elements provide the electronic heartbeats of today's time-ordered systems.

Although, from the standpoint of numbers of crystals used, the non-commercial market seems insignificant, let me paraphrase another statement made in the remarks of 1956: "Do not take too lightly today's accuracies. The military requirements will always be one step ahead of you". That statement is as true now as twenty-five years ago.

Consequently, I don't think it is at all inaccurate to say, that the low cost color TV crystal was made possible by developments stemming from the largely military R&D reported in earlier symposium proceedings.

Military requirements have always paced the state-of-the-art and have been the technology drivers, perhaps more so in this industry than in others because of its relatively small size - compared, let's say, with microelectronics.

Let me wind up with a few final comparisons regarding requirements:

In 1956, systems were fixed - carrier frequency; today spread-spectrum in the word, with consequent needs for stabilities 1000 to 10,000 times greater than then.

In 1956, radiation hardness was not a consideration; today it is, and getting frequency shifts less than 10^{-11} per rad translates to a need for better quartz.

In 1956, shock and vibration was of little consequence; today we want frequency shifts less than 10^{-11} per g or smaller, and units that can take the 20,000 g shock of cannon launch.

Meeting these requirements today will mean considerable commercial spin-offs tomorrow - that's the good news; the bad news is that when tomorrow comes, we will take the military requirements for granted.

Judging from your many successes to date, I know you will come through. And now I look forward to hearing what technical treats you have in store for us this year. Thank you.

Pr... 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

A HISTORY OF THE QUARTZ CRYSTAL INDUSTRY IN THE USA

Virgil E. Bottom

3441 High Meadows Dr.
Abilene, Texas
79605

Summary

With the aid of many of the people who were involved, this paper is an attempt to record the history of the quartz crystal industry in the USA. Although crystal units were in use at least a decade earlier, the beginning of the quartz crystal industry in the USA can logically be dated November 1941 when the Quartz Crystal Section was organized in the Office of the Chief Signal Officer.

The history of the industry is traced from its origin with the discovery of Piezoelectricity by the Curie brothers in 1880-81 to the present time. Special emphasis is placed upon the problems which were encountered and overcome in creating an industry capable of meeting the demands of the Armed Services of the United States and their Allies during WW II.

The record of cooperation among the large and small industries, governmental agencies, the armed services, universities and individual citizens inspires confidence that the most critical problem can be solved with cooperation, dedication and effort.

Introduction

The August 25, 1943 issue of LIFE Magazine carried the following letter to the editor:

> Sirs: In proportion to size these little glass like quartz wafers are perhaps the most remarkable of all the tools science has given to war. When the story of the almost incredible progress in research and manufacture of radio crystals can be told, it will prove to be a tale of one of the war's greatest achievements. No less significant will be the fruit of these advancements to a new world at peace where crystals will be the vibrating hearts of most telecommunication equipment. Gerald James Holton, Harvard University, Cambridge, Mass.

The "story" to which Prof. Holton so prophetically referred has not been told and perhaps can never be told in detail. MacLaurin, in the book INVENTION AND INNOVATION IN THE RADIO INDUSTRY, one of the books in the MIT studies of innovation, published in 1949, does not even mention piezoelectricity or the quartz crystal unit. But Prof. Holton's prophecy concerning the peace time uses of "those little glass like wafers" has been fulfilled far beyond anything that he could have foreseen.

The fortieth anniversary of the entrance of the United States into the great war known as World War II provides an appropriate occasion to review the "incredible progress" to which Dr. Holton referred even though quartz crystal units were known and in limited use more than a decade before that date. The 1939 decision of the Armed Services of the United States to convert its radio equipment to crystal control resulted in the creation of an industry which ultimately played an important role in the victory of the Allied Forces over the Axis Powers.

The years 1980-81 are also the centennial of the years in which the Curie brothers discovered the piezoelectric effect; the direct effect in 1880 and the converse effect in 1881.

It is fitting, therefore, that we should at this time look back upon the development of an industry which, in many respects, is unique; an industry staffed for the most part by people who were not trained scientists or engineers. On the contrary they were bartenders, lamp shade manufacturers, wood workers, stone cutters, mechanics and, above all, amateur radio operators or "hams". With a few professionals, most of whom knew little more than they about piezoelectricity and with a totally unsatisfactory basis of scientific information, these entrepreneurs tackled a job which would have taxed the ability of the most advanced and best equipped industrial concern in the nation. And they carried it off. But let us start at the beginning.

Early History of Piezoelectricity

Very soon after its discovery the Curies devised several instruments utilizing the piezoelectric effect. One of these was the piezoelectric voltmeter. Another was the piezoelectrometer which later became the basic instrument used by Pierre and Marie Curie in their work which led to the discovery of Radium. Otherwise, for more than three decades the piezoelectric effect remained little more than a laboratory curiosity. Further developments had to await the invention of the triode vacuum tube.

After the Curies the first application of the piezoelectric effect was made by Prof. P. Langevin in France in 1917. Langevin used X-cut plates of quartz to generate and detect sound waves in water. His object was to provide a means for detecting submarines and his work led to the development of SONAR and to the science of ultrasonics, the results of which are still making headlines (viz., the recent announcement of the successful use of ultrasonic imaging in mammography).

Langevin's work stimulated others to investigate

the phenomenon of resonance in piezoelectric crystals. Among those who became interested were A. M. Nicholson of the Bell Telephone Laboratories and Prof. W. G. Cady at Wesleyan University. Both men, working with Rochelle salt, observed the reaction of the resonant piezoid on the driving circuit and both applied for patents based upon their observations. Subsequent litigation resulted in a legal decision in favor of Nicholson who is therefore considered to be the inventor of the piezoelectric oscillator.

In 1919 Cady use a quartz piezoid to control the frequency of an oscillator and in a series of papers during the next three years he described the use of quartz bars and plates as frequency standards and wave filters. It is generally accepted that Cady was the first to use a quartz piezoid to control the frequency of an oscillator circuit.

Both Nicholson and Cady used devices which would today be called monolithic resonators having two sets of electrodes on the same crystal. It remained for Prof. G. W. Pierce of Harvard University to show, in 1923, that a quartz plate with only one set of electrodes could be made to control the frequency of an oscillator circuit using only one vacuum tube. Pierce's circuit has probably been used more than any other quartz crystal oscillator circuit.

Oddly enough, none of the circuits was understood by its inventor until Prof. K. S. Van Dyke, a student and colleague of Cady, showed in 1925 that the two electrode piezoelectric resonator is the electrical equivalent of a series resonant circuit shunted by a capacitor. He was able to relate the electrical parameters of the equivalent circuit to the physical properties of the piezoid. This information in the hands of Vigoreux, Wright, Terry and others led to the understanding of the piezoelectric resonator or piezoid and to its use both as a passive and as an active circuit element.

The importance of the discoveries of Cady, Nicholson and Pierce did not go unnoticed. In 1923 the Bell Telephone Laboratories established a quartz laboratory and the General Electric Company did likewise the following year. One of the individuals who recognized the potential of the quartz crystal unit was August E. Miller. In 1923 Miller left the optical business where he had become an expert in grinding quartz lenses to go into the business of making quartz crystal blanks for amateur radio operators or "hams"; the only market which then existed for the new device. It appears that "Augie" Miller may have been one of the first individuals to go into the business of making quartz crystal units.

In 1926 the A. T. & T. radio station WEAF in New York City became the first radio station in the United States to control its frequency with a quartz crystal unit. Within a few years all radio stations went to crystal control thus providing another small market for quartz crystal units.

Before 1926 all crystal units were X-cut bars or plates. In that year E. D. Tillyer of the American Optical Co. discovered the Y-cut. The temperature-frequency coefficient of the Y-cut is about +100 ppm per C^{o} whereas that of the X-cut is about -20 ppm per C^{o}. This fact naturally suggested the possibility of making piezoids, the frequencies of which would be independent of temperature and so work was started in several places to exploit the idea.

During the year 1927 Prof. Gerald Fox of the University of Iowa spoke to a convention of "hams" on the topic "The Piezoelectric Properties of Quartz". His talk stimulated some of the hams to try to make their own crystal units. One of them was E. L. (Al) Shideler of Manson, Iowa who not only succeeded in doing so but continued to do so until he retired forty years later. Another was Bill Peterson of Council Bluffs, Iowa who founded the Peterson Radio Co. and continued to make crystal units until his recent death. Still another was Herb Hollister of Merriam, Kansas who continued to make crystal units until the end of WWII. This list is doubtless incomplete.

The interest of radio amateurs and radio broadcasters created a mini-market for crystal units and several small businesses were established to satisfy the demand. One of these was the Miller Laboratories started by August Miller in the year 1928.

In January 1930 QST published an article by J. Herbert (Herb) Hollister entitled "Debunking Crystal Control" in which he pointed out "its utter simplicity" and described how a ham could make his own crystal unit. During the same year F. Dawson Bliley founded the Bliley Electric Co. to supply crystal units mainly to the amateur market. Today the Bliley firm is headed by sons of the founder and continues to be one of the most respected companies in the business.

In 1931 Herbert Blazier founded Monitor Piezo Products Co. and began to supply crystal units to hams, radio stations and two-way communication systems on the West Coast. Monitor continues today under the direction of John W. Blazier, son of the founder of the company. Also in 1931, Theodore S. Valpey established the Valpey Crystal Co. which still continues as the Valpey-Fisher Corp. under the direction of the son of the founder, Mr. Ted Valpey, Jr.

In 1931, Dr. W. A. Parlin, Professor of Physics at Dickinson College in Carlisle, PA and three of his students, Ed Minnich, Howard Bair and Charles Fagan decided to try to make crystal units for their ham rigs. They were assisted by Grover C. Hunt, an amateur lapidist who was employed by the college as an engineer. Not only were they successful but in the following year Hunt produced the first commercial units to be made in the Carlisle area. All of the quartz crystal operations in the Carlisle area are direct descendents of Hunt's original efforts.

In 1932 Leon Faber, who had taken up the hobby of amateur radio in 1913, began to make crystal units in Sandwich, IL. His first customers were other hams. When WW II broke out Faber formed a partnership with James Knights who owned and operated a battery shop in Sandwich. Their efforts produced the James Knights Co. which became one of the largest crystal manufacturers in the world. It continues today as the CTS-Knights Co.

The crystal units of the day were either X- or

Y-cut plates mounted loosely between two metal plates which served as electrodes. With suitable temperature control frequency stabilities of a few ppm could be achieved. This was a vast improvement over the free running oscillators of the time but the temperature control equipment was expensive, cumbersome and power consuming. Furthermore, the construction of the units made them susceptible to the effects of vibration and precluded their use in aircraft and other mobile systems.

Meanwhile work had continued to try to find ways of making units which would be less susceptible to the effects of temperature changes. In 1929 W. A. Marrison of the Bell Telephone Laboratories described the first quartz piezoid with a low temperature coefficient. It was made in the form of a doughnut and was later used as the isochronous element of the first crystal clock. It did not, however, prove practical for general use. Efforts to develop a practical unit having a low frequency-temperature coefficient were successful in 1934 when the AT- and BT-cuts were discovered independently by Koga in Japan, by Bechmann and Straubel in Germany and by Lack, Willard and Fair in the United States. Two years later Baldwin and Bokovoy of RCA introduced the V-cut.

Shortly thereafter G. W. Thurston of the Bell Telephone Laboratories discovered that a quartz plate of the rotated Y-cut family could be rigidly supported between two metal plates by clamping it at a few points near the edge and thus providing an air space between the quartz plate and the metal electrode. Although slightly more difficult to produce, the clamped, low temperature coefficient plates were suitable for mobile applications and this increased the market for crystal units.

In response to the new market several new companies were formed. Among them was Standard Piezo Co., the first crystal company in the Carlisle, PA area; organized by Grover Hunt and Linwood Gagne in 1935. In the same year the General Electric Co., which had established a quartz laboratory in 1924, decided that the crystal unit was ready for commercial production and assigned it to the G. E. Laboratories at Schenectady, NY under the direction of C. F. Baldwin. In addition to units produced for internal use, a few were sold on the open market.

One of the most important influences on the development of the crystal industry in the period 1935-40 was the Galvin Mfg. Co. (Motorola). Mr. Dan Noble, ex-professor of Electrical Engineering at the University of Connecticut, was convinced that crystal control was essential to effective two way radio communication. In order to obtain the crystal units which he needed for his new two-way systems he persuaded Mr. Galvin to place orders with and in some cases to subsidize a number of individuals who had some experience with crystal units. Among these were E. L. Ghideler of Fort Dodge, IA, Herb Hollister of Wichita, KS and "Ted" Tedford of Cincinnati, OH.

Production methods in use at the time were quite primitive. Orientation was done with reference to the natural faces of the crystal and consequently only faced quartz crystals could be used. The usual procedure was to determine the orientation with respect to a rhombehedral face using a protractor to measure the angular position. Sawing was done with a rotating blade in a slurry of silicon carbide and oil or water. The blades were usually of copper although other metals were used and at least one innovator of the day used old 78 rpm phonograph records. The crystal was held at the approximate angle by cementing it to a glass plate with wax or plaster Paris and a test cut was made. The resulting plate was lapped by hand until it would oscillate when placed between metal electrodes, the upper one of which was called a "penny plate" because of its shape and size. Once oscillation was obtained, the temperature was changed by means of a hair dryer or with dry ice and the direction of frequency change noted. Then the saw, which had remained idle during the process, was readjusted and another test plate cut. The process was repeated until a satisfactory angle was found and, if any quartz remained, the crystal was sawed into wafers. Although simple in principle, this method of orientation had two serious disadvantages; it was slow because the saw was idle for long periods of time and the test results were often ambiguous because of the effects of twinning and of coupled modes. Even at best, the accuracy was poor.

In a short time it was discovered that polarized light could be used in determining the angle. By passing a beam of plane polarized light through the blank in the X-direction the angle between the plane of the blank and the Z-axis can be determined with an accuracy of a few tenths of a degree. The use of polarized light greatly increased the productivity of the factory and the quality of the product.

As the market for crystal units expanded in the late thirties, the need for mechanization increased. The first crude saws were improvised from milling machines, surface grinders, and drill presses. Later these were turned on their sides, fitted with diamond blades and equipped with cross-feed compounds from milling machines. Eventually the Delta Mfg. Co., with the encouragement of Elmer Wavering of Motorola, produced 5000 drill press saws. Still later the Felker Corp. (now Felker Operations) produced the Felker Dimet saw, operating on the same principle but incorporating refinements making it a better production machine.

After a slice of quartz, called a wafer, was cut from the crystal it was etched with hydrofluoric acid to reveal the presence of twinning and other defects which were marked out and discarded. The remaining portion of the wafer (if any) was then diced into blanks ranging in size from 1.5 to 3 cm. The typical yield was 20 to 30 blanks per kilogram of quartz costing about thirty dollars. Unfortunately, many beautiful, fully faced crystals proved to be useless because of electrical and optical twinning which could not be detected until the crystal was sawed into wafers.

Blanks were lapped by hand on a rotating cast iron plate called a "stooging wheel" using silicon carbide and water as a slurry. After a few hours of work, the operator of a stooging wheel was usually

contributing blood to the slurry from his/her finger tips. Many strange and wonderful devices were used to replace hand lapping. Louis Patla of DX Crystal Co., the first to produce crystal units in the Chicago area, recalls making a lapping machine from an old bread dough mixer. But most of the early laps were made from drill presses. A crank was inserted into the chuck and used to drive a metal carrier between two flat, annular cast iron plates while abrasive slurry was fed in through holes in the top plate. Metal carriers were replaced by plastic ones when it was discovered that, if the two metal plates were insulated from each other, the quartz plates would announce their own frequencies in a radio receiver. Soon the Atlas Machine Co. began producing drill press laps and the Atlas lap became the most commonly used lapping machine in the industry. In 1938 the P. R. Hoffman Co. was organized to produce machinery for the expanding industry. One of its first products was the Hunt-Hoffman planetary lap which is still widely used today without substantial modification.

Frequency measurements were made by determining the difference between the frequency of the test blank and that of some reference frequency such as a secondary frequency standard consisting of a one MHz crystal unit controlling a frequency generator which developed a set of radio frequencies having integral multiples of 10 kHz. Measurement of the difference frequency, detected by a radio receiver, was a major problem which was solved in various ways. One of the early manufacturers, E. L. Shideler, used the family piano as an interpolation oscillator. More affluent manufacturers could procure an excellent analog frequency meter from General Radio Co. and within a short time David Packard and William Hewlett developed the HP-200 audio oscillator which not only met a need in the quartz crystal industry but also launched the Hewlett-Packard Corportaion.

Final adjustment to frequency was done using a piece of plate glass as a lap and silicon carbide or emery as an abrasive. The operator applied force at diagonally opposite corners of the blank while moving it in a figure 8 motion on the glass plate thereby unknowingly reducing the thickness at the edges and creating the contour which is essential in thickness shear plates in which the thickness is not small compared with the lateral dimensions. Finishing to frequency was an exasperating exercise. The frequency of the plate is determined primarily, but not exclusively, by the thickness. But the amplitude or "activity" depends upon the lateral dimensions. Frequently the blank has low activity when adjusted to the nominal frequency and edge grinding is required. But edge grinding has a second order effect on the frequency which often skips to a higher frequency making the blank unuseable for the desired purpose. This phenomenon, which is commonplace and well understood today, was in 1935 variously ascribed to demons or to the perversity of inanimate matter.

Nevertheless, with patience, experience and luck a skilled finisher could sometimes finish to frequency as many as twenty crystal units per day. And so with primitive equipment and methods and little or no understanding of the phenomena involved, the crystal industry in the United States produced an estimated 100,000 crystal units during the year 1939.

It is quite impossible to trace to their origins many of the innovations which were devised during this period. Many techniques, thought to be original by their inventors, were later reported to have been in use elsewhere at an earlier date. This situation was partly due to the desire to keep as a trade secret any innovation which reduced the cost or/and improved the product; and partly to the lack of any effective means of communication between the various entrepreneurs. After the United States became involved in the war patriotism took precedence over industrial secrecy and with only one or two notable exceptions cooperation rather than competition became the rule.

The Quartz Crystal Section

The decision to make large scale use of crystal control in military communication systems was made late in 1939. The Armed Services had considered crystal control before but the idea was always rejected on the bases of (1) lack of flexibility, (2) cost, and (3) unavailability. In the summer of 1939, during field maneuvers in Tennessee, comparative tests were made of FM vehicular radio equipment, with and without crystal control. About the same time some crystal controlled equipment, captured from the German Army by the British, was sent to Wright Field for evaluation. The result of these tests was the decision to convert all military radio equipment to crystal control. This was about two years before Pearl Harbor Day (Dec. 7, 1941) and after the war in Europe was already under way.

The wisdom of this decision was emphasized by Major General Roger B. Colton at the Chicago Conference on July 11-12, 1944 when, despite all of the problems which had by then been encountered, some of which were still unsolved, he could say:

> "Our decision to go into crystal controlled radios for widespread tactical use has been more than justified by the results obtained. The Army had radio before they had crystals. Now the Army has communications. That's the difference. Crystals gave us communications."

Before the United States had seriously started rearming representatives of the British Purchasing Agency came to the United States to obtain equipment for the Armed Services of the U. K. One piece of equipment which they ordered was a VHF transmitter-receiver for aircraft use employing 4 crystal units in the transmitter and 4 in the receiver. The equipment was made by the Bendix Corp. and the crystal units were the DC-11 and the DC-12 types. This set, with a few modifications became the SCR-522, the most important airborne set in the U.S. Air Force and its crystal unit became the CR-1.

As the war in Europe increased in intensity it became apparent that the United States might soon be involved. The pressure on the fledgling crystal industry increased as it was asked to supply crystal units for the Allied Forces as well as our own. It became clear during the latter months of 1941 that a serious bottleneck was developing and that something must be done to expedite the production of crystal units.

If any date can be ascribed to the beginning of the quartz crystal industry in the United States it is probably late October or early November 1941 when the Quartz Crystal Section (QCS) was organized in the Office of the Chief Signal Officer (OCSIGO). It was headed by Lt. Col. James D. O'Connel and house in Bldg. Temporary A at Buzzard's Point near the old Army War College in Washington. The mission of the QCS was to expedite the production of quartz crystal units to meet the requirements of our Armed Services and those of the Allies. The first problem was that of finding personnel who could solve the technical problems and guide the industry which was to be. The original group of civilians in the QCS included:

Dr. Clifford Frondel of Harvard University
Dr. Richard Stoiber of Dartmouth College
Mr. Sam Gordon of the Philadelphia Museum of Natural History, and
Dr. William Parrish of the Pennsylvania State University, all of whom were mineralogists.

Dr. Wally Richmond and Mr. Hugh Waesche, both of whom were geologists.

Dr. Carl Bertsch of Newark College of Engineering, a physicist and Mr. E. K. Woods of the Patent Agency.

Among the Officers initially assigned to the QCS were:

Capt. E. W. Johnson	Capt. Charles Miller
Lt. Les Atlass	Lt. Willie Doxey

The task of this group was to organize an industry, determine what facilities were required and to make them available, obtain the necessary priorities for materials and equipment and to expedite and schedule production. Furthermore it had to be done immediately. Although all three services had small groups working on crystal problems which were peculiar to their particular services, the QCS served as the focal point for the quartz crystal program in the U.S. during the next four years.

In July 1942 the QCS was moved to the unfinished Pentagon Building and Dr. K. S. Van Dyke joined the group as Chief Civilian. His knowledge of the basic theory of piezoelectricity added a much needed dimension to the staff of the QCS which had hitherto been made up largely of mineralogists and geologists. Early in 1943 Virgil E. Bottom joined the group. He was the first member who had actual experience in the manufacture of crystal units. A little later Dr. Allyn Swinnerton, a geologist and Vice President of Antioch College, joined the group.

Work of the Quartz Crystal Section

The first task of the QCS was to publicize the need for quartz crystal units in order to attract potential manufacturers. This was done so successfully that many letters were received from patriotic citizens offering to donate the crystal sets which had reposed in their attics for a decade or two. However the information did reach the desired places and would-be-manufacturers of crystal units whose regular activities had been halted by the War Production Board began to descend upon the QCS asking how they might use their idle facilities to make crystal units. This group which included machinists, blacksmiths, jewelry makers, and amateur radio operators turned in one of the most remarkable records in the annals of American industry. In less than three years Mr. William Halligan, the President of the Hallicrafters Co. was able to say:

> "We had no trained personnel to make crystals. We had no equipment with which to make crystals and out of this has been created what to my mind, is truly a miracle of American incentive and industry."

The information available to these neophyte crystal manufacturers consisted of a few articles in QST, an article or two in the Bell System Technical Journal and the patents which had been issued. A little later the Bell Telephone Laboratories came to the rescue with a collection of articles written by such authorities as W. P. Mason, Walter L. Bond, Roger Sykes, I. E. Fair, G. W. Willard, G. M. Thurston, A. R. d'Heedne, H. G. Weke and R. M. C. Greenidge. These papers were of inestimable value at the time and were later edited in book form by R. A. Heising and published by Van Nostrand in 1946 under the title QUARTZ CRYSTALS FOR ELECTRICAL CIRCUITS. These materials were made available to the new manufacturers who digested them, utilized them and, in many cases, made improvements upon them.

One of the first acts of the QCS was to expedite the development of X-ray diffraction equipment for use in orienting quartz blanks. This became the first commercial application of X-ray diffraction. A few companies had adapted old dental X-ray equipment for making Laue photographs of quartz blanks but the process was slow and the results scarcely better than those which could be obtained with polarized light. At the Western Electric Co. a primitive form of Bragg diffraction equipment was in use by Walter L. Bond but this information had not reached beyond the limits of the Bell Telephone Laboratories. Contracts for the development of X-ray equipment were placed with the General Electric Co. and with the North American Phillips Co. under the surveillance of Dr. William Parrish. Both companies produced equipment in such numbers that most of the crystal companies were equipped with X-ray machines by early 1943. Although the equipment designed and built by the two companies differed in many respects, both were so well designed and constructed that many of the original machines are still in use almost forty years later.

Much effort of the QCS went into expediting equipment, supplies and materials for the new industry. Because of the critical nature of the program it was assigned a very high priority which, along with the authority of the OCSIGO, was often invoked to obtain critical materials. At one time a certain piece of equipment was almost within our grasp when it was snatched away by an outfit called the "Manhattan Project" which had a higher priority. Some of us wondered briefly why Manhattan needed a General Radio Primary Frequency Standard. Two of the men who became expert in working with the War Production Board and other agencies were Tom Parrott and Wally Richmond and many crystal manufacturers owed their daily ration of quartz to the efforts of these two men.

Due to the critical nature of the industry the decision was made to locate the production facilities away from the East Coast which was within enemy range. Accordingly manufacturers in the Central region of the country were encouraged with the result that Kansas City and Chicago became manufacturing centers for the quartz crystal industry and remained so for years.

In May 1942 the WCS organized a training program in Kansas City under the auspices of the University of Kansas. The teacher of the course was John Zeigler, a young engineer, who had learned something about crystal units while working at RCA. The course was taught from copies of patents to a class of about 20 persons including Ernest O. Ruff and Edward Roper, both of whom are still active in the crystal business.

Soon thereafter Zeigler and George (Jack) McGrew organized a company called Crystal Products Co. Except for a small operation called Aerion, owned by the Aircraft Accessories Co., Crystal Products was the first crystal company in the Kansas City Area. All of the other factories in the area sprang from the company organized by McGrew and Zeigler.

By 1943 about 130 manufacturers were engaged in the production of crystal units. Twenty three of these were in the Chicago area, 20 in the New York area, 15 in the Carlisle area and 14 in the Kansas City area. The remainder were scattered over 20 states from Oregon to Florida and California to Massachusetts. The supervision of so many small plants, distributed over such an area was a major problem and the members of the staff of the WCS spent much time on travel duty instructing the new manufacturers, helping them with technical problems, coorelating test equipment and settling arguments between them and the Signal Corps Inspectors.

The Galvin Mfg. Co. (Motorola)

The task of coordinating the activities of 130 companies was clearly beyond the capability of the small staff of the WCS in Washington and so the Galvin Mfg. Co. (later Motorola) was induced to assist in the program. Galvin was heavily engaged in the production of personnel radio equipment known as the"handie-talkie" and "walkie-talkie" sets, all of which used the FT-243 crystal unit. Under the direction of Mr. Elmer Wavering, Galvin Mfg. Co. undertook the Herculean task of expediting and scheduling the production of FT-243 units, testing and assembling them into kits each of which contained 44,000 units on 4400 different frequencies. Mr. Nick Anton, an assistant to Mr. Wavering, prepared and distributed a Crystal News Sheet which supplemented the grapevine and helped contractors to keep abreast of new developments and information related to the production of FT-243 units.

One important contribution of the Galvin Mfg. Co. was the CES-1 Crystal Test Set. The CI-meter had not yet been invented and standardization of crystal units was unknown. It was necessary to provide each manufacturer with a test set consisting of a mock-up of the oscillator circuit of every piece of radio equipment. These test sets were known as "black boxes" from the color and shape of the cabinets which housed them. Each "black box" was periodically returned to the Signal Corps Laboratory at Fort Monmouth for recalibration. The CES-1 consisted of a Pierce oscillator circuit in which the critical parameters were variable. It could therefore be adjusted to simulate a number of different radio sets, thereby reducing the need for so many "black boxes". The CES-1 helped greatly with the correlation problem but did nothing to solve the standardization problem which awaited the invention of the CI-meter.

Growth of the Crystal Industry

Except for the Western Electric Co., all of the crystal unit manufacturers in existence before 1940 were very small. Bliley, one of the largest, had fewer than 15 employees. Monito Piezo had about the same number. It is doubtful if more than 100 persons were engaged in the production of crystal units in the year 1939.

During the period 1941-45 employment at Monitor reached over 2000 and at Bliley over 1400. Scientific Radio Products employed over 750 and a new plant, The Reeves Sound Laboratories in New York City which began production in the summer of 1942, soon employed nearly 1000. In Chicago the Ross Mfg. Co., headed by Ken Ross and the Beaumont Electric Co., operated by the Lieberman brothers, (who formerly manufactured lampshades) each employed over 500. Many smaller companies with fewer than 100 employees also made their contributions. One example was the Good-All Mfg. Co. of Ogallala, Nebraska which employed about 20 people.

In addition to the firms directly engaged in the production of crystal units, many many others were engaged in supporting activities. Diamond saw blades, abrasives, plastic holders, stainless steel electrodes, beryllium copper springs, neoprene gaskets, nickle plated screws and many other items were required to complete the assembly. Coordination of the production and distribution of these items was done by the War Production Board which attempted to match the limited supplies to the almost unlimited requirements.

This motley assemblage of manufacturers, coming from every part of the nation and from almost every type of activity turned in a most remarkable production record. Between 1941 and 1945 they produced an estimated 30 million crystal units.

The FT-241 Program

The radio equipment used by the armored cavalry and the field artillery divisions was designated the BC-608 series. This equipment used frequency modulation and operated in the 30-40 MHz range of frequencies. The crystal unit for the set was the FT-241, developed by the Bell Telephone Laboratories and produced by the Western Electric Co. Each set carried a complement of crystal units ranging in number from 72 to 120 thereby requiring an extremely large number of crystal units.

The crystal units, which operated in the frequency range 350-550 kHz were made with CT-cut quartz plates. They were made at the Hawthorne Plant of the Western Electric Co. in Chicago

where more than 10 million such units were produced between 1941 and 1945. The production of this plant was reported to be one million units per month when it was shut down in the summer of 1945. A complete account of the development of the FT-241 unit is given by W. F. Brew and A. E. Zwickard in Heising, Chapter XVI.

Although the Hawthorne plant was in a secure location militarily, it was early recognized that at least one more source for these units should be available. The technology involved in making the FT-241 was clearly beyond the capability of most of the small plants which were successfully making the FT-243 and CR-1 types. Accordingly a contract was awarded to the Federal Telephone and Telegraph Co. of Newark, a subsidiary of ITT, to set up a second source for the FT-241 unit. However, despite the expenditure of a large sum of money and much precious quartz and with the experience of the Western Electric Co. available, the war ended before any units of the FT-241 type were produced outside the Hawthorne plant in Chicago.

The Quartz Shortage

One of the most critical problems faced by the WPB was the shortage of raw quartz. Although quartz is found in a few other places including the United States, practically the only source of natural quartz of electronic grade is Brazil. At the beginning of 1939 the United States had a stockpile of about 10,000 pounds of quartz suitable for processing into crystal units. During 1939 about 30,500 kg (67,000 pounds) of quartz was imported from Brazil. The amount imported in 1940 was 57,500 kg. Approximately 10% of this quartz was considered to be of electronic grade. Dr. Richard Stoiber was assigned the problem of expediting the supply of quartz.

Before the end of 1942 the supply of faced quartz which could be processed without the use of X-rays was nearly exhausted and the industry was forced to begin to learn to process unfaced quartz of which a considerable quantity was still available. Various techniques and devices were developed to enable manufacturers to use this material. These are adequately described by G. W. Willard in Heising, Chap. IV.

By the end of 1942 X-ray diffraction equipment began to become available to the manufacturers enabling them to effectively utilize unfaced quartz crystals. This relieved the shortage somewhat but the supply was still inadequate and steps were considered to increase the supply. The quartz mining regions of Brazil are not in the "hot, steamy jungles" but rather in the highlands in the central states of Minas Gerais and Goias near the capital city of Brazilia (which, of course, did not exist at that time). The mining of quartz was not then, nor is it now, an organized activity. The 'garimpeiro' (diamond hunter) would dig or hunt until he found a crystal which he could trade to a local merchant for supplies. The merchant, in turn, exchanged it for goods and eventually the stone reached Rio de Janiero where it was finally sold to the purchasing agent of the Metals Reserve Corp. or some other purchaser.

This ad hoc system was incapable of meeting the needs of the growing crystal industry and the quartz shortage continued to become more acute. As an example, on Easter Sunday 1942 an executive order was obtained from the White House releasing 700 pounds of quartz which was needed for a special order for crystal units.

Various steps were considered for alleviating the shortage which was partly due to inadequate grading facilities in Rio. Consequently inspection teams were trained and sent to Brazil to inspect the quartz prior to shipment. Eventually several thousands of pounds of quartz were loaded on a ship bound for the Port of Newark. The ship was hardly out of the harbor before it was sunk by a German submarine. After that incident all quartz was flown to the United States in DC-3 cargo planes at a cost of $2.00/pd. (1941 dollars).

In an effort to encourage the mining of quartz the purchasing agents in Brazil were authorized to double the price of quartz. However the authors of this strategy had failed to take into account the economic system of the 'garimpeiro' who was not interested in building up a savings account and who only dug or hunted quartz when he needed to replenish his supplies of rice, beans and pinga. Consequently the result of the program was less, not more, quartz and the program was abandoned.

It was then proposed that mining machinery be sent to Brazil. Accordingly a cargo ship was loaded with bulldozers and other earth moving equipment and dispatched to Rio. It was sunk enroute. Another ship was loaded with similar equipment. It arrived safely but most of the machinery never reached the mining areas because of lack of roads and that which did reach the quartz mining area proved unsuitable because of lack of fuel and because the nature of the operation made mechanical mining impractical.

At this time it was considered uneconomical to process stones below about 200 grams in size. One day in 1943 Mr. Morris Hanauer, owner of American Gem and Pearl Co., a quartz importer, was trying to convince Lou Patla of DX Crystal Co. that he should use these small stones. "Lou", he said, "God made lots of small crystals but very few large ones. So why don't you go along with God and use the small ones ?" Soon ways were found to process the small stones with satisfactory results. This information spread quickly throughout the industry and for a time alleviated the quartz shortage.

Nevertheless, through the entire war period the quartz crystal industry lived on a hand to mouth basis and the shortage problem was eventually met, not so much by increasing the supply as by conservation. The yields were increased by making the blanks smaller and stones formerly considered unuseable were processed. Production records were maintained and with one or two notable exceptions, quartz was allocated only to the producers who could show that they were able to utilize it efficiently.

The lesson learned from the shortage of a critical raw material for which we were totally dependent on an offshore supply was not wasted.

Efforts were made to find a substitute for quartz. These were unsuccessful but at the end of the war it was learned that Germany, faced with a similar quartz shortage, had made some progress toward making man-made quartz. Consequently a major effort was mounted to make cultured quartz with such success that we are now almost completely independent of the natural material.

The Ageing Problem

By the middle of 1943 the task of setting up an industry was complete and crystal units were being produced in numbers adequate to meet the demand. It was then that the second, and even more serious crisis confronted the crystal program Reports began to filter in of extensive crystal failures both in service and in depot storage. The first responses to these reports ranged between indifference and disbelief. However the reports became more persistent and the Signal Corps Engineering Laboratories (SCEL) at Fort Monmouth were requested to investigate the situation. The progress was slow because the problem was treated as an academic question rather than a matter of the utmost urgency. Late in 1943 a telegram was received at OCSIGO which changed the situation. It read as follows:

COMMUNICATIONS EIGHTH AIR FORCE BASED
IN BRITAIN BROKEN DOWN LACK OF CRYSTALS
FINAL CALL E FIRE GAME BAKER

Reception of this communique caused a flurry of activity at both the OCC and at SCEL. A test facility euphemistically named "The Swamp" was constructed at Camp Coles Signal Laboratory. It subjected crystal units to temperatures of 100 F° in an ambient relative humidity of 100 % simulating the conditions reported from the Burma theatre. The tests quickly confirmed the field reports. All crystal units failed within a few days or at most a week or two. Some of the failures were due to the phenomenon of corrosion fatigue in which the ammonia released by the phenolic holders attacked the brass connectors at points of stress causing them to break. This problem was corrected by eliminating brass from the assembly. A much more serious problem was the ageing syndrome characterized by loss of activity and increased frequency. Many theories were advanced to explain the phenomenon and these had to be checked out experimentally. Meanwhile the manufacturers were producing crystal units at the rate of a million per month; nearly all of which were destined to be useless.

Before the end of 1943 it had become apparent that the most important factor in the ageing process was the surface of the quartz which was damaged in the process of lapping to frequency. Particles of quartz, partially loosened by abrasion, were further loosened by the effects of water vapor, eventually breaking away completely. The presence of these loose fragments on the surface of the blank reduced the activity by damping and their loss caused an increase in frequency because of the reduction of mass.

The problem of ageing was especially severe in the CR-1 unit used by the Air Force because these units operated at higher frequencies and required closer tolerances. As a temporary measure orders were issued by Wright Field that all quartz plates must be able to withstand the test of scrubbing with soap and water and a toothbrush. Soon each Signal Corps Inspector and each crystal finisher was equipped with a toothbrush and a dish of soapy water. The directive was a boon to the toothbrush industry but contributed little to the solution of the ageing problem.

By the end of 1943 it had been shown that quartz blanks which had been etched to frequency exhibited very small changes of frequency and activity even when subjected to tropical conditions. However considerable opposition existed to the idea of etching and much valuable time was lost in investigating other approaches before the decision was made to require etching. This was due to a reluctance to specify a manufacturing process which might require rewriting of the specifications and renegotiating of contracts and to a wise policy of specifying test results instead of manufacturing procedures. Yet it was impossible to depend upon inspection procedures to insure that crystal units would remain useable and a vast amount of work had failed to reveal any other manufacturing procedure which would do so.

Consequently a Conference of all Crystal Manufacturers was called on July 11-12, 1944 in the old Stevens Hotel in Chicago. The problem was explained in detail and the proposed remedy was presented. There was no time to rewrite the specifications or to renegotiate contracts so the manufacturers were asked to convert their production processes from hand lapping to etching. It was expected that units made by the etching process might be more expensive. Again the industry rose to meet a challenge. The manufacturers went back to their plants and converted to the new process with such success that they produced satisfactory units at a lower cost. As one example, Ken Ross, who had converted his coil winding plant in Chicago to a facility for making FT-243 crystal units, designed and built a continuous etching system which, with four operators, turned out as many finished crystal units as 20 operators finishing by hand lapping. The Ross system was soon widely copied throughout the industry.

Ironically, the ageing problem might have been avoided had better communication existed. The phenomenon of ageing had been noted as early as October 1940 at RCA. The work of H. E. LeRoy and V. E. Trouant led to a Company Confidential memorandum dated April 14, 1941 in which they said:

> "Etching... reduced the number of failures" and that "Etching will be incorporated in the processing as soon as details can be worked out and operators trained."

Unfortunately this work was unknown outside the RCA Laboratories until after the war and even there it does not appear to have been exploited.

To provide crystal units for special purposes in the field, twelve three-man crystal grinding teams were given a three months crash course in finishing crystal units. They were trained by Dr. D. G. McAA, a retired physician, at Camp Coles Signal Laboratory. These teams were provided

with the necessary equipment and supplies and shipped to various theatres of the war. Upon arrival they found that many of the crystal units in depot stocks were useless and much of their time was spent in opening, cleaning and grinding thousands of crystal units to new frequencies. Mr. H. J. Benedikter, who is today associated with the General Electric Crystal Manufacturing Facility, was a member of one of the original crystal grinding teams.

At home teams of GI's and civilians were trained to examine depot stocks of crystal units and to select out useable units which were often flown immediately to a war theatre. Later some attempts were made to salvage the defective units and at least one company, The Hudson American Co., was commissioned to do so. However most of the units were enclosed in phenolic plastic holders and many others contained brass contacts, both of which were considered unsatisfactory, leaving little besides the quartz blank to be salvaged. It soon proved to be uneconomical to salvage the units and ultimately millions of the unetched units were destroyed.

It may be worthwhile to examine the ageing episode; not to assess blame or responsibility but to see if, by doing so, we might avoid a similar debacle in the future. It appears that the ageing problem resulted from three basic mistakes; each understandable and perhaps forgivable. The first was to depend so heavily on an immature technology, the second was to disperse an industry so widely that it could not be properly supervised, and the third was failure to follow up with an adequate reliability test program. The recall programs of the United States automobile industry indicates that some of these lessons may not even yet have been learned.

Post War History of the Industry

It is difficult, if not impossible, to trace the histories of all of the firms which have been engaged in the business of manufacturing quartz crystal units. Only a few firms, notably Bliley, Monitor and Valpey still operate under their original (or similar) names and can show continuity of management. Most of the companies have either disappeared or have changed names and locations so many times that to trace their records is virtually impossible. It may be of some interest to trace the history of one company; noting that it is typical of the industry.

When Grover Hunt discovered that people were willing to pay for crystal units he formed a partnership with Linwood Gagne and they began to make crystal units in a small building adjacent to Gagne's home in Carlisle. In 1935 they formed Standard Piezo Co., the first crystal company in the Carlisle area. One of their employees was Wally Samuelson. In 1940 Hunt left to form his own company, the Hunt Corp. which is today Erie Frequency Control. At the end of the war Gagne sold Standard Piezo to a group of local business people headed by John Fowler. Later it was sold to Brown Oil Co. and Luther McCoy was made Sales Manager. In 1952 McCoy left to form his own company and Standard Piezo was sold to the Hupp Corp. which, in turn, sold it in 1958 to Herman Schall and Wally Samuelson who renamed it Piezo Crystal Co. In 1963 Schall and Samuelson sold out to the Renwell Corp. who sold it to the Sunshine Mining Co. The plant was subsequently sold to a holding company, Anchor Operating, Inc., which operates it today under the direction of Charles Jansik.

It is interesting to note that the 1980 edition of the QUARTZ CRYSTAL INDUSTRY GUIDE AND DIRECTORY states that Piezo Crystal Company was established in 1938. In a sense this is true, but it is in the same sense that an antique axe which has had several new handles and a few new heads is still the same axe.

A list of manufacturers of quartz crystal units dated 1 May 1945 includes the names of some 130 companies. One year later the number had dropped to fewer than half this number. After revival of the industry during the Korean War approximately 50 companies were engaged in the business of making crystal units. The number deceased again when the Korean War ended but has increased again in recent years.

Today the industry includes between 50 and 60 companies employing about 6500 people. It is impossible to say how many are engaged solely in the manufacture of crystal units since many of the companies also produce items such as crystal filters, TCXO's and VCXO's and in addition some of them produce only blanks while others only finish blanks.

The typical company, if it existed, would have about 100 employees including two engineers but the size of the companies varies by two orders of magnitude. The largest companies, CTS-Knights and Piezo Technology employ 580 and 515, respectively. The two smallest, American Crystal Co. and Defco employ 6 and 3 persons, respectively. Several companies have fewer than 10 employees.

The foregoing figures do not include the firms which support the crystal industry. These are more numerous than the actual crystal manufacturers and employ more people. Among the products supplied by these firms are cultured quartz, holders and bases, abrasives, chemicals, silver and gold, electronic instruments and machine tools, and supplies in an almost endless variety.

Problems of the Crystal Industry

The wartime expansion of the crystal industry came to an abrupt halt with the end of the war. Contracts were cancelled, plants were converted back to civilian production and equipment was dispersed. The number of producers dropped from 130 to below 50 in a few months. When the Korean War came in 1950, bringing with it a new requirement for crystal units it was necessary to subsidize the companies to obtain the crystal units needed by the Armed Services.

For its entire forty years the quartz crystal industry has been plagued by the feast/famine cycle. In times of military rearmament the demand has escalated only to be followed by abrupt deescalation when the crisis was over. Until only recently

the civilian requirements for quartz crystal units have been small compared with military requirements in times of crisis making continuous production impossible.

One problem in the quartz crystal business has been lack of research and development. It is true that in the decade following WW II the Armed Services did support an R & D program which yielded some useful results. A notable example is the program which led to the development of a cultured quartz industry. But for the most part the results have not been effectively utilized; partly due to the fluid nature of the industry. Outside a few large companies, notably the Hewlett Packard Corp., very little work in quartz crystal technology has been supported with industrial funds.

The field of piezolectricity has always been, and continues to be ignored by the Schools of Electrical Engineering. Quartz crystal technology has been treated as an art, or even worse, as black magic. With the exception of the students who have received formal training at Northern Illinois University under Dr. W. E. Newell and at Colorado State University and McMurry College under Dr. Virgil E. Bottom, most of the people in the industry have learned from one another going back to the handful of men who started the industry who had, themselves, learned in the School of Experience.

Academic neglect of the field is also shown by the dearth of literature available to the newcomer. The classic work of Cady, the books of W. P. Mason and the compliation of 1942 papers by Heising include practically everthing which has been published in this country. The book publishers naturally have been reluctant to make the necessary expenditures to publish books for a nonexistent clientele.

The industry has had its share of "get rich quick" entrepreneurs who jumped at the opportunity to make a quick profit during times of emergency but these have been a very small minority. Most of the men who entered the crystal industry, particularly during the war, were motivated by the challenge of a difficult job and by the desire to make a patriotic contribution as much as by the profit motive. Many continued in the business as long as they were financially able to do so; often bidding on production contracts at prices below the cost of production in order to hang on a little bit longer. They have done so at great financial sacrifice to themselves and to the detriment of the industry.

The industry has always been too small to be of interest to Wall Street and the large corporations and technically too difficult for the small entrepreneur without adequate financial backing and scientific capability. In those instances where large corporations have taken over small crystal companies the results have been almost uniformly disastrous. Generally speaking, the management of the large companies have attempted to apply the same production management techniques which work successfully in the production of most standard components; failing to recognize the great complexity of the apparently simple crystal unit. When they have found out, they have been unable or unwilling to make the capital investments needed to put the production of crystal units on a firm financial and technical basis.

The crystal industry has experienced four bursts of activity during the past decade. These involved crystal units for color television sets, citizen band radio, watches and clocks, and lastly microprocessors. Each of these products has resulted in a temporary expansion of the industry followed by a collapse when prices fell below the cost of production in the United States. In order to compete successfully for these mass markets it appears that the crystal industry needs more capital and more trained people. Unfortunately both seem to be in short supply.

Consequently the quartz crystal industry which produces the device that W. P. Mason has called "The cornerstone of the communication art", has remained, like the North American Indian, a ward of the government, dependent upon it for financial and technical support, and struggling to stay alive to be in position to meet the challenge of the next military crisis.

EPILOGUE

The author wishes to thank each of the persons who contributed material for this account. Without their help it could not have been prepared. He is fully aware of the deficiencies of the paper. Many important topics have been omitted and numerous persons who made important contributions have not been mentioned. Despite every effort to make the report factual, it is almost certain that errors have been made. The author accepts full responsibility for these and solicits corrections. He also asks forgiveness of those who have been omitted and of those who may have been unfairly represented.

Those having further information and/or corrections are urged to communicate with the author. Perhaps at a later date the story can be told in greater detail and more precisely.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

THE ROLE OF ANALOG DEVICES IN A DIGITAL AGE

RICHARD C. WILLIAMSON

Massachusetts Institute of Technology
Lincoln Laboratory
Lexington, MA 02173

ABSTRACT

Advances in surface-acoustic-wave, and other analog technologies occurring over the last decade have yielded a variety of components which provide efficient means for processing wide-bandwidth signals. For example, SAW devices are becoming widely used in connections, radar, and signal-analysis equipment. At the same time, advances in digital technology continue at an impressive rate and now almost every electronic system has a mix of analog and digital components. Both digital and SAW-based techniques exist for performing critial signal-processing functions such as fixed and programmable matched filtering and Fourier transformation. This talk reviews the operation of wide-bandwidth SAW signal processors including pulse compressor, convolvers and chirp-transform units. The critial issue of processing flexibility, of analog/digital interfaces, and of matching the bandwidth and coherence time of SAW devices to incoming signals are addressed. A comparison of digital and SAW signal processors is made by analyzing the required size and computation rate of hypothetical digital systems which might potentially replace SAW devices. It becomes clear that for many types of wide-bandwidth signal processing functions, SAW devices will provide significant advantages for the foreseeable future.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703. May 1981

NONLINEAR PROPERTIES OF QUARTZ CRYSTAL AND QUARTZ RESONATORS : A REVIEW

J.J. Gagnepain

Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25000 Besançon - France

Summary

The nonlinear behavior of quartz crystal is at the origin of almost all the frequency instabilities of resonators and oscillators. Intrinsic and induced nonlinearities can be distinguished, and both are of the same order of magnitude. The nonlinearities lead to two different classes of phenomena, which correspond to the propagation of a finite amplitude wave in a nonlinear medium or to the propagation of a small amplitude wave in a nonlinear strained medium.

Harmonic generation, amplitude-frequency effect, intermodulation are related phenomena of the first type. They depend on the crystal anisotropy and geometry, and on the wave structure.

The sensitivities to external or internal perturbations are a consequence of the nonlinear coupling between the high frequency wave and the quasistatic deformation induced by the perturbation. Temperature, force, pressure, acceleration, electric field sensitivities are studied for applications to high stability oscillators and to sensors.

A second approach for describing the vibrations of a crystal consists in using the concept of lattice waves and phonons. Such a microscopic model enables to describe by means of phonon interactions, due to the crystal anharmonicities, phenomena like finite thermal conductivity, thermal expansion, acoustic attenuation, velocity variation, in a unitary description.

Introduction

Acoustic waves do follow nonlinear propagation laws as soon as their amplitude is finite or the propagation medium not isolated from the environment. Therefore they cannot entirely be described by the mean values of their characteristics as given by the linear laws. Because of the important progresses achieved in the field of resonators, oscillators, and frequency control during the two last decades, a linear description was not sufficient, and it became necessary to take into account the nonlinear behavior of the acoustic waves for reducing their instabilities or sensitivities to perturbations. Almost all the variations of the characteristics of the waves have their origin in the nonlinearities, and one almost could say that a perfectly linear device would be perfectly stable. Nonlinearities generally are described by macroscopic models. In this case intrinsic and induced nonlinearities can be distinguished. The first ones are peculiar to each medium and are represented by elastic, piezoelectric, dielectric, etc..., constants of higher orders. However, even in a perfect crystal which would not exhibit any higher order constants, the propagation of the wave itself deforms the medium, and therefore modifies the conditions of the propagation ; this corresponds to induced nonlinearities, which are of the same order of magnitude as the previous ones. All these nonlinearities are at the origin of harmonic generation and related phenomenon such as amplitude-frequency effect and intermodulation. The nonlinearities also are responsible of couplings between the ultrasonic wave and external or internal quasistatic perturbations. This second class of phenomena corresponds to the sensitivities of the acoustic devices to temperature, force and pressure, acceleration, vibrations, fields, diffusion processes, etc...

The microscopic theory of the vibrations of a crystal lattice leads to the concept of phonons. In the harmonic approximation the lattice waves are independant, and the energy of each given mode is constant ; there is no sharing between the different modes. This representation enables to introduce the notion of specific heat, but a more realistic approach must go beyond the harmonic approximation and take into account anharmonicities. As a consequence of anharmonicities are interactions between phonons, which give to the crystal the properties of thermal expansion, finite heat conductivity, acoustic attenuation, acoustic velocity shifts.

In anisotropic media the intensities of the nonlinear phenomena depend on the crystal orientation and on the wave propagation direction. Particular configurations can be found either with minimized nonlinearities for resonator, oscillator and filter applications, or at the opposite with maximized nonlinearities for devices used in signal processing. Acoustic waves also are attractive for sensor applications when they can be made selectively sensitive to a particular physical quantity.

Propagation of a finite amplitude wave

The exact, and therefore nonlinear, equations of elasticity were presented by Murnaghan[1] in 1951. A more general description of an electrically polarized elastic solid, or continuum, was given in 1956 by Toupin[2]. In more recent papers Thurston[3] and Tiersten[4] gave the elastic and electroelastic rotationally invariant equations in a form including mechanical inertia, which are more directly applicable to the description of wave propagation. The complete set of electroelastic equations can be found in ref. 4, and in the present paper only a brief description of the purely elastic problem will be given.

When refered to the convenient material (Lagrangian) coordinate system the stress equation of motion takes the form

$$\rho_o \ddot{u}_j = P_{ij,i} \tag{1}$$

On a free surface with a unit normal of components N_i the adjoined boundary condition is

$$N_i P_{ij} = 0 \tag{2}$$

ρ_o is the specific mass of the body at rest, u_j are the mechanical displacement components and P_{ij} is the Piola-Kirchoff stress tensor, which can be written

$$P_{ij} = c_{ijk\ell} u_{k\ell} + \gamma_{ijk\ell mn} u_{k,\ell} u_{m,n} + \delta_{ijk\ell mnpq} u_{k,\ell} u_{m,n} u_{p,q} \tag{3}$$

The $c_{ijk\ell}$'s are the regular second order elastic constants.

$\gamma_{ijk\ell mn}$ and $\delta_{ijk\ell mnpq}$ are related to the third and fourth order elastic constants $c_{ijk\ell mn}$ and $c_{ijk\ell mnpq}$ following the relations

$$\gamma_{ijk\ell mn} = c_{ink\ell}\,\delta_{jm} + (\tfrac{1}{2})c_{ijn\ell}\,\delta_{km} + (\tfrac{1}{2})c_{ijk\ell mn} \tag{4}$$

$$\delta_{ijk\ell mnpq} = (\tfrac{1}{2})c_{iqn\ell}\,\delta_{km}\,\delta_{jp} + (1/4)c_{ijk\ell qn}\,\delta_{pm} + (1/4)c_{ijq\ell mn}\,\delta_{kp} + (\tfrac{1}{2})c_{iqk\ell mn}\,\delta_{pj} + (1/6)c_{ijk\ell mnpq} \tag{5}$$

The first term in Eq. (3) is the linear term. $\gamma_{ijk\ell mn}$ and $\delta_{ijk\ell mnpq}$ introduce quadratic and cubic nonlinearities in the propagation equation and boundary condition. System (1)-(2) can be solved by using a method of successive approximations, and the solutions lead to the determination of the second and third harmonic levels[5-8], of amplitude-frequency effect[8,9], and of intermodulation[10,11].

Amplitude-frequency effect

The amplitude-frequency effect corresponds to the dependence of the frequency of a resonator on the oscillation level and appears as a distorsion of the amplitude and phase resonance curves. The nonlinear resonances of quartz resonators were investigated by Bottom[12], Warner[13] and Hammond[14], essentially on AT-cut resonators by using amplitude resonance curve methods. BT-cut quartz resonators were also studied by Seed[15]. A free oscillation method was used by Smolarski[16] and applied to length extensional low frequency quartz resonators. Typical amplitude and phase nonlinear resonance curves[17] are shown on fig. 1.

The inconvenient of the resonance curve method or of the free oscillation method is not to enable to separate the thermal effect and the purely nonlinear effect. Therefore they must be used only at a temperature corresponding to the turn over point temperature. However a new method was proposed by Planat[18] for measuring dynamically the A-F effect. In this method the driving signal is amplitude modulated and the corresponding resonance frequency modulation is detected with a phase bridge. The additional thermal effects can be completely cancelled out when the modulation frequency is large enough compared to the thermal relaxation frequency of the crystal.

With regards to the curves of fig. 1, it is obvious that the resonance frequency of a resonator driven at high level is not so well defined, and all the amplitude variations will be transformed into frequency shifts. In order to reduce the A-F effect doubly rotated quartz cuts were investigated[19] and it was pointed out that the A-F effect could be minimized for a 20° rotation angle around the Z axis, as shown on fig. 2.

A one-dimensional model gives a simple analytical expression of the A-F effect for rotated Y-cut thickness shear resonators

$$\frac{\omega-\omega_o}{\omega_o} = \frac{3\pi^2}{256h^2} U_o^2 \left(\frac{3C_{22} + 6C_{266} + C_{6666}}{\overline{C_{66}}}\right) \tag{6}$$

$2h$ is the quartz thickness and U_o the oscillation amplitude. This relation shows the influence of the third and fourth order elastic constants. If the former ones are well known for quartz[20] and other materials[21], the last ones still are unknown. Nevertheless equation (6) and the experimental data offer a possibility of evaluation of the c_{6666} constant. Values obtained on 5, 10 and 100 MHz AT-cut quartz resonators are given in table I.

f_o (MHz)	5	10	100
C_{6666} (N/m²)	8.10^{13}	3.10^{13}	5.10^{13}

Table I

The previous relation does not involve the energy trapping in the resonator. A more complete relation was proposed by Tiersten[7], which takes into account the influence of the electrodes. The particular plano-convex shape which is very often used for trapping has not been considered yet.

The propagation of surface acoustic waves (SAW) obeys the same nonlinear laws of equations as bulk waves. The boundary condition corresponds to a free surface limiting a semi-infinite medium. The determination of the A-E effect is more difficult because of the complex structure of the surface wave with its three coupled modes. When comparing with bulk wave resonators a main difference appears : the SAW propagates over a large number of wave lengths. In this case, and at the opposite of bulk waves, the nonlinearities are essentially due to the third order elastic constants. The influence of the fourth order ones becomes negligible for a propagation length larger than the wave length. This property enabled to evaluate the values of the SAW A-E effect from the fundamental constants of the crystal for ST, X and Y-cuts[8], and the study as a function of the crystal anisotropy is in process[9]. The preliminary theoretical studies show a much lower A-E effect of SAW when compared to BAW.

This was confirmed experimentally (fig. 4). But as already pointed out it can be difficult with low Q-factor resonators to distinguish between nonlinear elastic effects and thermal effects, and both generally are included in the measurements [10]. The pure elastic A-E effect was measured by using the A-M method[11] on ST-cut 100 MHz SAW resonators. Comparison is made with BAW resonators on table II and it can be observed that SAW A-E effect is about 1000 times lower.

	Cut	Frequency	Q factor	A-E coef. k
SAW (for comparison)	ST	11 MHz	20 000	$1.1\ 10^{-3}$/A
BAW (for comparison)	AT	10 MHz	6 000	$2.5\ 10^{-1}$/A
SAW (for comparison)	AT	5 MHz	2 000	$2\ 10^{-1}$/A

Table II

The A-E coefficient k is defined by the relation $\Delta f/f = kI^2$ where I is the driving current amplitude.

Intermodulation

When the driving signal is composed of two nearby angular frequencies ω_1 and ω_2, intermodulation is generated by the crystal nonlinearities if the driving powers are moderately high. It can be easily seen from equation (3) that the quadratic nonlinear terms will generate spectral lines at the harmonic frequencies $2\omega_1$, and $2\omega_2$ and also at $\omega_1+\omega_2$ and $\omega_1-\omega_2$. These spectral lines being largely out-of-band are filtered out by the resonator. However the cubic nonlinear terms generate harmonic frequencies $3\omega_1$ and $3\omega_2$, also frequencies at ω_1 and ω_2 which modify the fundamental components and indeed are at the origin of the A-E effect, and intermodulation frequencies at $2\omega_1+\omega_2$, $2\omega_1-\omega_2$, $2\omega_2+\omega_1$ and $2\omega_2-\omega_1$. If the fundamental frequencies ω_1 and ω_2 are located within the resonator bandwidth, the intermodulation spectral lines $2\omega_1-\omega_2$ and $2\omega_2-\omega_1$ can also be within the bandwidth as shown on fig. 4. The other ones are filtered out. This effect is important in front-end filters when submitted to high powers.

The complete analysis of the intermodulation due to bulk effects was achieved for quartz by Tiersten [7] and for lithium tantalate by Planat [12].

For test tones symetrically located with respect to the resonator resonance frequency and with equal powers, the intermodulation level defined by the ratio P_{TT}/P_{IM} of the test tone power P_{TT} (at ω_1 or ω_2) over the third order intermodulation power P_{IM} (at $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$) can be written[7]

$$\frac{P_{TT}}{P_{IM}} = \frac{A^2}{\delta_c^2 P_{TT}^2} \qquad (7)$$

where A is a factor depending on the resonator geometry and on the electrical circuit parameters. δ_c is the effective fourth order nonlinear constant, which is to be calculated from equation (5) for the crystal cut under consideration. Its expression for a Y-cut thickness shear quartz resonator is

$$\delta_c = \frac{C_{222}}{2} + C_{2666} + \frac{1}{6} C_{6666} \qquad (8)$$

and involves the same coefficients as for the A-E effect. Therefore the fourth order elastic constant C_{6666} can also be evaluated from intermodulation measurements. Values are given in table III.

	Frequency (MHz)	C_{6666} (N/m^2)
Tiersten & Smythe	11.7	7.4×10^{12}
	111.4	8.0×10^{12}
Planat	10	4×10^{12}
	100	

Table III

For lithium tantalate the intermodulation level ratio is given by a lengthy expression which involves a large number of elastic constants. Therefore reduction of the data and evaluation of fourth order elastic constants was not performed. However the δ_c nonlinear constant was evaluated and compared with quartz[12]. For this purpose inter-

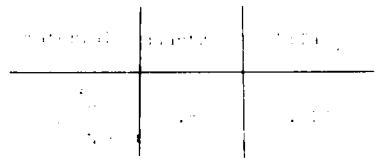


Table IV

temperature coefficient from the fundamental mode of the crystal. The third order temperature relation is :

$$f = f_0 \left[1 + a_1 (T - T_0) + a_2 (T - T_0)^2 + a_3 (T - T_0)^3\right] \quad (1)$$

where a_1, a_2, a_3 are the static T.C.'s measured at the turnover temperature T_0, and T is the dynamic T.C.

For AT-cut crystals $T_0 \approx 25°C$, and for BT-cut $T_0 \approx 65°C$.

This shows that achieving frequency stability over wide temperature ranges requires a temperature stable cut of the crystal to avoid the a_1 with a_2 and a_3 with BT-cut.

Initial work was processed with surface wave resonators, preliminary results on regular ST-cut SAW resonators, which are all the same order of magnitude as for BAW AT-cuts[12], but strongly dependant on the mounting and the crystal fixation.

Pressure-sensitivity

The influence of diametric forces or bending forces applied at the periphery of circular plates has been studied experimentally and theoretically in both cases of doubly rotated quartz cuts[44-46]. For the theoretical models based on isotropic[47] and anisotropic[48] plate bending stress distribution[49], the results are in good agreement with the measured values, as shown on fig. 8 for AT and SC cuts. Minimum and maximum of sensitivity can be obtained by positioning the plate fixation points, either for desensitizing the resonator or at the opposite increasing the sensitivity for sensor applications[47,48].

Frequency shifts also are caused by stresses in the thin metal films deposited on the quartz substrate. The long term aging of resonators is partly attributed to the relaxation of these stresses. The SC-cut orientation corresponds to a compensation of the influence of such planar stresses[50,51,52]. The three new doubly rotated TS, TTC, TTTC correspond to stress compensation and it is not so surprising that their orientation angles are very close ; the differences between thermal and electrode stresses, anisotropic and isotropic stresses, being rather small.

Force-sensitivity of SAW delay lines was also investigated in the case of compressional and bending forces on rectangular plates, and of diametric forces on circular plates[70,71].

Comparable sensitivities are observed (fig. 9). Because of their planar structure SAW devices seem to be well adapted to sensor applications.

Pressure-sensitivity

For stable quartz resonators pressure sensitivity does not appear as long as the crystal is under vacuum, as expected with good glass or metal enclosures. Therefore the studies of pressure-sensitivity were achieved rather for optimizing pressure transducers.

With bulk waves resonators the hydrostatic pressure is not directly applied to the crystal, but transformed into diametric forces by means of particular enclosures[68]. Therefore in this case force-sensitivity must be considered first.

With surface waves the simplest device could consist in applying the hydrostatic pressure on the different faces of a crystal held on a rigid plane. At 100 MHz the corresponding sensitivity is 1 Hz/kPa, and is to low for sensor application. The solution consists in using a thin anisotropic circular plate clamped along its periphery, which is bended by the hydrostatic pressure applied on one side[72,73]. The sensitivity as a function of the position of the wave on the diaphragm is shown on fig. 9. This calculation was made by using a threedimensional model for determinating the static stress distribution[70,71]. A differential pressure sensor is obtained by using two delay lines on the same diaphragm. This enables to compensate the influence of temperature, and to increase the sensitivity to 400 Hz/kPa.

Acceleration-sensitivity

A quartz plate subjected to an acceleration field experiences internal forces (the body forces) and external forces (the reaction forces at the fixation points). Because of the implications of g-sensitivity for spatial and tactical quartz oscillators many studies were undertaken on this subject.

Valdois[74] presented experimental results on the dependance of the sensitivity to acceleration direction, as shown on fig. 10. The origin of the g-sensitivity is attributed by Lee[75,76] to angular imperfections in the position of the supports on the crystal periphery, and by Janiaud[77] to disymetry of the resonator composed of the crystal blank and of the mounting. This last hypothesis was verified with B.V.A. resonators, the technology of which enables to make mounting and fixation bridges much more symetrical. A reduction of the g-sensitivity from 10^{-9}/g to 5.10^{-11}/g was achieved[78]. Other possibilities of improvement are offered by compensation techniques, which utilize two resonators connected electrically in series with opposite sensitivities[78,79,80], or enantiomorphous stacked crystals[81] or auxiliary accelerometers for open-loop compensation[82,83].

Even if the g-sensitivity can be reduced by one order of magnitude, such an improvement is still not sufficient, and a combination of appropriate mounting, less sensitive doubly rotated

cuts and compensation effect is to be used to desensitize the bulk wave resonators by one more order of magnitude.

For SAW devices the situation is worse, and the first measurements, performed by Levesque, show a sensitivity hundred times large, i.e. of the order of $10^{-7}/g$[84,85]. Indeed the contribution of the mounting is dominant. This is shown on fig. 11, where two or four supports were used. The worst configuration corresponds to a plate fixed with epoxy on a rigid plane.

c. Electric-field sensitivity

The application of a DC field on a quartz resonator induces frequency shifts through different mechanisms.

- If the DC field is applied in a direction for which the crystal is piezoelectric a deformation follows which can modify the plate thickness and the specific mass, and therefore the resonance frequency. This is a direct effect, but which occurs only for particular directions.

- An indirect, but non negligible, effect is the nonlinear coupling between the static strains and the high frequency wave through the third order nonlinearities as for a mechanical bias. It is of the same order of magnitude as the previous one.

- Electrostriction is present even in non-piezo material. It corresponds to quadratic terms in the strain (or stress) -electric field relation. It appears as a correcting quantity in the static strain generation.

- More important is the electroelastic effect. It corresponds in the strain-field relation to terms with products of a strain component by a field component. It is equivalent to a direct modification of the crystal elastic constants by the DC field.

The study of these effects was undertaken in length and thickness extensional resonators and in singly and doubly rotated thickness shear resonators[86-91]. Essentially effective electroelastic constants or combinations of constants were calculated by reduction of the experimental results. But the complete set of values of the fundamental electroelastic constants is still to be obtained.

Static measurements of some nonlinear third and fourth order piezoelectric constants and of the complete set of dielectric constants up to the fourth order of quartz were achieved by Besson[79].

A secondary effect due to the crystal ionic conductivity can be observed. The initial frequency shift due to the previous effects is followed by a relaxation phenomenon as shown on fig. 12. This relaxation corresponds to the migration of ionic impurities (Na^+, Li^+, K^+ ...), under the DC field. These alkalin ions are trapped by Al^{+++} atoms and occupy an interstitial position. The relaxation can be characterized by a relaxation time τ, according to the relation

$$\tau = \tau_o \exp (E/k_o T) \qquad (10)$$

where E is the activation energy of the ion, k_B the Boltzmann constant and T the absolute temperature (typically $\tau_o \approx 10^{-13}$ s $E < 1$ eV).

This relaxation phenomenon opens a new possibility for studying at room temperature the impurities of quartz crystal by separating the exponential components of the relaxation and calculating the different activation energies. In account of the differences in the ion mobility each value of the activation energy is characteristic of one type of impurity[92].

If the static frequency-electric field effect is of interest in quartz for the fundamental points of view of determination of the electroelastic constants and of characterization of impurities, new possibilities are offered by electron-phonon interactions in piezo-semiconductors.

In a crystal, which at the same time is piezoelectric (the elastic wave goes along with an electric field) and semiconductor (charge carriers can move under an electric field) it is possible to speed up or down the elastic wave by applying a DC field in the propagating wave direction or in the reverse direction. This phenomenon was used for acoustic amplification[93], but with moderate success. More recently, Grudkowski showed the feasability of high-Q SAW resonators with Gallium Arsenide (Q-factors almost as good as with quartz were achieved) and the possibility of controlling the resonance frequency with an external DC voltage by varying the depletion depth beneath a Schottky barrier electrode[94]. Velocity or frequency shifts as large as 500 ppm can be obtained for a reverse bias of -30 V. These new devices must be pointed out (even if not made of quartz) because they will open interesting possibilities for TCXO, VCXO, and phase shifter applications.

Crystal anharmonicities and phonon interactions

The vibrations of a solid can also be described by a microscopic model. The lattice vibrations are represented by elementary lattice waves or phonons, and if the interatomic forces are considered to be linear, harmonic approximation, the different lattice modes are independant. Therefore the energy stored in each given mode is constant and there is no sharing between different modes.

According to quantum mechanics the energy of each mode takes only discrete value and the total mean energy of the lattice is

$$\bar{E} = \sum_{jk} \hbar\omega(kj) \left[n(k_j) + \tfrac{1}{2}\right] \qquad (11)$$

where the summation is performed over all normal modes of wave number k and frequency ω and over all branches j. $n(k_j)$ is the mean number of pho-

...tons with energy $\hbar\omega(kj)$ stored in mode kj and $h\nu_0$ (where ν_0 is the zero point energy).

The mean number of phonons n(kj) follows the Bose-Einstein distribution law

$$n(kj) = \frac{1}{\exp\left(\hbar\omega(kj) / k_B T\right) - 1} \qquad (12)$$

This representation is sufficient to introduce the concept of specific heat, given by the mean energy temperature derivative in the absence of stresses

$$C = \frac{\partial W}{\partial T} = \frac{\hbar}{VT} \sum_{kj} \omega^2(kj)\, n(kj)\, [n(kj)+1] \qquad (13)$$

with $\beta = 1/k_B T$. V is the total volume of the crystal.

A more realistic approach has to go beyond the harmonic approximation and to consider cubic and quartic terms in the atomic displacements for defining the potential energy. As a consequence of these anharmonicities are the possible interactions between phonons. The actual presentation refers to the works of Klemens[97], Maradudin[98], and Maris[99].

In the Landau-Rumer theory[98] these interactions are represented by means of higher order Hamiltonians composed of products of creation a^+_k and annihilation a_k operators. Thus for three phonon interaction processes it is written :

$$\sum_{kk'k''} a(kj)\, a(k'j')\, a(k''j'')\, a^+(kj)\, a^+(k'j')\, a^+(k''j'').$$

The interaction must verify energy conservation law : $\hbar\omega(kj) - \hbar\omega(k'j') + \hbar\omega(k''j'')$. According to the involved process the quasimomentum is conserved (N processes) or not (Umklapp processes). These interactions give a finite life time to the phonons.

If N(kj) is the number of phonons of mode k when out of equilibrium, the collisions between phonons will force mode k to come back to equilibrium. A relaxation time $\tau(kj)$ is defined by the relation

$$\left(\frac{\partial N}{\partial t}\right)_{\text{collisions}} = -\frac{N(kj) - n(kj)}{\tau(kj)} \qquad (14)$$

The sound wave is considered as a beam of low energy phonons (kj) and frequency $\omega(kj)$. Its number of phonons is largely in excess with respect to the equilibrium number. On account of the uncertainty principle, the finite life time of a phonon makes its energy uncertain by $\hbar/\tau(kj)$. Beyond the Landau-Rumer method should be valid when $\hbar/\tau \gg \hbar\omega$, or $\omega\tau \ll 1$. This corresponds to high frequency sound wave or to large thermal phonon lifetimes when temperature is low.

A more classical approach is due to Akhiezer[99] which is valid in the limit $\hbar\omega \ll k_B T$, when the phonon mean free path is smaller than the sound wave length. This condition is verified at high temperature. In this way the sound wave is treated macroscopically and the anharmonicities are introduced by stating that the frequency $\omega(kj)$ of a phonon depends upon the state of strains of the crystal

$$\omega(kj) = \omega_0(kj) + \frac{\partial\omega(kj)}{\partial\eta_{\alpha\beta}}\,\eta_{\alpha\beta} \qquad (15)$$

$\eta_{\alpha\beta}$ is the Lagrangian strain component.

Thermal conductivity

The heat flow in a given direction α is

$$q_\alpha = \frac{1}{V} \sum_{kj} N(kj)\, \hbar\omega(kj)\, v_\alpha(kj) \qquad (16)$$

where $v_\alpha(kj)$ is the phonon group velocity. The difference $\Delta N(kj) = N(kj) - n(kj)$ is calculated with the Boltzmann transport equation, which at the steady state and without strains takes the form

$$\left(\frac{\partial N}{\partial t}\right)_{\text{coll}} = v_\alpha(kj)\, \frac{\partial N(kj)}{\partial T}\, \frac{\partial T}{\partial x_\alpha} \qquad (17)$$

$\partial T/\partial x_\alpha$ represents the temperature gradient.

With the approximation $\partial N/\partial T \simeq \partial n/\partial t$ this leads to

$$q_\alpha = -\frac{1}{V} \sum_{kj} \tau(kj)\, v_\beta(kj)\, \frac{\partial T}{\partial x_\beta}\, \frac{\partial n(kj)}{\partial T}\, \hbar\omega(kj)\, v_\alpha(kj) \qquad (18)$$

The thermal conductivity coefficient is deduced from eq (18)

$$\kappa_{\alpha\beta} = \frac{\hbar}{VT} \sum_{kj} \tau(kj)\, v_\alpha(kj)\, v_\beta(kj)\, \omega^2(kj)\, n(kj) \times [n(kj)+1] \qquad (19)$$

The finite thermal conductivity of the crystal entirely results on the anharmonicities and mainly on Umklapp interaction processes, without those interaction heat would propagate as acoustic waves.

If the mean value of a given quantity, let us say f, is defined by the relation

$$\langle f \rangle = \frac{\sum_{kj} f\, \omega^2(kj)\, n(kj)\, [n(kj)+1]}{\sum_{kj} \omega^2(kj)\, n(kj)\, [n(kj)+1]} \qquad (20)$$

With the help of eq (13) one can write

$$\sum \omega^2 n(n+1) = \frac{CVT}{\hbar k^2} \langle f \rangle \qquad (21)$$

Thus the thermal conductivity coefficient takes the simpler form

This expression is to be compared to the relation obtained from the kinetic theory of gases

Thermal expansion

Thermal expansion has its origin in the anharmonicities. The free energy A of the crystal is the sum of the energy due to elastic deformations and of the thermal phonon energy

The strain derivatives, at constant temperature, give the thermodynamic tension

By the same expression the thermal expansion coefficient is obtained by deriving with respect to temperature at constant tension.

The Grüneisen constant $\gamma(k_j)$ refered to a phonon of polarization e_j and wave vector k_j is defined by

Then the thermal expansion is written

Acoustic attenuation

Again because of the anharmonicities, a sound wave propagating in the solid will exchange energy with the heat reservoir constituted by the cloud of thermal phonons. In average this exchange will correspond to a loss of energy of the sound wave, given by the attenuation factor

where γ^2_{eff}, an effective Grüneisen constant, is a complex function of $\omega\tau$ and depends on the involved interaction processes. v_{ij} is the sound wave phase velocity and ρ the crystal specific mass.

A simultaneous velocity shift is observed related to the same Grüneisen constant

These results show that thermal conductivity, thermal expansion, acoustic attenuation and related velocity (or frequency) shifts all have their origin in the crystal lattice anharmonicities. The microscopic theory enables to relate these phenomena, which generally are considered as independant in the linearized macroscopic approach. Connection can be made between microscopic and macroscopic descriptions and for instance the Grüneisen constant was related to the third order elastic constants by Brugger [14].

Conclusion

Notewortle improvements of the performances of quartz resonators have been achieved during the last decade, which largely are a result of the studies on nonlinearities. Solutions of most of the different problems were found, at least in principle, and the important contribution of doubly rotated cuts, and particularly of SC-cut, must be pointed out. The doubly rotated cuts enable to reduce amplitude-frequency effect, and therefore also intermodulation, they exhibit fewer activity dips, they minimize the planar-stress sensitivity, and especially they improve the dynamic thermal behavior, which is one of the main causes of short term instabilities of oscillators. Quartz is the crystal with the main qualities. However for some applications other materials can be advantageously used, such as lithium tantalate which is more linear. For berlinite the studies are still on process and no conclusion can be given yet. Also some new possibilities are offered by surface waves, which are mostly interesting with regard to their planar structure ; bulk and surface waves could find complementary applications.

The improvements are also due to new designs of resonators, like BVA resonators[15], which allow better control of the electrical and geometrical characteristics, to new technology for very high frequency resonators[16], and to new electronics system with lowered noise[17].

The characterisation of the quartz material is the next important step. It has been shown how the anharmonicities contribue to the thermal and acoustical properties of the crystal. The actual studies on frequency fluctuations and flicker noise sources try to demonstrate that their origins must also be found in the anharmonicities and phonon interactions.

References

(1) F.D. Murnaghan, "Finite deformation of an elastic solid", Wiley and Sons Inc., (1951).

(2) E.A. Toupin, "The elastic dielectric", J. Ration. Mech. Anal., 5, 849, (1956).

(3) R.N. Thurston, "Wave propagation in fluids and normal solids", Physical Acoustics, W.P. Mason Ed., vol. 1, part A, Academic Press (1964).

(4) H.F. Tiersten, "On the nonlinear equations of thermo-electroelasticity, Int. J. Eng. Sci., 9, 587, (1971).

(5) P.B. Thompson, H.F. Tiersten, "Harmonic generation of longitudinal elastic waves", J. Acoust. Soc. Am., 62, n° 1, 33, (1977).

(6) H.F. Tiersten, J.C. Baumhauer, "Second harmonic generation of surface waves in isotropic elastic solids", Ultrasonics Symposium, IEEE cat. n° 80 CH1578-3, (1980).

(7) E.J. Vella, J.L. Stegeman, P.M. Bridge, "Analysis of parametric mixing and harmonic generation of surface acoustic waves", J. Appl. Phys., 50, 82, (1979).

(8) M. Planat, G. Théobald, J.J. Gagnepain, "Propagation non linéaire d'ondes élastiques dans un solide anisotrope"
I Ondes de volume Onde Electrique, 60, n° 8-9, 33, (1980).
II Ondes de surface Onde Electrique, 60, n° 11, 61, (1980).

(9) H.F. Tiersten, "Analysis of nonlinear resonance in rotated Y-cut quartz thickness-shear resonators, 29th AFCS, (1975).

(10) H.F. Tiersten, "An analysis of intermodulation in rotated Y-cut quartz thickness-shear resonators", 28th AFCS, (1974).

(11) M. Planat, G. Théobald, J.J. Gagnepain, P. Siffert, "Intermodulation in Xcut lithium tantalate resonators", Electronics Letters, 16, n° 5, 174, (1980).

(12) V.E. Bottom et al, "Mounting techniques for improved heat dissipation in quartz crystal units", Contract DA 36-039 sc-5485, Final Report, (1952).

(13) A.W. Warner, "Design and performance of ultraprecise 2-5 mc quartz crystal units", Bell System Technical Journal, 39, 1193, (1960).

(14) D.L. Hammond, C. Adams, L. Cutler, "Precision crystal units", 17th AFCS, (1963).

(15) A. Seed, "Nonlinear effects in quartz resonators", 4th Int. Congress on Acoustics, Copenhagen, Aug. (1962).

(16) A. Smolarski, "Free oscillation-frequency method for investigating quartz resonators", Bull. Acad. Pol. Sci., XII, n° 6, 11, (1964).

(17) J.J. Gagnepain, R. Besson, "Nonlinear effects in piezoelectric quartz crystals", Physical Acoustics, W.P. Mason Ed., vol. XI, 245, Academic Press, (1975).

(18) M. Planat et al., "Nonlinear propagation of surface acoustic waves on quartz", 34th AFCS, 255, (1980).

(19) J.J. Gagnepain, J.C. Poncot, C. Pégeot, "Amplitude-frequency behavior of doubly rotated quartz resonators", 31st AFCS, 17, (1977).

(20) R.N. Thurston, J.J. McSkimin, P. Andreatch, "Third order elastic coefficients of quartz", J. Appl. Phys., 37, (1966).

(21) Landolt-Bornstein, "Elastic, piezoelectric and related constants of crystals, III, vol. 11, Springer-Verlag, (1979).

(22) H.F. Tiersten, "Analysis of nonlinear resonance in thickness-shear and trapped-energy resonators", J. Acoust. Soc. Am., 59, n° 4, 866, (1976).

(23) M. Planat et al, "Nonlinear characteristics of SAW grooved resonators", Ultrasonics Symposium, IEEE 80 CH 1602.2, (1980).

(24) M. Planat, L. Penavaire, "Influence of quartz anisotropy on the nonlinear characteristics of surface acoustic waves", to be published.

(25) W.J. Tanski, "High performance SAW resonator filters for satellite use", Ultrasonics Symposium, IEEE 80 CH 1602.2, (1980).

(26) H.F. Tiersten, "Analysis of intermodulation in rotated Y-cut quartz thickness-shear resonators", 28th AFCS, 1, (1974).

(27) M. Planat et al, "Intermodulation in X-cut lithium tantalate resonators", Electronics Letters, 16, n° 9, 174, (1980).

(28) R.C. Smythe, "Intermodulation in thickness-shear resonators", 28th AFCS, 9, (1974).

(29) A.F. Wood, A. Seed, "Activity dips in AT-cut crystal", 21st AFCS, 420, (1967).

(30) C. Franx, "On activity dips of AT crystals at high level of drive", 21st AFCS, 436, (1967).

(31) J. Birch, D.A. Weston, "Frequency/temperature, activity/temperature anormalies in high frequency quartz crystal units", 30th AFCS, 32, (1978).

(32) J. Balbi, J. Duffaud, R. Besson, "A new nonlinear analysis method and its application to quartz crystal resonator problems", 32nd AFCS, 162, (1978).

(33) J.C. Baumhauer, H.F. Tiersten, "Nonlinear electroelastic equations for small fields superposed on a bias", J. Acoust. Soc. Am., 54, n°4, 1017, (1973).

(34) H.F. Tiersten, "Perturbation theory for linear electroelastic equations for small fields superposed on a bias", J. Acoust. Soc. Am., 64, n° 3, 832, (1978).

(79) A. Kusters, C.A. Adams, "Production statistics of SC (or TTC) crystals", 34th AFCS, 167, (1980).

(80) P. Hauden, S. Rousseau, J.J. Gagnepain, "Sensitivities of SAW oscillators to temperature, forces and pressure", 34th AFCS, 312, (1980).

(81) J.F. Dias, F. Karrer, "Stress effects in acoustic surface wave circuits and applications to pressure and force transducers", IEEE Int. Solid-state Conf., 166, (1974).

(82) T.M. Reeder, D.E. Cullen, M. Gilden, "SAW oscillator pressure sensors", Ultrasonics Symp., IEEE 75 CHO 994-4SU, (1975).

(83) S. Rousseau, "Sensibilité à la pression de lignes à ondes de surface : étude d'un capteur à quartz", Thesis n° 104, Besançon, LPMO, March (1980).

(84) M. Valdois, J. Besson, J.J. Gagnepain, "Influence of environment on a quartz resonator", 28th An. Freq. Cont. Symp., 19, (1974).

(85) E.C.Y. Lee, Kuang-Ming Wu, "Effects of acceleration on the resonance frequencies of crystal plates", 30th AFCS, 1, (1976).

(86) E.C.Y. Lee, Kuang-Ming Wu, "The influence of support-configuration on the acceleration sensitivity of quartz resonator plates, 31st AFCS, 29, (1977).

(87) J. Janiaud, "Modélisation de l'influence d'une accélération sur la fréquence des résonateurs à quartz", Thesis report. ONERA, Chatillon/Bagneux, March (1978).

(88) R. Besson et al, "Design of a bulk wave quartz resonator insensitive to acceleration", 33rd AFCS, 347, (1979).

(89) J.J. Gagnepain, F.L. Walls, "Quartz crystal oscillators with low acceleration sensitivity", Techn. Report. NBSIR 77-855.

(90) R. Besson, "Quartz crystal and superconductive resonators and oscillators", Proc. 10th PTTI, Nasa Techn. Mem. 80250, 101, (1978).

(91) A. Ballato, "Resonators compensated for acceleration fields", 33rd AFCS, 322, (1979).

(92) J.M. Przyjemski, "Improvement in system performance using a crystal oscillator compensated for acceleration sensitivity", 32nd AFCS, 429, (1978).

(93) J.A. Emmons, "Acceleration sensitivity compensation in high performance crystal oscillators, Proc. 10th PTTI, Nasa Techn. Memo. 80250, 55, (1978).

(94) M. Valdois, P. Levesque, P. Hartemann, "Acceleration dependence of the surface acoustic wave oscillator frequency", Ultrasonics Symp. Proc. IEEE 77 CH 1264-1SU, 936, (1977).

(95) P. Levesque et al., "Theoretical and experimental analysis of SAW quartz oscillator acceleration sensitivity", Ultrasonic Symp., Proc. IEEE 79 CH 1482-9, 896, (1979).

(86) K. Hruska, V. Kazda, "The polarizing tensor of the elastic coefficients and moduls of quartz", Czech. J. Phys., 18, 500, (1978).

(87) J. Kusters, "The effect of static electric fields on the elastic constants of a quartz", 24th AFCS, 46, (1970).

(88) K. Hruska, A. Khogali, "Polarizing effect with alpha-quartz rods and the electroelastic tensor", IEEE Trans. Sonics and Ultrasonics, SU-18, n° 3, 171, (1971).

(89) K. Hruska, "The electroelastic tensor and other second order phenomena in quasilinear interrelation of polarizing effect with thickness vibrations of a quartz plate", 31st AFCS, 159, (1977).

(90) K. Hruska, "Polarizing effect with doubly rotated alpha-quartz plates vibrating in thickness", IEEE Trans. Sonics and Ultrasonics, SU-25, n° 6, 390, (1978).

(91) R. Besson, "Measurement of nonlinear elastic, piezoelectric, dielectric coefficients of quartz crystal. Applications", 28th AFCS, 8, (1974).

(92) R. Brendel, "Characterization of ionic impurities in quartz crystal by frequency relaxation under DC electric fields", to be published.

(93) A.P. Hutson, D.L. White, "Elastic wave propagation in piezoelectric semiconductors", J. Appl. Phys., 33, 40, (1962).

(94) J.W. Grudhowski et al, "GaAs monolithic SAW devices for signal processing and frequency control", Ultrasonics Symp., 80 CH 1602.2, 88, (1980).

(95) P.G. Klemens, "The thermal conductivity of dielectric solids at low temperatures", Proc. Roy. Soc. A 208, 108, (1951).

(96) A.A. Maradudin, "Thermal expansion and phonon frequency shifts", Phys. Stat. Sol., 2, 1493, (1962).

(97) A.J. Maris, "Interaction of sound waves with thermal phonons in dielectric crystals", Physical Acoustics, vol. VIII, 279, Academic Press, N.Y. (1971).

(98) L. Landau, G. Rumer, Phys. Z. Sowjet Union, 11, 18, (1937).

(99) A. Akhieser, "On the absorption of sound in solids", J. Phys. (USSR), 1, (1939).

(100) R. Brugger, "Generalized Grüneisen parameters in the anisotropic Debye model", Phys. Rev., 137, n° 6A, (1963).

(101) R. Besson, "A new electrodeless resonator design", 31st AFCS, 147, (1977).

(102) M. Berte, "Acoustic bulk wave resonators and filters operating in the fundamental mode at frequencies greater than 100 MHz", 31st AFCS, 122, (1977).

(103) S.R. Stein et al. "A systems approach to high performance oscillators", 32nd AFCS, 527, (1978).

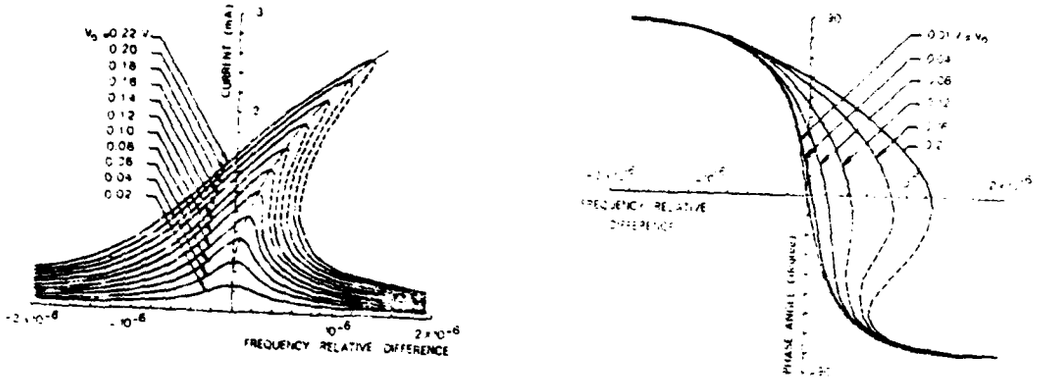


Fig. 1

a) Amplitude resonance curves (5 MHz, 5th overtone)

b) Phase resonance curves

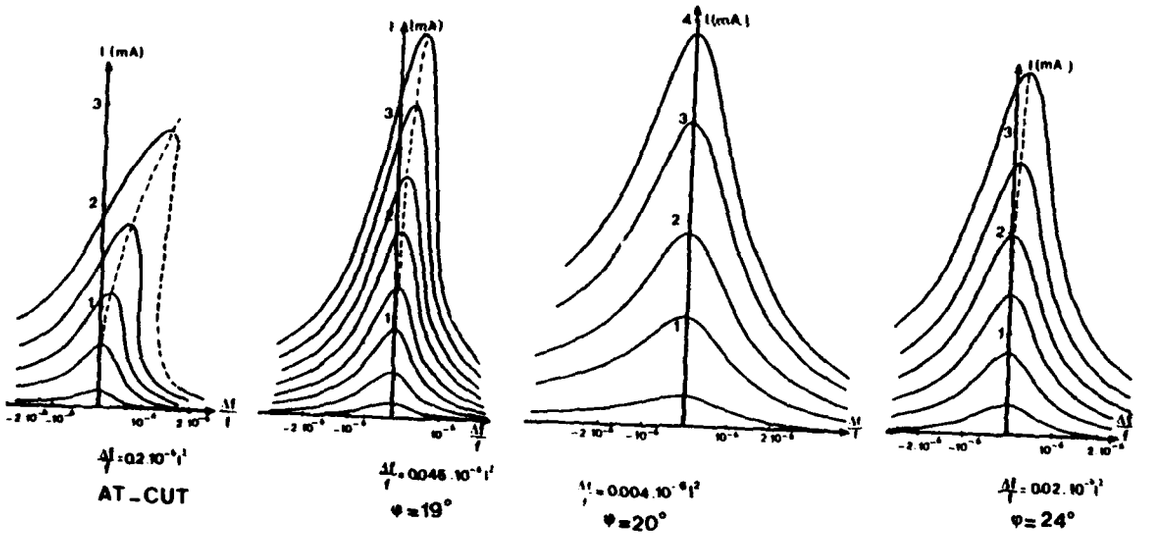


Fig. 2

Amplitude-frequency effect of doubly-rotated cuts (5 MHz, 3rd overtone)

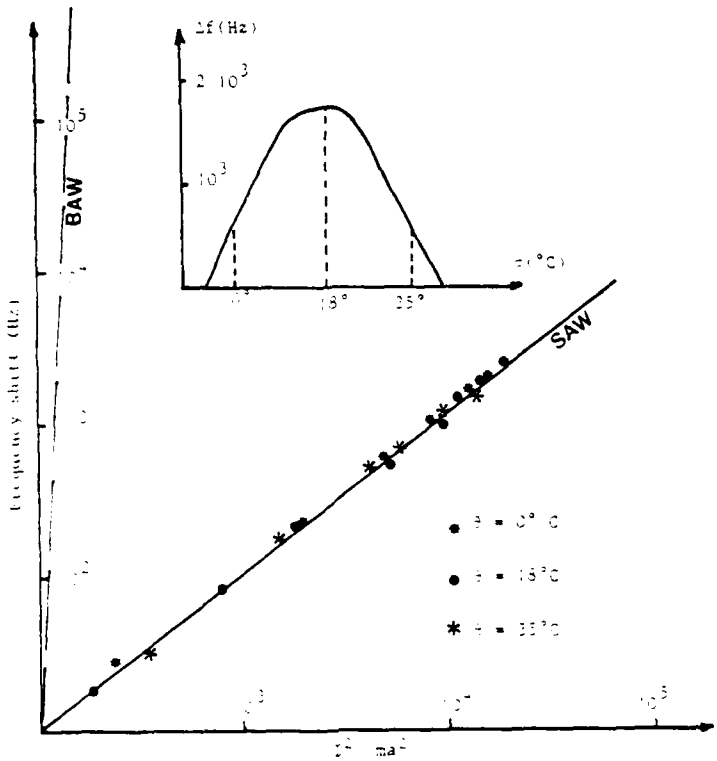


Fig. 3

Amplitude-frequency effect :
Comparison between BAW and SAW at 100 MHz

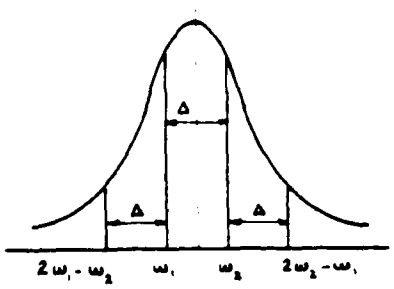


Fig. 4

Test tone and intermodulation frequencies

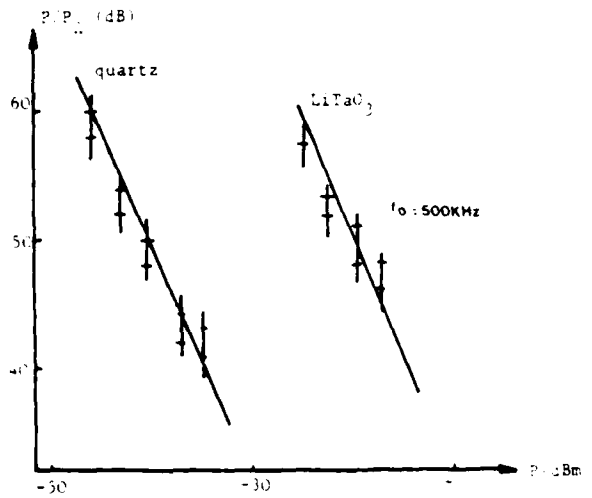
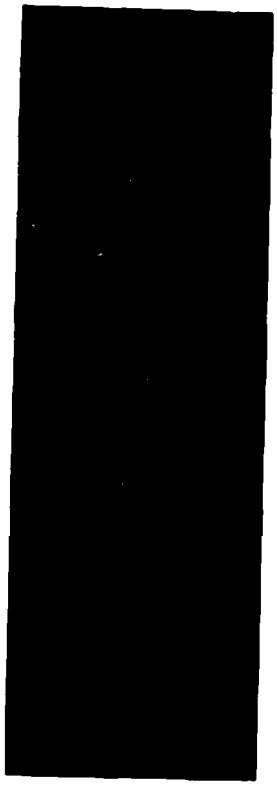


Fig. 5

Intermodulation :
Comparison between quartz and Lithium Tantalate



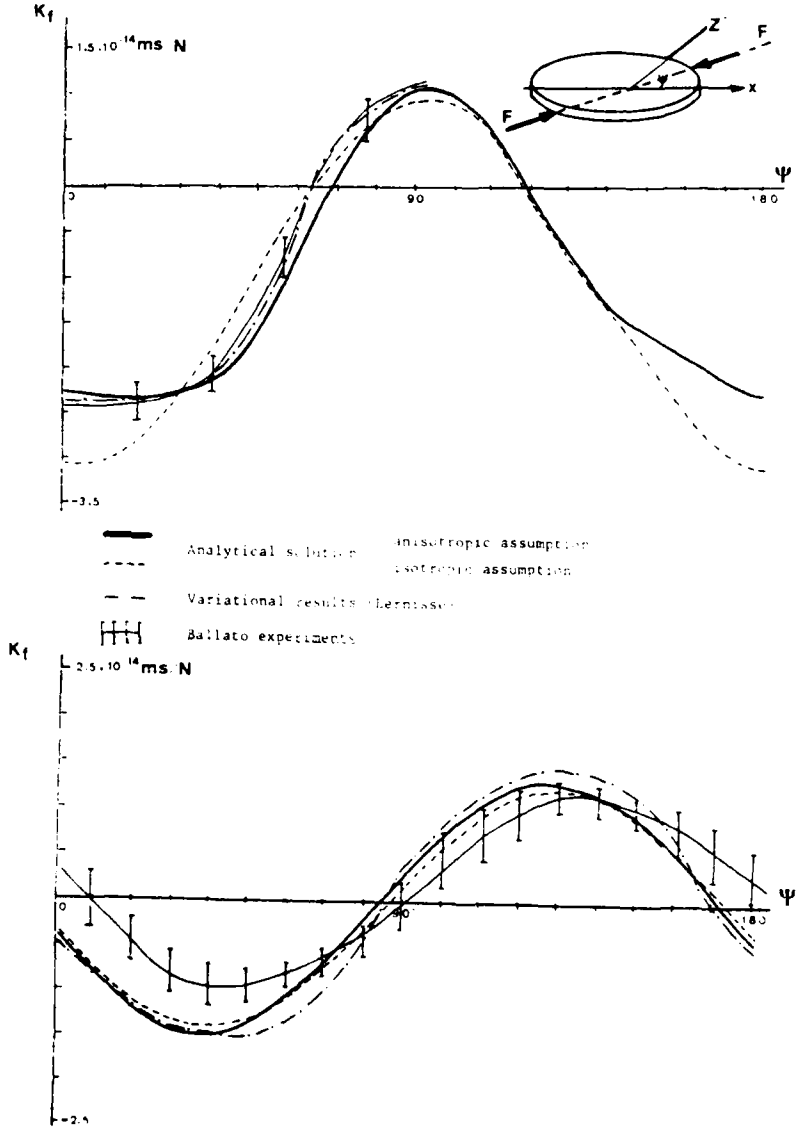


Fig. 6

5 MHz subharmonic generation observed in a resonator driven at 15 MHz

Fig. 7

a) Force sensitivity AT cut

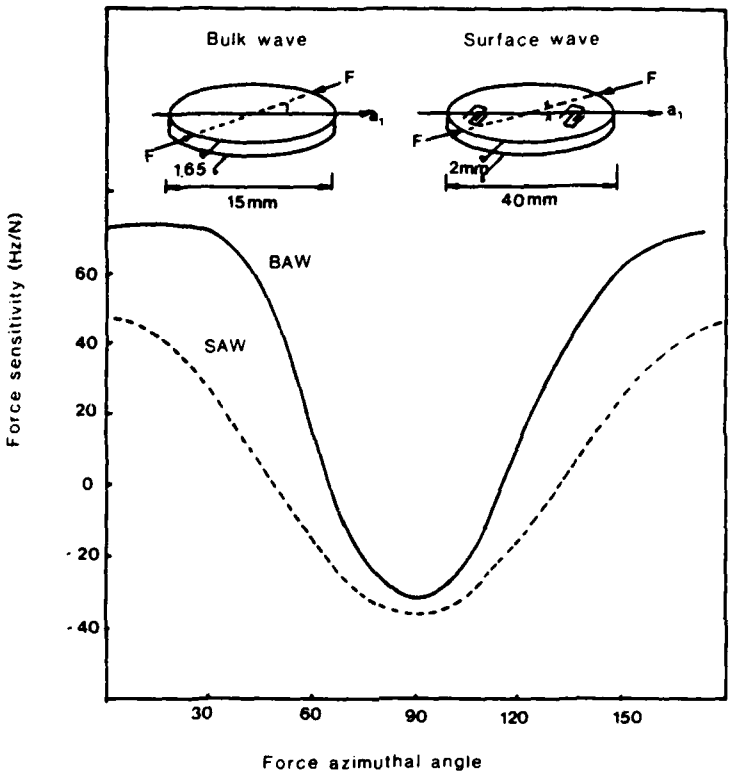
b) Force sensitivity : SC cut

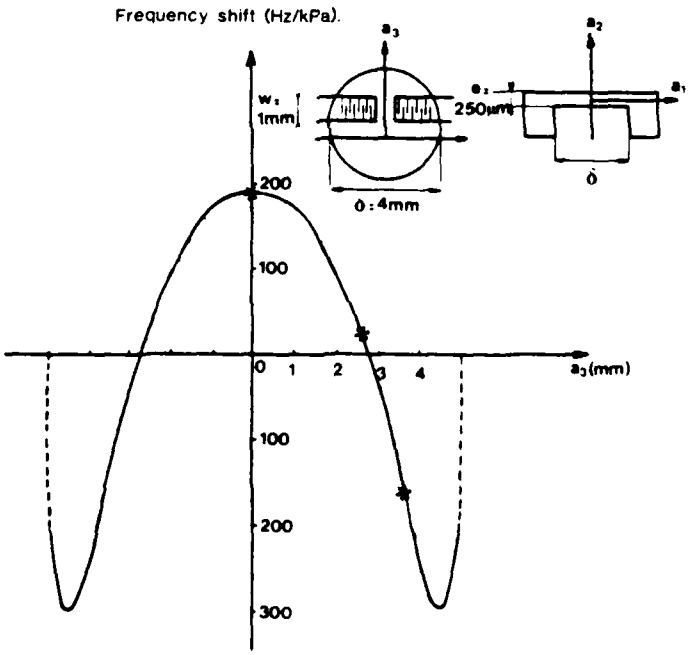
Fig. 8

Comparison between bulk and surface waves

Fig. 9

Pressure sensor


Bulk wave
Surface wave
F
F
F
F
1.65
15mm
2mm
40mm
BAW
SAW
Force sensitivity (Hz/N)
60
40
20
0
-20
-40
30
60
90
120
150
Force azimuthal angle


Frequency shift (Hz/kPa).
1mm
250μm
0:4mm
200
100
100
200
300
0
1
2
3
4
a3(mm)

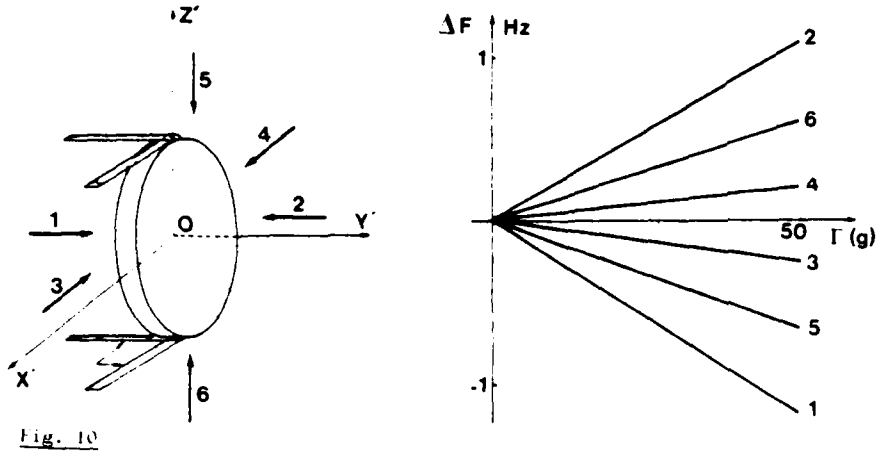
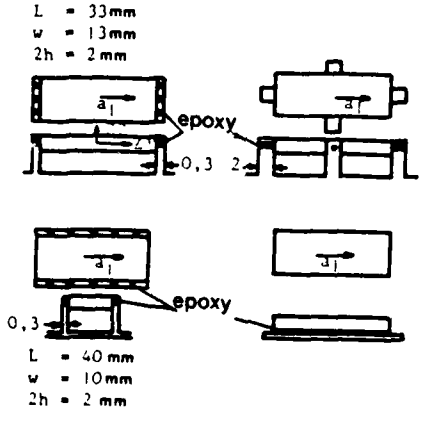


Fig. 10

Frequency variations as a function of acceleration



mounting	crystal cut	sensitivities (Hz G) at 105 MHz	
		measured	calculated
2 supports along Z_2 (fig. a)	AT	18	20
4 supports (fig. b)	ST	7,3	8,2
2 supports along Z_1 (fig. c)	ST	4,5	2,2
Epoxy (fig. d)	ST	45	

Calculated and measured acceleration sensitivities (at 105 MHz, propagation along Oa_1) as a function of the mounting configuration

Fig. 11

Calculated and measured acceleration sensitivities (at 105 MHz, propagation along Oa_1) as a function of the mounting configuration

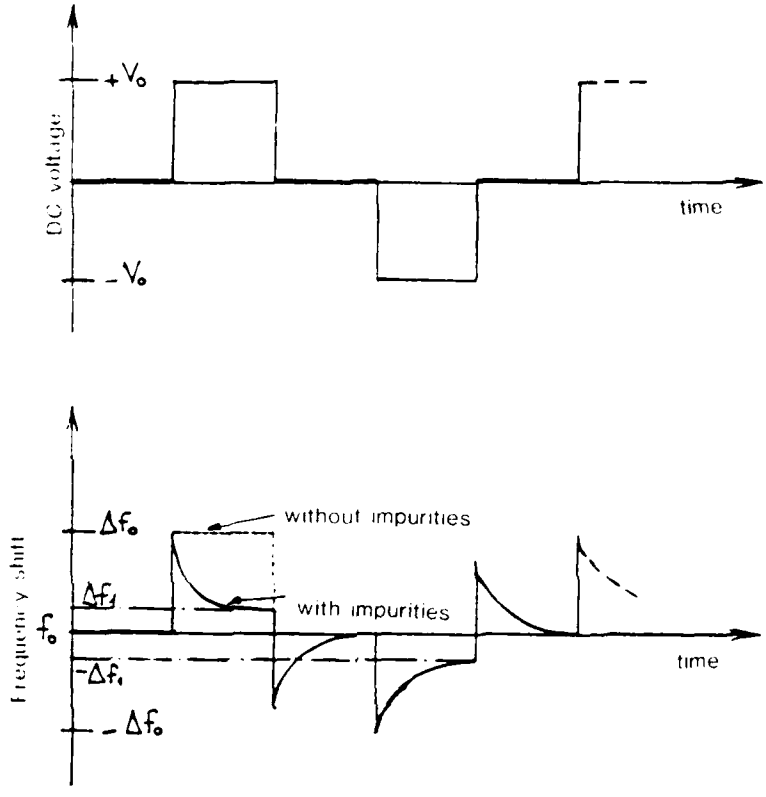


Fig. 12

Influence of an external DC field on the resonance frequency and relaxation phenomenon

[illegible] 1981 IEEE [illegible] Symposium, [illegible], May 1981

[illegible] OF ACCELERATION ON FREQUENCY STANDARDS AND CLOCKS

[illegible] Filler

[illegible] Army Electronics Research and Development Laboratory (ERADCOM)
Fort Monmouth, New Jersey 07703

Abstract

The acceleration sensitivity of frequency standards and clocks is assuming increasing significance in modern communications, navigation, and identification systems. A tutorial description is presented on how acceleration, particularly vibration, affects the frequency of the most acceleration sensitive component in modern frequency standards, the crystal oscillator. A simple mathematical model describing the effect of harmonic motion on the oscillator output is presented. The coefficient of acceleration sensitivity, [illegible], is shown to have vector properties. The acceleration sensitivity vector, [illegible], is introduced. The behavior of the oscillator output in both the time domain and the frequency domain is presented.

The simple model is expanded to include 1) the consequences of frequency multiplication which can lead to the loss of the carrier, and 2) random vibration. The phase noise resulting from a given vibration spectrum is calculated.

A method for the measurement of the vibration sensitivity coefficient is shown. A review is given of the current state-of-the-art in 1) the minimization of the sensitivity to vibration and 2) acceleration compensation.

Key words: Quartz, quartz crystals, quartz resonators, quartz oscillator, crystal oscillator, g-sensitivity, acceleration, vibration, phase noise, acceleration compensation.

Introduction

This is a tutorial paper on the effects of acceleration on frequency control devices. The major emphasis is on simple harmonic acceleration, i.e. vibration, although unidirectional accelerations and random vibrations are also treated. The effects of vibration are becoming more and more important as other noise sources are reduced and system performance requirements under extreme environmental conditions are made more stringent. Elsewhere in these Proceedings there are several papers describing research aimed at extending the state-of-the-art in acceleration resistant devices. The goal of this paper is to review the terminology and develop the mathematical foundation which will enable the frequency source vendor to properly specify his product and the user to understand the consequences of [illegible] operation.

Frequency vs. Acceleration

The most acceleration-sensitive component in precision frequency and time sources is the quartz crystal resonator. All of the mathematical development in this paper is dependent on the assumption that the acceleration sensitivity which arises from improper mechanical resonant design, such as loose components or wires, has been avoided.

It has been shown that the fractional frequency shift of the output frequency of a crystal oscillator is proportional to the change in acceleration magnitude that the oscillator experiences.[1] The proportionality constant is a function of the direction of the acceleration as referenced to a set of axes fixed to the oscillator. For a given direction, therefore,

$$\frac{F_1 - F_0}{F_0} = \frac{\Delta F}{F_0} = \Gamma A_1 - \Gamma A_0 \qquad (1)$$

where F_1 is the frequency when the resonator is experiencing an acceleration of magnitude A_1 and F_0 is the corresponding frequency for some other acceleration of magnitude A_0, in the same direction as acceleration A_1. For simplicity, A_0 is assumed to be zero. F_0 is therefore the static frequency. This equation can be generalized to include directional dependence by defining a direction dependent coefficient of acceleration sensitivity, $\Gamma(\theta, \phi)$. Equation (1) then becomes

$$\frac{\Delta F}{F_0} = \Gamma(\theta, \phi) A(\theta, \phi) \qquad (2)$$

where θ and ϕ are the spherical coordinates of the acceleration direction relative to a set of axes fixed to the oscillator. $A(\theta, \phi)$ is the magnitude of the acceleration in the direction defined by θ and ϕ.

The "Tipover" Test

The directional dependence of Γ can be demonstrated by performing the simple experiment depicted in Figure 1. An oscillator is rotated about a horizontal axis, e.g. axis 1, in the earth's gravitational field. (The magnitude of acceleration given in this paper will be in units of the magnitude of the earth's gravitational acceleration at sea level, which is 980 cm/sec^2

and is denoted by "g".) The frequency is recorded at various angular positions relative to the initial position. Each pair of positions separated by 180^0 of rotation corresponds to a change in acceleration of twice the earth's gravitational acceleration (up = +1g and down = -1g, therefore +1g - (-1g) = +2g) hence the term "2g tipover" test.

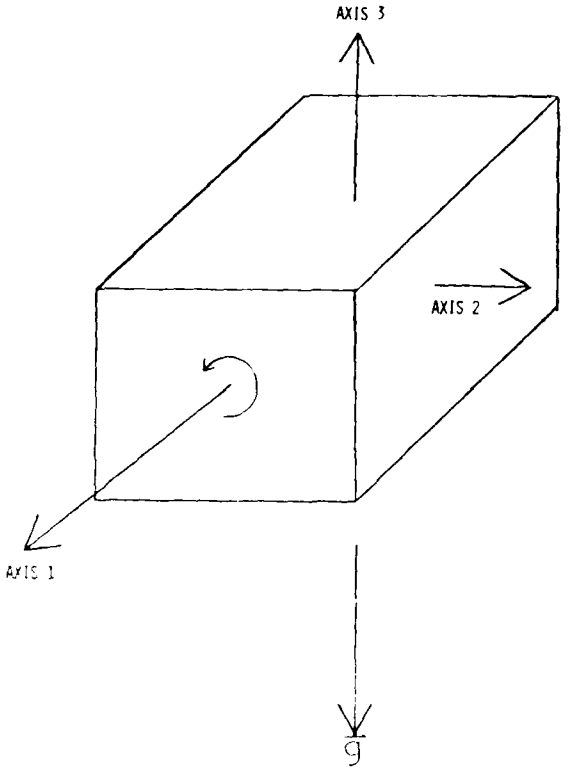


Fig. 1 - "2-g" Tip-over Test

Figure 2 shows actual data recorded for rotations about three mutually perpendicular axes, with data points taken every 22.5^0 in both the clockwise and counterclockwise directions. The solid line is a least-square fit to the function

$$Y(\theta) = A\sin(\theta + \delta) \qquad (3)$$

where $Y(\theta)$ is $(F(\theta) - F(0))/F(0)$, A and δ are parameters determined from the fit. To determine the 2g-tipover response one must subtract points separated by 180^0, to give

$$\Delta Y_{2g}(\theta) = Y(\theta) - Y(\theta - 180^0) \qquad (4)$$

Using equation (3) in equation (4), we have

$$\Delta Y_{2g}(\theta) = A[\sin(\theta + \delta) - \sin(\theta + 180^0 + \delta)] \qquad (5)$$

Since $\sin\theta = -\sin(\theta + 180^0)$, equation (5) becomes

$$\Delta Y_{2g}(\theta) = 2A\sin(\theta + \delta) \qquad (6)$$

where $\Delta Y_{2g}(\theta)$ is now the fractional frequency shift for a 2g change in acceleration.

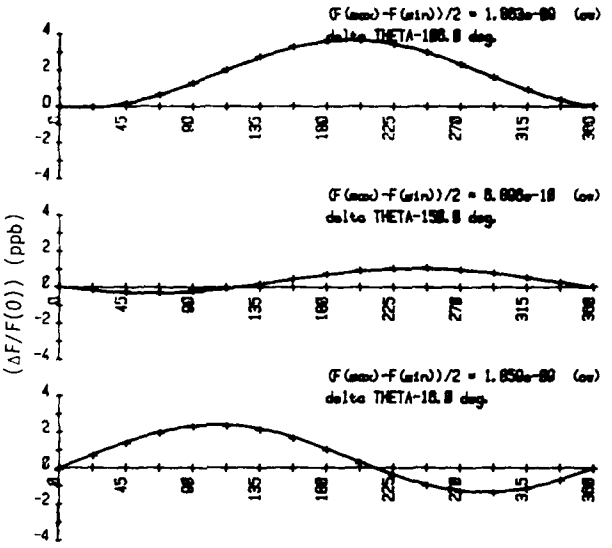


Fig. 2 - "2-g" Tip-over Test Results

To normalize this result to the per unit g result, one would divide by 2 and, therefore, arrive back at the original relationship given by equation (3). Since $\Delta Y_{2g}(\theta)$ is the same as $\Delta F/F$ in equation (2) and the magnitude of the acceleration is 1g, it can be seen that $\gamma(\theta)$ for a given ϕ is proportional to $\sin(\theta)$. From the sinusoidal dependence of $\gamma(\theta, \phi)$ it can be deduced that the acceleration sensitivity is a vector with magnitude and direction.[2,3] If this vector is denoted by $\bar{\Gamma}$, we can rewrite equation (2) as

$$\frac{\Delta F}{F_0} = \bar{\Gamma} \cdot \bar{A} \qquad (7)$$

where A is the acceleration vector. The magnitude and direction of Γ can be determined by measuring the scalar components along each of three mutually perpendicular axes (γ_1, γ_2, and γ_3). The magnitude is then given by

$$|\bar{\Gamma}| = \sqrt{\gamma_1^2 + \gamma_2^2 + \gamma_3^2} \qquad (8)$$

while the direction is given by

$$\hat{\Gamma} = \frac{\gamma_1}{|\bar{\Gamma}|}\hat{i} + \frac{\gamma_2}{|\bar{\Gamma}|}\hat{j} + \frac{\gamma_3}{|\bar{\Gamma}|}\hat{k} \qquad (9)$$

where $\hat{i}$, $\hat{j}$ and $\hat{k}$ are the unit vectors along axes 1, 2, and 3, respectively.

A significant consequence of the vector nature of the acceleration sensitivity is that for acceleration in the plane normal to Γ the scalar product and therefore the acceleration sensitivity

is always zero.

Simple Harmonic Motion

In most applications, some component of the motion experienced by electronic equipment is vibratory in nature. These vibrations can be expressed as a sum of sine and cosine terms. For simplicity, this discussion will be limited to simple harmonic motion at a single vibration frequency.

The displacement of an object undergoing simple harmonic motion can be expressed as

$$X(t) = \frac{D}{2} \cos (2\pi f_v t) \qquad (10)$$

where $X(t)$ = time dependent displacement, D = peak-peak amplitude, and f_v = vibration frequency.

The acceleration is given by the second derivative with respect to time of the displacement. Therefore, using equation (10) we have:

$$A(t) = A_p \cos (2\pi f_v t) \qquad (11)$$

where A_p = peak accelertation = $-(\frac{1}{2})(2\pi f_v)^2 D$

It can be seen from equation (11) that the peak acceleration, A_p, is dependent on the peak-to-peak displacement and the square of the vibration frequency. For a constant displacement the acceleration falls off rapidly with decreasing vibration frequency.

The variation of frequency with time for a frequency source undergoing simple harmonic motion can be determined by combining equation (11) with equation (2) and eliminating the scalar product by defining ΔF as $\Delta F \cdot \hat{A}$, where $\hat{A}$ is the unit vector in the direction of the acceleration. Then

$$\frac{\Delta F}{F_0} = \frac{\Delta A_p}{} \cos (2\pi f_v t). \qquad (12)$$

This can be rewritten as

$$F(t) = F_0 (1 + \Delta F_p \cos (2\pi f_v t)) \qquad (13)$$

where $\Delta F_p = \Delta A_p F_0$.

It can be seen from equation (13) that the output frequency fluctuates about the center frequency, F_0, between the limits $F_0 + \Delta F_p$ and $F_0 - \Delta F_p$, in a sinusoidal fashion at a rate of one cycle every $2\pi/f_v$ seconds. This variation is shown schematically in Figure 3. Figure 4 is an example of an oscilloscope trace of a vibrating crystal oscillator output. The upper curve is the vibration waveform and the lower curve is the oscillator output showing, in a much exaggerated fashion, the variation in output frequency with acceleration amplitude.

Frequency Domain

In many applications the frequency deviation due to vibration must be measured. It is difficult to obtain quantitative information from the time domain analysis above. It is very useful to

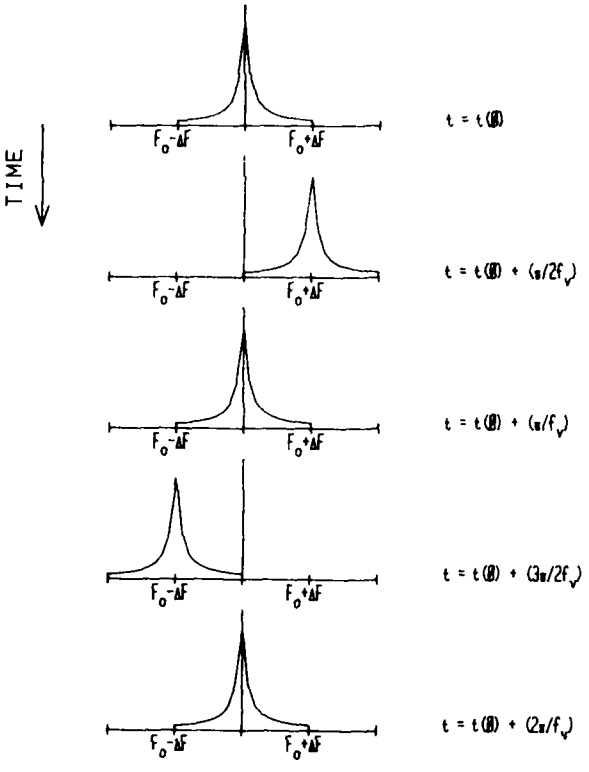


Fig. 3 - The "Instantaneous" Output Frequency of a Frequency Modulated Carrier

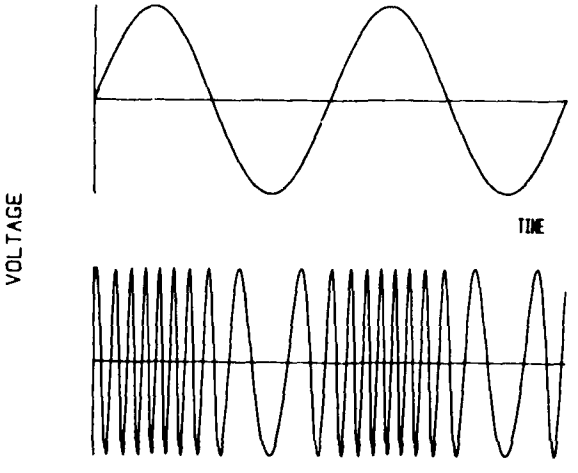


Fig. 4 - The Output Voltage of a Frequency Modulated Carrier and the Modulating Signal

transform equation (13) into the frequency domain where spectral analysis can be performed.[4] This will also help to explain the effect of vibration on systems such as doppler radar.

The voltage appearing at the output of an oscillator is given by

$$V(t) = V_0 \cos(\phi(t)) \qquad (14)$$

where the phase, $\phi(t)$, is the integral of the frequency over time. If the frequency is fluctuating, due to vibration, the phase in equation (14) becomes the integral of equation (13). That is

$$\phi(t) = \int f(t)\,dt = \int F_0 (1 + F_p \cos(2\pi f_v t))\,dt \qquad (15)$$

or

$$\phi(t) = 2\pi F_0 t + \left(\frac{F_p}{f_v}\right)\sin(2\pi f_v t). \qquad (16)$$

When this result is combined with equation (14), we get

$$V(t) = V_0 \cos\left(2\pi F_0 t + \left(\frac{F_p}{f_v}\right)\sin(2\pi f_v t)\right) \qquad (17)$$

This equation is in the form of a cosine of a sine which can be expanded in an infinite series of Bessel functions.[5] The inclusion of the Bessel function expansion into equation (17) results in

$$\begin{aligned} V(t) = {} & V_0 \cdot J_0(\beta) \cos(2\pi F_0 t) + \\ & J_1(\beta)\left[\cos(2\pi (F_0 + f_v)t) - \cos(2\pi (F_0 - f_v)t)\right] + \\ & J_2(\beta)\left[\cos(2\pi (F_0 + 2f_v)t) + \cos(2\pi (F_0 - 2f_v)t)\right] + \ldots \end{aligned} \qquad (18)$$

where $\beta = \frac{F_p}{f_v}$ and is called the modulation index.

The first term in equation (18) is the original carrier signal with an amplitude, relative to its quiescent state, of $J_0(\beta)$. $J_0(\beta)$ is less than 1 for all β's greater than zero. The other terms are the "vibration induced sidebands". There are sidebands at $F_0 + f_v$, $F_0 - f_v$, $F_0 + 2f_v$, $F_0 - 2f_v$, etc. The ratio of the amount of power carried by the nth sideband to the power remaining in the carrier is denoted by $\mathcal{L}_v^n$ (pronounced script L).

where

$$\mathcal{L}_v^n(f_v) = \frac{V_0 \cdot J_n(\beta)}{V_0 \cdot J_0(\beta)} = \frac{J_n(\beta)}{J_0(\beta)} \qquad (19)$$

or converted to decibels

$$\mathcal{L}_v^n(f_v) = 20 \log \frac{J_n(\beta)}{J_0(\beta)} \text{ dBc} \qquad (20)$$

where dBc refers to dB above or below the carrier.

Small Modulation Index

It is now useful to investigate the behavior of the J's and $\mathcal{L}_v$ for typical values of the modulation index, β. The specification of γ for the best currently available off-the-shelf oscillator is 2×10^{-9}/g. If the carrier frequency, F_0, is 5 MHz and the peak acceleration, A_p, is 5g, then $\beta = \gamma A_p F_0 = 0.05$ Hz. Several approximations can be made for Bessel functions if β is less than 0.01.[5] Therefore, for this oscillator, the following approximations are valid for all f_v above 5 Hz:

$$\begin{aligned} & J_0(\beta < 0.01) \approx 1 \\ & J_1(\beta < 0.01) \approx \beta/2 \\ & J_n(n > 2, \beta < 0.01) \approx 0 \end{aligned} \qquad (21)$$

The definition of $\mathcal{L}_v$ from equation (20), using the definition of β from equation (18) and the approximations of equation (21) becomes

$$\mathcal{L}_v(\beta < 0.01) = 20 \log\left(\frac{\gamma A_p F_0}{2 f_v}\right) \qquad (22)$$

Since all of the J_n's ($\beta < 0.01$) with $n \geq 2$ are equal to zero, we have, for all $\mathcal{L}_v^n$ with $n \geq 2$ that $J_n/J_0 = 0$ and

$$\mathcal{L}_v^{n \geq 2}(\beta < 0.01) = 20 \log(0) = -\infty \text{dBc} \qquad (23)$$

That is, for small β, most of the power is in the carrier and a small amount is in the first upper and lower sideband.

Using the values of the previous example of $\gamma = 2 \times 10^{-9}$/g, $F_0 = 5$ MHz and $A_p = 5$g, the value of $\mathcal{L}_v^1$ for several vibration frequencies becomes

$$\begin{aligned} \mathcal{L}_v^1(5\text{Hz}) &= -46\text{dBc} \\ \mathcal{L}_v^1(25\text{Hz}) &= -60\text{dBc} \\ \mathcal{L}_v^1(50\text{Hz}) &= -66\text{dBc} \\ \mathcal{L}_v^1(500\text{Hz}) &= -86\text{dBc} \end{aligned}$$

It can be seen that the sideband power ratio falls off at 6 dB per octave or 20dB per decade.

Measurement of γ

To determine the value of γ for a given direction, a measurement must be made of $\mathcal{L}_v^1$ for a known vibration frequency and a known peak acceleration amplitude in the direction of interest. The acceleration sensitivity coefficient along that direction is then

$$\gamma = \left(\frac{2 f_v}{A_p F_0}\right) 10^{\mathcal{L}_v^1/20} \qquad (24)$$

$\mathcal{L}_v^1$ can be measured using the system shown schematically in Figure 5. The essential components are the shake table, with accompanying drivers for applying a pure sinusoidal vibration, and a spectrum analyzer. The local oscillator, in this example, is used to translate the frequency of the oscillator under test into the acceptance range of the spectrum analyzer. The programmable controller and plotter are added for operator convenience. The role of the frequency multiplier will be discussed in the next section. Figure 6

is a trace from the spectrum analyzer when a 5MHz oscillator is vibrated at a level of 10g and at a frequency of 100 Hz. It can be seen that $\mathcal{L}_v$ is -60dBc, which corresponds to a γ of 4.0×10^{-9}/g along the direction of acceleration.

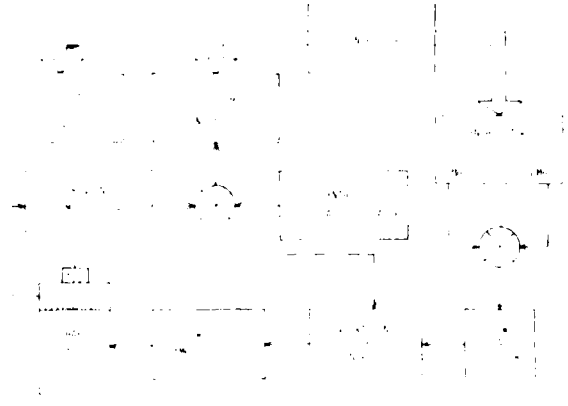


Fig. 5 - Acceleration Sensitivity Measurement System

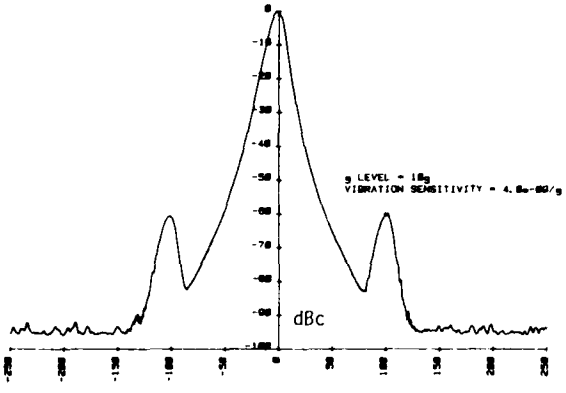


Fig. 6 - Power Spectrum of a Resonator Undergoing 100Hz, 10g Vibration

When this experiment is performed for several different frequencies, and the $\mathcal{L}$'s are plotted as a function of frequency, f_v, a curve like the one in Figure 7 is obtained. It can be seen that γ is independent of vibration frequency. It is also independent of peak accelerations below about 50g. If the value of γ is measured along 3 mutually perpendicular axes, the vector, Γ, can be constructed, which defines the acceleration sensitivity for that device.

Resonance Phenomena

The acceleration sensitivity coeffecient must be measured over a range of vibration frequencies to reveal any resonance behavior in the oscilator or resonator structure. Figure (8) is a plot of γ vs. vibration frequency for an oscillator with a mechanical resonance at 425Hz. The effective acceleration level at the resonator is increased, due to the resonance, by nearly a factor of 17 over the applied acceleration level. The resonance in this example was due to mechanical support structure of the circuitry. Similar results have also been observed due to resonances in the crystal mounting structure.

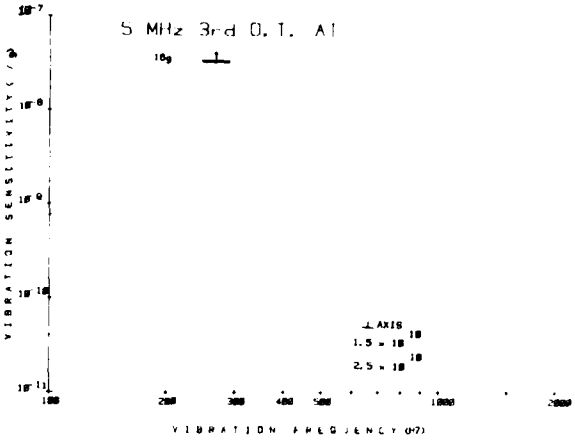


Fig. 7 - γ vs. F_v

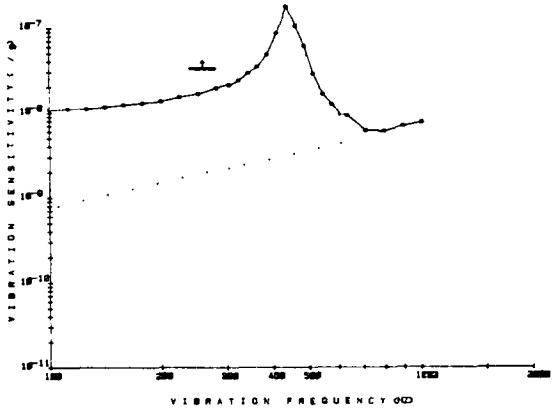


Fig. 8 - Oscillator with a Mechanical Resonance

Large Modulation Index

The discussion so far has been confined to the case where the modulation index β, given by $\Delta F/f_v$, is small. In that case several simplifying approximations were made and the sideband level was given by $\mathcal{L}_v = 20 \log (\gamma A_p F_o/2f_v)$ dBc. The question remains as to what happens when β is not small enough for the approximations to be valid. This can occur if γ, F_o or A_p is very large or if f_v is small. One such occurence commonly takes place in systems whose frequency is multiplied to high values.

Many systems have frequency sources in the megahertz range and require carrier signals in the gigahertz region. Frequency multiplication of several hundred is not uncommon. If N is the frequency multiplication factor, we can perform an analysis similar to that in equations (12) through (16) where the frequency of the carrier signal is now multiplied by N.

$$F_{carrier} = NF(t) = N(F_o + F_p \cos(2\pi f_v t)) \quad (25)$$

As in eq. (14), we have

$$V_{carrier}(t) = V_o \cos\left(2\pi NF_o t + \left(\frac{NF_p}{f_v}\right) \sin(2\pi f_v t)\right) \quad (26)$$

and, as in equation (16), we have the result

$$V_c(t) = V_o \cdot J_o(N\beta) \cos(2\pi NFot) + J_1(N\beta)\,[\cos(2\pi(NFo + f_v)t) - \cos(2\pi(NF_o - f_v)t)] \quad (27)$$

where $\beta = A_p F_o / f_v$, as before. The relative level for the first sideband now becomes

$$\mathcal{L}_v^1(fv) = 20 \log \left|\frac{J_1(N\beta)}{J_o(N\beta)}\right| \text{ dBc} \quad (28)$$

If $N\beta$ is still less than 0.01, the approximations made above are still valid and equation (28) becomes

$$\mathcal{L}_v^1(f_v) = 20 \log (N\beta/2) \quad (29)$$

or

$$\mathcal{L}_v^1(f_v) = 20 \log (\gamma A_p F_o / 2fv) + 20 \log (N) \quad (30)$$

The first term in equation (30) is identical to equation (22). The second term shows an increase in the sideband level by 20 dB for every 10X multiplication of the frequency.

In the measurement system depicted in Figure 5, the sensitivity of $\mathcal{L}_v$ to frequency multiplication is exploited through the use of a 10X frequency multiplier. The effective dynamic range of the spectrum analyzer is thereby increased by 20 dB. Figure 9 is the display from a dual trace spectrum analyzer showing the spectrum of an oscillator being vibrated at 100 Hz. The curve labelled 1X is the direct output. The curve labelled 10X is the output after frequency multiplication by ten. As is predicted by the above analysis, the sideband level increased by 20 dB.

If $N\beta$ is larger than 0.01, the approximations made earlier are invalid and the actual values of the Bessel functions must be computed. An example is provided by the GPS system. The reference source is a 5.115 MHz oscillator and the carrier frequency is 1.575 GHz, i.e., the multiplication factor is 308. For $\gamma = 2 \times 10^{-9}$/g, Fo = 5.115MHz and, for example, an A_p of 5g, the values of the resulting $\mathcal{L}$'s are listed in Table I.

TABLE I

N = 1, NFo = 5.115 MHz	N = 308, NFo = 1575 MHz
$\mathcal{L}_v^1$ (52 Hz) = -66 dBc	$\mathcal{L}_v^1$ (52 Hz) = -16 dBc
$\mathcal{L}_v^1$ (10 Hz) = -52 dBc	$\mathcal{L}_v^1$ (10 Hz) = -0.9 dbC
$\mathcal{L}_v^1$ (6.55 Hz) = -48 dBc	$\mathcal{L}_v^1$ (6.55 Hz) = + ∞
$\mathcal{L}_v^1$ (5 Hz) = -46 dBc	$\mathcal{L}_v^1$ (5 Hz) = +2.3 dBc

At 52 Hz the multiplication affects only slightly the value which would be predicted by the approximations, i.e. (20 log (308) ~ + 50 dB). At 10 Hz the deviation from 20 log (N) is more marked. A plus sign indicates that the amount of power in the first sideband is greater than that in the carrier. At 6.55 Hz, which corresponds to an $N\beta$ of 2.4, the carrier power goes to zero! The ratio of the power in the first sideband to the carrier power continues to fluctuate as $N\beta$ is increased as a consequence of the oscillatory nature of the Bessel functions.

It must be stressed that it is not the frequency multiplication which causes the redistributation of the power from the carrier to the sidebands, but rather the increase in the modulation index. For a given set of vibration conditions, the sideband levels are the same for a device with F_o = 5.115 MHz and N= 308 (e.g. crystal oscillators), as they are for a device with F_o = 1575 MHz and N = 1 (e.g. a SAW oscillator).

Vibration induced sidebands degrade the performance of Doppler radar systems. These systems rely on detecting small signals reflected from moving targets. The signals are shifted in frequency from the carrier by an amount proportional to the velocity of the target. Sidebands produced by vibration are indistinguishable from sidebands received from targets, causing possible false detections. As is seen in the next section, broadband vibration will lower the useful range of a Doppler radar by raising the background noise level at the carrier.

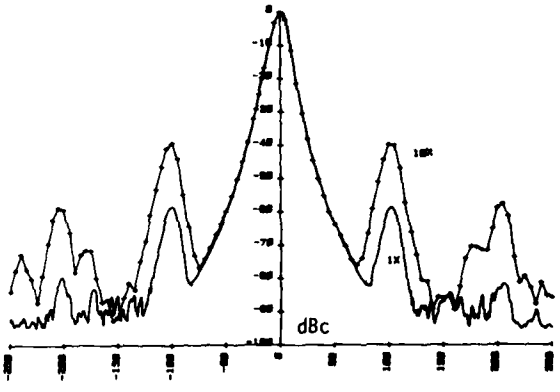


Fig. 9 - The Result of Frequency Multiplication

Random Vibration

In many situations, crystal oscillators are subject to random vibrations, i.e. the vibratory power is distributed over a range of frequencies. The worst case condition is usually described by a power spectral density plot of the vibration envelope.

The power spectral density of a vibration is the mean square value of the acceleration in a 1 Hz bandwidth.[8] The power spectral density is expressed in units of g^2/Hz. The rms value of the acceleration within a frequency band of 1 Hz around f is,

$$A^2_{rms} = \int_{f-.5Hz}^{f+.5Hz} G(f')\, df' \qquad (39)$$

where G(f) is the power spectral density of acceleration at frequency f.

The value of the effective peak acceleration at frequency f to correspond to the pure sine value is

$$A_p(f) = \sqrt{2}\ A_{rms}(f);\ (1\ Hz\ bandwidth) \qquad (40)$$

The ratio of the power in one of the 1 Hz wide sidebands to the carrier power, using equation (22), and assuming small vibrations, is

$$\mathcal{L}^1_v(f) = 20 \log(\tfrac{1}{2}\sqrt{2}\ A_{rms}(f) \cdot F_o/f) \quad (1\ Hz\ Bandwidth) \qquad (41)$$

Again, if $\Gamma = 2 \times 10^{-9}/g$, $F_o = 5.115MHz$, and G(f) is as in Figure 10, using equation (41) the power ratio becomes

$$\mathcal{L}^1_v(f) = 20 \log (7.23 \times 10^{-3}\ A_{rms}(f)/f) \qquad (42)$$

and is plotted in Figure 11.

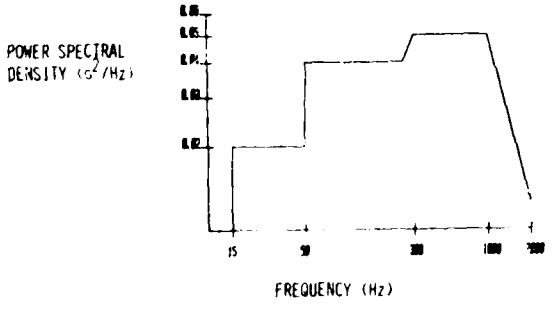


Fig. 10 - An Example of a Random Vibration Envelope

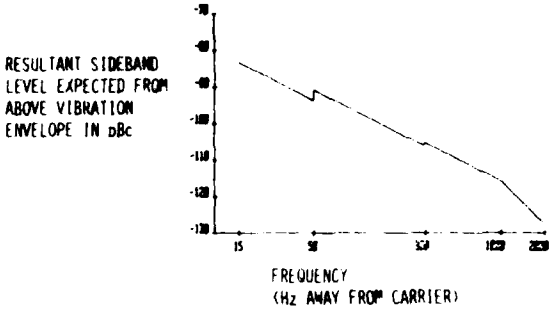


Fig. 11 - Noise Expected From Vibration Envelope Given in Figure 10

Clock Error

Acceleration can also lead to clock errors. Because typical vibrations have an average acceleration of zero, i.e. as much positive acceleration as negative acceleration, timing errors would be appreciable only as a consequence of an unidirectional acceleration. Examples would be a sharp turn by a high performance aircraft or a residual half-cycle from a sinusoidal vibration.

Let us assume that a constant acceleration A takes place during the time interval between t_1, and $t_1 + t_a$. The frequency is, therefore, F_o at all times except during the acceleration interval when it is $F_o + \Gamma F_o A$. Since phase change is a measure of the time interval, and phase is the integral of frequency, the phase change is

$$\phi(t) = 2\pi \int_0^t F(t')\, dt' \qquad (43)$$

As stated above,

$$F(t) = Fo\ ;\ t < t_1 \text{ and } t > t_1 + t_a \qquad (44)$$

and

$$F(t) = Fo + \Gamma F_o A\ ;\quad t_1 < t < t_1 + t_a \qquad (45)$$

Using eqs. (43) and (44) in (42), the phase change is

$$\phi(t) = 2\pi F_o t + 2\pi \Gamma F_o A t_a \qquad (46)$$

The apparent time interval, t', is related to the actual time interval, t, by

$$2\pi F_o t' = 2\pi F_o t + 2\pi F_o \Gamma A t_a. \qquad (47)$$

Simplifying eq. (47), we have

$$t' = t + \Gamma A t_a. \qquad (48)$$

The time error Δt, is t' - t or

$$\Delta t = \Gamma A t_a \qquad (49)$$

Using the previous example of $\Gamma = 2 \times 10^{-9}/g$ and an acceleration of 5g, a typical value for Δt would be

$$t \approx 10 \text{ nsec/second of acceleration} \qquad (50)$$

This error can be significant, since modern systems, e.g. in high performance aircraft, can require nanosecond synchronization.

Acceleration Sensitivity Reduction

There are two classes of methods for reducing the coefficient of acceleration sensitivity: (1) passive and (2) active. In passive methods, no attempt is made to sense the vibration or dynamically change the output frequency. In active methods, an acceleration sensor and a feedback network is used to alter the oscillator frequency and thereby compensate for the acceleration-induced frequency shifts.

Table II lists several of the most actively investigated methods for reducing the coefficient of acceleration sensitivity, separated into the two classes.

TABLE II

Passive Methods	Active Methods
SC-cut Resonators	Varactor Feedback
Improved Mounting	Direct-to-Crystal
Resonator Pairs	
Vibration Isolation	

The most actively investigated method today for reducing the acceleration sensitivity is the use of the (doubly rotated) SC-cut of quartz.[9,10] The SC-cut is less sensitive to many types of stresses than the more commonly used AT-cut. Another method which is receiving attention is improved mounting.[13,14] Some success has been achieved, most notably with the BVA design.[13] A third passive method which can be used together with other methods, is to use two resonators, with their acceleration vectors, Γ, aligned antiparallel.[11,12] The resultant vector is then the difference between the two magnitudes. When two devices with identical Γ vector magnitudes are aligned such that the vectors are antiparallel, the resultant acceleration sensitivity is zero.

The least desirable passive method is mechanical vibration isolation. Mechanical isolation adds to the size and weight and, as is seen in Figure 12, actually amplifies the acceleration at some low frequencies. Vibration isolation can be used, with or without the other methods, in cases where the low frequency response of the isolation system is inconsequential.

Both active methods consist of an accelerometer situated along the acceleration sensitivity vector to sense the magnitude of the scalar product of Γ and A and an operational amplifier to adjust the accelerometer output level and feed the accelerometer signal back to modulate the output frequency. The difference is in the method of feedback. In the varactor method, an electronically variable capacitor is placed in series with the resonator. The accelerometer derived feedback signal is phase shifted, amplified and applied to the varactor in such a way as to vary the reactance of the oscillating loop to cancel the vibration induced frequency changes. In the direct-to-crystal method,[16] the processed accelerometer signal is applied directly to the resonator electrodes. This method exploits the voltage sensitivity of doubly rotated resonators.[17,18] A more detailed discussion of the direct-to-crystal method and a comparison of it with the varactor method is given elsewhere in these proceedings.[16]

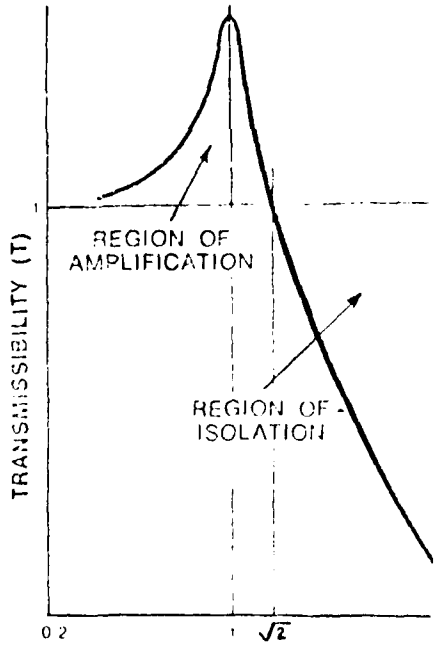


Fig. 12 - Transmissibility vs. Vibration Frequency for Mechanical Isolation

Conclusion

We have shown how acceleration effects frequency standards and clocks. Vibration effects are significant even for atomic standards, because the servo-loop time constants for these devices are usually on the order of a few seconds. Therefore, for vibration frequencies above a few Hz, the servo-loop cannot compensate and the vibration sensitivity of the atomic standard will be the same as that of the crystal oscillator in the standard. Measurements of the vibration sensitivity of atomic standards must therefore be performed dynamically. For all frequency standards, we have seen that it is sufficient to measure the acceleration sensitivities along three mutually perpendicular directions. From the three quantities, the acceleration sensitivity vector, Γ, which characterizes the device, can be calculated.

The effect of internal mechanical resonances on oscillator performance has also been discussed. These resonances must be considered in system design. Frequency standards must be

mounted in systems in such a way as to preclude degradation of laboratory performance in the field due to external resonances.

It is hoped that paper will promote a standardization of data reporting methods and specifications. It is recommended that the acceleration sensitivity of frequency standards be specified by the magnitude and direction of Γ, both while undergoing a constant acceleration, and undergoing vibration at several frequencies. The location and Q of any mechanical resonance must also be reported.

References

1. W.L. Smith and W.J. Spencer, "Precision Crystal Frequency Standards," Proceedings, 15th Annual Symposium on Frequency Control, US Army Electronics Command, Fort Monmouth, NJ, pp. 139-155, (1961). National Technical Information Service Accession Nr. AD265455.

2. J.M. Przyjemski, "Improvement in System Performance Using a Crystal Oscillator Compensated for Acceleration Sensitivity," Proceedings 32nd Annual Symposium on Frequency Control, Fort Monmouth, NJ, pp. 426-431, (1978). Copies available from Electronics Industries Association, 2001 Eye Street, N.W. Washington, DC 20006.

3. M. Valdois, J. Besson and J.J. Gagnepain, "Influence of Environment Conditions on a Quartz Resonator," Proceedings 28th Annual Symposium on Frequency Control, US Army Electronics Command, Fort Monmouth, NJ, pp. 19-32, National Technical Information Services Accession Nr. AD AD11113.

4. A more detailed discussion of this topic may be found in R.L. Filler, "The Effects of Vibration on Quartz Crystal Resonators," R&D Technical Report DELET-TR-80-10, US Army Electronics Research and Development Command, Fort Monmouth, NJ 07703, NTIS Acc. Nr. AD A085642.

5. M. Abramowitz, and I.A. Stegun, "National Bureau of Standards, Applied Mathematics Series, No. 55," Government Printing Office, Washington, DC, 1964. Revised by Dover, New York, 1965.

6. A. Ballato and M. Nakazawa, these Proceedings.

7. J. Moses, "Navstar Global Positioning System Oscillator Requirements for the GPS Manpack," Proceedings, 30th Annual Symposium on Frequency Control, US Army Electronics Command, Fort Monmouth, NJ, pp. 390-400 (1976). National Technical Information Service Accession Nr. AD A046089.

8. A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw-Hill, NY, 1965.

9. A. Ballato, "Doubly Rotated Thickness Mode Plate Vibrators," in Physical Acoustics: Principles and Methods, (W.P. Mason and R.N. Thurston, eds.), Vol. 13, Chapter 5, Academic Press, New York, (1977), pp. 115-181.

10. A. Warner, B. Goldfrank, M. Meirs, and M. Rosenfeld, "Low 'g' Sensitivity Crystal Units and Their Testing," Proceedings, 33rd Annual Symposium on Frequency Control, US Army Electronics Command, Fort Monmouth, NJ, pp. 306-310, (1979). Copies available from the Electronics Industries Association, 2001 Eye Street, N.W., Washington, DC 20006.

11. J.J. Gagnepain, and F.J. Walls, "Quartz Crystal Oscillators with Low Acceleration Sensitivity," Technical Report NBSIR 77-855, National Bureau of Standards Washington, DC 20234, March 1977, 15 pages.

12. A.D. Ballato, "Resonators Compensated for Acceleration Fields," Proceedings, 33rd Annual Symposium on Frequency Control, US Army Electronics Command, Fort Monmouth, NJ, (1979), pp. 327-336. Copies available from Electronics Industries Association, 2001 Eye Street, N.W. Washington, DC 20006. (This reference has an extensive bibliography.)

13. R. Besson, J.J. Gagnepain, D. Janiaud, and M. Valdois, "Design of a Bulk Wave Quartz Resonator Insensitive to Acceleration," Proceedings 33rd Annual Symposium on Frequency Control, US Army Electronics Command, Fort Monmouth, NJ, (1979), pp. 337 - 345. Copies available from Electronics Industries Association, 2001 Eye Street, N.W. Washington, DC 20006.

14. T.J. Lukaszek and A.D. Ballato, "Resonators for Severe Environment," Proceedings of the 33rd Annual Symposium on Frequency Control, (1979), US Army Electronics Command, Fort Monmouth, NJ pp. 311-321. Copies available from Electronics Industries Association, 2001 Eye Street, N.W. Washington, DC 20006.

15. D.A. Emmons, "Acceleration Sensitivity Compensation in High Performance Crystal Oscillators," Proceedings 10th PTTI, NASA Tech Memorandum. 80250, Greenbelt, MD, Nov 1978, pp. 55-82.

16. V.J. Rosati and R.L. Filler, these Proceedings.

17. J.A. Kusters, "The Effect of Static Electric Fields on the Elastic Constants of Quartz," Proc. 24th Annual Symposium on Frequency Control, US Army Electronics Command, Fort Monmouth, NJ, 1970, pp. 46-54. National Technical Information Service Accession Nr. AD 746 210.

18. C.K. Hruska, "Zero Polarizing Effect with Doubly Rotated Quartz Plates Vibrating in Thickness," IEEE Trans. Sonics and Ultrasonics, Vol. SU-27, March 1980, pp. 87-89.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

VACUUM PROCESSING SYSTEM FOR QUARTZ CRYSTAL RESONATORS*

J. M. Frank
General Electric Company
Neutron Devices Department
St. Petersburg, Florida 33733

Summary

An ultra-high vacuum system has been developed to process quartz crystal resonators through electroding and sealing of the enclosures without air exposure. The functions to be performed in the system are ultraviolet/ozone cleaning, vacuum bake, deposition of gold electrodes to a specified resonant frequency, and thermocompression sealing of the enclosure. Unplated resonators, mounted in flatpack ceramic frames, are loaded into the system from a controlled argon atmosphere and the completed units are removed from the system through the same argon atmosphere. During normal operation, the inside of the vacuum chamber will not be exposed to air.

The major criteria in the design of the system were established to permit laboratory fabrication of small batches of resonators that have been subjected to a minimum of processing variables, have the lowest attainable internal pressure, and are sealed in the cleanest environment that can be provided.

Key Words (for information retrieval) Ultrahigh vacuum system, Non-air exposed process, Ceramic flatpack crystal unit, Ultraviolet/ozone cleaning, Electroding, Thermocompression sealing

Introduction

This paper describes a vacuum system which has been developed to process the ceramic flatpack crystal unit. Figure 1 shows a typical flatpack resonator that has been described previously[1-4].

The functions to be performed in the system are ultraviolet/ozone cleaning, vacuum baking, deposition of gold electrodes to a specified resonant frequency, and thermocompression sealing of the enclosure.

The design goals included a mechanism capable of handling the resonator components during electroding and sealing under vacuum with a minimum of friction between the moving parts. This eliminates the need for dry lubricants and the generation of metal particles that could contaminate the crystal. Another goal was to provide an entrance/exit chamber that would allow the vacuum chamber to be opened without exposing the inside of the chamber and internal fixturing to the atmosphere.

System fixtures include frame carriers, cover carriers and mask assemblies that have provisions for alternately processing either three 5-MHz units or five 20-MHz units during a run†.

Discussion

In order to achieve the design goals, a single tubular vacuum chamber with an adjoining glovebox was developed. Figure 2 is an external view of the vacuum system and Figure 3 shows the special glovebox attached to the vacuum chamber. The chamber is 8 inches in diameter and 20 inches long. Rough pumping of the system is provided by two molecular sieve pumps. A high vacuum is produced with a 300-liter/second ion pump or a 6-inch cryopump. The high vacuum pumps may be isolated from the system by two 6-inch all-metal gate valves. There are no organic materials located in the high vacuum part of the system. In addition to the normal vacuum measuring equipment, an Inficon 200 mass analyzer is used to monitor processes. The glovebox has a clamshell opening in one end whereby it may be joined to the vacuum chamber during loading and unloading of resonators and for servicing of the evaporation sources.

*The system has been designed and constructed at the General Electric Neutron Devices Department under ERADCOM-sponsored Contract No. 78-09408. The Neutron Devices Department is a Department of Energy contractor operating under Contract DE-AC04-76DP00656.

†References to 5-MHz units and 20-MHz units in this paper are to identify the large ceramic flatpack configuration (HC-6 equivalent) and the small ceramic flatpack configuration (HC-18 equivalent).

Figure 4 is a cross-sectional schematic of the chamber and includes the fixturing that provides for frictionless movement of the parts within the chamber. The key fixture elements are the frame carrier, the cover carrier and the sealing ram assembly. The frame carrer is essentially a beam with the appropriate number of cavities for the resonator frames; it is attached to a 10-inch linear motion feedthrough. The cover carrier is a saddle-like device that fits over the frame carrier to align each pair of covers with each frame assembly prior to sealing. The frame and cover carriers are equipped with tungsten wire springs to secure the resonator parts during processing. The two halves of the cover carrier are held together with screws and must be separated for the insertion of the covers and for removal of the resonator assemblies. The sealing mechanism consists of a pneumatic diaphragm operated ram that is bellows-sealed with the ram shaft bearing located external to the vacuum. The ram housing includes bakeout heaters located near the area in which the sealing takes place. Figure 5 shows how the key fixture elements are brought together for the sealing operation.

A cover carrier lifter is located in the sealing area and is used to lift and support the cover carrier while the crystals are being electroded. After the frames are retracted into the electroding position, a linear motion vacuum feedthrough located in the side of the chamber is advanced to make electrical contact to each of the three crystals being electroded. The electronic equipment used for frequency measurements during tuning include a gain/phase measuring system that employs high temperature thick film ceramic pi networks (Figure 6); those networks are located close to the crystals being processed. Three pairs of gold charged tungsten filaments are used for the electrode deposition sources.

After the electroding operation is completed the frames are advanced to horizontally realign the frames with the covers. The cover lifter is advanced to lower the covers and establish vertical alignment. The 10-inch linear feedthrough is then adjusted to a predetermined value that allows thermocompression sealing of the first assembly. The procedure is repeated for the remaining assemblies. The sealing ram is equipped with a swivel type plunger to allow for slight variations in parallelism between resonator components being assembled.

The chamber may be backfilled with oxygen for ultraviolet/ozone cleaning[5,6]. Three mercury vapor lamps located inside the chamber are designed to emit ultraviolet light to generate ozone and a fourth lamp emits ultraviolet light of a longer wavelenth to decompose the ozone.

The access end of the chamber is equipped with a 10-inch diameter vacuum flange upon which a platform is mounted. Figure 7 shows an oblique view of the platform assembly with the frame carrier extended to expose the three cavities for the frames.

Figure 8 is an end view of the platform assembly and shows the basic equipment used for electroding. This includes three mask assemblies, three movable pi network contact assemblies for measuring frequency, six gold evaporation sources, and bakeout heaters. The access flange also contains five ports for electrical feedthroughs, a shutter actuator and the 10-inch linear motion vacuum feedthrough.

The 10-inch linear motion feedthrough supports the frame carrier that moves the frame/blank assemblies betwen the electroding zone and the sealing zone. The feedthrough has a revolution counter that provides the capability of controlling linear positions during electroding or sealing to within 0.002 inch. The bearings for the feedthrough are located external to the vacuum.

Figure 9 shows the frame carrier and cover carrier assembly with parts removed from two of the three positions to expose one cover and one electroded frame assembly. A spring loaded retainer at the top of each cavity holds the cover in place.

Figure 10 shows the frame and cover carriers with one side of the cover carrier removed. This exposes the spring loaded retainers located at the top of each frame cavity for holding the frames in place. The cover carrier rests on top of the frame carrier to provide the vertical alignment between the resonator components and horizontal alignment is achieved by using a tab and slot at the upper right end of the cover carrier.

The system fixtures include two sets of frame carriers, cover carriers and mask assemblies that have provisions for alternately processing either three 5-MHz units or five 20-MHz units during a run.

Additional studies are planned for the ultraviolet/ozone cleaning process using the Inficon 200 mass analyzer. These will provide information to optimize the equipment and processes in this system.

CONCLUSIONS

1. The system has been operated for approximately twenty processing cycles without any mechanical failures.

2. A plating-to-frequency resolution of 1 ppm is regularly achieved on 5-MHz fundamental resonators.

[illegible] The [illegible] Mbr resonator is 1 part in [illegible]. This value is expected to improve with UV/Ozone cleaning experiments.

[illegible] The lowest pressure obtained to date, but no electroplating at 5.0 [illegible] or sealing at 300 °C, is [illegible] torr.

[illegible] The lowest system pressure to date, after bakeout and ultraviolet exposure, is [illegible] torr at room temperature.

[illegible] Small shifts in frequency of approximately [illegible] have been observed subsequent to ultraviolet cleaning of electroded crystals which [illegible] the vacuum system.

References

[illegible] J. Walton and J. R. Vig, "Crystal Resonator [illegible] Migration," U.S. Patent Nr. [illegible], 388, Jan 6, 1976.

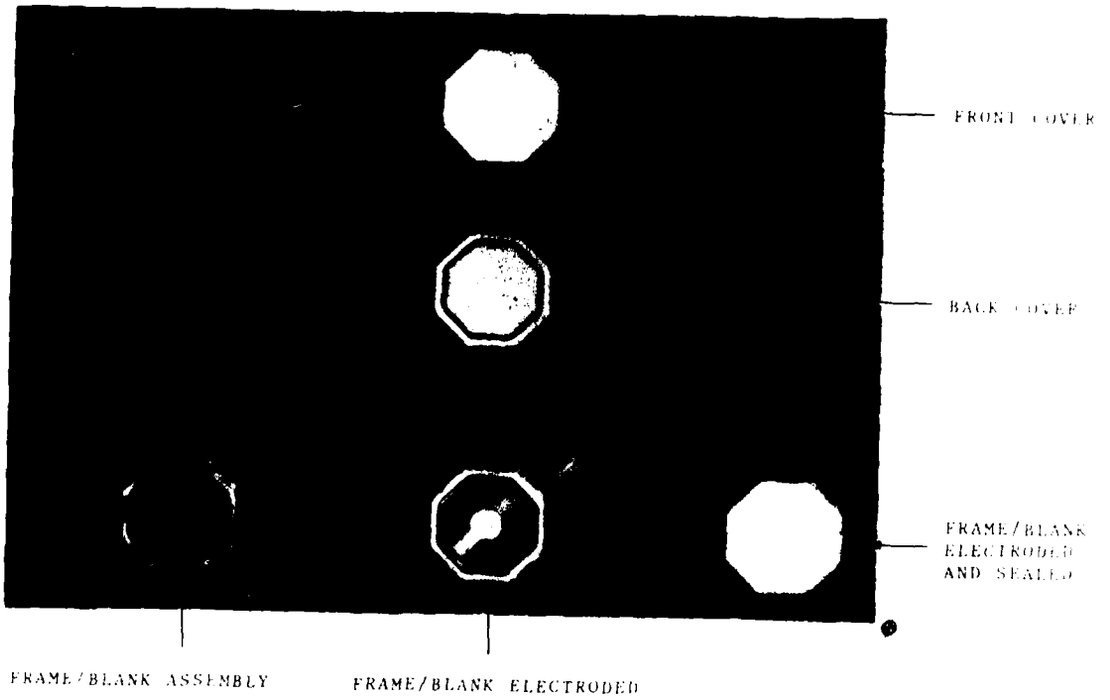
[illegible] J. R. Vig and E. Hafner, "Packaging Precision Crystal Resonators," Technical Report ECOM-4154, U.S. Army Electronics Command, Fort Monmouth, N.J., July 1973.

E. Willcox, J. Knox, E. Hafner and J. R. Vig, "A New Ceramic Flatpack for Quartz Resonators," Proc. AFCS 28, June 1975, pp [illegible].

R. L. Peters, "Ceramic Flatpack Enclosures for Precision Quartz Crystal Units," Proc. AFCS 30, June 1976, pp 22-231.

J. R. Vig, J. W. LeBus, R. L. Filler, "Further Results on UV Cleaning and Ni Electrobonding," Proceedings, 29th Annual Symposium on Frequency Control, copies available from Electronic Industries Association, 2001 Eye Street, NW, Washington, DC 20006, 1975, pp 220-229.

J. R. Vig, J. W. LeBus, "UV/Ozone Cleaning of Surfaces," IEEE Trans on Parts, Hybrids, and Packaging, Volume PHP-12, 4 (1976), pp [illegible].


FRONT COVER
BACK COVER
FRAME/BLANK
ELECTRODED
AND SEALED
FRAME/BLANK ASSEMBLY
FRAME/BLANK ELECTRODED

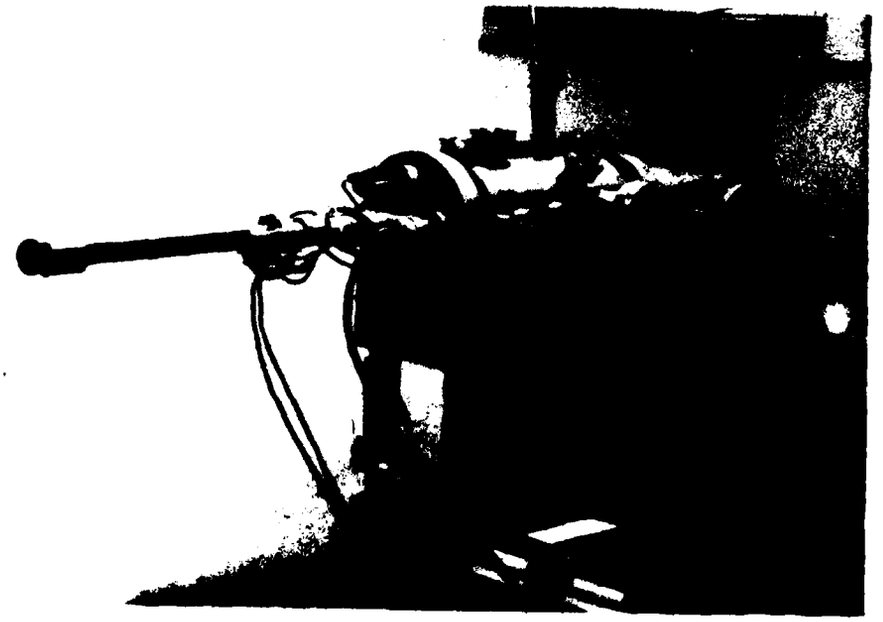




Figure 4. Vacuum System Glovebox

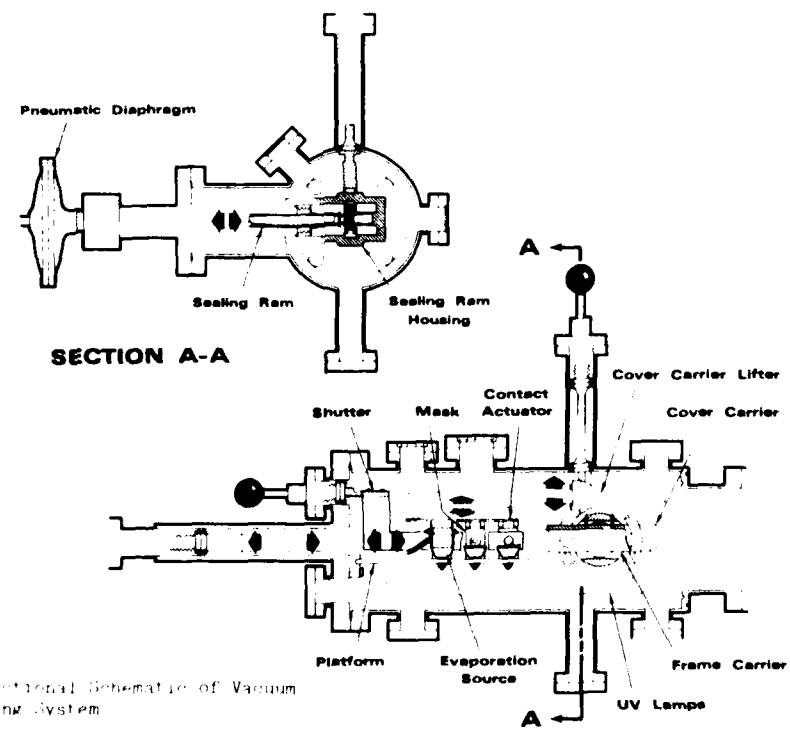


Figure 5. Cross-sectional Schematic of Vacuum Processing System

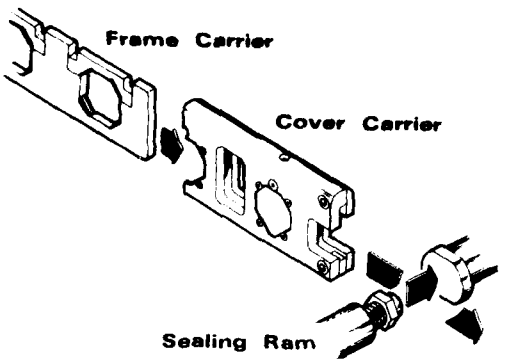


Figure [illegible]. Fixture Elements for Sealing



Figure [illegible]. Petri-Net Models





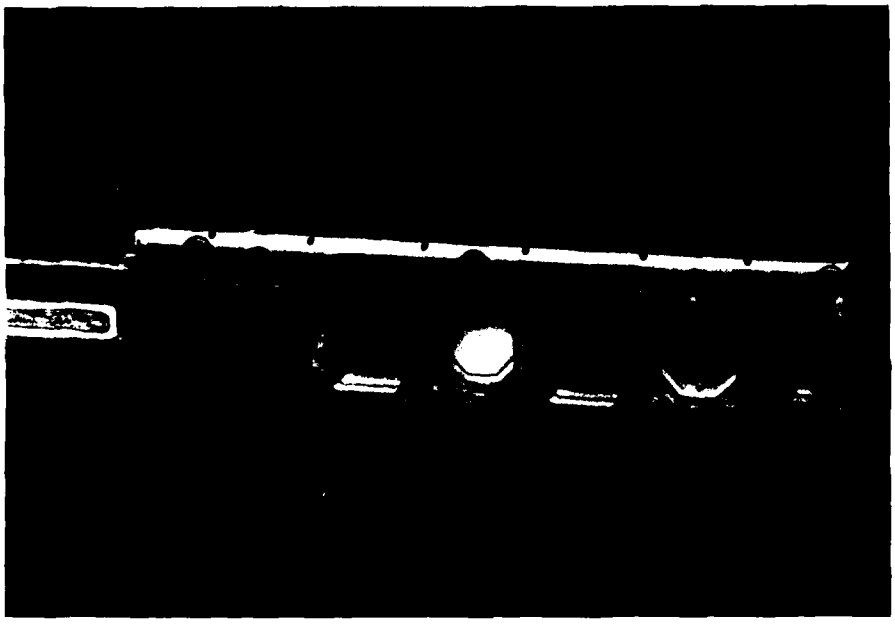


Figure 9 Frame Carrier and Cover Carrier Assembly

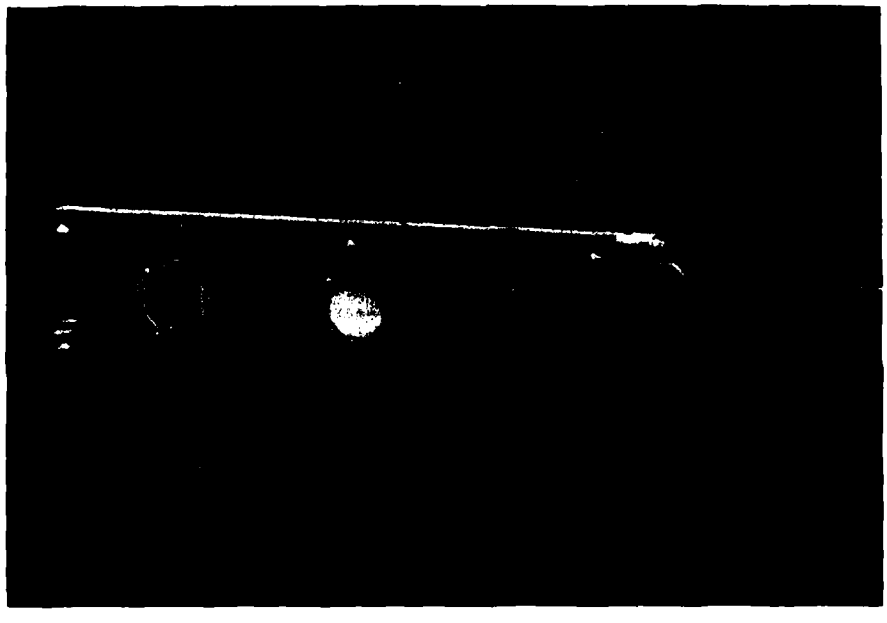


Figure 10 Frame and Cover Carriers with One Side of Cover Carrier Removed

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

METALLIZATION OF QUARTZ OSCILLATORS

Arthur T. Lowe

Process Technology Laboratory, SRDL
Motorola, SC
5005 E. McDowell Road
Phoenix, Arizona 85008

Summary

Utilizing a miniature, rectangular, 4MHz quartz resonator, various oscillator properties such as frequency, R_s, and Q were investigated as a function of base metal thickness, metal type (gold and non-gold), adhesion layer type and deposition technique.

For this crystal, the linear proportionally between the thickness of CrAu and the lowering of the frequency (mass loading) deviated significantly above 4000 Å of metal. Also, for this CrAu metallization system an R_s minimum was found near 3800 Å of metal. In addition to gold, oscillator properties and temperature characteristics were examined for copper and aluminum metallization. Average R_s values for the crystals using copper and gold were 50 to 70 ohms and for aluminum 150 to 200 ohms. Resistance measurements and Auger analysis of these metals on quartz indicated a thermal stability to 300°C. Additionally, the concept of batch processing to frequency (no 'plate to frequency') is evaluated relative to these studies.

Introduction

The basic process of electrode formation[1] and frequency adjust[2] by metallization of quartz is an integral part of the design and fabrication of most oscillators. For optimization of oscillator properties and processing variables, a thorough understanding of the interactive film properties is an absolute requirement.

To evaluate metallization parameters, it is necessary to distinguish them from the numerous other design and fabrication parameters that effect oscillator properties such as R_s. In addition to sensitive variables such as metal pattern type[3], bonding technique,[4] and package design,[5,6] there are processing parameters that directly effect thin film formation such as surface morphology[7] and surface cleanliness[8] that must be well controlled and characterized. To accomplish this, the frequency, R_s and Q must be monitored after each processing step and correlated to physical measurements made on the films themselves.

It is the objective of this study to empirically examine the effects of metal deposition on quartz oscillator properties. A miniature 4 MHz oscillator was utilized for studying gold metallization and directly comparing it with other metal electrode alternatives such as copper and aluminum. Much of the data obtained, however, is extendable to metallization of quartz in general.

In addition to considering metal type the adhesion layer type and the deposition technique are evaluated. If the uniformity of deposition is good and the frequency specification range is fairly wide, the process of batch coating to frequency (no individual 'plate to frequency') can be established. Considered in this study are the limitations of this batch processing concept for various metal types.

Fabrication and Testing

Crystal Design The crystal utilized as a process development vehicle was a vertically mounted, AT cut, rectangular crystal similar in design to that described by Zumsteg.[9] The crystal is mounted on a header as shown in Figure 1. For ease of processing, the metal was not patterned; the metal fully covered both sides of the crystal. The lead was attached to the electrode with a conductive epoxy cured near 100°C and the crystal was additionally supported at the base with a nonconductive epoxy. The crystal vibrated in a thickness shear mode coupled with a third overtone flexure mode at a frequency of 4.194 MHz.

Processing In this study, oscillator properties were measured after each processing step. The initial blank frequency was first measured by driving the crystal across an air gap with equipment made by Conner Electronics. The blank was then etched with ammonimum biflouride as a means of improving surface quality[10] and the frequency was measured again. The metal films were then deposited by either resistive heating evaporation (a Varian 3120 system) or sputtering (a Sloan sputter gun). After metallization the crystals were lightly clamped into a fixture (for metal contact) and the frequency and R_s were measured with a Saunders 150B crystal monitor. Additionally, the fixture was pumped down to 100 millitorr in order to simulate an evacuated package environment. After mounting the crystal to the header, the frequency and R_s are again measured under vacuum. An average lot size represents the characterization of 40 to 60 oscillators.

Gold Metallization

Oscillator properties for a set of quartz

devices with Cr-Au electrodes 4335 Å thick (600 Å Cr + 3735 Å Au) are shown in Table I. Each one of the values represents an average of 50 oscillators with a standard deviation of 0.002 MHz for frequency and 10 ohms for R_s. As shown, the frequency is raised by 28 KHz from the thinning of the crystal by the etch and the frequency is lowered by 50 KHz from the mass loading of the metal. With the typical R_s requirements of less than 100 ohms for a miniature crystal of this type, 50 ohms is well within specifications. When connecting the leads to the crystal with epoxy, there is a dampening effect which raises the R_s from 10 to 30%. Again, the metallized crystal measurements were made at a pressure of less than 100 millitorr, and for this particular crystal the R_s lowers by a factor of 3 or 4 when going from one atmosphere to 100 millitorr pressure.

OSCILLATOR PROCESSING WITH 4335Å Cr-Au METALLIZATION

	f(MHz)	R_s(OHMS)	C_1(pf)	Q
BLANK	4.220			
BLANK-AFTER ETCH	4.248			
*METALLIZED (UNMOUNTED)	4.197	52.9	0.0033	216,450
*MOUNTED	4.195	59.2	0.0036	178,063

*TESTED UNDER VACUUM

Table I

The lot shown in Table I was evaporated gold; however, a run with a similar lot size using sputtered gold gave an R_s value within the standard deviation. There is nothing inherent in improving R_s with sputtered films; however sputtering is the preferred technique for electrode formation due to better adherence properties and better run to run reproducibility.[11,12] The more common intermediate layers used for better adherence of gold to quartz are chrome[13] and titanium-palladium.[14] In this study it was found that a thin (50 Å) sputtered layer of TiW also gave excellent adherence properties.

Mass Loading For a crystal that vibrates in the thickness shear mode there is a classical relationship by Sauerbrey[15] for change in frequency with coating of the crystal. This shift is proportional to the ratio of the thickness of the metal to the thickness of the quartz times the ratio of the density of the metal to the density of the quartz:

$$\Delta f = \frac{T_{fl}}{Tq} \cdot \frac{d_{fl}}{dq} \cdot fq$$

Δf = frequency shift

fq = frequency of quartz with no metal

T_{fl} = thickness of metal film

Tq = thickness of quartz

d_{fl} = density of metal

dq = density of quartz

Using this equation for Cr-Au the frequency shifts 10 KHz per 1000 Å of film deposited, as shown in Fig. 2 (marked theoretical). Sauerbrey's equation has been refined[16] to include the acoustic impedance, Z, of the metal and correct for nonlinearity found when depositing thicker films. Figure 2 also shows empirical data when starting at one blank frequency and metallizing to different frequencies by depositing different thicknesses of metal. Again each one of these points represents 50 to 60 oscillators with a standard deviation of 10 KHz. The simple linearity appears to hold fairly well below 4000 Å of film. There is a complication in addition to acoustic impedance of the metal, that is, different modes of vibration (other than thickness shear) will have different sensitivities to mass loading. With a coupled flexure mode for this particular cyrstal there could be nonlinearities in frequency shift associated with this mode of vibration.

A second set of experiments is shown in Figure 3 in which the blank frequencies varied in such a way that for each increment of metal thickness, a frequency of 4.194 ± .003 MHz was obtained. Again the deviation from linearity above 4000 Å was observed.

R_s Minimum For each of the oscillators that was metallized to 4.194 MHz, an R_s and Q was measured. A least squares fit of metal thickness vs R_s gives the classic R_s minimum shown in Figure 4. With the many factors that effect R_s other than metallization, the minimum shifts easily from lower to higher values of R_s; however, for this particular crystal the thickness minimum was consistently between 3500 Å and 4000 Å of CrAu.

As the films increase in thickness, the R_s increases due to the mechanical dampening from the weight of the metal becoming a dominant factor. As the film becomes thinner, the theory[17] is that the resistance of the film contributes significantly to raising R_s. Even with a film thickness of 300 Å; however, the resistance is less than an ohm and will not contribute significantly to the R_s. A discontinuous film would have significant resistance, yet the films in this study were continuous down to 1000 Å. Additionally, nucleation and growth studies of gold[18] have shown that continuous films are grown after only a few hundred angstroms. This depends on substrate, temperature, etc., but a continuous film is normally seen well below 500 Å. Further studies of the microstructure of the films, such as grain size and grain directionality, surface morphology and film uniformity are necessary for evaluating the localized electrode variations which cause a rise in R_s.

Non-gold Metallization

In considering alternatives to gold, the metals which have both the lowest resistivity and the lowest modules of elasticity are copper, aluminum and silver. These properties are shown in Table II. The conductivities of these metals are fairly similar, however, the densities of these materials vary widely resulting in advantages and disadvantages for each. For this particular study, the less expensive metals, copper and alum-

inum were characterized for direct comparison with gold.

ELECTRODE PROPERTIES

METAL	DENSITY (g cm^3)	RESISTIVITY (MICROHM-CM)	MODULUS OF ELASTICITY (PSI x 10^6)
GOLD	19.32	2.35	11.6
COPPER	8.92	1.67	16
ALUMINUM	2.70	2.65	9
SILVER	10.49	1.59	11
TITANIUM	4.50	42	16.8
CHROMIUM	7.20	12.9	36

Copper Three sets of oscillators with copper electrodes are shown in Table III. Again, each one of these values represents an average of 50 to 60 measurements. The copper was sputter deposited; using a TiW adherence layer. With a deposition of 4400 Å of copper the frequency was linearly proportional to the density relative to gold (half the density, half the shift). In other words, Sauerbrey's relation was applicable at this thickness.

SPUTTERED TiW-Cu COATINGS — 4400Å

	f(MHz)	R_s(OHMS)	C_1(pf)	Q
1)BLANKS	4.293			
UNMOUNTED	4.265	48.5	0.0033	235,000
MOUNTED	4.264	65.0	0.0033	174,000
2) UNMOUNTED	4.191	63	0.0029	215,630

SPUTTERED Cu — 10,000Å

3) UNMOUNTED	4.221	108	0.0029	120,075

One lot fabricated on frequency (4.194 MHz) and one slightly off frequency, both gave good R_S values. As can be seen, the R_S went up significantly with mounting. A deposition made with no intermediate layer resulted in poor adherence (peeling, etc.) and the average R_S was higher. An additional contribution to the high R_S for this lot was the thicker film contributed to mechanical loading - the film was not in the R_S minimum 'well' as described with gold.

The oscillator properties of copper were found to be equivalent to gold in many respects. The exception to this being coppers stability in air. Copper will continuously absorb water[19]; if tight frequency control is necessary copper is difficult to process. The copper coated oscillators in this study were vacuum cured, stored in vacuum and built into evacuated packages; with these processing precautions the devices were quite stable.

Aluminum The resonators with sputtered aluminum coatings consistently gave high R_S values, as shown in Table IV. Shown are three lots at different width to thickness ratios. The best R_S values obtained were at a w/t of 2.9. An oxide cap forms on aluminum which increases the contact resistance from the lead to the film by a few ohms. This is not, however, a significant addition to the R_S value. A possible explanation, is that a stress point is created in the bonding area where punch through occurs in the oxide, resulting in localized differences in motional resistance. The R_S minimum well does broaden with aluminum, yet thickness loading is a partial contributor to the high R_S values.

SPUTTERED ALUMINUM COATINGS — 9800Å

	f(MHz)	R_s(OHMS)	C_1(pf)	Q
1) BLANK	4.224			
UNMOUNTED	4.208	138.7	0.0031	88,500
MOUNTED	4.208	183.6	0.0031	66,030
2) BLANK	4.315			
UNMOUNTED	4.300	183.7	0.0036	55,970
MOUNTED	4.300	237.7	0.0037	42,010
3) BLANK	4.215			
UNMOUNTED	4.199	204.8	0.0052	35,540
MOUNTED	4.199	322	0.0051	23,070

Aluminum has many desirable features including being inexpensive, having a good acoustic impedance match to quartz, and having the option of being anodized for final frequency adjust.[20] With the proper bonding techniques aluminum is quite useful as an electrode material and one such technique is ultrasonic bonding which breaks through the oxide and leaves a very low resistance bond. Packages are being designed utilizing this bonding concept.

The R_S values obtained with these three metals for a 4.194 MHz ± .003 crystal is shown in Table V. Again these values represent 50 to 60 oscillators and the R_S standard deviation is 10 to 20 ohms. Gold and copper are nearly equivalent with aluminum being less desirable with this bonding type.

PACKAGED OSCILLATORS — 4.194 MHz

METAL	R_s (OHMS)
EVAPORATED Cr-Au (3000Å)	60
SPUTTERED TiW-Au (3000Å)	73
SPUTTERED TiW-Cu (4700Å)	65
SPUTTERED AL (9800Å)	183

Thermal Stability The temperature characteristics in air of these metals on quartz are shown in Figure 5. The sheet resistance of the copper goes up dramatically above 300°C and it goes up

significantly for gold also above 350°C. Both aluminum and silver are stable to 500°C. Explanation of these results can be seen in the analysis that follows. Figure 6 is a depth profile of an aluminum film obtained with Auguer Electron Spectrocopy (combined with incremental sputtering into the film). To the left is the surface of the film where an oxide cap on the film is evident. This keeps aluminum films stable in air through 500°C. Silver is stable relative to oxygen and showed no interesting reactions or diffusion up to 500°C.

In Figure 7 it can be seen that at 300°C the copper begins to take in oxygen and at 350°C it is completely converted into the copper oxide which results in high sheet resistance. The analysis of the Cr-Au sample as shown in Figure 8 indicates the diffusion of chrome through gold at 350°C and the subsequent formation of CrO on the surface. This is the reason for the gradual increase in sheet resistance of the film and can result in bonding reliability problems. Studies of the Cr-Au system have shown chrome to diffuse through gold, even at temperatures well below 300°C.[21] For long term stability, if the crystal is going to see temperature variations, TiW-Au or Ti-Pd-Au are the preferred metallization systems to use.

R_S Minimum A positive aspect of utilizing metals lighter than gold is shown in Figure 9. For a first order approximation, these metals have similar conductivities and similar nucleation and growth characteristics. The R_S vs thickness curve will then have a minimum which broadens proportional only to density change. For example, aluminum will have a mechanical dampening effect (that will raise the R_S) at much thicker films relative to gold.

Batch Processing

Batch processing, within the context of this study, is defined as metallizing a large number of crystals to frequency with one deposition. In other words, there is no individual 'plate to frequency' steps. The limits to batch processing are established in Figure 10. This concept, of course, depends on the initial frequency specification range (shown in ppm), the uniformity of the deposition system and the density of the film. The best sputtering systems available today have non-uniformities (from run to run or across the deposition chamber) to within ± 4%. This means that with gold, the best that can be obtained is a 40 ppm range. For aluminum; however, with a variation of ± 4%, the crystals will be within a few parts per million. The process then becomes more of an etch to frequency (of the blank) which is also a batch process. It was found that an etching frequency range could be compensated for by the metal thickness distribution in the deposition system and even do better than Fig. 10 implies.

Conclusions

For the 4.194 MHz crystal under study, the frequency shift with respect to thickness had a significant deviation from linearity above 4000 Å of Cr-Au, and the R_S minimum was found to be at a thickness of 4000 Å of Cr-Au. Directly comparing the oscillator properties of gold, copper and aluminum electrodes, it was determined that copper and gold gave equivalent R_S values, as long as the copper was processed properly. The aluminum coated oscillators consistently gave high R_S readings. The temperature stability of these metals on quartz was found to be good up to 300°C. Also, considering the improvements in deposition systems, the concept of batch processing of quartz oscillators was established.

Acknowledgements

The author would like to thank Ray Wells for providing blanks, Bill Wadlin for the Auger analysis, Frank Garcia for the metal depositions and Craig Thornton for the measurements made thoughout the studies.

References

1. H.W. Weinhart and H.G. Wehe, "Metal Electrodes Deposited on Quartz Crystals By The Evaporation Process", Quartz Crystals For Electrical Circuits, ed. Raymond A. Heising, D. Van Nostrand, New York (1946).

2. R. Fischer, L. Schulzke, "Plating to Frequency -A Powerful Fabrication Method For Crystals With Closely Controlled Parameters", Proceedings of the 30th Annual Symposium on Frequency Control, p. 209 (1976).

3. A.E. Zumsteg and P. Suda, "Properties of a 4MHz Miniature Flat Rectangular Quartz Resonator Vibrating in a Coupled Mode", Proceedings of the 30th Annual Symposium on Frequency Control, p. 196 (1976).

4. Raymond L. Filler and John R. Vig, "The Effect of Bonding on the Frequency vs Temperature Characteristics of AT-Cut Resonators", Proceedings of the 30th Annual Symposium on Frequency Control, p. 264 (1976).

5. P.D. Wilcox, et.al., "A New Ceramic Flat Pack for Quartz Resonators", Proceedings of the 29th Annual Syposium on Frequency Control, p. 202 (1975).

6. P.C.Y. Lee and K.M. Wu, "The Influence of Support-Configuration on the Acceleration Sensitivity of Quartz Resonator Plates", Proceedings of the 31st Annual Syposium on Frequency Control, p. 29 (1977).

7. R.N. Castellano, et.al., "The Relationship Between Quartz Surface Morphology and the Q of High Frequency Resonators", Proceedings of the 31st Annual Syposium on Frequency Control, p. 126 (1977).

8. R.K. Hart, W.H. Hicklin, and L.A. Phillips, "Methods of Cleaning Contaminants From Quartz Surfaces During Resonator Fabrication", Proceedings of the 28th Annual Symposium on Frequency Control, p. 89 (1974).

9. See Reference 3.

10. John R. Vig, J.W. Le Bus and R.L. Filler, "Chemically Polished Quartz", Proceedings of the 31st Annual Symposium on Frequency Control, p. 131 (1977).

11. L. Holland, Vacuum Deposition of Thin Films, Chapman and Hall, London, p. 264 (1961).

12. John L. Vossen, Thin Film Process, Academic Press, New York (1978).

13. D. Chaabra, N. Ainslie, D. Jepson, Proc. of the ECS Meeting, Dallas, Texas (1967).

14. G.L. Dybwad, "Aging Analysis of Quartz Crystal Units With Ti-Pd-Au Electrodes", Proceedings of the 31st Annual Symposium on Frequency Control, p. 144 (1977).

15. Gunther Sauerbrey, "The Use of Quartz Oscillators for Weighing Thin Layers and for Microweighing", Zeitshrift Physik, 155, p.206 (1959).

16. Owen Lewis and Chih-shun Lu, "Relationship of Resonant Frequency of Quartz Crystal to Mass Loading", Proceedings of the 29th Annual Symposium on Frequency Control, p.5 (1975).

17. Roger E. Bennett, Quartz Resonator Handbook, p. 107 (1960).

18. Constantine A. Neugebraur, "Condensation, Nucleation and Growth of Thin Films", Handbook of Thin Film Technology, ed. Leon Maissel and Reinhard Glang McGraw-Hill, N.Y. (1970).

19. S.P. Sharma, "Absorption of Water on Copper and Cuprous Oxide", J. Vac. Sci. Technol., 16 (5), p. 1557 (1979).

20. Virgil Bottom, "A Novel Method of Adjusting the Frequency of Aluminum Plated Quartz Crystal Resonators", Proceedings of the 30th Annual Symposium on Frequency Control, p. 249 (1976).

21. See reference 13.

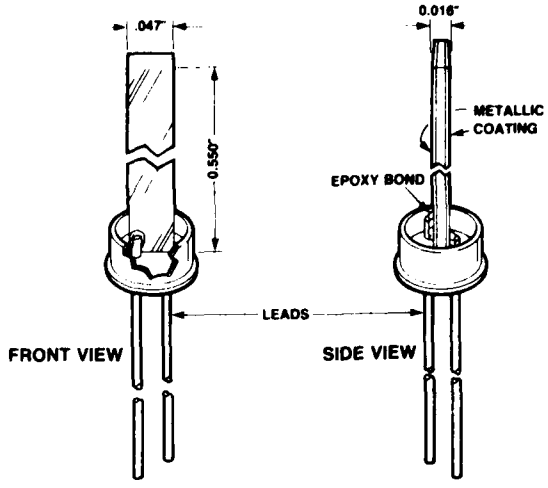
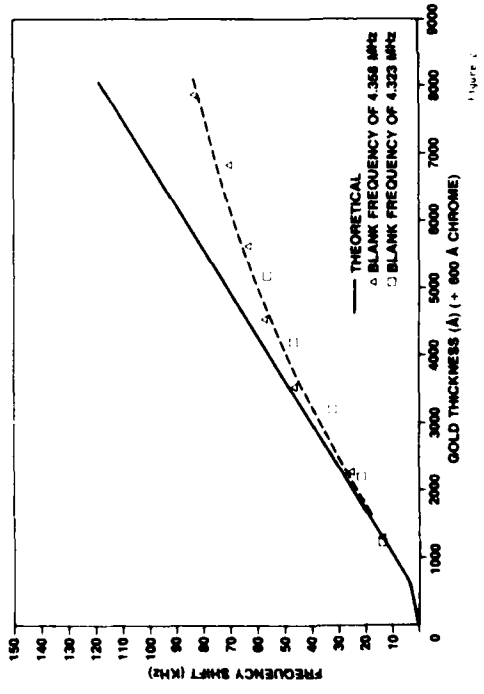


Figure 1



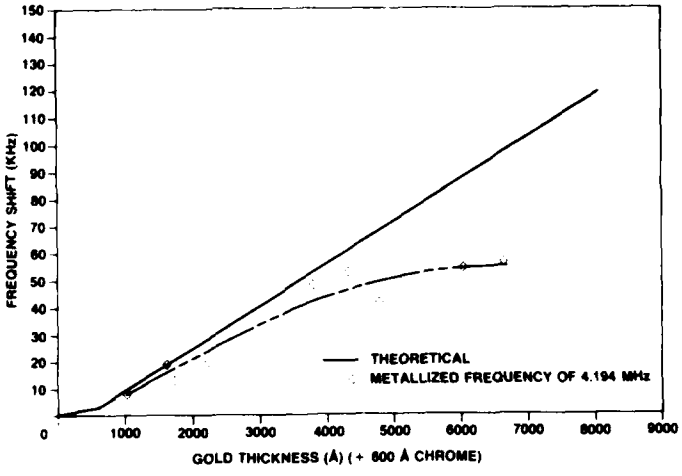


Figure 3

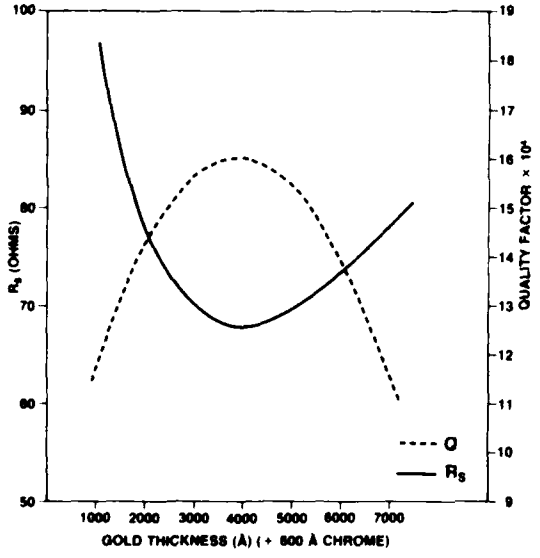


Figure 4

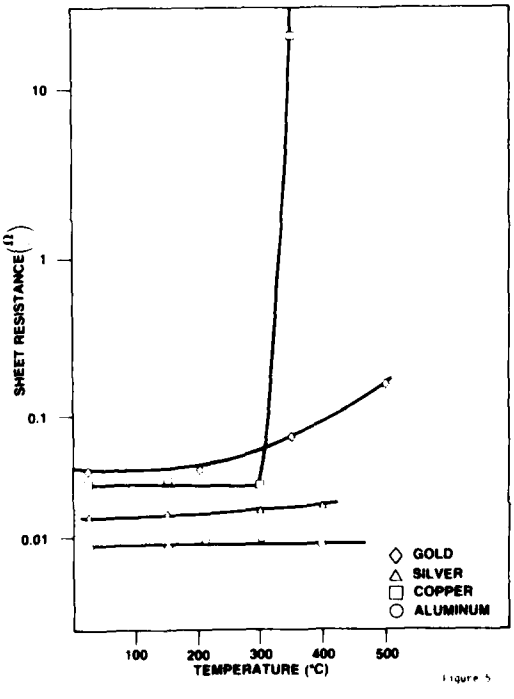
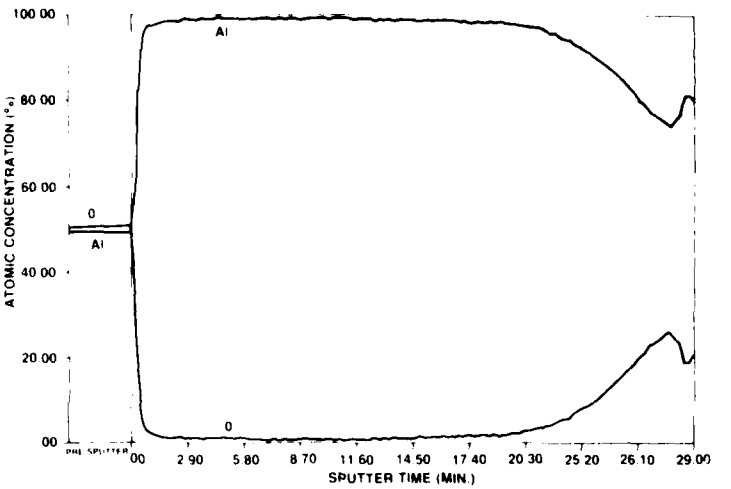
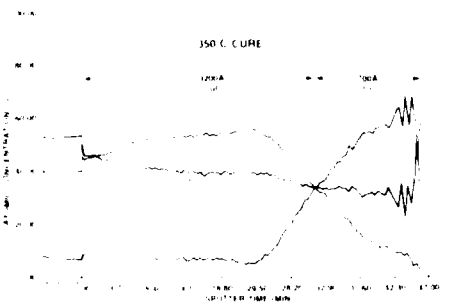
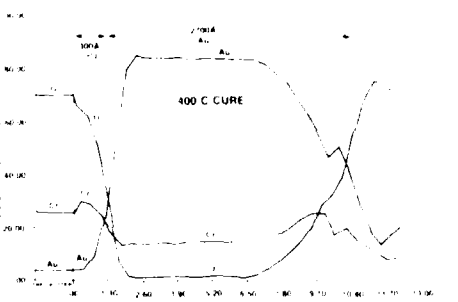
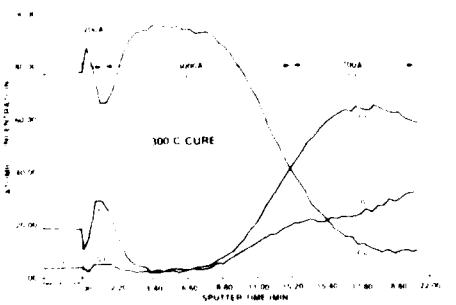


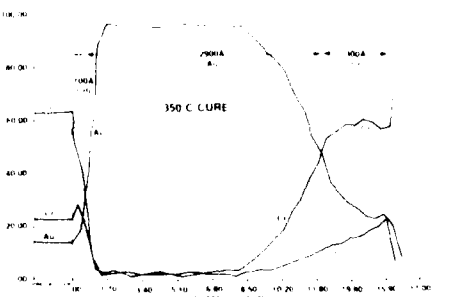
Figure 5

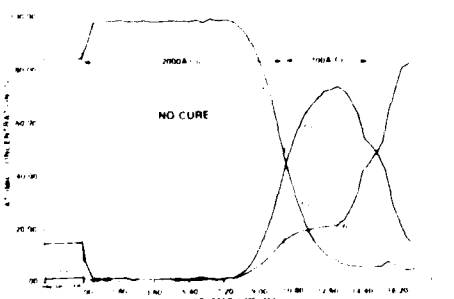

100 00
80 00
60 00
40 00
20 00
00
ATOMIC CONCENTRATION (%)
Al
O
Al
O
00
2 90
5 80
8 70
11 60
14 50
17 40
20 30
25 20
26 10
29 00
SPUTTER TIME (MIN.)

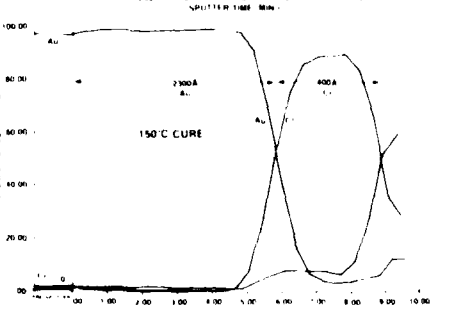

350 C CURE

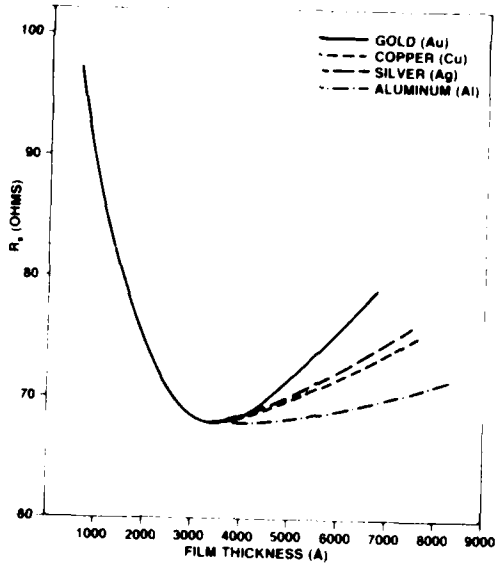

400 C CURE

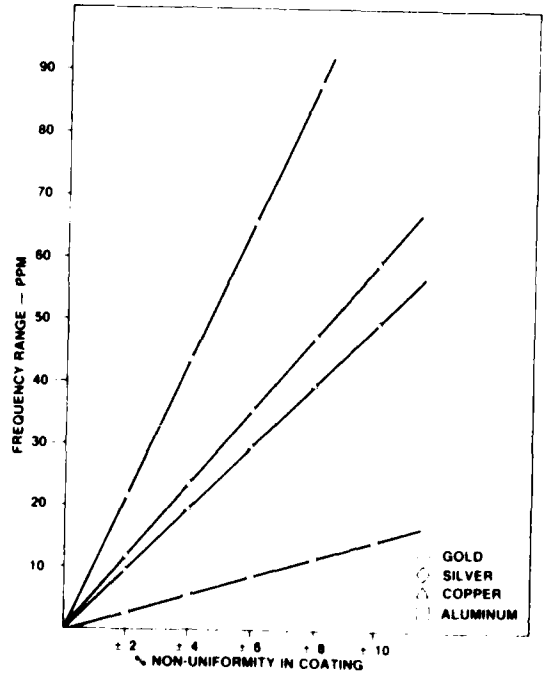

300 C CURE


350 C CURE


NO CURE


150°C CURE


GOLD (Au)
COPPER (Cu)
SILVER (Ag)
ALUMINUM (Al)
R_s (OHMS)
100
90
80
70
60
1000
2000
3000
4000
5000
6000
7000
8000
9000
FILM THICKNESS (Å)


FREQUENCY RANGE – PPM
90
80
70
60
50
40
30
20
10
GOLD
SILVER
COPPER
ALUMINUM
± 2
± 4
± 6
± 8
± 10
% NON-UNIFORMITY IN COATING

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

X-Ray Goniometry of the Modified Doubly Rotated Cuts

Ernst Knolmayer

Quarzkeramik GmbH, Stockdorf, West-Germany

SUMMARY

The first rotation for the cuts RT, FC, IT and SC is made slightly different from usual. After the first slightly different rotation one gets netplanes parallel to the X'-axis. Quarter-circle diagrams were made for the netplanes and cuts similar to the half-circle diagrams for the singly rotated cuts. To get similar temperature frequency relations as for the original cuts it is necessary to correct the second rotation appropriately. By making a section with the faces Z and X' we have similar relations for the X-ray goniometry as by making a section with Z and X faces for singly rotated cuts. These modified cuts have about the same properties as the unmodified. It is easier to make these cuts.

INTRODUCTION

The fast warmup behavior of an OCXO with the modified SC (MSC) resonator is not obviously different from the behavior of an OCXO with SC resonators. The MSC and MIT cuts are to propose for TCXO's (better transient temperature behavior). The inductance of the modified RT cut (MRT cut) is about 30 times higher and the quality factor is about 4 times higher than for AT. The temperature frequency behavior for this cut is good. All the properties mentioned above predestinate this cut for very narrow band filters.

The Modified Cuts

The following table contains the unmodified doubly rotated cuts (2,3,4,5,6) and the modified doubly rotated cuts with the angles of first rotation.

The Cuts

Unmodified		Modified	
Cut	First rot.	Cut	First rot.
RT	± 15°	MRT	±13°54'
FC	± 15°	MFC	±13°54'
IT	± 20°	MIT	±19°06'
SC	± 22°	MSC	±23°25'

Fig.1, Fig.2, Fig.3 and Fig.4 demonstrate the geometrical relations for the modified doubly rotated cuts in the XY plane.

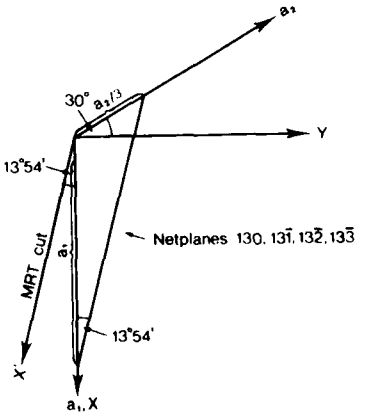


Fig.1 Geometrical relations in the XY plane for the MRT cut.

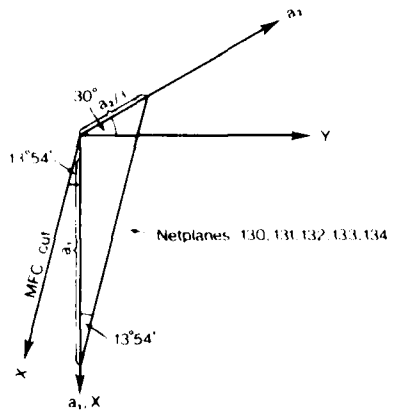


Fig.2 Geometrical relations in the XY plane for the MFC cut.

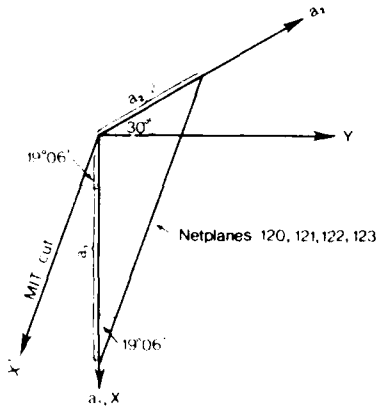


Fig.3 Geometrical relations in the XY plane for the MIT cut.

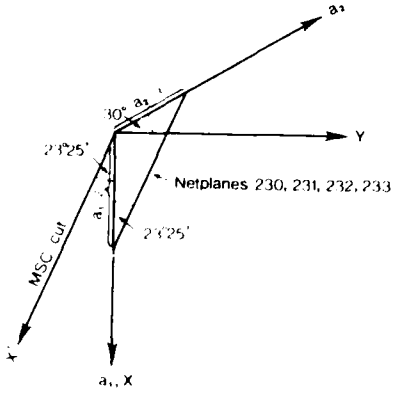


Fig.4 Geometrical relations in the XY plane for the MSC cut.

As can be seen in the table, the first rotation has two signs. That means that there are two geometrically different cuts with the same properties. Figs. 1 to 4 demonstrate the case with negative first rotation. The X'-axis is parallel to the netplanes. The first rotation is round the Z-axis. The X'-axis is the intersection line of the cut and the XY plane. The intersection line of the netplanes with the XY plane is parallel to the X'-axis. The angle between this line and the X-axis is also the angle of first rotation. The second rotation is round the X'-axis. The netplane data are taken from a paper of the author presented at the Symposium 1980 (7).

Acknowledgement

The author thanks Mr. F. Meilhammer and Mr. L. Misković for the preparation of the resonators and Mr. E.O. Kunte for his drawing and Mrs. B. Panzer for her typing.

References

1. R.A. Heising; Quartz Crystals for Electrical Circuits.

2. R. Bechmann, A.D. Ballato and T.J. Lukaszek; "Frequency-Temperature Behavior of Thickness Modes of Double-Rotated Quartz Plates", Proc. 15th Annual Symposium on Frequency Control, 1961 pp.22-48

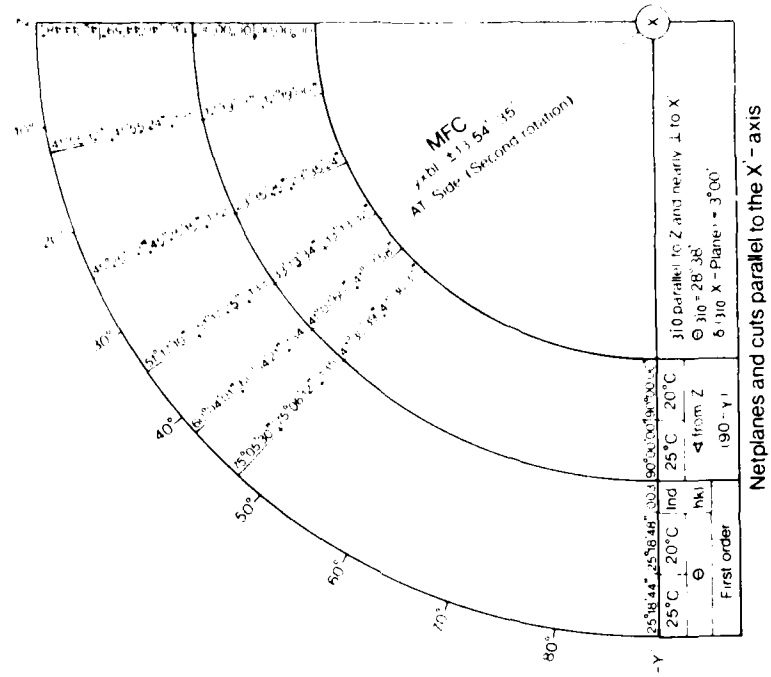
3. R. Bechmann, A.D. Ballato, and T.J. Lukaszek; "Higher-order Temperature Coefficients of the Elastic stiffnesses and Compliances of Alpha-Quartz", Proc. IRE, Vol.50, No.8, August 1962 pp. 1812-1822

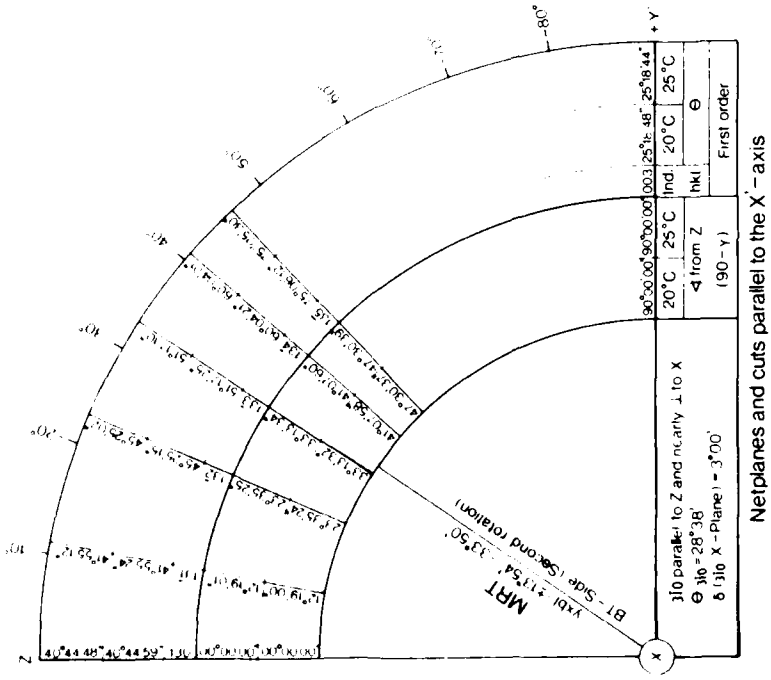
4. R. Bechmann, "Thickness-Shear Mode Quartz Cut with Small Second- and Third-Order Temperature Coefficients of Frequency", Proc. IRE, Vol. 49, No.9, Sept. 1961, p. 1454

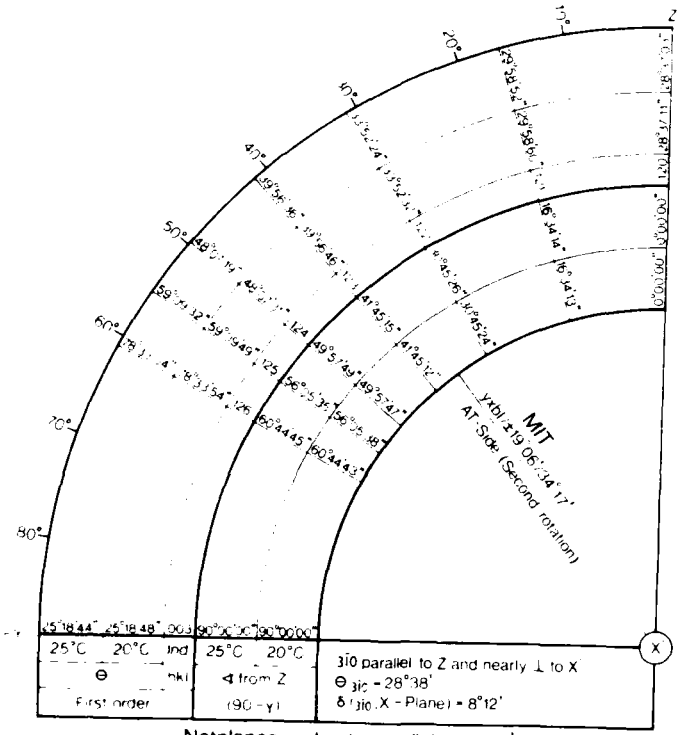
5. F. Lagasse, J.Ho, and M. Bloch; "Research and Development of a New Type of Crystal - The FC Cut", Proc. 26th Annual Symposium on Frequency Control, June 1972, pp.148-151

6. E.P. EerNisse; "Calculations on the Stress Compensated (SC-cut) Quartz Resonator", Proc. 30th Annual Symposium on Frequency Control, 1976 pp. 8-12

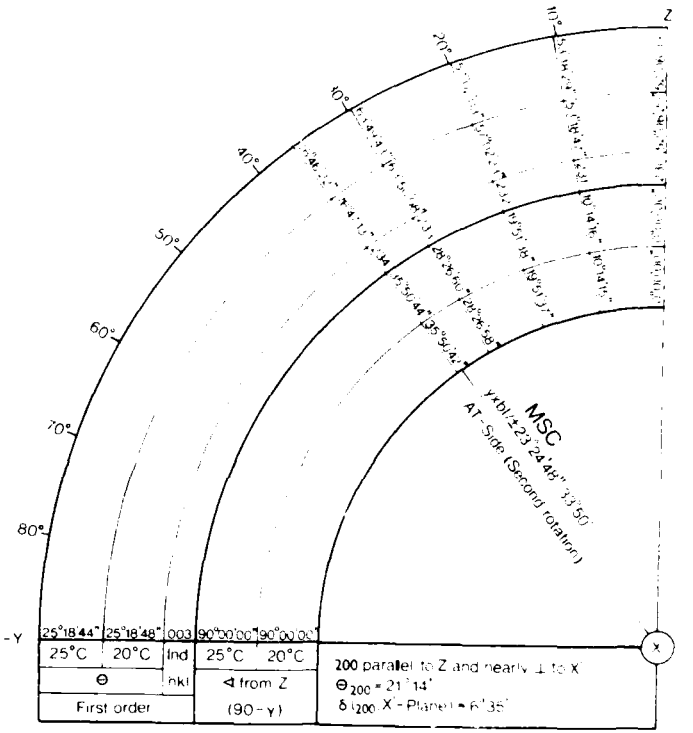
7. E. Knolmayer; "Netplane Date for Quartz, Berlinite, $LiTaO_3$ and $LiNbO_3$,
Proc. 34th Annual Symposium on Frequency Control", 1980 pp. 102-111.







Netplanes and cuts parallel to the X'− axis
Quartz MIT − cut
Cu $K\alpha_1$ − line



Netplanes and cuts parallel to the X'− axis
Quartz MSC − cut
Cu $K\alpha_1$ − line

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

AN INSTRUMENT FOR AUTOMATED MEASUREMENT OF THE ANGLES OF CUT OF DOUBLY ROTATED QUARTZ CRYSTALS

J. L. Chambers, M. A. Pugh and S. T. Workman
Advanced Research and Applications Corporation
1223 E. Arques Avenue
Sunnyvale, CA 94086

R. W. Birrell and R. J. Valihura
P. R. Hoffman Co.
321 Cherry Street
Carlisle, PA 17013

Summary

The ongoing development of an automated x-ray orientation system for accurate measurement of the angles of cut of doubly rotated (e.g., SC-cut) quartz crystal plates is described. The goal of this development is a prototype instrument capable of determining the conventional θ/ϕ angles of cut to within $\pm$ 10 arc seconds of accuracy for the SC cut, $\pm$ 6 seconds for the AT cut, and $\pm$ 30 seconds for the other cuts on the $\theta>0^{\circ}$ bulk wave zero temperature coefficient locus. Determinations are to be made at a rate of a least 50 plates per hour. These plates are assumed to be precut to within about 1° of the correct orientation.

The design chosen is that of a laser-assisted Laue diffractometer under control of a dedicated minicomputer. Use of the Laue method requires only a single degree of freedom in the motion of the crystal plate leading to a flexible and relatively open design.

Introduction

In this paper we describe the ongoing development of a laser-assisted Laue diffractometer for accurate automated measurement of the angles of cut of quartz crystals, including double rotated cuts.[1] The device to be described utilizes the x-ray diffraction pattern from the crystal in a different manner than conventional orientation equipment, by employing a continuous x-ray spectrum. This Laue approach[2] gives rise to a number of design advantages, as outlined below.[3]

Design Goals

The task to be performed is, given a blank cut to within about $\pm$ 15' of the desired orientation angles, θ, ϕ, to measure these angles to a few arc-seconds accuracy in order to determine the corrections required to yield the desired angle of cut. These angles are shown in Fig. 1. The orientation of the face of the blank with respect to the crystal lattice is determined by θ and ϕ, as defined in terms of a thickness vector, t, which is to be oriented along the Y-axis before the blank is cut in the X-Z plane. The axis normal to the face of the cut blank is denoted by ψ; thus, rotation about ψ is rotation of the blank about the normal to its face. Note that by specifying the orientation of two planes within the crystal, in this case the 001 and 010, the default orientation of the crystal has been determined completely. In general, the orientation of a crystal lattice with known unit cell constants can be completely specified by the normals to any two non-colinear crystal planes.

The major objectives of the development to be described are as follows:

1. Prototype System. A prototype instrument is to be delivered upon completion of the project.

2. Types of Blanks. The system should be able to measure both singly and doubly rotated cuts of general orientation, and in particular any cut on the bulk-wave zero temperature coefficient locus. It must be adaptable to a variety of blank sizes, and thicknesses ranging from 55 microns to 2 mm.

3. Accuracy and Reproducibility. The accuracy/reproducibility of the measured angles of cut required are 10"/4" for SC-cut blanks, 6"/3" for AT cuts, and 30"/15" for other cuts on the $\theta > 0$ branch of the bulk-wave zero temperature coefficient locus.

4. Measurement Rate. These determinations are to be made at a rate of 50 blanks/hour, for doubly rotated cuts.

Thus, greater flexibility and accuracy than commonly found in conventional x-ray orientation equipment, while maintaining rates suitable for a production environment have been the major design considerations.

The Laue Method

The approaches investigated for this application fall into two classes: those utilizing conventional Bragg diffraction of monochromatic x-ray sources, and those utilizing Laue diffraction of "white" radiation such as that obtained from a tungsten-target x-ray tube operated below the characteristic K-line potential.

The basic differences between these two methods are shown in Fig. 3. The diffraction from the set of planes represented by the reflecting plane in the figure serves to "sift out" a narrow band of the incident x-ray spectrum at the position of the diffraction maximum (or "reflection"). As the crystal orientation is changed the position of the reflection changes, "tuning" through the incident spectrum in the process. In the monochromatic case, the reflection is observable only over that very narrow range of crystal orientation where it is tuning through the characteristic line. Thus, orientation of a crystal by the Bragg method in the general case requires three degrees of freedom in crystal motion, in order to produce the diffracting condition. In contrast, in the Laue case, the position of the spot can be followed as the orientation of the crystal is changed, thus permitting the desired measurements without any crystal motion geared specifically toward producing a diffracting condition. The Laue method can therefore lead to an open, relatively inexpensive, and easily automated design because of the less complex crystal handling required. In practice, many blanks we must deal with are round, and an initial search in ψ is required to locate the desired reflections. Thus, the Laue method can be implemented with only a single rotary motion of the blank about ψ.

The major disadvantage of the Laue method is that the diffracted intensities are weaker due to the absence of the characteristic K-line. As discussed below, this problem can be largely alleviated by careful choice of instrument geometry.

The manner in which the measured Laue spot positions arise from the crystal lattice orientation is shown more clearly in Fig. 3, which shows one of the many reflections in a Laue pattern in traditional back-reflection geometry. The angles α, β, and γ represent small incremental changes in the crystal orientation. The resulting positional change in the diffraction maximum is the same as if reflection were occurring from a mirror. At a 500 mm crystal-detector distance the x-ray spot moves about 5 microns for each arc-second change in crystal orientation. For example, to achieve 10 arc-second accuracy the spot position must be measured to within 50 microns.

A conceptual method for completely determining the crystal lattice orientation was thus arrived at as shown in Fig. 4, again in traditional back-reflection geometry. The disc-shaped blank is mounted on a rotary stage. Two x-ray detectors with apertures large enough (e.g. $\pm 1^{\circ}$) to accommodate mounting errors as well as the error in the angles of cut are positioned so that only at some unique ψ will two reflections enter the counters simultaneously. The exact position of each reflection within its counter is then determined. From this information the normals to the two sets of diffracting planes, and thus the orientation of the crystal lattice, can be computed.

Assuming that the ψ-axis of the blank coincides exactly with the stage rotation axis, the angles of cut have been determined. However, due to errors in mounting, this may not be the case, and a method of correction for such an error is needed. Reflection of a laser beam from the face of the blank as it is rotated about ψ provides this information. The method of laser-assisted diffractometry has been described by Vig.[4]

The measurements are complicated by the fact that blanks are not necessarily supplied along with information concerning which face is, for example, the +X face. Turning the blank over gives rise to a mirror image diffraction pattern, as shown in the simulated SC-cut patterns in Fig. 5a, b, which correspond to the detector plane in Fig. 4. A minimum of three detectors is thus needed to account for this front/rear ambiguity, with one detector positioned on the mirror plane.

Pre-Design Experiments

Pattern Simulations

The requirement that detectors be positioned such that reflections enter the counters simultaneously at only one unique ψ prompted development of techniques for simulating the Laue diffraction patterns, thus enabling suitable pairs of reflections to be determined. Fig. 5 is an example of such simulations in traditional back-reflection mode for an SC-cut, although the simulation program is capable of working with an aribtrary instrument geometry and any specified angles of cut (an example is shown later). While it is not obvious simply from looking at the patterns, suitable pairs can be found and have been located both in the simulations and by experiment. The simulations have been verified experimentally, as shown by the observed SC-cut back-reflection pattern in Fig. 6. In this geometry, rotation about ψ produces a simple rotation of the pattern, and the agreement between the observed and predicted patterns is thus apparent.

Breadboard Instrument

A breadboard device was constructed to enable these initial experiments to be performed. A schematic of this device is shown in Fig. 7. The system consisted of a tungsten-target x-ray source, a rotating stepping-motor stage, an x-ray detector, a translating stepping-motor stage capable of moving a scanning slit across the detector to localize the spot within it, and a dedicated minicomputer and CAMAC interface to the breadboard to perform the automatic data acquistion and control functions.

The early measurements performed with this device were highly informative, and involved determination of the reflection intensities of SC-cut blanks in the back-reflection region. The strongest reflections were only about 1000 counts/sec. Relatively long counting times (over 30 seconds) were required to adequately map the spot profile in a single dimension. Ways of increasing the intensities were therefore sought. Optimization of the type of tungsten-target x-ray tube used, and installation of a helium path between the crystal and detector improved intensities by up to 50%; however, a more substantial improvement was clearly required. By far the greatest improvement in intensities occurred through use of non-perpendicular,

glancing-angle geometry, as shown in Fig. 8. This geometry permits observation of reflections at lower diffraction angles, where the peaks are on the average considerably more intense. An order-of-magnitude improvement was obtained, yielding intensities adequate for the required measurements. The drawback of working in such a mode is the more complex geometry involved as described later.

A number of techniques were explored for mapping the intensity profile within the detector. The greatest precision was obtained where the reflections were scanned with a slit in relatively coarse (50-100 micron) steps and fit by least squares to a Gaussian model profile. The three parameters refined were the amplitude, A; the standard deviation, S, corresponding to about 2/3 of the spot radius; and the center position, M. Fig. 9 shows several sets of determinations for different slit widths, scan widths, step sizes, and counting times. The best reproducibilities were obtained when the slit was about as wide as the spot radius. The reproducibilities were excellent, in the 4-8 micron range. Even so, the reproducibilities were still largely counting-statistics limited, as determined by variation of the source intensity. Typical scan times were in the 15-20 second range, with translating stage motion accounting for about ½ of the time. (Stages selected for the prototype are considerably faster). Since this agreement is about 5-10 times better than required for 10 arc-second precision in the measurement of a single reflection, the inherent precision in the Laue method appears to be quite high.

Least-Squares Orientation

In order to determine the angles of cut, more than one reflection must be measured. Because of measurement errors, the geometrical relationships among the reflections (for example, the angles between them) may not be exactly those predicted on the basis of the known crystal lattice. The method chosen for determining the lattice orientation from the measured reflection positions was therefore a least-squares fit of a model lattice to the observations based on rigid-body rotation of the model. This approach has the advantage of fully utilizing any number of reflections, permitting for example later addition of more detectors for increased accuracy. A program was implemented to perform the least-squares orientation and used to investigate the relationship between the errors in spot position and the resulting error in the computed crystal orientation. Fifty-micron random errors in spot position at a 500 mm sample-detector distance gave rise to 2.5-5 arc-second errors in the direction normal to the desired face of an SC-cut blank, depending on the particular reflections used. Thus, it appears that the design goals can be met if measurement errors can be kept less than 50 μ at 500 mm (which the experiments indicated is highly feasible).

Prototype Design

These highly promising results have led to the design and fabrication of a prototype instrument. The design incorporates a 20° glancing-angle geometry. An example of the pattern produced by an SC-cut blank in this configuration is shown in Fig. 10. The small circles with cross hairs indicate the positions of the x-ray spots at the ψ-value of interest (218.12°). The dotted lines show the path traced by each spot as the crystal is rotated about ψ. Triangles indicate those positions where a spot "tunes" to a tungsten L-line (it is highly undesirable to utilize these L-lines to boost intensity, since the spot profile becomes highly non-uniform and sensitive to changes in orientation). The $1\ \bar{2}\ \bar{1}$ reflection, shown at the top of its path at $\psi = 218.12^{\circ}$, and the $0\ \bar{1}\ 0$ reflection are a good candidate pair for measurement of SC-cut crystals. The geometry is thus more complex than simple back-reflection, but is nevertheless readily predictable.

To account for any irregularities in intensity distribution across the x-ray spot, and for any systematic errors in the alignment of the instrument, it was decided that measurements would be made relative to a calibration crystal. The calibration is performed once at the beginning of each batch of blanks. A calibration crystal is thus required for each type of cut to be measured. Jack Kusters and Charles Adams of Hewlett-Packard have very kindly provided such a crystal for SC-cut blanks.

The design incorporated three standard scintillator/photomultiplier x-ray counters each equipped with a single stepping-motor translator. Mapping of the intensity distribution within the detector in two dimensions is accomplished with a slit arrangement as shown in Fig. 11. The horizontal slit is translated across the detector aperture to yield the vertical coordinate. The vertical slit is then moved to that position and the profile mapped in the horizontal dimension by turning the blank about ψ with the rotating stage.

Prototype Instrument

The prototype instrument is shown in Fig. 12. At the center of the instrument is the rotating stage on which the crystal is mounted. The crystal is held in place by a three-point vacuum chuck. The stage height is adjustable to accommodate blanks of different thickness. The x-ray source and 20° incident beam collimator mount are apparent. Three arches serve as the detector supports. All three detectors can be set to any inclination between 15° and 65° at a fixed crystal-to-detector distance of 400 mm. The azimuth of the central detector is fixed; those of the two outside detectors are variable from 20° to 90°, allowing accessibility to a wide range of the diffraction pattern. Each detector arch supports a translating stepping motor stage and a slit assembly similar to that in Fig. 11. These stages are capable of a 10 micron step size and a 2 cm/sec slew rate, thus significantly decreasing the contribution of stage positioning time in the profile mapping.

Preliminary x-ray measurements made with the prototype indicate that the intensities are slightly higher than for the breadboard device, probably arising from the greater number of degrees of

freedom in the x-ray source alignment. Because the instrument will be calibrated with reference crystals the mechanical alignment of the components is not required to be in the arc-second range. Nevertheless, to enable the initial search for reflections to be successfully conducted alignment errors must be kept small with respect to the angle subtended by a detector aperture (e.g. in the 1 arc-minute range). The critical alignment of the instrument is currently under way, and it is expected that performance measurements will soon be available.

References

1. This work was funded under Contract No. DAAK-20-79-C-0254 from the US Army Electronics Research and Development Command, Fort Monmouth, NJ. P. R. Hoffman Co. is the prime contractor and has been responsible for design of the mechanical components of the prototype system. Advanced Research and Applications Corp. (ARACOR) is a subcontractor to P. R. Hoffman Co.

2. The use of the Laue method for this application was first suggested to us for this application by Roger W. Ward.

3. R. W. Birrell, R. J. Valihura, J. L. Chambers, M. A. Pugh, and S. T. Workman, Research and Development Technical Report DELET-TR-79-0254-1, U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ, 1980.

4. J. R. Vig, "A High Precision Laser Assisted X-Ray Goniometer for Circular Plates," Proc. 29th Annual Symp. on Frequency Control, pp. 240-247, 1975.

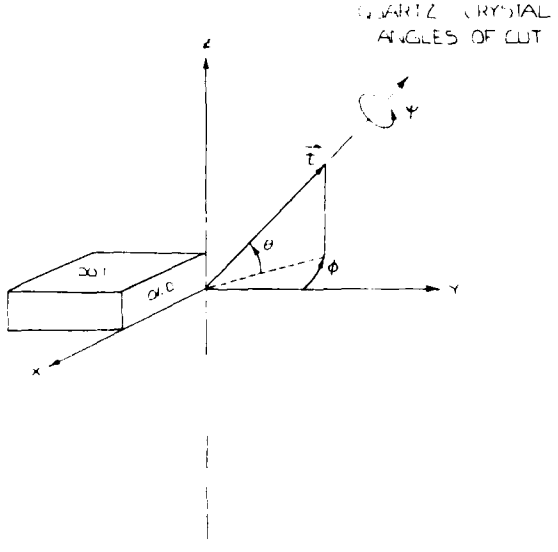


Figure 1. Convention for the angles of cut of a quartz crystal. The thickness of the cut blank is to be along t.

LAUE vs. MONOCHROMATIC TECHNIQUE

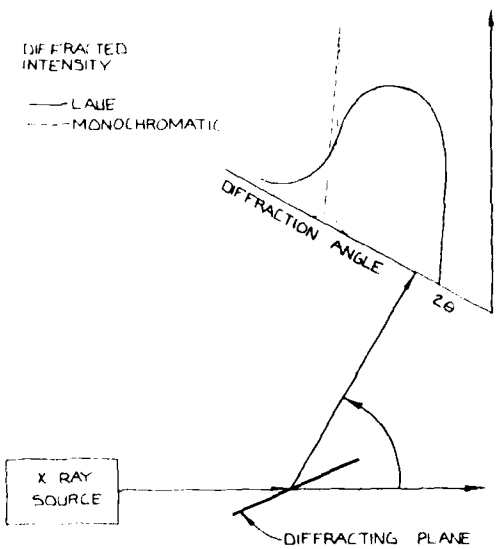


Figure 2. Laue vs. Bragg diffraction. Because a continuous x-ray source is used in the Laue case, the diffraction maximum is observable over a wide range of crystal orientations.

ANGULAR SENSITIVITIES

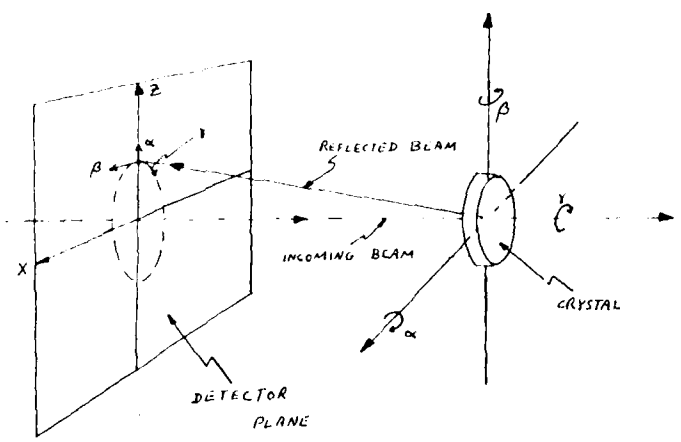


Figure 3. Sensitivity of Laue spot motion to changes in crystal orientation. Only one of many reflections in the pattern is shown, in traditional back-reflection geometry.

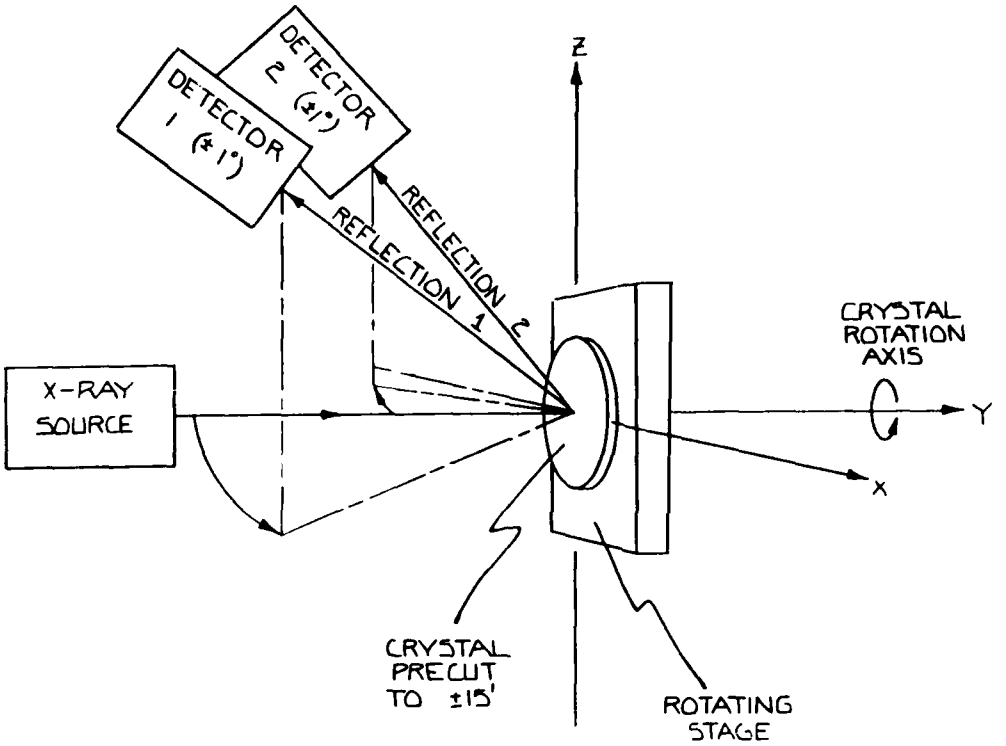


Figure 4. Concept for determination of the angles of cut by the Laue method. In practice a laser system will be present to measure the orientation of the front face of the blank, and a third x-ray detector will be added to resolve the front/rear side ambiguity described in the text.

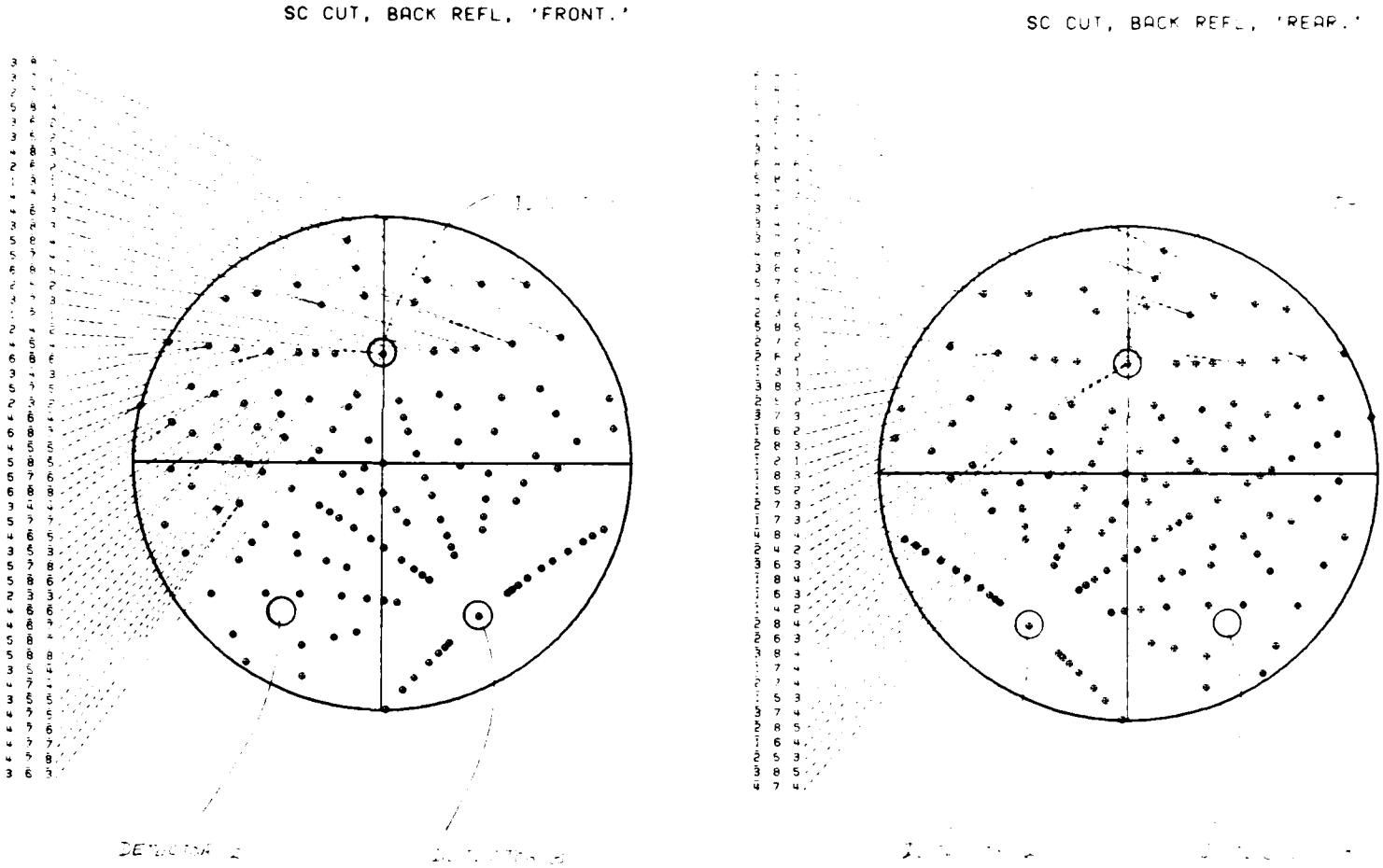


Figure 5.

Simulated SC-cut Laue back-reflection patterns. Fig. 5a corresponds to the pattern produced when the face nearest +Y faces the incoming beam. In Fig. 5b, the face nearest -Y is toward the incident beam.

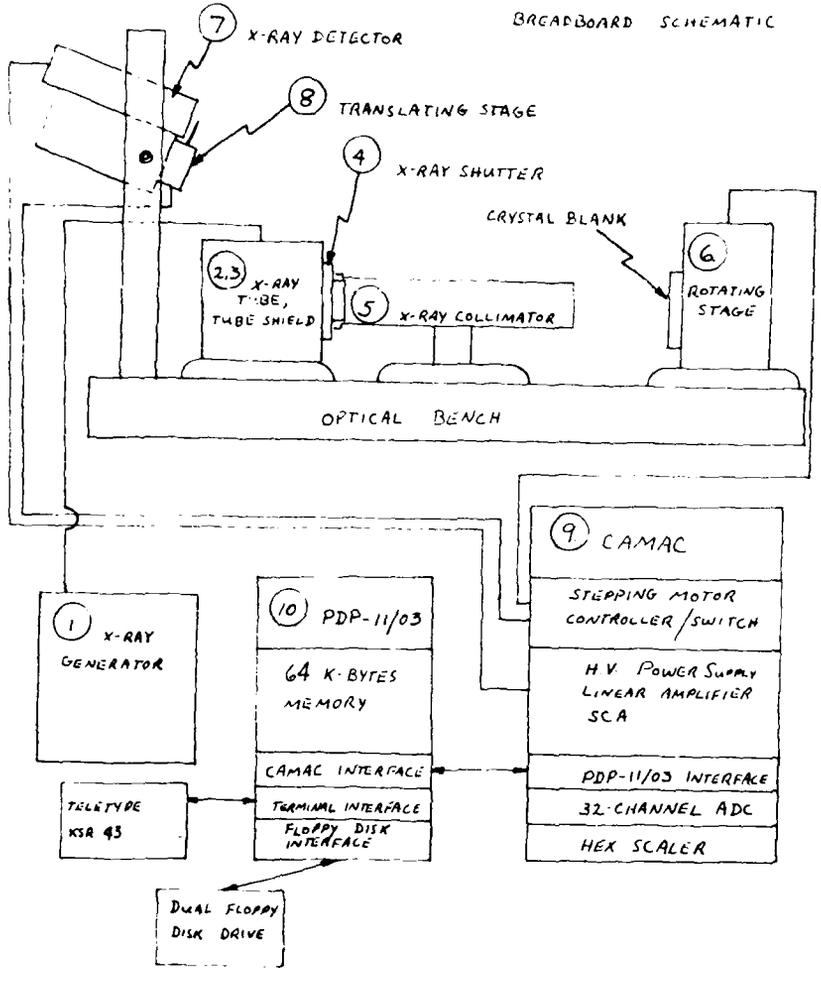


Figure 7.

Schematic of the breadboard device used for the early measurements.

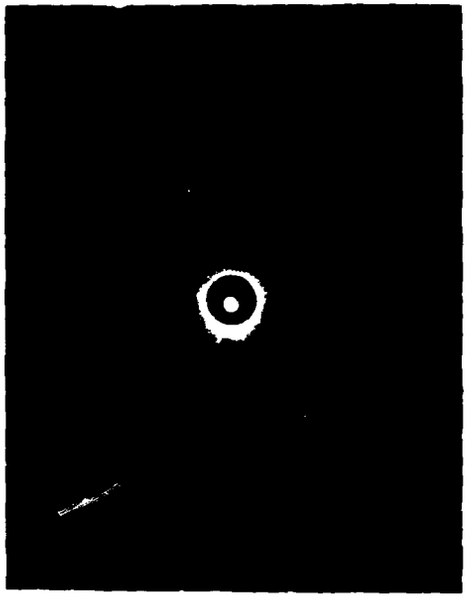


Figure 6.

Observed SC-cut back-reflection pattern. This is a "front" side pattern, rotated approximately 30° counterclockwise from that in Fig. 5a.

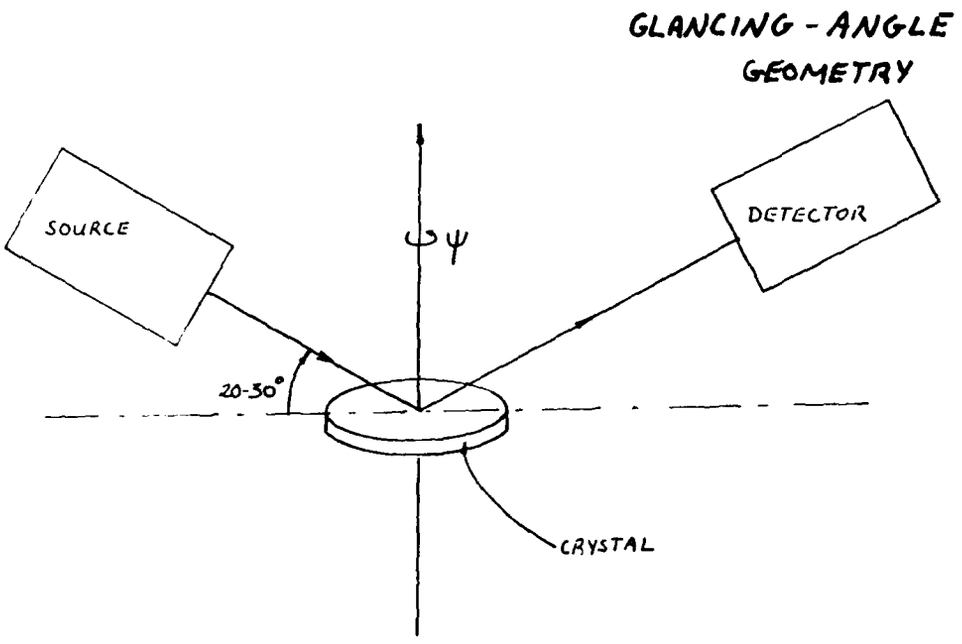


Figure 8. Glancing-angle geometry. The instrument is "unfolded" from the traditional back-reflection mode, permitting reflections to be observed at lower diffraction angles.

$1\ \bar{2}\ \bar{1}$ Reflection Scan Summary

Slit	Scan Width	Step Size	Sec/Step	(Sec/Scan)	(A)	(s)	(m)	(σ_m)
0.7	2S	0.2	1.0	15	2902	0.647	9.58	(6.2)
0.7	2S	0.1	0.5	13	2862	0.657	9.37*	(5.5)
0.7	2S	0.4	2.0	14	2868	0.659	9.37	(7.3)
0.7	3S	0.4	2.0	24	2951	0.621	9.42*	(4.1)
0.7	3S	0.4	2.0	24	2968	0.620	9.42	(3.9)
0.7	3S	0.2	1.0	20	2985	0.619	9.43	(4.1)
0.7	3S	0.2	1.0	20	2930	0.634	9.42	(6.2)
0.7	3S	0.4	1.0	12	2989	0.632	9.89*	(6.7)
0.3	3S	0.4	1.0	12	1453	0.604	9.53*	(8.6)
1.0	3S	0.4	1.0	12	3986	0.653	11.908*	(4.9)
2.0	3S	0.4	1.0	16	7697	0.886	14.58*	(4.3)
3.0	3S	0.4	1.0	22	8757	1.101	10.66*	(3.9)

Figure 9. Table of scan results for the $1\ \bar{2}\ \bar{1}$ reflection in glancing-angle mode, under a variety of conditions. Mean values of the peak amplitude (A), width in mm (s), and central position in mm (m) obtained by fitting the observed scans to a Gaussian are presented, averaged over five determinations. The standard deviation of the central position in microns over the five scans are given by σ_m. Values of m marked with an asterisk represent scans taken after physically repositioning the slit, so that no agreement with the previous value of m is expected. The column labelled "Slit" is the slit width in mm.

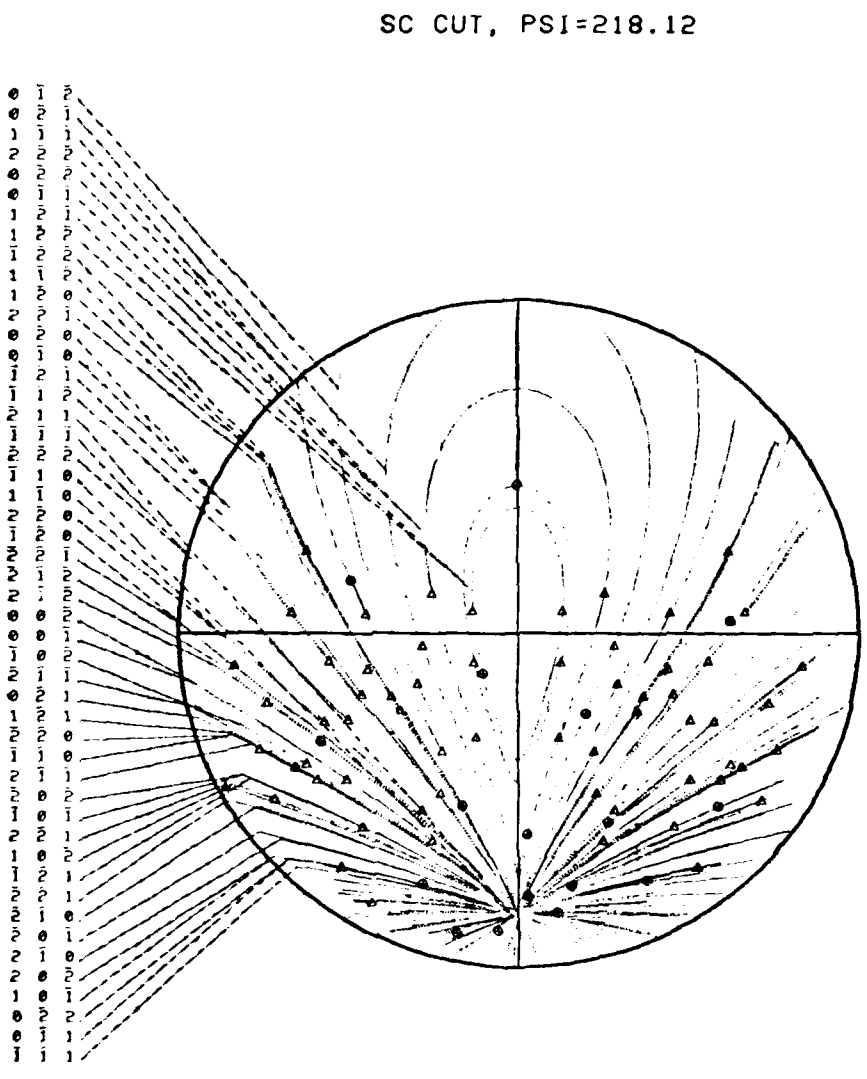


Figure 10. SC-cut Laue pattern in glancing-angle mode. This simulation shows the paths traced by the reflections as the blank is rotated about psi, as described in the text. The simulation was performed with a 20° glancing angle, as in the prototype. The center of the pattern is at that point where the incident x-ray beam, if reflected from the surface of the blank, would strike the detector plane. The plane of the pattern is normal to this imaginary reflection.

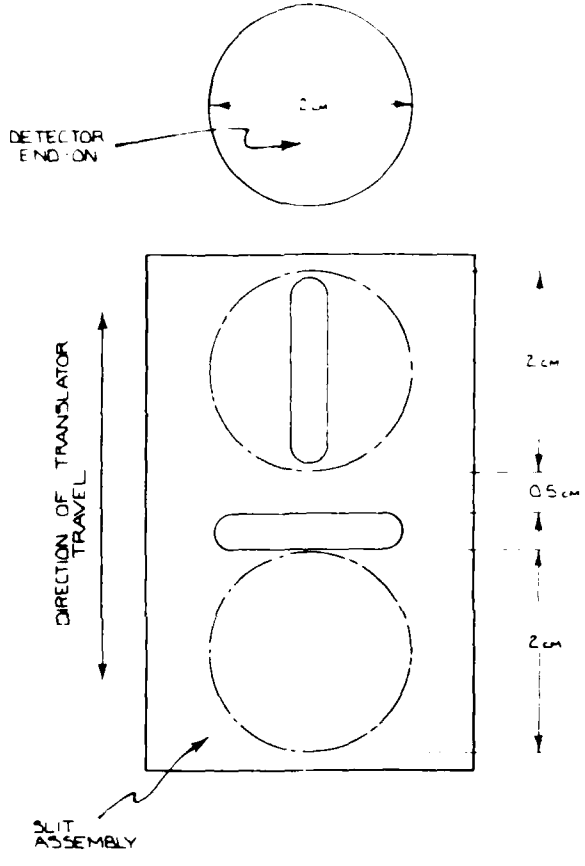


Figure 11.

Schematic of the slits used to map the intensity profiles within the x-ray detectors. The horizontal slit is used to determine the vertical coordinate; the vertical slit in conjunction with the rotator is used to gather the data needed for computing the horizontal coordinate.



Figure 12.

Photograph of the prototype Laue diffractometer, before alignment. The rotating stage is at the instrument center. The three arches support the x-ray detectors. The laser system for measurement of the orientation of the face is not shown.

Proc. 35th Ann Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

FORCE- AND ACCELERATION- FREQUENCY EFFECTS IN GROOVED AND RING- SUPPORTED RESONATORS

Mitsuo Nakazawa*
Princeton University
Princeton, NJ 08540

Theodore Lukaszek & Arthur Ballato
US Army Electronics Technology & Devices Laboratory
USAERADCOM, Fort Monmouth, NJ 07703

Abstract

We have investigated certain novel plate resonator configurations that are referred to as grooved- and ring- supported resonators. These are relatively inexpensive to fabricate because of their simplicity, are amenable to high frequency operation with certain parameters not achievable with surface wave resonators, and give good promise of providing stable frequencies irrespective of the static boundary forces and the shock and vibration forces of accelerations encountered in field use.

In this paper we report preliminary results of our studies made on AT-cut resonators of various designs. The rationale for choosing these designs is given, along with details of the measurement apparatus. Results are given of mode spectra, force-frequency, and acceleration-frequency coefficients.

Key Words: Quartz, crystal resonators, frequency control, frequency stability, acoustic waves, bulk acoustic waves, surface acoustic waves, filters, signal processing, frequency synthesizers, wrist watch crystals.

Introduction

Precision crystal resonators are required to provide secure, nonjammable, and reliable command, control, communication, and intelligence (C³I) voice and data links that remain operable in very severe environments. Insensitivity to static and dynamic forces is a primary consideration in the design of such resonators, as well as such commercial applications as wrist watches. The problem of external forces applied to resonators has received increasing attention during the past few years, with good results. Some of the solutions proposed involve considerable cost penalties, manufacturing difficulties or both. Furthermore, certain desirable future features such as miniaturization and direct production of UHF frequencies do not appear to be readily possible with these solutions.

We have, therefore, investigated certain novel plate resonator configurations that are referred to as grooved- and ring- supported resonators. These are developments of earlier work by the authors and others that are relatively inexpensive to fabricate because of their simplicity, are amenable to high frequency operation with certain parameters not achievable with surface wave resonators, and give good promise of providing stable frequencies irrespective of the static boundary forces and the shock and vibration forces of acceleration encountered in field use.

The resonator is made from a single piece of quartz or other material, and the ring that supports the central vibrating portion provides the means both of stiffening the inner resonating member against flexure and of carrying most of the mounting stress. It is experimentally shown that both the static force-frequency and the dynamic acceleration-frequency effects are reduced in magnitude for these designs.

Frequency Shifts

A great deal of effort has gone into the design of crystal resonators[1-20] and into characterizing the types of effects that produce frequency shifts and changes in vibrator behavior.[21-81] Some of the causes of frequency shifts and changes in behavior of resonators are:

(1) Static and dynamic thermal-frequency effects[20,25,30-35,71]
(2) Static force-frequency effect[36-39,41,45,46,48,50,71]
(3) Dynamic acceleration-frequency effect[21,40,44,44,47,75-79]
(4) Drive level- frequency effect[42,54]
(5) Mode coupling/activity dips[51,52,53,55,59,60]
(6) Intermodulation effect[62]
(7) Film stress-frequency effect[63,81]
(8) Aging, from contamination, diffusion, and mounting stress relief, etc.[56,57,64,65,67,69]
(9) Other nonlinear and linear processes[70,72,73,74,80]

The various effects and their interrelationships are figuratively represented in the drawing of Fig. 1. Figure 2 shows a plot of the idealized frequency-time behavior of a quartz oscillator. The

* On leave from Shinshu University, Nagano 380, Japan.

disturbances stem largely from mechanical and thermal effects. In Fig. 3 is given a chronology of the most important advances in crystal technology since 1945; most of these are directed toward a lessening of the sensitivities indicated in Fig. 2.

The Structures of Ring-Supported & Grooved Resonators

Particular examples of both ring-supported and grooved crystal resonators are shown in Fig. 4. We shall report on results of experiments made with plano-convex ring-supported, and single-slot grooved resonators in this paper. The concavo-convex ring-supported resonator shown in the figure is useful for applications having accelerations directed toward the centers of curvature. Energy-trapping is obtained by making the radius of the concave surface larger than that of the convex surface. Figure 5 supplies the rationale for using ring-supported (r-s) resonators for out-of-plane acceleration fields; frequency shift is proportional to deflection of the plate at the center, and is less for the r-s resonators than for the conventional variety. An analogous case of stress relief occurs with the use of flying buttresses (Fig. 6). The thrust of the roof is carried outward and downward by these ancillary supporting structures.

Figure 7 defines the relevant parameters describing plano-convex resonators, and tabulates measured values of some of the resonators used in our studies. Figure 8 defines and tabulates the bi-convex parameters.

Scanning electron microscope (SEM) photographs of plano-convex resonator No. 1 are shown in Fig. 9 at magnifications of 28X (Top) and 190X (bottom). These show, respectively, the ring support and convex floor regions of the plate, and a more detailed view of the side wall leading out to the convex floor.

Figure 10 defines the parameters relevant to single-slot grooved resonators, and gives measured values for two units. In Fig. 11, parameters and measurements of double-slot grooved resonators are given. SEM photos of resonator T-1 at magnifications of 14X (top) and 44X (bottom) are seen in Fig. 12; the groove was cut ultrasonically.

Frequency Spectra

Energy-trapping considerations[2,3,4,5] will dictate in large measure the structure of the crystal mode spectra at different harmonics. In the designs utilized in our study the units were not optimized for fundamental operation; in fact, at the fundamental harmonic the vibratory motion was largely "untrapped" and spread into the ring and support system, leading to a poor mode spectrum and poor Q. The spectra at the third and fifth harmonics indicate more nearly optimal degrees of trapping (higher Q values and cleaner structure). Figures 13 and 14 show examples of r-s and grooved resonators, respectively.

Force- Frequency Effects

The traditional form of the force-frequency curve for AT-cut plates of conventional plano-plano design is shown in Fig. 15.[6,7,8] The force azimuth angle ψ is measured from the X axis; the ordinate scale given is not the normalized parameter $K_f(\psi)$ of Ratajski,[9] but an earlier form.[7] The reason for this more primitive form is that comparison is now to be made with r-s and grooved units having a variety of geometrical parameters, and suitable definitions of analagous $K_f(\psi)$ scales are not completely without ambiguity.

The force-frequency measurement apparatus is seen in Figures 16 and 17. Experimental results for B-3 are given in Fig. 18. It is seen that the maximum shift is about one third that of Fig. 15; the unexplained but consistently measured drop to negative values for $|\psi| \sim 25°$ is a feature not encountered in the other units, and likewise not seen with conventional resonators. Figure 19 compares the third and fifth harmonics of unit B-4; the reduction in amplitude with respect to Fig. 15 is again found. Figure 20 compares the grooved resonators T-1 and T-2 at the third harmonic; both have a three-fold improvement over the conventional units.

Mode spectrographs at fundamental and third harmonics of a sample reference crystal of conventional design is shown in Fig. 21. The main mode at each harmonic was measured as function of drive level, and no dependency was noted. However, when the first unwanted mode of the fundamental harmonic was measured at various drive levels, a dependence was noted. Figure 22 shows the result for voltages of 0.03 and 0.003 across the crystal. As the excitation is decreased, the zero points are seen to move away from the X axis.

Acceleration- Frequency Effects

The centrifugal accelerator pictured in Fig.23, and shown in schematic in Fig. 24, was used to subject the resonators to constant accelerations. The apparatus, which is shown with ancillary measurement equipment in Fig. 25, is capable of subjecting units to 1000 g of acceleration. The rotor arms are fitted with jigs that receive the resonators mounted in HC-6 holders. The acceleration effect is measured by a phase shift technique.

Sample results are given in Fig. 26 and Fig. 27. One important aspect of the curves shown is the apparent nonlinearity of the results with increasing g values; this would indicate a γ_o value which is a function of the g level used, in distinction to the constant value used to date.

Equivalent Network

The bisected transmission line equivalent network for a conventional thickness mode resonator is given in Fig. 28.[10] In Fig. 29 is given the corresponding network for a grooved resonator.

Acknowledgement

The authors wish to thank R. Tilton, G. Le"eune, J. Gualtieri, E. Hakim, and R. Sartore of ETSD, Fort Monmouth; Roger Ward of Colorado Crystal Co., R. Besson of ENSMM, Besançon, and N. Oura of Tokyo Institute of Technology.

Referencesª

1. T. Shinada, "Holder of quartz crystal resonator," Japanese patent 42-10494, filed 28 Dec. 1964, issued 9 Jun. 1967.

2. T. Hiruta, "Holder of highly stable quartz crystal," Japanese patent 44-1496, filed 6 Apr. 1965, issued 23 Jan. 1969.

3. T. Hiruta, "Holder of quartz crystal resonator vibrating in thickness-shear modes," Japanese patent 44-17194, filed 21 Jun. 1965, issued 29 July 1969.

4. D.L. Hammond and L.S. Cutler, "Crystal resonators," U.S. Patent 3 339 091, patented Aug. 29, 1967.

5. L.S. Cutler and D.L. Hammond, "Crystal resonators", Reissue 26,707 of U.S. Patent 3 339 091, reissued Nov. 4, 1969.

6. T.J. Lukaszek, "Mode control and related studies of VHF quartz filter crystals," IEEE Trans. Sonics Ultrason., vol. SU-18, pp. 238-246, Oct. 1971.

7. G.K. Guttwein, A. Ballato, and T.J. Lukaszek, "VHF-UHF piezoelectric resonators," U.S. Patent 3 694 677, patented Sept. 26, 1972.

8. M. Nakazawa, "Analysis of vibrational mode in concaved quartz crystal plate with probe method," Paper of Tech. Group on EA-76-43, Inst. Electron Commun. Eng. Japan, pp. 23-28, 1976.

9. M. Nakazawa and S. Wakui, "On the grooved AT-cut quartz crystal," submitted to J. Inst. Electr. Comm. Eng. Jpn.

10. R.Besson, "Résonateur à quartz à électrodes non adhérentes au cristal," French patent 76 010 35, issued 19 Jan. 1976, and extended by patent 76 162 89, issued May 1976, US Patent 4 135 108, issued 16 Jan. 1979, and pending extensions.

11. R. Besson, "Résonateur piézoélectrique à suspension améliorée," French patent 77 173 09, issued 7 Jun. 1977; US patent application S/N 920 675.

12. R. Besson, "Résonateur piézoélectrique à cristal autosuspendu," French patent 78 022 61, issued 27 Jan. 1978, and pending extensions.

13. R. Besson, "Bi-résonateur piézoélectrique," French patent 78 28 728, issued Sep. 1978, and pending extensions.

14. R. Besson & P. Maitre, "Résonateur piezoélectrique à tiroir," French patent 78 35 631, issued Dec. 1978, and pending extensions.

15. R. Besson & J.P. Valentin, "Oscillateur haute frequence autothermostaté," French patent 79 18 553, issued Jul. 1979.

16. M.Berté, "Acoustic bulk wave resonators and filters operating in the fundamental mode at frequencies greater than 100 MHz," Proc. 31st AFCS, June 1977, pp. 122-125.

17. M. Berté, & P. Hartemann, "Quartz resonators at fundamental frequencies greater than 100 MHz," IEEE Ultrasonics Symposium, Sep 1978, pp. 148-151.

18. C. Pegeot, "UHF oscillator using SC cut quartz crystal, with low noise performances and high long term stability," Proc. 34th AFCS, May 1980 pp. 233-236.

19. R.J. Besson & U.R. Peier, "Further advances on B.V.A. quartz resonators," Proc. 34th AFCS, May 1980, pp. 175-182.

20. M. Nakazawa, "Improving frequency-temperature characteristics of grooved AT-cut plates," Proc. 34th AFCS. May 1980, pp. 152-159.

21. A. Ballato, T.J. Lukaszek, & G.J. Iafrate, "Subtle effects in high stability vibrators," Proc. 34th AFCS, May 1980, pp. 431-444.

22. L. Bidart & J. Chauvin, "Direct frequency crystal oscillator," these proceedings.

23. I. Koga, "Thickness vibrations of piezoelectric oscillating crystal," Physics, vol. 3, 1932, pp. 70-80.

24. R.D. Mindlin, "[illegible] and flexural vibrations of [illegible] Appl. Phys., vol. 22, 1951, pp. [illegible]

25. R. Bechmann, "[illegible] der Eigenschwingungen [illegible] Quarzplatten [illegible] Elektroak., vol. [illegible]

26. R. Bechmann, A. Ball[illegible], "Frequency-temperature characteristics of quartz resonators derived from the temperature behavior of the elastic constants," Proc. 16th AFCS, April 1962, pp. 77-109.

27. W.G. Cady, Piezoelectricity, New York: McGraw-Hill, 1946; New York: Dover: 1964, pp.1-199.

28. R.A. Heising, ed., Quartz Crystals for Electrical Circuits, Van Nostrand, NY, 1946, pp. 1-33.

29. I. Koga, M. Aruga, & Y. Yoshinaka, "Theory of plane elastic waves in a piezoelectric crystalline medium and determination of elastic and piezoelectric constants of quartz," Phys. Rev.,

vol. 109, 1958, pp. 1467-1473.

30. W.P. Mason, "Low temperature coefficient quartz crystals," Bell Syst. Tech. J., vol. 19, 1940, pp. 74-93.

31. A. Ballato & J.R. Vig, "Static and dynamic frequency-temperature behavior of singly and doubly rotated, oven-controlled quartz resonators," Proc. 32nd AFCS, May-June 1978, pp. 180-188.

32. A. Ballato, "Frequency-temperature-load capacitance behavior of resonators for TCXO application," IEEE Trans. Sonics Ultrason., vol. SU-25, July 1978, pp. 185-191.

33. M. Nakazawa, Y. Nakamura, & Miyashita, "Frequency-temperature characteristics of quartz crystal flexure bars and quartz crystal tuning forks," IEEE Trans. Sonics Ultrason., vol. SU-25, Sept. 1978, pp. 369-376.

34. T. Kawashima, H. Sato, & H. Ochiai, "New frequency temperature characteristics of miniaturized SC-cut quartz resonators," Proc. 34th AFCS, May 1980, pp. 131-139.

35. T. Bowden and A. Theobald, "Dynamic Thermal Sensitivity of SAW quartz oscillators," Proc. IEEE Ultrasonics Symp., November 1980, pp. 264-267.

36. A. Ballato, "Effects of initial stress on quartz plates vibrating in thickness modes," Proc. 34th AFCS, May-June 1980, pp. 89-114.

37. E.G. Gerber & M.H. Miles, "Reduction of the frequency-temperature shift of piezoelectric resonators by mechanical stress," Proc. IRE, vol. 49, 1961, pp. 1650-1654.

38. R.W. Meyer and E.W. Blair, "Stress dependence of the frequency of quartz plates," Proc. IEEE, vol. 55, pp. 565-566, Apr. 1967.

39. J.M. Matajski, "Force-frequency coefficient of singly rotated vibrating quartz crystals," IBM J. Res. Dev., vol. 12, pp. 92-99, Jan 1968.

40. M. Valdois, J. Besson, and J.J. Gagnepain, Influence of environment conditions on a quartz resonator, Proc. 28th AFCS, May 1974, pp. 19-32.

41. P.C.Y. Lee, S. Wang, & Y.Markenscoff, "Effects of initial bending on the resonance frequencies of crystal plates," Proc. 28th AFCS, May 1974, pp. 14-18.

42. H.F. Tiersten, "Analysis of nonlinear resonance in thickness-shear and trapped-energy resonators," J. Acoust. Soc. Am., vol. 59, 1976, pp. 966-978.

43. P.C.Y. Lee & K.-M. Wu, "Effects of acceleration on the resonance frequencies of crystal plates," Proc. 30th AFCS, June 1976, pp. 1-7.

44. P.C.Y. Lee, K.M. Wu, and Y.S. Wang, "Effects of acceleration on the resonance frequencies of crystal plates," J. Acoust. Soc. Amer., vol.63, pp. 1039-1047, Apr. 1978.

45. A. Ballato, E.P. EerNisse, & T.J. Lukaszek, "The force-frequency effect in doubly rotated quartz resonators," Proc. 31st AFCS, June 1977, pp. 8-16.

46. A. Ballato, T. Lukaszek, & E.P. EerNisse, Force-frequency and other effects in doubly rotated vibrators," Technical Report ECOM-4536, US Army Electronics Command, Ft. Monmouth, NJ 07703, Sep. 1977, 45 pp.

47. T. Lukaszek & A. Ballato, "Resonators for Acceleration Environments," Technical Report DELET-TR-79-10, US Army Electronics R&D Command, Fort Monmouth, NJ 07703, June 1979, 33pp.

48. P.C.Y. Lee & K. -M. Wu, "Nonlinear effect of initial stresses in doubly rotated crystal resonator plates," Proc. 34th AFCS, May 1980, pp. 403-411.

49. D.L. Hammond, "Precision quartz resonators," in Proc. 15th AFCS, May-June 1961, pp. 125-138.

50. C.R. Mingins, L.C. Barcus, & R.W. Perry, "Effects of external forces on the frequency of vibrating crystal plates," Proc. 16th AFCS, April 1962, pp. 47-76.

51. E. Hafner, "Some phenomena in VHF Crystal Units," Proc. 10th AFCS, May 1956, pp. 183-189.

52. W. Shockley, D.R. Curran, and D.J. Koneval, "Trapped-energy modes in quartz filter crystals," J. Acoust. Soc. Amer., vol. 41, no. 4, Part 2, 1967, pp. 981-993.

53. H. Fukuyo, H. Yoshie, & M. Nakazawa, "The unwanted responses of crystal oscillator controlled by AT-cut plate," Proc. 21st AFCS, April 1967, pp. 402-419.

54. D.L. Hammond, C.A. Adams, & L. Cutler,"Precision crystal units," Proc. 17th AFCS, May 1963, pp. 215-232.

55. I. Koga, H. Fukuyo, & J.E. Rhodes, "Modes of vibration of quartz crystal resonators investigated by means of the probe method," Proc. 13th AFCS, May 1959, pp. 54-70.

56. Y. Oomura, "Miniaturized circular disk AT-cut crystal vibrator," Proc. 30th AFCS, June 1976, pp. 202-208.

57. E. Hafner, "Crystal resonators," IEEE Trans. Sonics Ultrason., vol. SU-21, October 1974, pp. 220-237.

58. A. Ballato, "Doubly rotated thickness mode plate vibrators," in Physical Acoustics: Principles and Methods, W.P. Mason and R.N. Thurston, Eds., vol. 13. New York: Academic,

1977, ch. 5, pp. 115-181.

59. A.F.B. Wood & A. Seed, "Activity dips in AT-cut crystals," Proc. 21st AFCS, April 1967, pp. 420-435.

60. A. Ballato and R. Tilton, "Electronic activity dip measurement," IEEE Trans. Instrum. Meas., vol. IM-27, March 1978, pp. 59-65.

61. M. Nakazawa, "Specific directions of plane elastic waves in thin quartz crystal plates," Int. J. Electronics, vol. 48, no. 3, 1980, pp. 275-282.

62. H.F. Tiersten, "Analysis of intermodulation in rotated Y-cut quartz thickness-shear resonators," Proc. 28th AFCS, May 1974, pp. 1-4.

63. E.P. EerNisse, "Quartz resonator frequency shifts arising from electrode stress," in Proc. 29th AFCS, May, 1975, pp. 1-4.

64. T. Lee, H. Wasshausen, C. Cook, M. Katz, & E. Hafner, "Surface preparation and characterization techniques for quartz resonators," Proc. 27th AFCS, June 1973, pp. 98-112.

65. H. Fukuyo & N. Oura, "Surface layer of polished crystal plates," Proc. 30th AFCS, June 1976, pp. 254-258.

66. T. Lukaszek & A. Ballato, "What SAW can learn from BAW: implications for future frequency control, selection, and signal processing," Proc. 1980 Ultrasonics Symposium, pp. 173-183.

67. N. Oura, N. Kuramochi, I. Matsuda, H. Kono, & H. Fukuyo, "Effect of the worked layer in quartz crystal plates on their frequency stabilities," IEEE Trans. Instrum. Meas., vol. IM-30, June 1981, pp. 139-143.

68. N. Kuramochi, I. Matsuda, N. Oura, & H. Fukuyo, "Analysis of the temperature dependence of ^{198}Hg lamp profiles," J. Opt. Soc. Amer., vol. 70, no. 12, 1980, pp. 1504-1507.

69. R.L. Filler & J.R. Vig, "The effect of bonding on the frequency vs. temperature characteristics of AT-cut resonators," Proc. 30th AFCS, June 1976, pp. 264-268.

70. M. Planat, D. Hauden, J. Groslambert, & J.J. Gagnepain, "Nonlinear propagation of surface acoustic waves on quartz," Proc. 34th AFCS, May 1980, pp. 255-261.

71. D. Hauden, S. Rousseau, & J.J. Gagnepain, "Sensitivities of SAW oscillators to temperature, forces and pressure: application to sensors," Proc. 34th AFCS, May 1980, pp. 312-319.

72. M. Planat, G. Theobald, & J.J. Gagnepain, "Propagation non linéaire d'onde élastiques dans un solide anisotrope. I. Ondes de volume," L'Onde Electrique, vol. 60, no. 8-9, August-September 1980, pp. 33-40.

73. M. Planat, G. Theobald, & J.J. Gagnepain, "Propagation non linéaire d'ondes élastiques dans un solide anisotrope. II. Ondes de surface," L'Onde Électrique, vol. 60, no. 11, November 1980, pp. 61-67.

74. J.J. Gagnepain, "Nonlinear phenomena in bulk and surface acoustic wave resonators," V Konferencja Piezoelektronici, Varsovie, Polska, November 1980, 8 pp.

75. P.C.Y. Lee & C.S. Lam, "Nonlinear effect of transverse acceleration on the resonances of doubly rotated, rectangular crystal plates," these proceedings.

76. V. Rosati, "Reduction of the effects of vibration on SC cut quartz crystal oscillators," these proceedings.

77. R.L. Filler & J.R. Vig, "The acceleration and warmup characteristics of four-point-mounted SC- and AT-cut resonators," these proceedings.

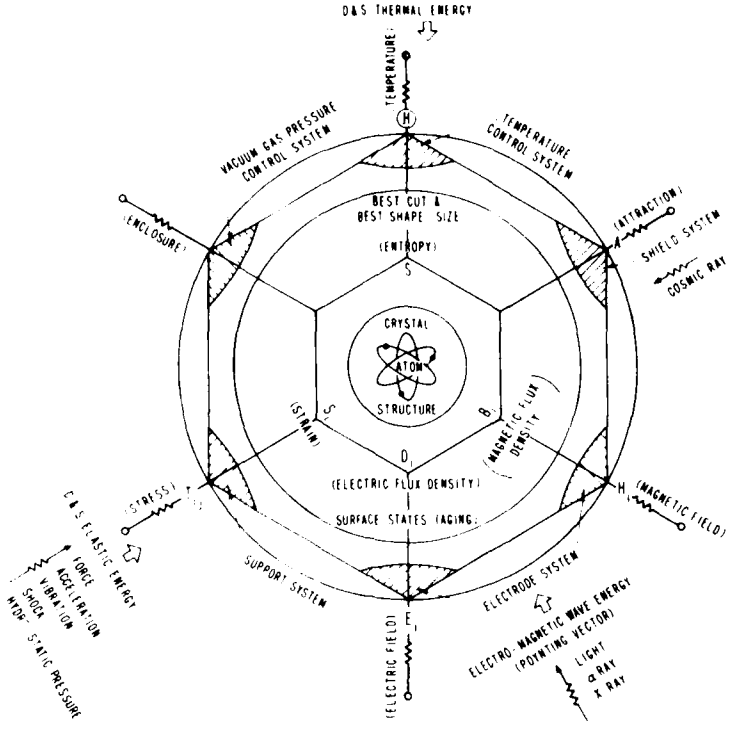
78. B. Goldfrank, J. Ho, & A.W. Warner, Update of SC cut crystal resonator technology," these proceedings.

79. R.L. Filler, "The effect of vibration on frequency standards and clocks," these proceedings.

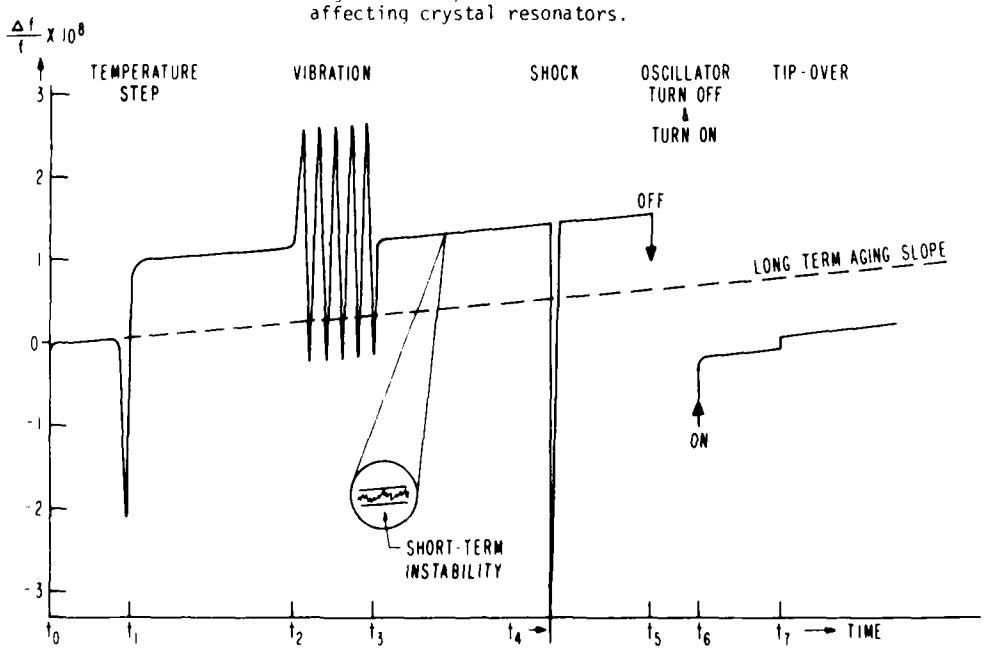
80. J.J. Gagnepain, "Nonlinear properties of quartz crystal and quartz resonators: a review," these proceedings.

81. A. Ballato, E.P. EerNisse, and T.J. Lukaszek, "Experimental verification of stress compensation in the SC cut," Proc. IEEE Ultrasonics Symposium, September 1978, pp. 144-147.

*A number of the references were presented at the Annual Frequency Control Symposia, US Army Electronics R&D Command, Fort Monmouth,NJ 07703. They are cited here as AFCS for brevity.



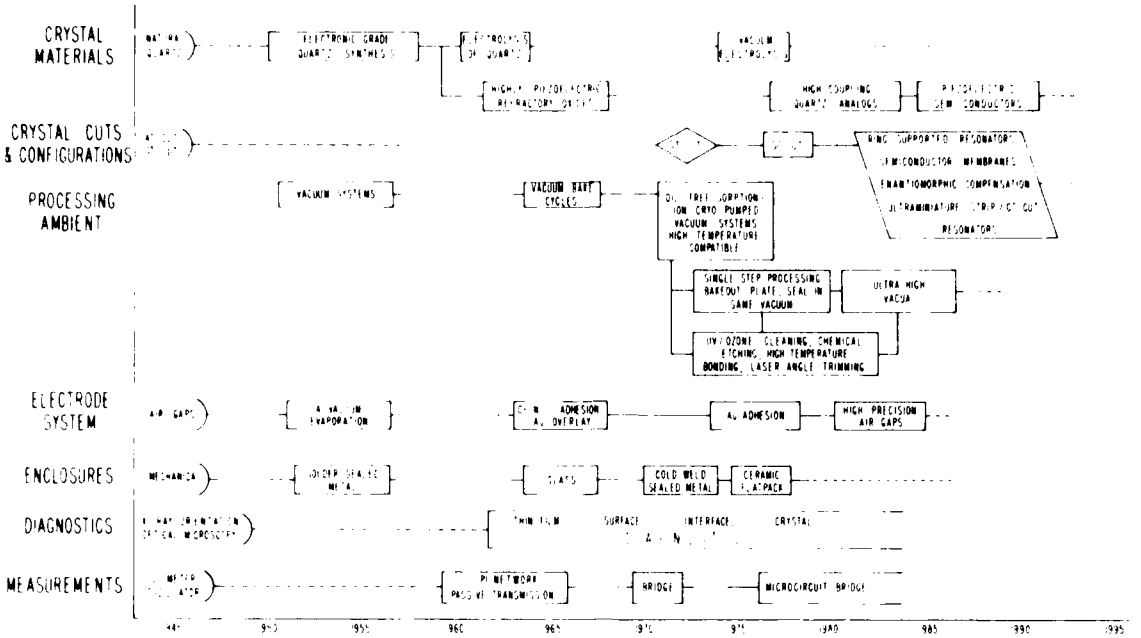
1. Figurative representation of influences affecting crystal resonators.



IDEALIZED FREQUENCY-TIME BEHAVIOR OF A QUARTZ OSCILLATOR

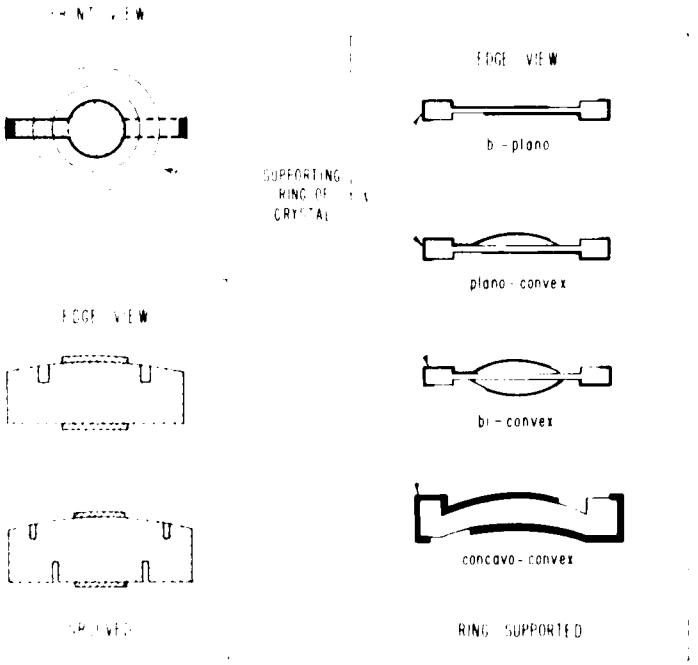
2. Idealized frequency-time behavior of a quartz oscillator.

PROGRESSION IN CRYSTAL TECHNOLOGY

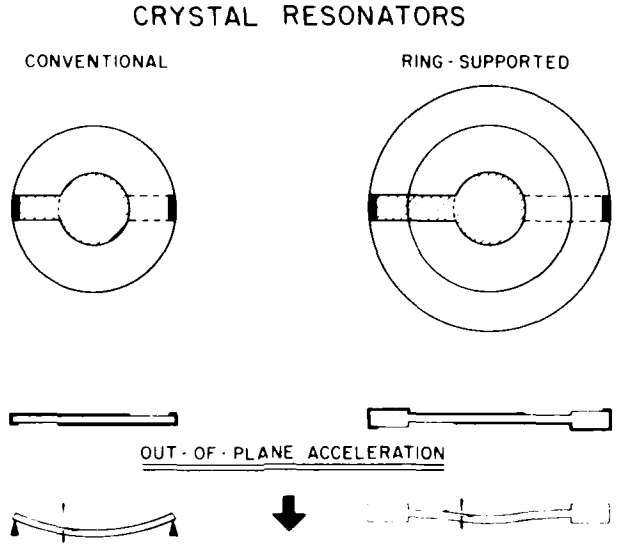


3. Progression in crystal technology.

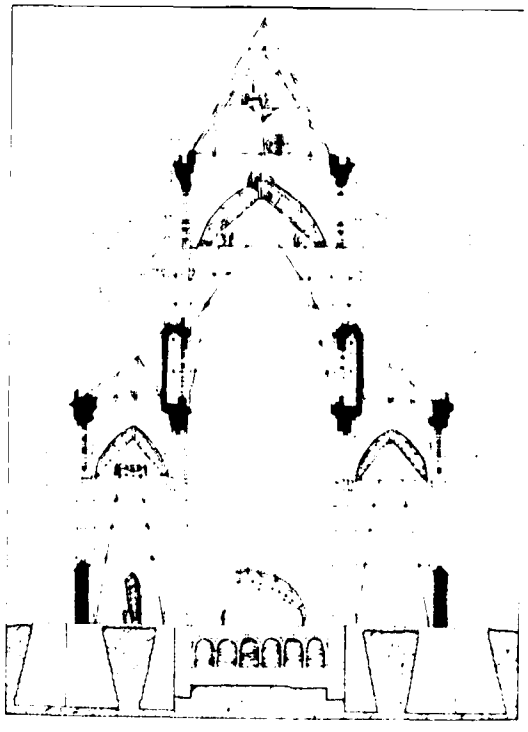
RING SUPPORTED AND GROOVED CRYSTAL RESONATORS



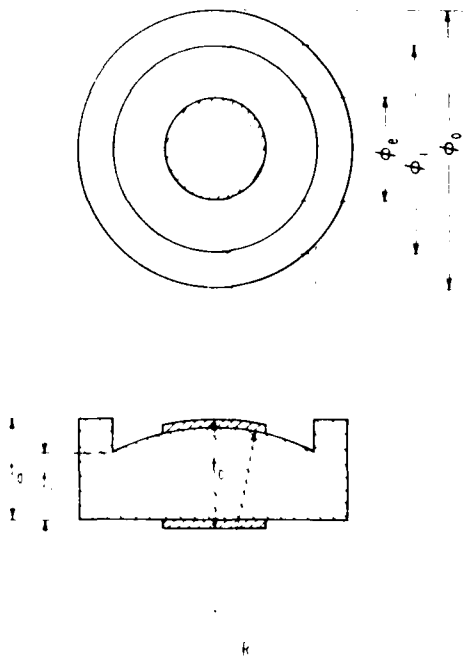
4. Ring-supported and grooved crystal resonators.



5. Conventional and ring-supported crystal resonators.

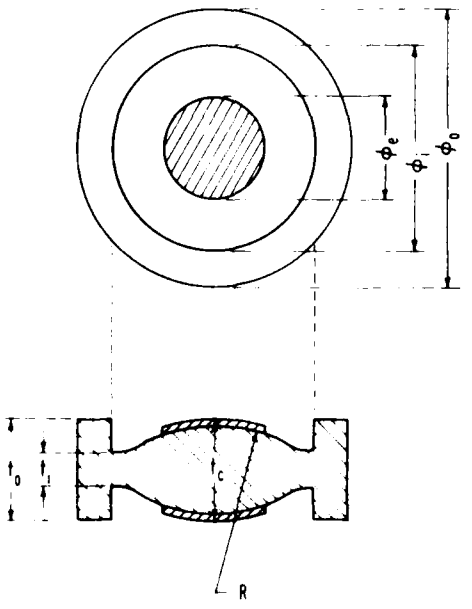


6. Cathedral cross-section showing flying buttresses.



TYPE	UNIT [mm] ϕ_e	ϕ_i	ϕ_0	t_c	t_i	t_0	t_c/t_i	R
REF	4 496		12 929	0 184	0 184	0 184	1 000	∞
No 1	4 140	6 274	8 001	0 484	0 412	0 568	1 175	68 5
No B-3	4 013	5 842	8 001	0 776	0 740	1 005	1 049	120
No B-4	4 064	6 020	7 950	0 700	0 662	1 000	1 057	120
No B-6	4 572	5 512	7 950	0 681	0 618	1 003	1 103	≃ 60
No B-7	4 343	6 756	8 000	0 670	0 575	1 000	1 166	≃ 60
No B-8	3 759	5 182	7 620	0 665	0 609	1 002	1 092	≃ 60
No B-9	3 581	5 613	7 976	0 676	0 610	1 006	108	≃ 60
No B-10	3 607	5 512	7 976	0 686	0 623	1 011	1 102	≃ 60
No B-11	3 429	5 486	8 026	0 711	0 648	1 010	1 097	≃ 60
No B-12	4 204	5 639	8 001	0 699	0 633	1 007	1 105	≃ 60
No B-13	3 835	5 537	7 950	0 688	0 624	1 000	1 102	≃ 60

7. Ring-supported plano-convex resonator geometry and measured parameters.



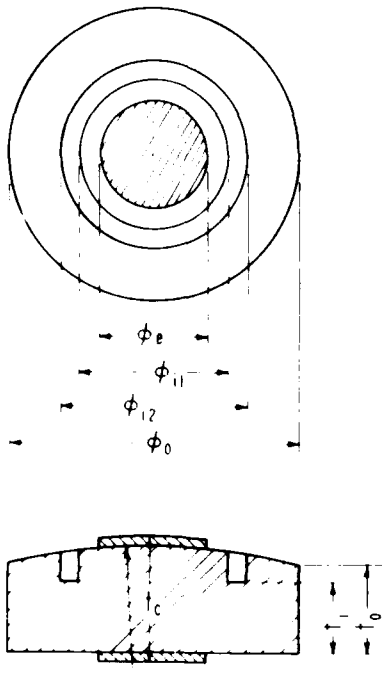
TYPE	UNIT [mm] ϕ_e	ϕ_i	ϕ_0	t_c	t_i	t_0	t_c/t_i	R
No Bi-1	3 480	5.575	7 938	0 701	0 571	1 006	1 227	≃60
No Bi-2	3 327	5 663	7 976	0 707	0 573	1 005	1 234	≃ 60

8. Ring-supported bi-convex resonator geometry and measured parameters.



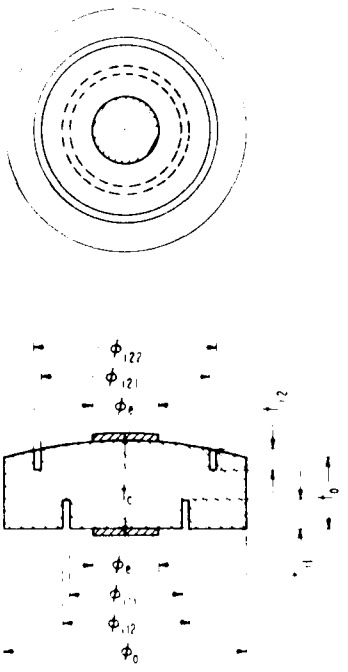


9. SEM photos of ring-supported resonator.



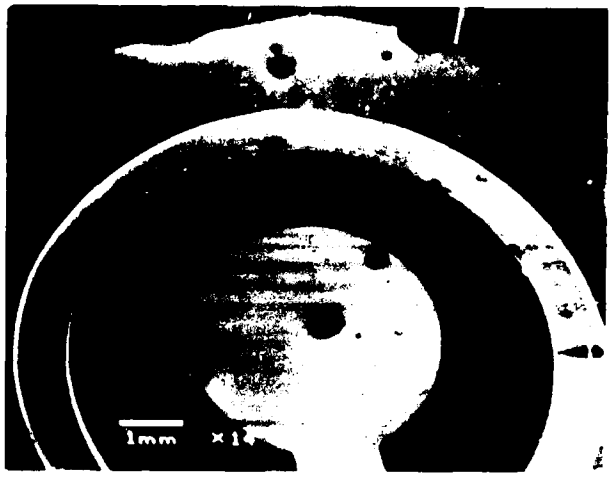
TYPE	UNIT [mm]								
	ϕ_e	ϕ_{i1}	ϕ_{i2}	ϕ_o	t_c	t_i	t_o	t_c/t_o	R
No T-1	4 115	7 823	9 576	13 386	0.514	0.034	0 209	2.459	≃ 74
No T-2	4 953	7 925	9 652	13 869	0 515	0 127	0 260	1 981	≃ 94

10. Single-slot grooved resonator geometry and measured parameters.



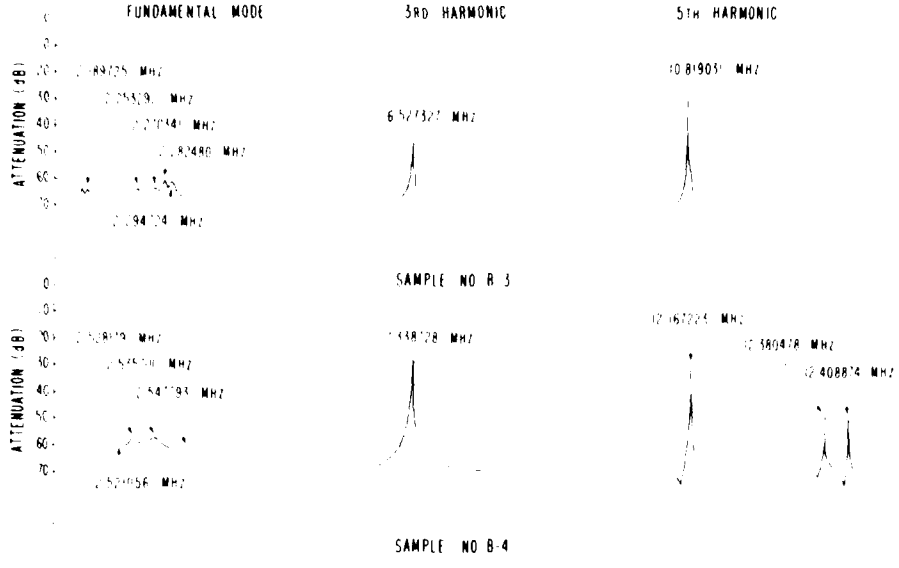
TYPE	UNIT [mm]											
	ϕ_e	ϕ_{i11}	ϕ_{i12}	ϕ_{i21}	ϕ_{i22}	ϕ_o	t_{i1}	t_{i2}	t_o	t_c	t_c/t_o	R
No T-3	5 029	8 026	8 865	11 481	12 421	15 875	0 114	0 152	0 194	0 458	2 361	≃ 120
No T-4	5 704	8 052	8 839	11 455	12 421	14 981	0 108	0 143	0 218	0 412	1 890	≃ 145
No T-5	5 029	8 026	8 839	11 455	12 420	14 985	0 127	0 127	0 233	0 432	1 854	≃ 140
No T-6	4 216	7 493	8 331	10 947	11 913	13 158	0 142	0 142	0 254	0 348	1 370	≃ 230

11. Double-slot grooved resonator geometry and measured parameters.

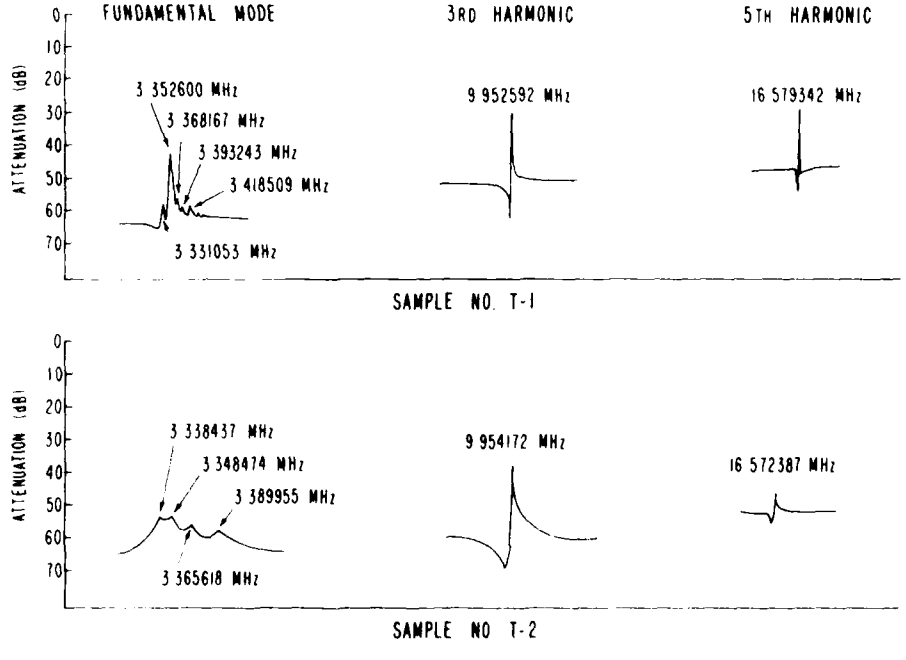




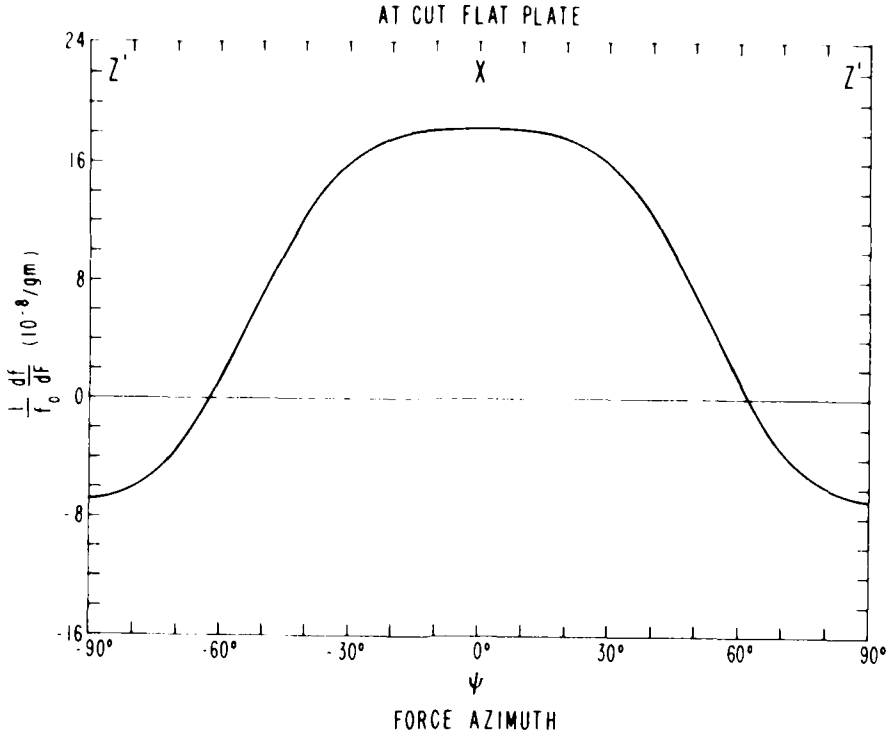
12. SEM photos of grooved resonator.



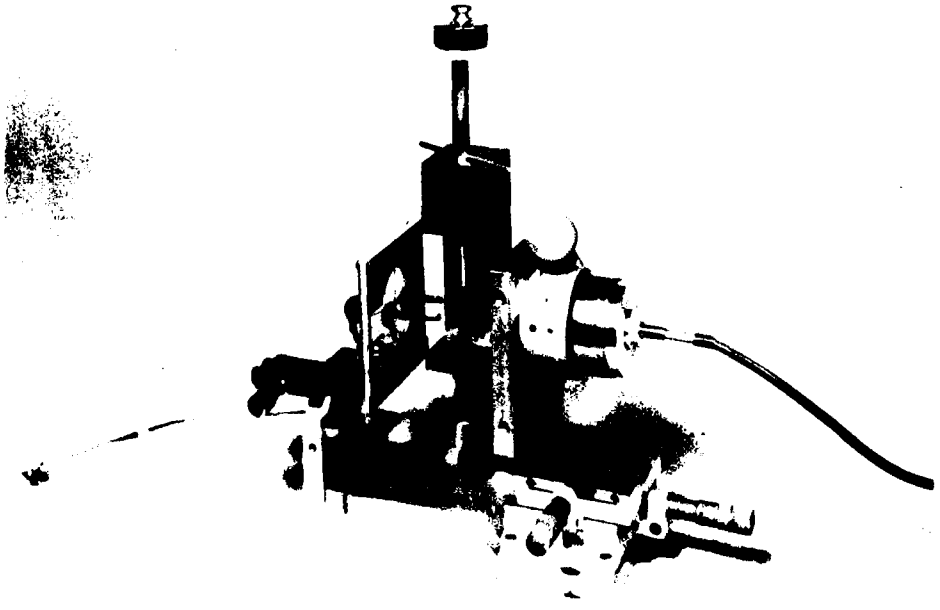
13. Spectrographs of samples B-3 and B-4 at first, third, and fifth harmonics.



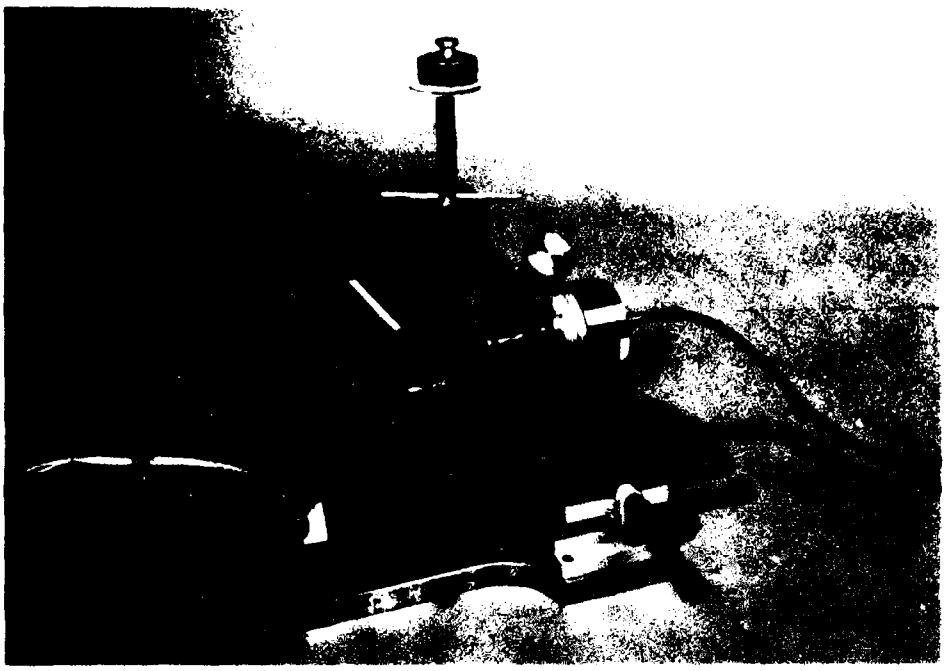
14. Spectrographs of samples T-1 and T-2 at first, third, and fifth harmonics.



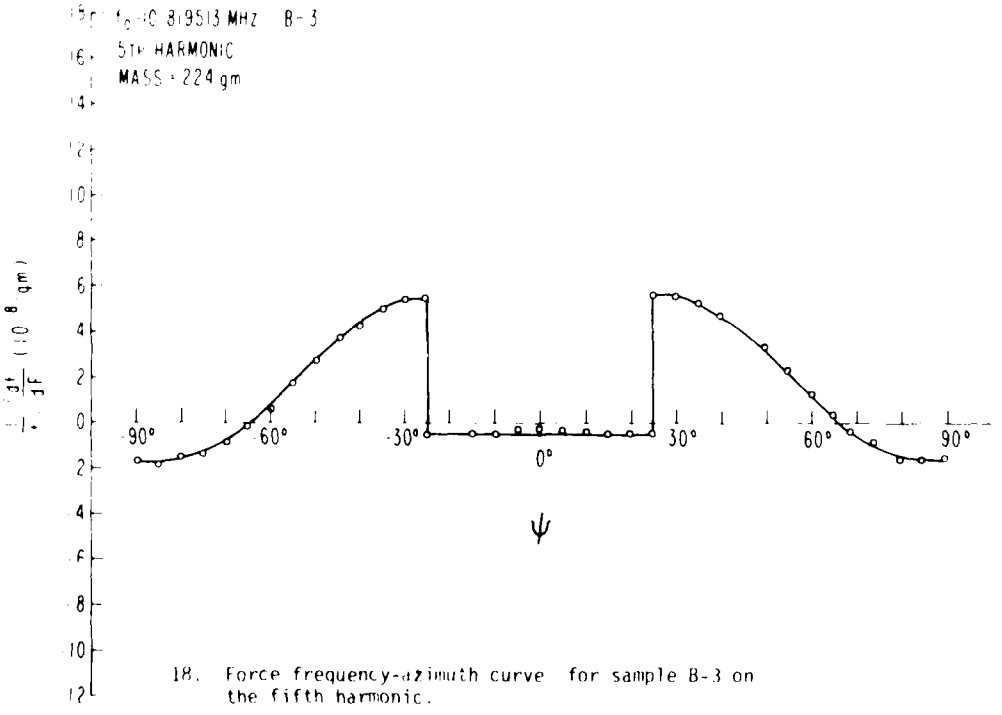
15. Force-frequency-azimuth curve for conventional AT-cut plate.



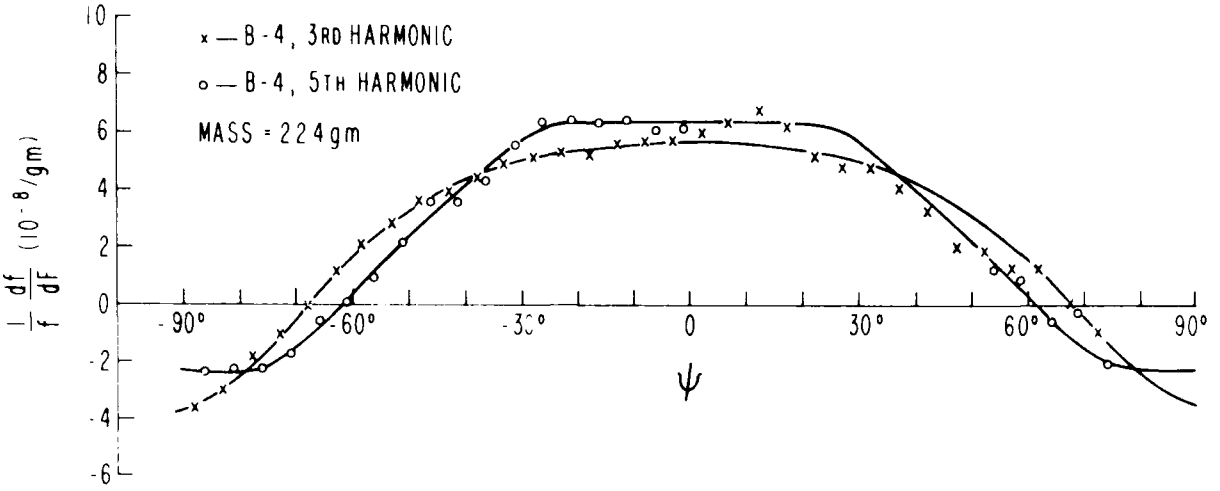
16. Force-frequency effect apparatus-- view from right.



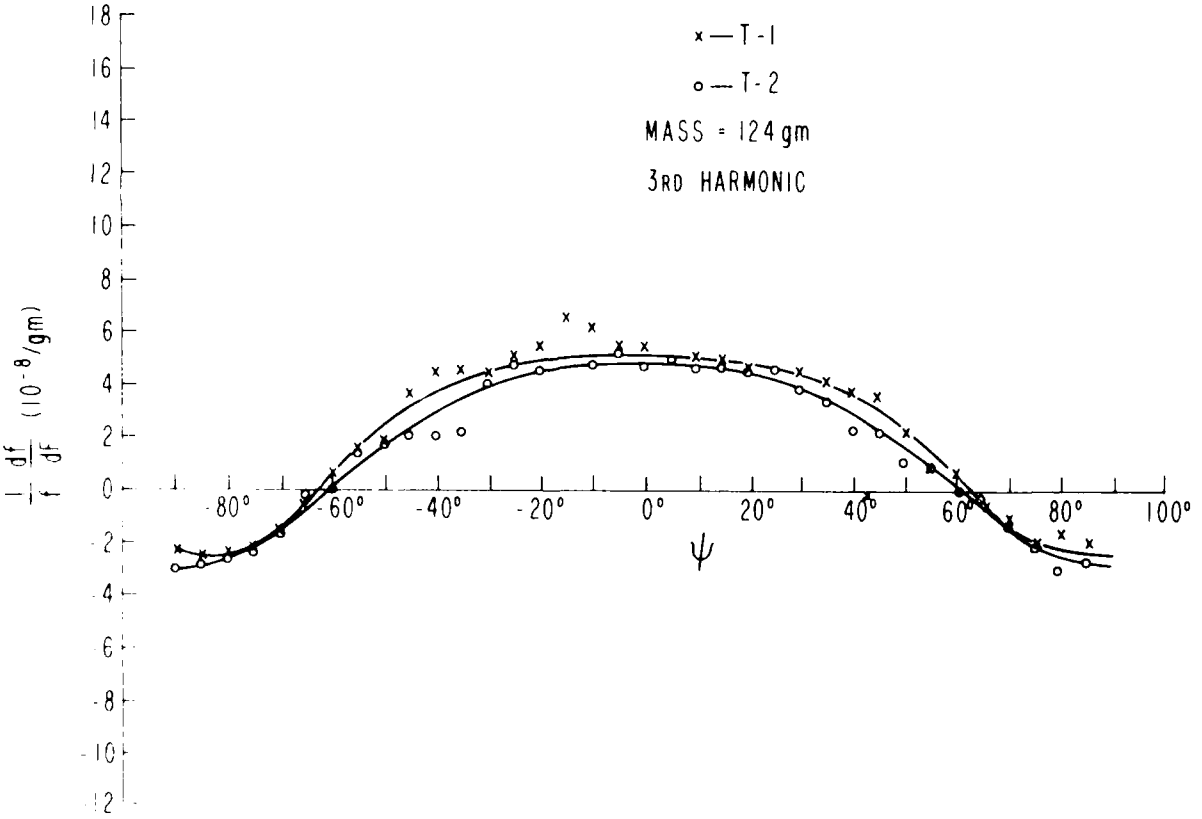
17. Force-frequency effect apparatus-- view from left.



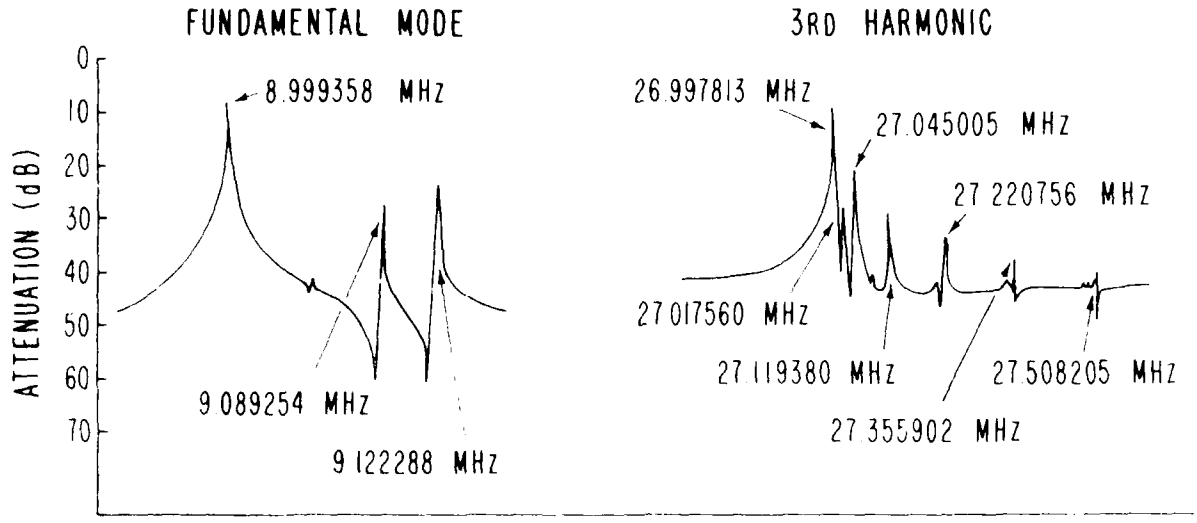
18. Force frequency-azimuth curve for sample B-3 on the fifth harmonic.



19. Force-frequency-azimuth curves for sample B-4 on the third and fifth harmonics.

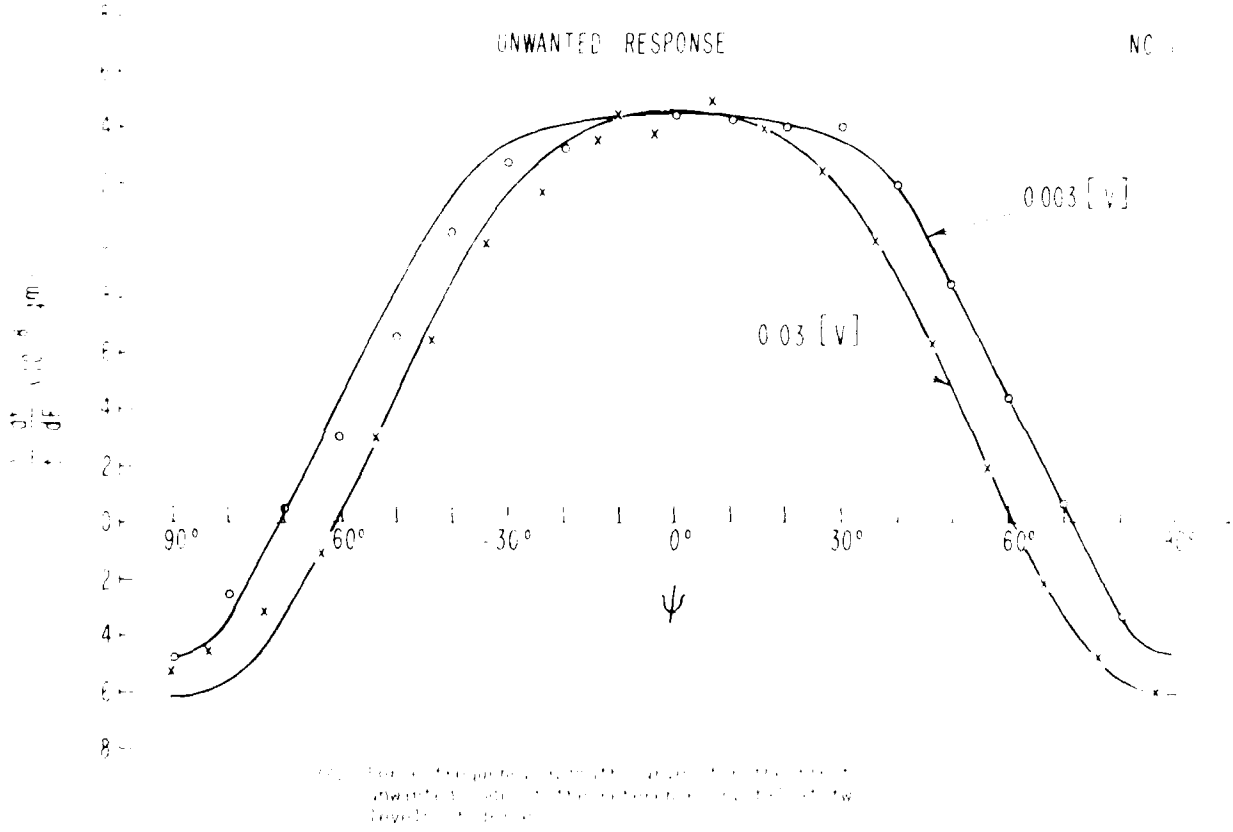


20. Force-frequency-azimuth curves for samples T-1 and T-2 on the third harmonic.



SAMPLE REFERENCE CRYSTAL

[illegible] a conventional, flat AT-cut crystal at the first and third harmonic. Reference crystal.



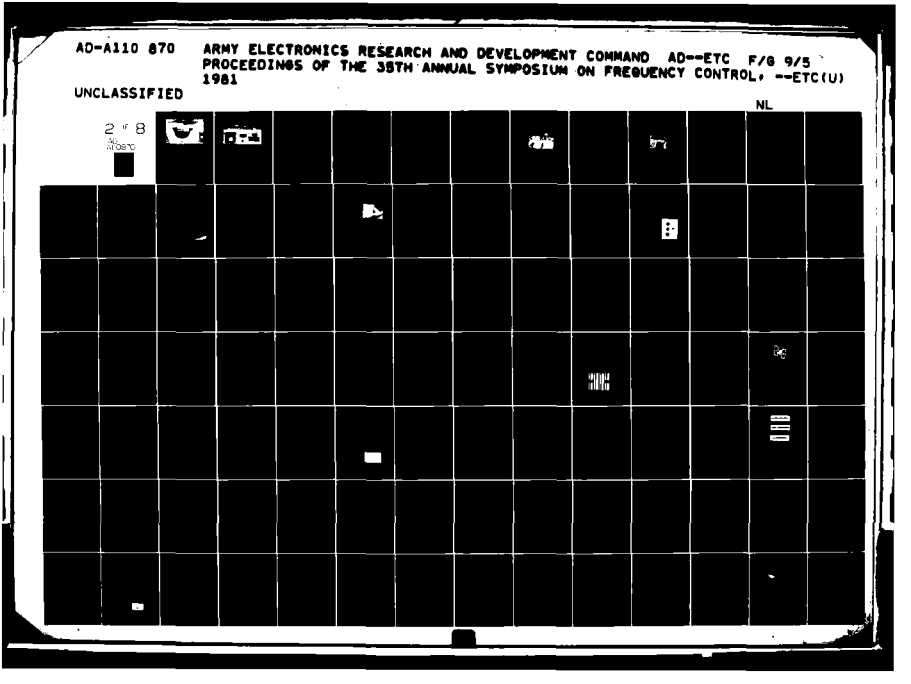
[illegible] unwanted [illegible] crystal [illegible]

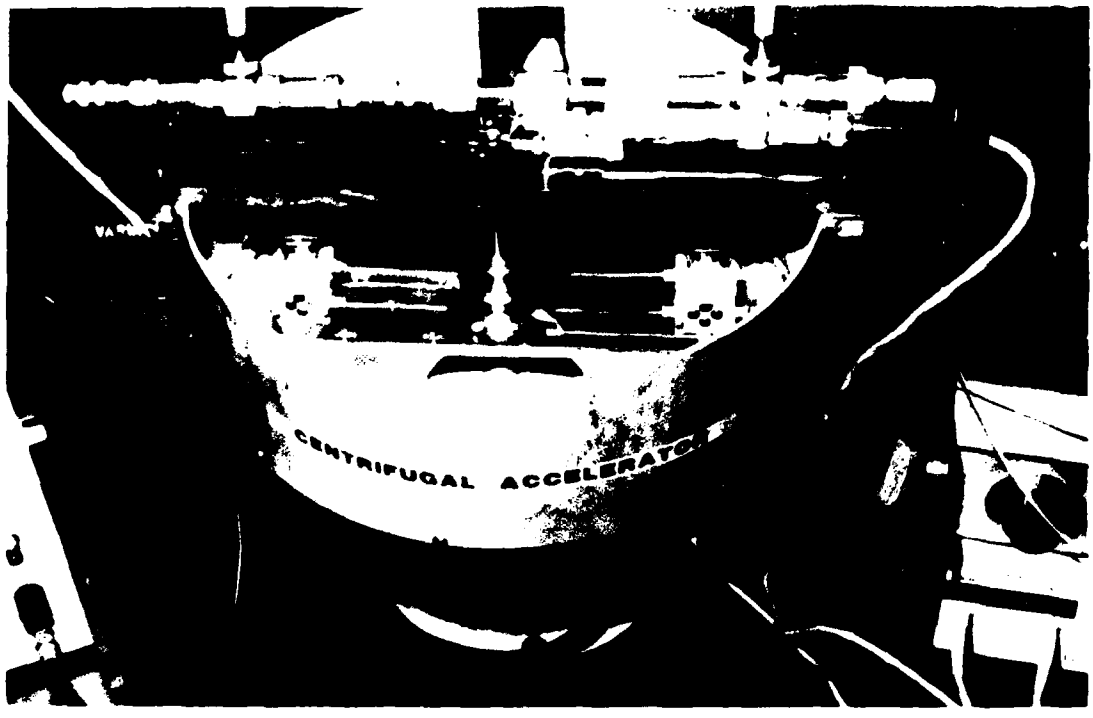
AD-A110 870 ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND AD--ETC F/G 9/5
PROCEEDINGS OF THE 35TH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL, --ETC(U)
1981

UNCLASSIFIED

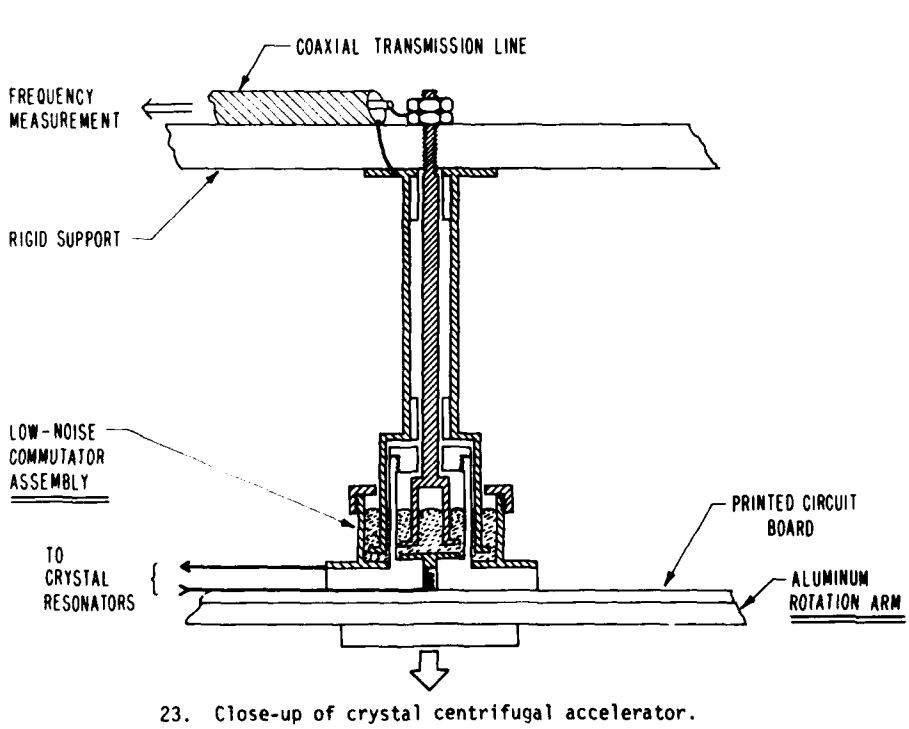
NL

2 OF 8
AD A110870

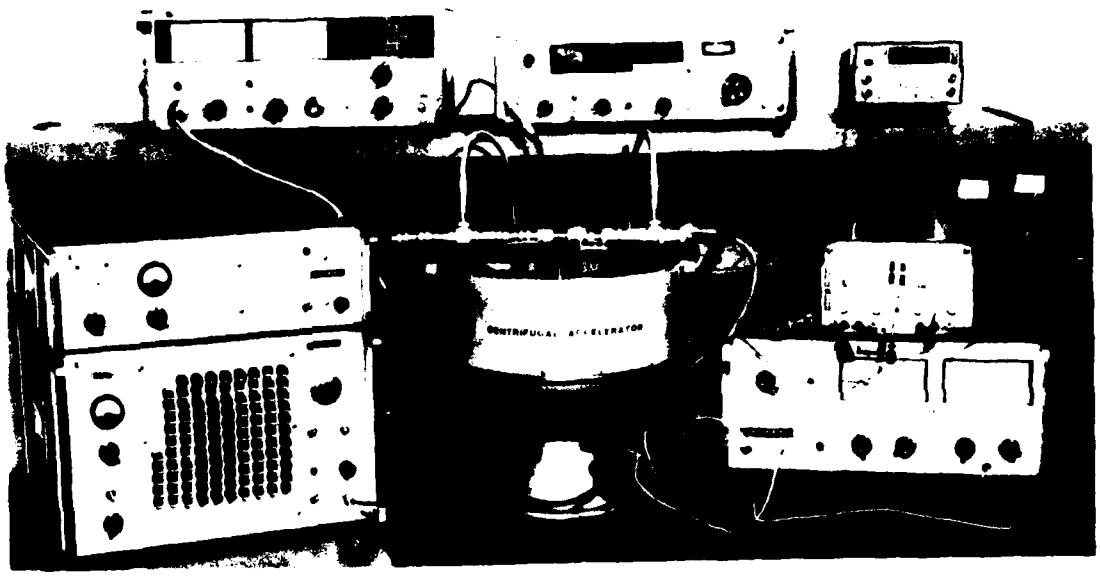




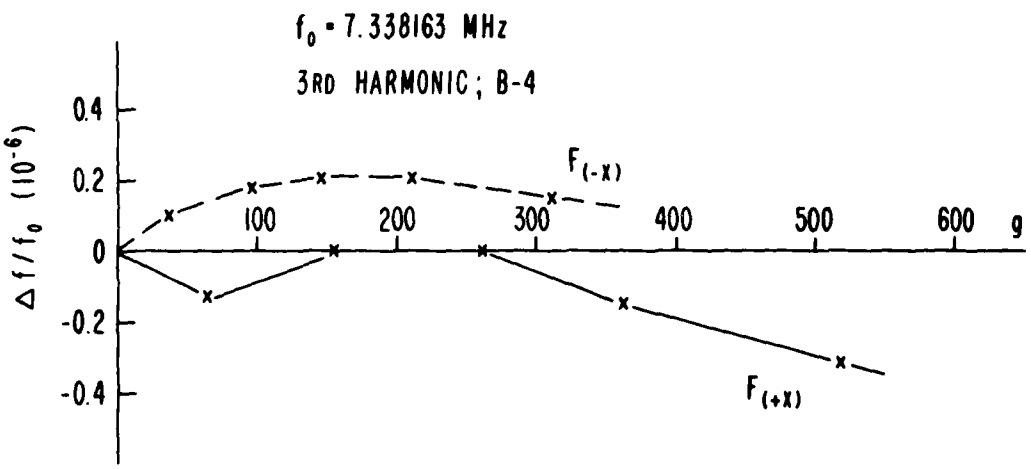
24. Schematic drawing of centrifugal accelerator.



23. Close-up of crystal centrifugal accelerator.

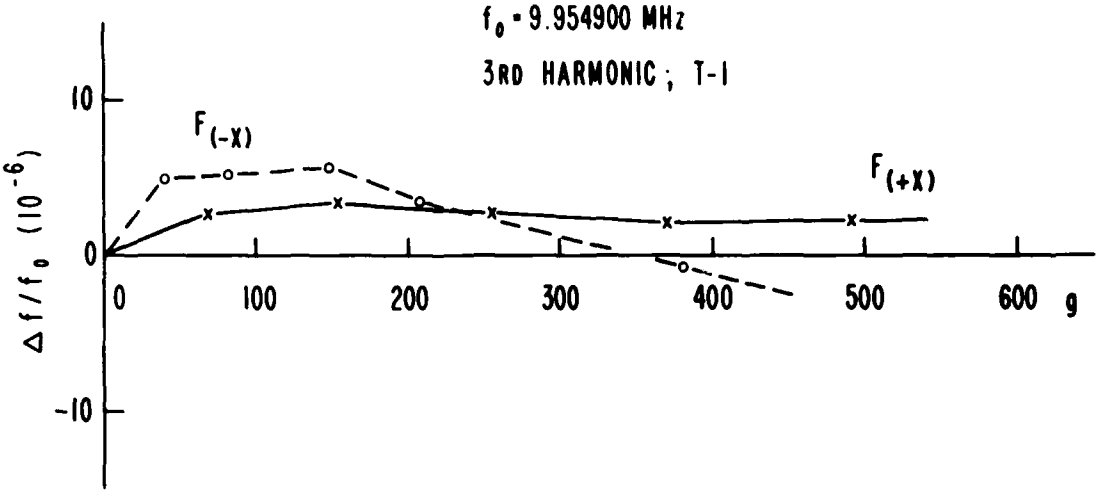


25. Centrifugal accelerator and phase measurement apparatus.



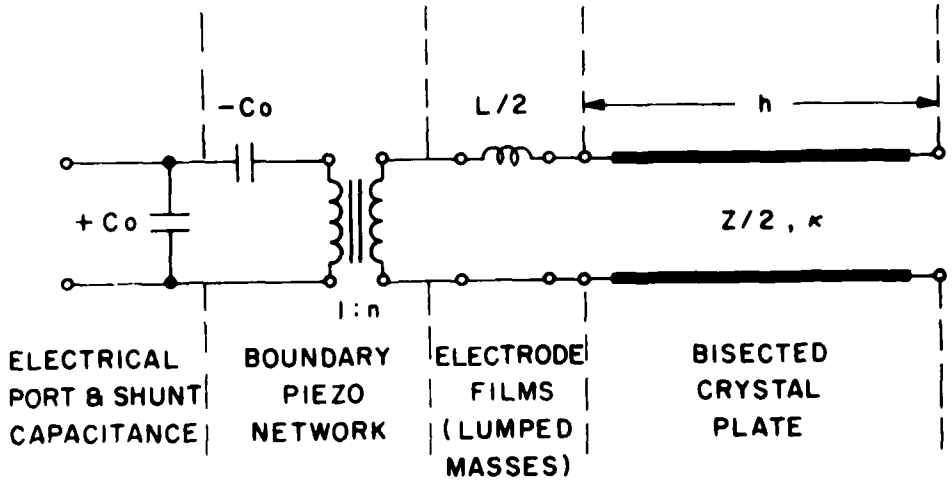
ACCELERATION-FREQUENCY EFFECTS OF X-AXIS DIRECTION FOR SAMPLE NO. B-4

26. Acceleration-frequency effect for the x-axis direction of sample B-4 at the third harmonic.



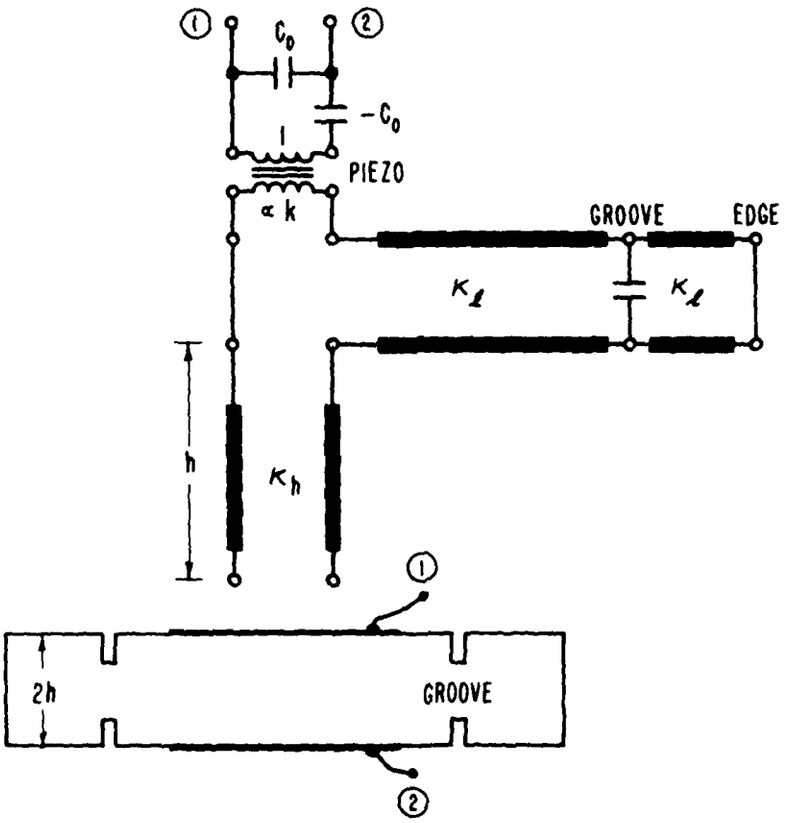
ACCELERATION-FREQUENCY EFFECTS OF X-AXIS DIRECTION FOR SAMPLE NO. T-1

27. Acceleration-frequency effect for the x-axis direction of sample T-1 at the third harmonic.



28. Bisected equivalent network of a conventional, flat resonator.

GROOVED RESONATOR BISECTED NETWORK



29. Bisected equivalent network of a grooved resonator.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

UPDATE OF SC CUT CRYSTAL RESONATOR TECHNOLOGY

B. Goldfrank, J. Ho and A. Warner

Frequency Electronics, Inc.
New Hyde Park, N.Y. 11040

Abstract

Previously we have discussed the fabrication of the SC cut crystal, the mechanics of cutting the doubly rotated blanks from quartz bars and the test methods used to determine the 'g' sensitivity. In this paper we will discuss the following areas. First, a simplified x-ray orientation using a rotated x-cut method; second, the frequency change with acceleration as you vary Psi; and third, a presentation of test results on finished oscillators using SC cut crystals.

Simplified X-ray Orientation Measurements, Using A Rotated X Cut Method

In the past[1] we have used a method of x-raying SC cut quartz blanks, which made use of an existing high precision goniometer, set for measuring AT blanks, within a few seconds of arc. There were advantages and disadvantages.

Advantages

1. The method used an existing AT set-up.
2. The angle measured, Theta, was identical to the angle set on the saw table, making angle correction direct and simple.
3. It was capable of making absolute measurements, as well as ones using a reference standard.

Disadvantages

1. The tilt necessary to bring the AT x-ray plane (01.1) vertical to the goniometer table was about 15°. This made the data measurement very sensitive to the flat angle Psi.
2. Calculating Theta from Theta prime required a knowledge of Phi and calculation of "turn over" required both Theta and Phi. (By the term "turn over" we mean the temperature at which the temperature coefficient of frequency is zero.

The disadvantages were dealt with by the construction of an accurate crystal holding jig and by highly correcting the -X surface of the quartz bar before cutting.

Since the time of this work, we have assembled a second very precise, universal, double crystal x-ray goniometer, which is dedicated to SC measurements. The accuracy is ten (10) seconds of arc. We believe we have developed some new approaches to simplifying the measurement of Phi, Theta and Psi, and what may be even more significant we have found a way of relating the "turn over" of the finished crystal to a single angular measurement with no calculations necessary other than a simple graph.

Figure 1 shows an overall view of the GE x-ray goniometer table, with an added, extremely stable second crystal and a carefully constructed, directly calibrated micrometer screw.

In the new method of measuring, the crystal blank is considered as a rotated X-cut, making Phi about 8° rather than 22°. This derives, of course, from the three-fold symmetry of quartz. Theta is still near 34° and Psi ranges from 0 to -15° as before.

For the measurement of Phi and Theta, we use the 22.3 plane, whose normal lies in the X-Z plane at 34 degrees 17 minutes from X. This is very near Theta and only 8° away for Phi. The Bragg angle is 48°, so the angle between the x-ray beam and the ionization chamber is 84°. The cosine of 34 degrees 17 minutes is .82626, and $\phi = \sin^{-1}(1.2103 \sin \phi')$. Phi prime is the angle measured by the x-ray with the crystal blank mounted on a normal vacuum chuck barrel with the Y' (the old X') axis parallel to the table. One surface of the uncut quartz bar is usually corrected to provide the Y' direction in the blank, which in this case, is not very critical. For measurement of Theta, a mild tilt which is not critical, is provided to compensate for the Phi angle and $\theta = \tan^{-1}(\tan \theta' \cdot (\cos \phi)^{-1})$. Phi is the angle measured as described above and Theta prime is the angle measured by the x-ray. The measurement is made using a non-critical 6° tilt back and with the Z' axis parallel to the table. The calculations to arrive at the specified Phi and Theta are now small enough to present on a simple sheet of graph paper. A calculator is not necessary once the plots are made.

Figure 2 shows the relationship between Phi and Theta for a constant turn over temperature.

The upper curve is taken from a paper by Ballato and Iafrate[2] (1976 30th Annual Frequency Control Symposium). As a Phi departs from the AT and goes through FC, ITC, etc. to the SC, the angle Theta must be changed accordingly to give the same turn-over temperature. On the same graph, we have plotted the values of Theta prime which would result in the Theta values shown. Note especially that over the range of interest the data is invariant with the value of Phi as shown in Figure 3. For specified Phi's from 23° to 24°, Theta prime varies only 0.6 minutes and the Phi's between 22° and 23° Theta varies only .02 minutes.

Figure 4 shows the values of Theta prime, the angle measured directly by the x-ray, for SC turn-over temperatures between 50° and 85°C, for any reasonable Phi value.

This chart, of course, is for a particular design. For other contours[3] and other frequencies the curve would be the same but the left hand scale will be shifted up or down slightly.

Should it be desirable to measure the Psi angle by x-ray, either for measuring existing flats or for determining the place to generate flats, this can be done at either the X' or Z' axes, using the rotated Y cut nomenclature. Using a Y axis reflection, the 02.0 plane, the X' direction of the plate will be some "A" degrees away from the x-ray indication, where $A = \tan^{-1}(\tan\phi \sin\theta)$. For a ϕ_x of 6.25° (ϕ_y = 23.75°) and a θ of 33.91°, A = 3.496°. Of course, a prepared crystal standard can also be used to verify this relationship. For reflections from the Z' edge, we can use an x-ray plane whose normal is in the X-Z plane and is 90° from the 22.3 plane used for measuring Phi. Such a plane is the 11.3 plane at 53.75° and Bragg angle 32°.

The same approach can be used, and the angle the x-ray will make with the Z' axis is now $B = \sin^{-1}(\cos 53.75° \sin \phi_x)$. For ϕ_x = 6.25°, "B" = 3.69°. This is summarized in Figure 5.

Frequency Change With Acceleration

Charts[4] showing the worst case radial acceleration coefficients to be expected for various mounting angles (Psi), using a three point 90° mount, have been supplied by Professor Peter Lee. The coefficients range from one to twenty five times 10^{10} per g.

Data taken on ten 5 MHz 3rd overtone blanks supplied by ERADCOM and randomly mounted have shown about 80% compliance with this calculated curve. Further work is in progress to find the optimum mounting.

To date, the 5 MHz 5th overtone crystal units, mounted by thermo-compression bonding at ψ =-15° are still the best. When very accurately mounted yields of 50% are possible with acceleration cofficients less than $3PP10^{-10}/g$.

Thermo-compression (TC) bonding tests were conducted using a pure nickel ribbon with a 0.015 inch wide triangular cross-sectional gold stripe coined to one side. Figure 6 shows a quartz blank so mounted. The bonding fixtures and carbide tools were all designed and fabricated at FEI. The average pull strengths obtained were 1.5 kilograms or 3.3 pounds. The force applied during the bonding was 16,000 psi.

These results compared well with those obtained using aluminum clad nickel ribbon. In all cases, quartz was present on the ribbon after pull testing.

The two main advantages of this ribbon are 1) the removal of an intermetallic bond, and 2) the potential for bonding crystals with thin edges.

Test Results[5] From Completed Oscillators Using SC Cut Crystals

Figure 7 shows aging data. Depending on process conditions, the 5th overtone crystal ages at 3 to 4 $PP10^{10}$/day. Also we have crystals that age less then 2 X 10^{11}/day. The most important point is that the aging is monotonic and predictable, unlike the fundamental crystal seen in Figure 8. At the same frequency of oscillation, the frequency of the fundamental varies as much as $5PP10^{9}$.

Figure 9 shows 'g' sensitivity results for fundamental crystals. Figure 10 shows the 5th overtone crystal. Fifth overtone crystals have better repeatability, and fundamental crystals suffer because of sensitivity to circuit capacitance change.

Figure 11 shows the short-term stability of the 5 MHz 5th overtone oscillators. We have achieved a 1 to 10 second stability of approximately 7 $PP10^{13}$. Figure 12 shows a measurement of phase noise.

Acknowledgments

The authors would like to thank Mr. James Tsaclas, for his contributions to the x-ray programs. This work is supported in large part by a contract with the U.S. Army ERADCOM.

References

1. B. Goldfrank and A. Warner, 34th Proc. Annual Symposium on Frequency Control, U.S. Army Electronics Command, Ft. Monmouth, N.J., pp. 183-186. Copies available from Electronic Industries Association, 2001 Eye St., N.W. Washington, D.C. 20006.

2. Ballato and Iafrate, 30th Proc. Annual Symposium on Frequency Control, U.S. Army Electronics Command, Ft. Monmouth, N.J., pp. 141-156 (1976). National Technical Information Service Accession No.AD A046089.

3. Vig, Washington and Filler, 35th Proc. Annual Symposium on Frequency Control, U.S. Army Electronics Command, Ft. Monmouth, N.J. Copies available from Electronic Industries Association, 2001 Eye St., N.W. Washington, D.C. 20006.

4. Professor Peter Lee, Princton University, private communications.

5. Many of these tests were carried out at ERADCOM using oscillators supplied by Frequency Electronics, Inc.



Figure 1. Universal, Double Crystal, X-Ray Goniometer

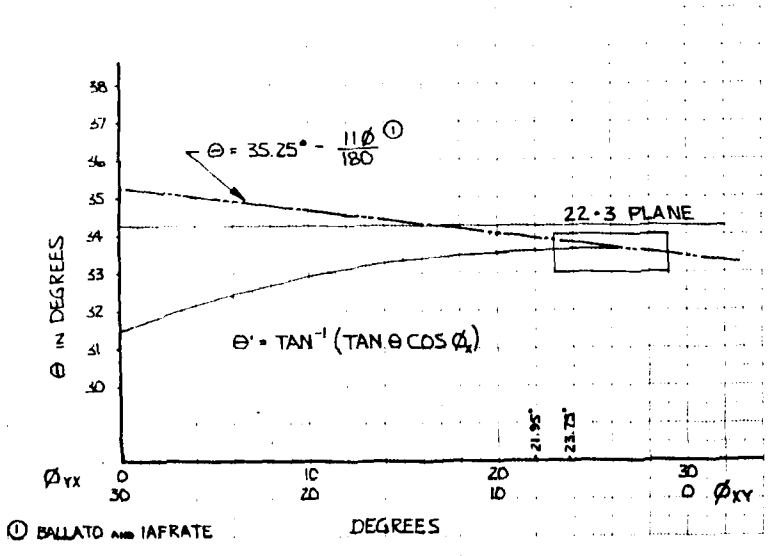


Figure 2. Relationship between θ, θ', and ϕ For a Constant Turn-Over Temperature.

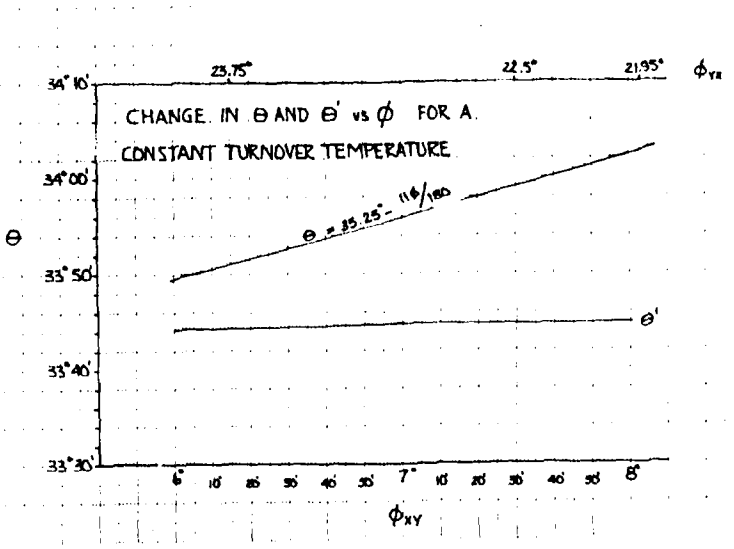


Figure 3. Enlarged Section of Figure 2.

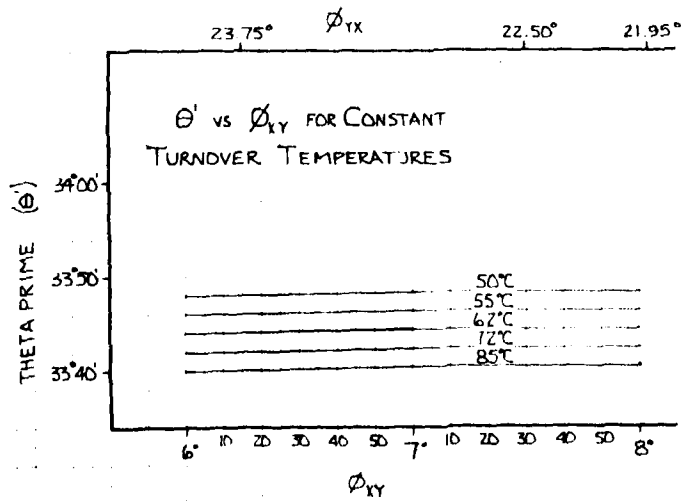


Figure 4. Θ' Angles for Various Turn-Over Temperatures

YX-ℓt	XY-ℓt	Plane	Angle Between X-Ray and Axis
X'	Y'	02·0	$\tan^{-1}$ (tan ϕ_X sin θ)
Z'	Z'	11·3	$\sin^{-1}$ (cos 53.75° sin ϕ_X)

For ϕ_X = 6.25°
θ = 33.91°

Then X' is 3.50°
Z' is 3.69°

Figure 5. Summary of Information for Edge X-Ray Measurement of Crystal Blanks



Figure 6. Crystal Blank Mounted by Gold Stripe Ribbon

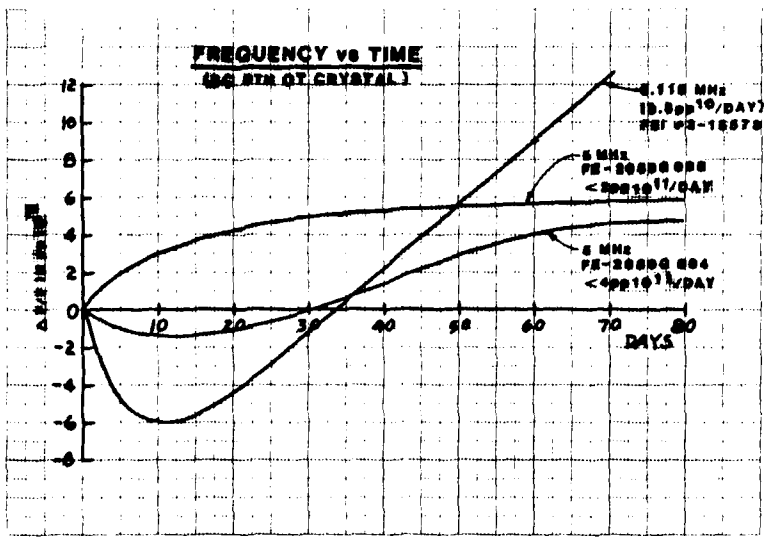


Figure 7. Aging Data on 5th Overtone Crystal Oscillators

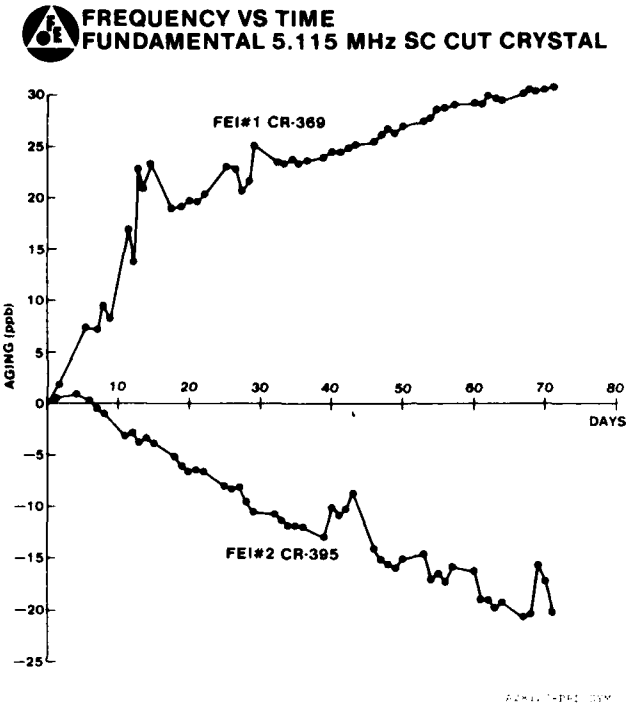


Figure 8. Aging Data on Fundamental Mode Crystal Oscillators

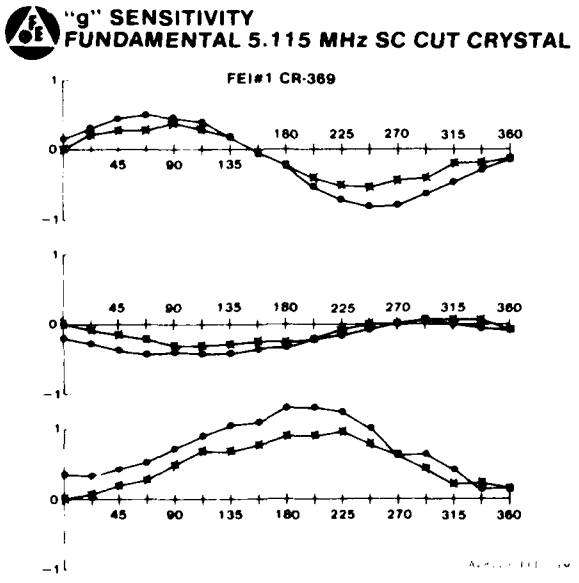


Figure 9.
"g" Sensivitiy of Fundamental Mode Crystal Oscillators

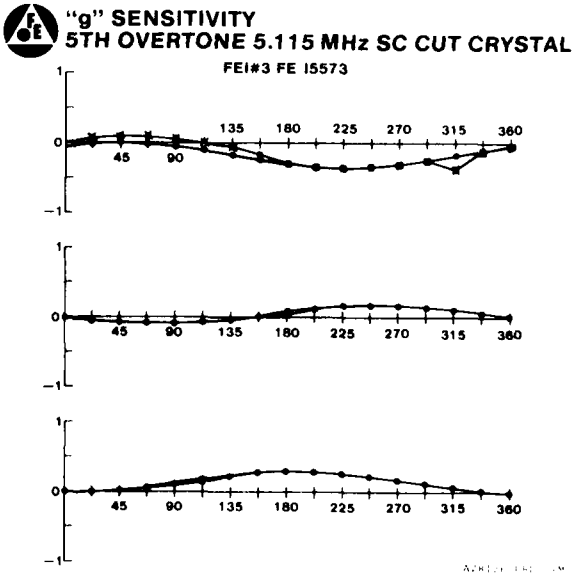


Figure 10.
"g" Sensitivity of 5th Overtone Crystal Oscillators

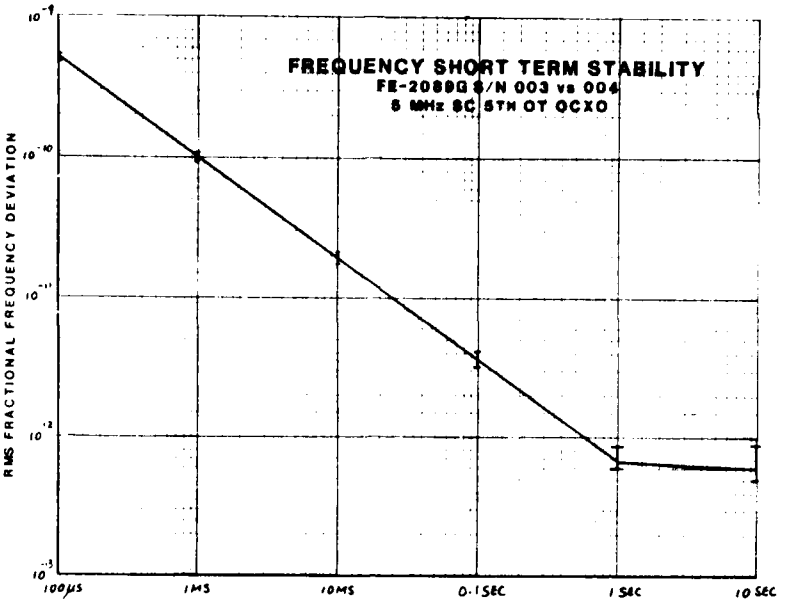


Figure 11. Short Term Stability of Crystal Oscillators

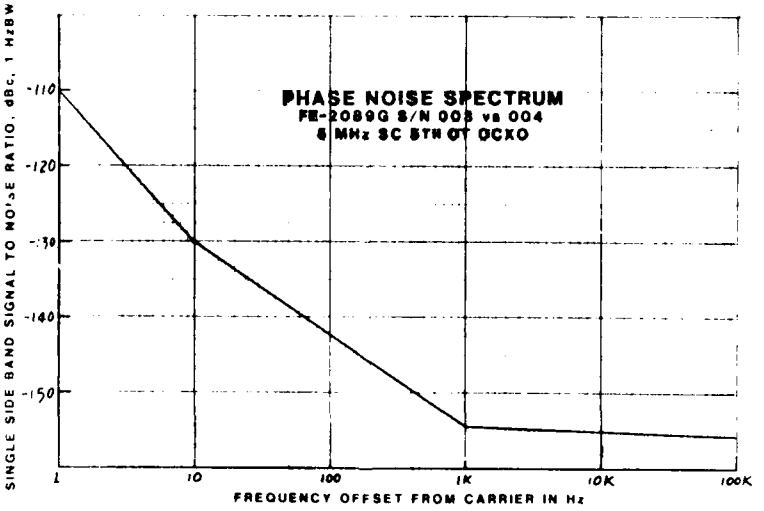


Figure 12. Phase Noise of Crystal Oscillators

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

DESIGN OF HIGH PERFORMANCE SC RESONATORS

Roger W. Ward

Colorado Crystal Corporation
2303 West 8th Street
Loveland, CO 80537

SUMMARY

SC cut alpha quartz resonators have been fabricated using fundamental, third, and fifth overtone designs, covering the frequency range of 4 MHz to 100 MHz. Data on several designs indicates the theoretical maximum Q for SC resonators is 15-20% higher than for AT cut resonators[1]. For example, a 60 MHz third overtone SC cut has been fabricated with a Q of 250,000, which is the theoretical maximum Q for an AT at that frequency.

"Zero" angles have been determined for several SC designs (plano-plano, plano-convex, bi-convex, fundamental and overtone).

Vig's method for adjusting the turning point of 5 MHz fundamental SC's by contour control[2] has been utilized in production.

Chemical polishing of SC's is being performed on a production basis. Studies indicate that the 1:4 $HF:NH_4F$ solution proposed by Vig[3] is very forgiving of surface contamination on SC's, whereas his proposed 1:2 $HF:H_2O$ solution results in surface degradation on the + X surface of the SC cut.

All contoured SC designs are processed such that the same crystallographic side of the blank is contoured. This is accomplished by squeeze testing each blank in a circuit described by Bond[4].

Aging studies on AT's and SC's processed using identical manufacturing techniques show that the SC typically ages into its final aging rate 5-10 times faster than the AT, and the SC typically exhibits a 2-3 times improvement in ultimate aging rate.

INTRODUCTION

Much has been written about the SC cut since it was first predicted by Holland[5] in 1974. However, most of the reported studies deal only with 5 MHz fifth overtone[6], 10 MHz third overtone[7], and 5-15 MHz fundamental[8] units. As a result, many of the design tools that exist for the AT-cut crystal designer do not exist for the SC-cut designer. For example, how the zero angle for the SC changes with contour and overtone, how the turnover temperature changes with orientation angle Θ, how the design of the SC affects Q. This study provides partial answers to these items.

ORIENTATION

Since the SC-cut is a "double-rotated" cut it is important to know how various properties of the crystal change with the orientation angles ϕ, Θ, and Ψ.

EerNisse[9] has calculated the stress coefficients for C-mode thickness-shear resonators along the zero temperature coefficient locus. He determined that the coefficient goes to zero at a calculated value of ϕ=22.5°. This zero stress coefficient determines the location of the SC (Stress-Compensated) cut. Calculations made by EerNisse[10] indicate that the change in Θ required to achieve stress compensation is a second order function of the change in Θ required to achieve a given frequency-temperature performance. In other words, for all practical purposes, the SC cut is defined as ϕ=22°30' (Kusters'[11] experimentally determined ϕ to be 21°56'), independent of Θ for changes in Θ up to possibly ± 1°. The slope of the stress coefficient taken from EerNisse's curve is about 3.3×10⁻⁶ Dyn⁻¹cm²/min ϕ. If one assumes an "error" in ϕ of ± 1' from the SC "zero" stress angle, then Figure 1 results, which shows the relative stress sensitivity of various cuts along the zero temperature coefficient locus. At ± ½ degree from the ϕ angle for the SC-cut the relative stress sensitivity is 10 times that of the SC at ± 1' error.

The zero angles for several SC designs are shown in Table I. The zero angle is defined as the angle for which the upper and lower turning points (df/dT = 0) and the inflection temperature (d^2f/dT^2 = 0) coincide. The zero angle varies by about 34 minutes for various designs. In contrast, the AT would vary about 12-13 minutes.

Ballato's[12] temperature coefficients were used to generate Figure 2, which shows the lower turning point (df/dT = 0) versus Θ for the SC. The equations used were:

$f = A\Delta T + B\Delta T^2 + C\Delta T^3$, where

$A = 0.8665 \times 10^{-6}/°C$

$B = 12.3 \times 10^{-9}/°C^2$

$C = 58.2 \times 10^{-12}/°C^3$

$$A = A + \frac{dA}{d\Theta} d\Theta$$

$\Delta T = T - 25°C$

$dA/d\Theta = -3.78 \times 10^{-8} °C^{-3} \text{ min}\Theta^{-1}$

T = Temperature in °C

dA/dΘ was determined experimentally — Ballato's value for dA/dΘ is too large. For each 1 minute change in Θ the lower turning point was calculated by setting df/dT = 0.

Note from Table 1 that the inflection temperature is a function of the design. In general, as the active area of the blank becomes smaller (the plate "looks" larger) the inflection temperature increases. This is well illustrated by the fact that the 20 MHz fundamental SC blanks were operated at their third overtone to provide the 60 MHz third overtone data point in Table I, resulting in a 5 °C shift in inflection temperature. Hence, the inflection temperature alone does not determine if one has an SC-cut; the mechanical design must also be known.

Since the "2g" effect on the SC cut is a function of the mounting location, Ψ, it is necessary to keep track of the crystallographic orientation of the blank. To accomplish this, all contoured SC's are contoured on the same crystallographic side — specifically the side that is negative on compression. This side is identified by using a squeeze-tester voltmeter circuit described by Bond[4]. By means of this convention one of the two vectors needed to locate oneself around the blank is known. The other vector is kept track of by grinding a flat on the YY' edge of the blank. This location is known from how the blank was cut out of the original bar of quartz. Then one must identify the minimum 2g mounting location by means of polarized light, x-rays, or force-frequency testing. The minimum 2g mounting location must be found experimentally for each blank mounting geometry.

It is not necessary to keep side-to-side identity of the AT, because the AT is symmetrical upon rotation about Ψ.

All blanks are cut from Sawyer Premium Q quartz (Q greater than 2.2 million) using the following technique, developed by the author, and as illustrated in Figure 3: right-hand Y-bar quartz has three Y-axes, which results in 3 possible "first cut" locations for the SC. If one choses orientation II in Figure 3, the first SC saw cut is made at about + 8° (clockwise) from the + X-axis, thereby obtaining a bar that has the correct SC ϕ rotation and is ready to be placed onto a slurry saw for the Θ rotation cut, similar to cutting AT's. By using this technique the SC is rather easy to produce. If the Y-bar is grown with a large + X height the yield of SC wafers can be quite high, approaching the yield of AT's from a Y-bar. There will be a small amount of + X growth in a few wafers, but this does not seem to cause any problems.

The blanks are sorted into ½° increments on Θ. ϕ is not measured, except before the first cut. Typical distribution is ± 3° on Θ. ϕ appears to be controlled to ± 5°.

Bottom[11] describes a fast, simple, accurate iterative method for measuring ϕ and Θ on SC's, which will enable anyone with an x-ray goniometer to measure the orientation of SC blanks.

By processing 100 blanks, all starting out within a ½° group, a spread of turning points equivalent to ± 1½° (1 sigma) results, probably due to angle shifts resulting from lapping and contouring. For ovenized applications above 75°C this angle control is not adequate to achieve high yields. Angle correction would improve the yields considerably.

GEOMETRY

Vig[1] et al has determined that the zero angle for 5MHz fundamental mode SC's is very contour sensitive — more so than a similarly designed AT. He has utilized this effect to adjust the slope of the frequency-temperature curve to be zero at the crystal's inflection temperature. We have utilized this effect to extend the effective coverage of our blank inventory. For example, by the proper selection of contour a blank's angle can be "adjusted" by about ± 2 minutes without adversely affecting the crystal's electrical parameters. This allows greater flexibility in the turning point range a given blank inventory can cover, as well as enabling one to fill in any holes that might exist in an inventory, thereby allowing faster turn-around for a customer's custom part.

Q

The maximum Q obtainable for the AT cut is given by Warner[1] to be Qf = 15×10^{12} Hz. For a 10 MHz AT-cut crystal the maximum Q would be 1.5 million. In production a maximum Q of 1.3 million (87% of theoretical maximum) has been achieved at 10 MHz using a 5th overtone AT-cut. The maximum Q achieved in production for SC's at 10 MHz is 1.2 million for a plano-convex, 14mm diameter third overtone crystal. Resistance is 40 ohms. A Q of 1.4 million has been obtained on a 2 diopter bi-convex design, with a resistance of 95 ohms. In addition, a 60 MHz, third overtone 7mm diameter plano-plano blank has been produced with a Q of 250,000 and a resistance of 75 ohms. An AT would have a maximum Q of 250,000 at this frequency, and a producible Q of perhaps 200,000. This indicates 1) that SC's have a potential Qf of greater than 15×10^{12}Hz (probably 17.5×10^{12}Hz) and 2) that the loss mechanisms of the SC require more contour and/or a larger diameter-to-thickness ratio than an equivalent AT to achieve its maximum Q.

SURFACE FINISH

Predominately, the SC blanks have been final lapped with 1 micron aluminum oxide. During lapping the operators experience no difficulties with the B-mode. The C-mode remains quite strong throughout lapping and contouring. The blanks are then chemically polished using a 4:1 mixture of 40% ammonium fluoride (NH_4F) and 48% hydrofluoric acid (HF)[3]. The chemically polished blanks are quite transparent — it is possible to read print thru them at a distance of several feet.

Due to the difficulty in obtaining commercially prepared ammonium fluoride solution, it became necessary to prepare the 40% by weight solution using NH_4F flakes and DI water. No etch differences have been observed between etch solutions made using this solution and the commercial solution.

In general, poor results were obtained when a 2:1 solution of water and HF, as suggested by Vig, was used to chemically polish SC's. One side was well polished while the other was blotchy. Indications are that the blotchiness was due to inadequate cleaning of the blanks prior to etching. Several other dilutions were used, including 4:1[3], with little or no improvement. Since these blanks were cleaned the same as others which polished well with HF/NH_4F, one must conclude that the HF/NH_4F solution's etching characteristics are very forgiving of surface contamination, whereas the HF/H_2O solution is not.

A group of mechanically polished SC's were fabricated and compared to similarly designed chemically polished resonators. No differences were noted between the two groups. Resistance, Q, and aging was comparable. Therefore, due to the difficulty of mechanically polishing contoured blanks, all contoured blanks are now chemically polished. High frequency resonators are mechanically polished for overtone operation, since some resistance degradation has been observed on chemically polished high frequency overtone designs.

FINISHING

All low frequency blanks, less than 30 MHz third overtone, are plated with pure gold in a cryogenically pumped vacuum system. The blanks are typically bonded using a conductive devitrifying glass frit which is cured at 450°C. The resonators are then final plated, again using pure gold. Each resonator is heated to its operating temperature prior to final plating to insure adequate frequency precision at the crystal's turning point. The resonators are sealed in an all glass bulb using an RF induction heated system.

High frequency resonators are plated with aluminum and are usually sealed in coldweld packages.

TESTING

All SC's are measured in a Saunders 2000 Test System from 0° to 120°C. Third order curve-fit coefficients are computed by the system for each crystal. No spurious modes to a resolution of 2×10^{-7} have been observed in any SC resonators of any design.

Aging studies are performed in precision, oven-controlled oscillators. Aging rates on 10 MHz third overtone SC's is typically less than 5×10^{-10}/day after 2 weeks. Similar AT's exhibit aging rates of about 2×10^{-9}/day to 5×10^{-10}/day after 90 days.

A system has been set up to measure the 2g effect on crystals, similar to the one used by Warner[8]. The system utilizes a Hewlett-Packard 3585A Spectrum Analyzer to measure the sidebands generated when the crystal is vibrated on a vibration table. With the 3585A the system resolution is less than 1×10^{-11}/g without using any multipliers or amplifiers. Figure 4 shows a block diagram of the system and Figures 5A-C show typical plots of sidebands generated as the crystal is vibrated. The most obvious advantages to this method of measuring 2g is that a highly stable oven and long warm-up times are not required. Hence, every crystal manufactured can be checked for 2g. Typically, the 10 MHz third overtone SC's have g-sensitivities less than 5×10^{-10}/g on 2 axes, and approximately 2×10^{-9}/g on the axis thru the center of the blank and the mounting points.

The poorer performance on the axis thru the mounts is possibly explained as follows: this is the only axis upon which one can impress diametrically opposed forces upon the blank, thereby "exciting" the force-frequency effect in the blank. Using the coefficients measured by EerNisse, et al[14], and by calculating the mass of the blank, one can verify that the force-frequency effect will indeed cause 2g effects on the order of 10^{-9}/g when applied along the most force sensitive axis. Therefore, it would follow that the least g-sensitive crystal would be one that is mounted at least near the zero force-frequency sensitive point. More studies will be done on this important parameter.

Vig measured the g-sensitivity of the 60 MHz third overtone crystal at 3.5×10^{-10}/g, worst axis. The crystal is in a 3-point mount TO-5 package.

PHASE NOISE

Przedpelski[15] measured phase modulation under vibration on the 60 MHz TO-5 units mentioned above to be less than - 60dBc from 100 Hz to 3200 Hz in his oscillator package. Similar AT's (except using a 2-point mount) had sidebands as poor as - 25dBc. This suggests the possibility of using the SC in low phase-noise applications where the crystal frequency must be multiplied up to GHz frequencies.

FREQUENCY SPECTRUM MODE STUDIES

Spectrum analyzer studies on SC's reveal a family of C-mode inharmonics as well as the B-mode with its inharmonic family. This is illustrated in Figure 6. A few oscillator designs have had difficulties with the B-mode, but with proper filtering these problems have been eliminated.

CONCLUSION

In conclusion, the SC cut is capable of providing greatly improved aging, 2g, Q, and phase noise performance when compared to the AT cut — if the SC's surface is properly prepared and it is properly oriented and mounted. As our understanding of the SC has increased, we are now able to design SC's with the same skill as AT's and over the same frequency range, thereby providing improved performance for the precision frequency control market. The SC has come out of the laboratory and into the marketplace.

REFERENCES

1. A. Warner, "Design and Performance of Ultraprecise 2.5-MC Quartz Crystal Units", **Bell Systems Technical Journal**, Sept. 1960. P. 1198.
2. J. Vig, W. Washington, R. Filler, "Adjusting the Frequency vs. Temperature Characteristics of SC-cut Resonators by Contouring", These Prodeedings.
3. J. Vig, R. Brandmayr, R. Filler, "Etching Studies on Singly and Double Rotated Quartz Plates", 33rd AFCS*, 1979, pp. 351-358.
4. W. Bond, **Crystal Technology**, (John Wiley & Sons, NY, 1976), p. 16.
5. R. Holland, "Nonuniformly Heated Anistropic Plates: I. Mechanical Distortion and Relaxation", IEEE Trans. Sonics and Ultrasonics, Vol. SU-21, July 1974, pp. 171-178.
6. A. Warner, B. Goldfrank, M. Meirs, and M. Rosenfeld, "Low 'g' Sensitivity Crystal Units and Their Testing", Proc. 33rd AFCS*, 1979, pp. 306-310A.
7. J. Kusters, C. Adams, and J. Leach, "TTC's — Further Developmental Results", Proc. 31st AFCS*, 1977, pp. 3-7.
8. R. Filler and J. Vig, "Fundamental Mode SC-cut Resonators", Proc. 34th AFCS*, 1980, pp. 187-193.
9. E. EerNisse, "Quartz Resonator Frequency Shifts Arising From Electrode Stress", Proc. 29th AFCS*, 1975, pp. 1-4.
10. E. EerNisse, Private communication, Nov. 1979.
11. J. Kusters and J. Leach, "Further Experimental Data on Stress and Thermal Gradient Compensated Crystals", Proc. IEEE, Vol. 65, Feb. 1977, pp. 282-284.
12. A. Ballato, "Double Rotated Thickness Mode Plate Vibrators", Phys. Acoustics, Vol. XIII, 1977, pp. 115-181.
13. V. Bottom, to be published, Van Nostrand Reinhold, 1981.
14. E. EerNisse, T. Lukaszek, and A. Ballato, "Variational Calculation of Force-Frequency Constants of Doubly Rotated Quartz Resonators", IEEE Trans. on Sonics and Ultrasonics, Vol. SU-25, No. 3, May 1978, pp. 132-138.
15. Andrzej Przedpelski, A.R.F. Products, Boulder, CO. Private Communication, Nov. 1980

*AFCS: Annual Symposium on Frequency Control, US Army Electronics Command. Ft. Monmouth, NJ, 07703. Copies available from Electronics Industries Association, 2001 Eye Street, NW, Washington, DC 20006

TABLE I
SC-CUT ZERO ANGLES
$\phi = 21°56'$

f(MHz)	OVERTONE	NOMINAL INF. TEMP. (°C)	θ
5	1	100	33°40'
5	3	107	34°00'
5	5	115	33°55'
10	1	98	33°30'
10	3	96	34°04'
20	1	98	33°50'
60	3	103	33°57'

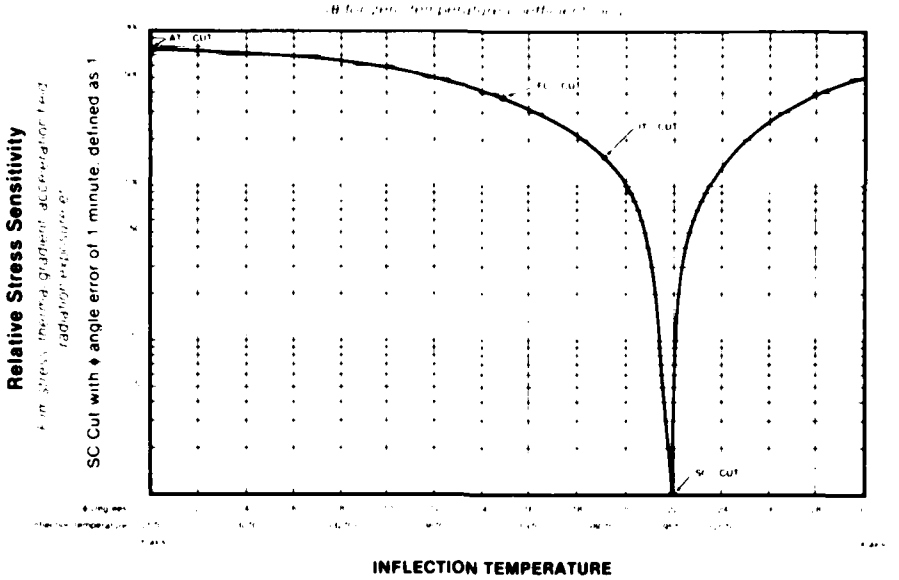


Figure 1: Thermal Transient Effect versus ϕ

Cut	Orientation		Inflection Temp (°C)	Stress, Effects
	ϕ	θ		
SC	21.9°	33.9°	95	0
SC'	21.9° ± 1 min	33.9° ± 1 min	95	1 (Defined as reference)
AT	0°	35.3°	25	830 times more sensitive to stress than SC'
FC	15°	34.3°	50	420 times more sensitive to stress than SC'
IT	19.1°	34.1°	80	140 times more sensitive to stress than SC'

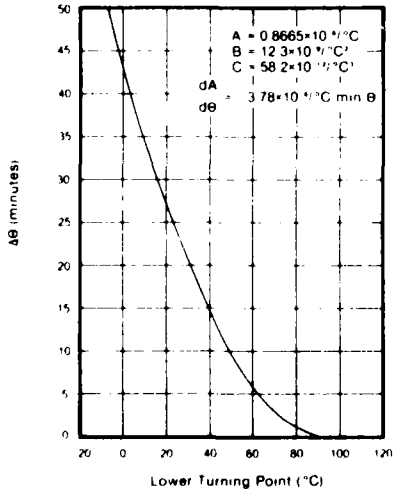


Figure 2: Lower Turning Point vs. Δθ for SC-Cut Quartz

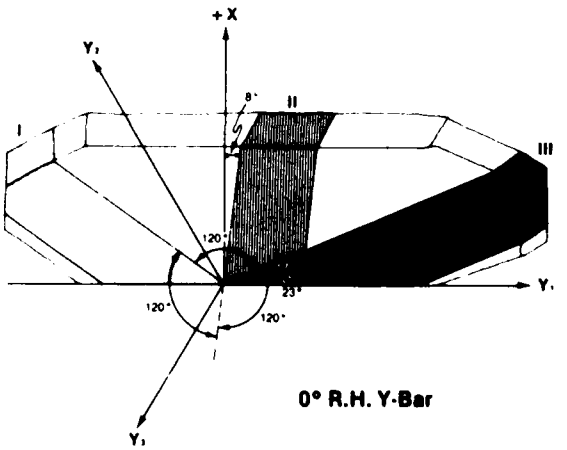


Figure 3: Three locations for making first cut on 0° R.H. Y-bar. For maximum utilization of quartz, orientation II is preferred.

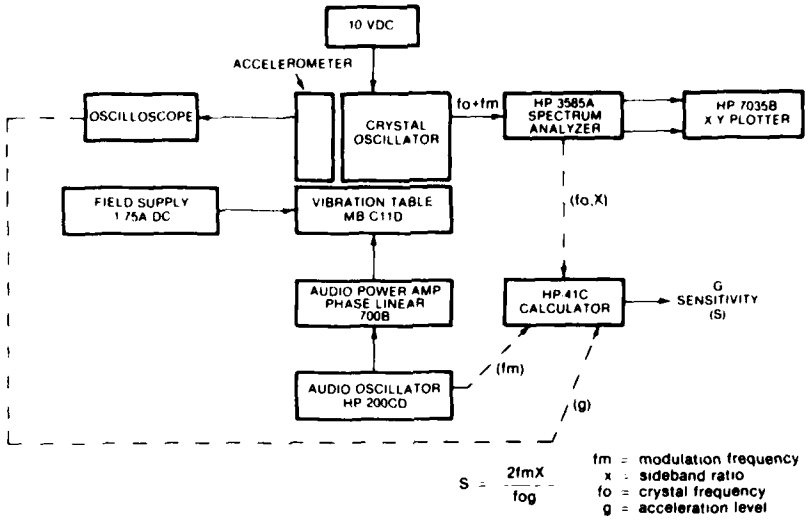


Figure 4: Block Diagram of "2G" Test System

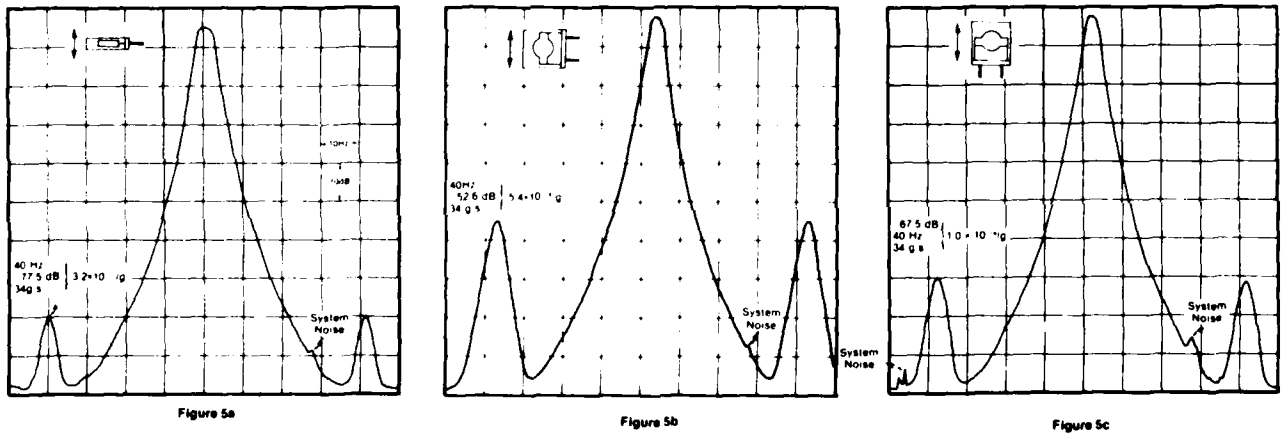


Figure 5a Figure 5b Figure 5c

Figure 5: 2g vibrational tests on a 10 MHz 3rd overtone SC cut crystal.

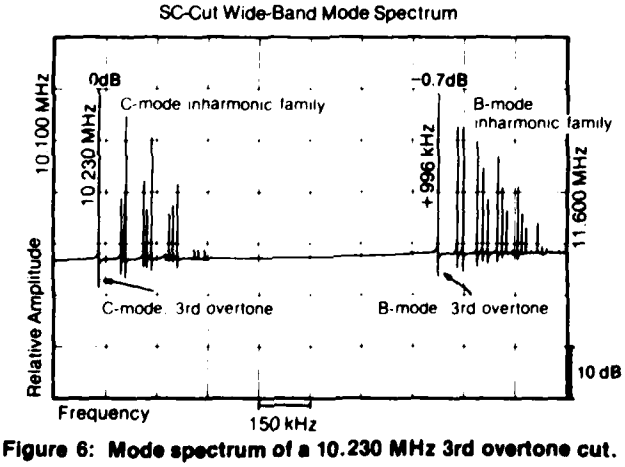


Figure 6: Mode spectrum of a 10.230 MHz 3rd overtone cut.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

ADJUSTING THE FREQUENCY VS. TEMPERATURE CHARACTERISTICS OF SC-CUT RESONATORS BY CONTOURING

John R. Vig, William Washington and Raymond L. Filler

US Army Electronics Technology & Devices Laboratory (ERADCOM)
Fort Monmouth, NJ 07703

Abstract

The effects of contouring on the frequency vs. temperature characteristics of fundamental mode and third overtone SC-cut resonators have been investigated. It has been found that the frequency vs. temperature characteristics are highly sensitive to blank contour, and that when a blank is recontoured, the slope at the inflection temperature changes approximately linearly with the change in contour.

For AT-cut crystals, the contour is a critical part of the design. For similar SC-cut resonators, acceptable performance can be achieved over a much broader range of contours. The contour dependence of the temperature characteristics has been successfully applied to "angle correct" SC-cut resonators. Examples of this "angle correction" are presented.

Preliminary results on the effects of electrode dimensions on the frequency vs. temperature characteristics are also reported.

Key Words: Quartz, quartz resonators, quartz crystal, SC-cut, resonators, crystal resonators, contour, frequency control, frequency vs. temperature, angle correction.

Introduction

The advantages of SC-cut resonators are well known by now. The SC-cut has proven to be superior to other cuts especially in properties that are of significance to high stability oven controlled crystal oscillators (OCXO's). These properties include thermal transient behavior, drive level sensitivity, electrode stress sensitivity, incidence of activity dips, and frequency vs. temperature characteristics.

The major disadvantage of the SC-cut is that making such resonators to a specified frequency vs. temperature characteristic is difficult. For several important applications of OCXO's, for example, SC-cut resonators with a turnover temperature between 90° and 100°C are desirable. The "ideal" SC-cut resonator frequency vs. temperature characteristic for OCXO's is shown in Figure 1. For such resonators, the slope of the frequency vs. temperature characteristic remains small over a wide temperature range about the inflection temperature.

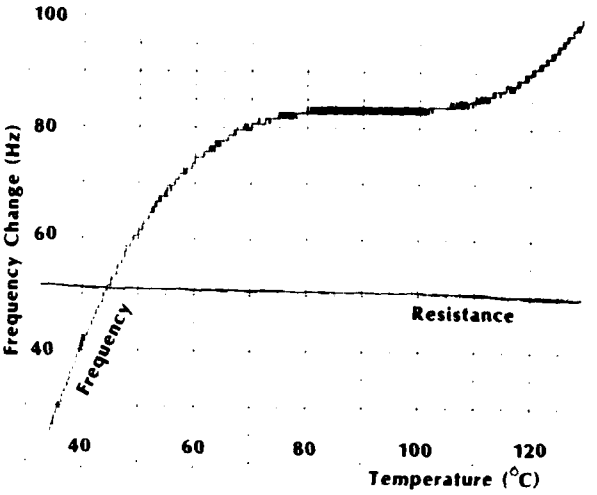


Fig. 1 - "Ideal" Frequency vs. Temperature Characteristic

Kusters and Adams[1] have described how the SC-cut's turnover temperatures vary with the angles of cut, ϕ and θ. Their results are reproduced in Figures 2 and 3, for convenience. Figure 2 shows the turnover temperature vs. ϕ; Figure 3 shows the turnover temperature vs. θ. From Figure 3 one can see that even if ϕ is controlled perfectly, in order to place the turnover temperatures within 5°C of the inflection temperature, θ must be within about a ±4 second range. Of course, if instead of specifying turnover temperature one specified the slope of the frequency vs. temperature characteristic at the operating temperature, the angle tolerances generally become easier to obtain[1-2]. For example, for the slope at inflection temperature to be within the range $\pm 1 \times 10^{-8}$ per °C, a θ-angle tolerance of nearly ±20 seconds is sufficient.

To obtain precise control of the frequency vs. temperature characteristics, precise control of the angles of cut is necessary. However, from experience with other cuts of quartz, angle control alone is not sufficient. Other parameters, such as blank contour and electrode dimensions, can also affect the frequency vs. temperature. This paper is a progress report on investigations

aimed at determining the effects of such fabrication parameters on the frequency vs. temperature characteristics of SC-cut resonators.

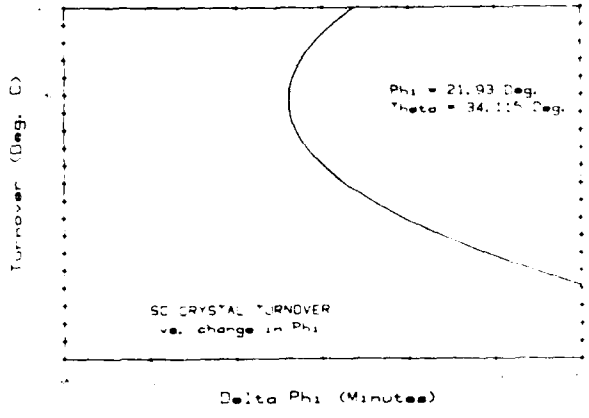


Fig. 2 -Turning Points vs. ϕ

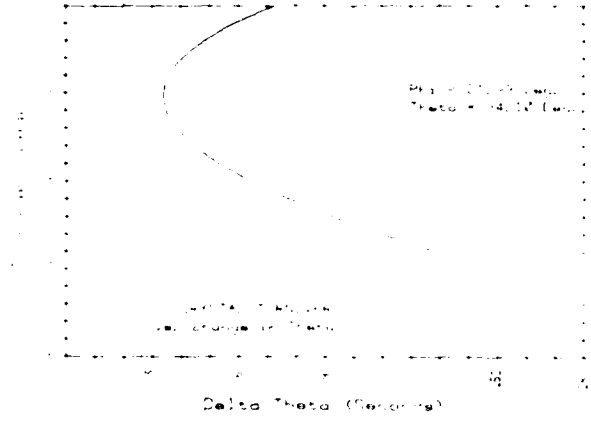


Fig. 3 - Turning Points vs. θ

The Effects of Blank Contour

The radius of curvature, i.e. the contour, of a resonator blank has been found to have a significant effect on the frequency vs. temperature characteristics of previously investigated cuts of quartz resonators. For example, the apparent θ - angles of plano-convex AT-cut crystals are higher than those of the same angle-of-cut plano-plano resonators[3-4]. Similarly, the turnover temperatures of 5MHz 3rd overtone BT-cut resonators were found to decrease significantly with increasing blank curvature.[5] In the current study, the effects of blank contour have been investigated for 5 MHz fundamental mode, 5 MHz 3rd overtone and 10 MHz 3rd overtone plano-convex SC-cut resonators.

Experimental Procedure

All blanks had a diameter of 14mm. The blanks were cemented into "buttons", then the spherical contours were generated, using commercially available optical diopter cups[6], on the Elgin Model 1-121 lapping machine shown in Figure 4. As the diopter cup rotates, the button holding the blank is moved back-and-forth radially from the center of the diopter cup to near the edge. The button is free to rotate, so that each time the button reaches the center of the diopter cup, it is spun part-way around by the rotation of the cup.



Fig. 4 - Machine Used For Contouring

Since the diopter cup's curvature is marked on the cups in units of diopters, for convenience, the blank curvatures throughout these experiments are described in units of diopters, rather than in meters or reciprocal meters. The units are related by[7]: (radius of curvature, in meters) x (contour, in diopters) = 0.530.

A water slurry of 3 micron aluminum oxide abrasive was used in the contouring operation. Since during the study each blank was recontoured several times, care was taken to remove only enough material each time to generate a full contour. Prior to each contouring operation, the sides to be contoured were painted with a water insoluble ink, and the contouring was monitored by observing the decrease of the ink-spot size. The contouring was terminated shortly after the ink-spot had been completely removed.

Subsequent to contouring, the blanks were etched lightly, mounted and bonded into HC-6 enclosures, plated with pure gold electrodes, then solder sealed. After measuring the frequency vs. temperature characteristics, the resonators' enclosures were opened, the blanks were stripped, recontoured, remounted along the same orientation as before, plated the same as before, and resealed. The flat sides of the blanks were left untouched during reprocessing, i.e. the angles of cut of the flat sides remained constant throughout the experiments.

Fundamental Mode 5 MHz SC-cut Resonators

The convex sides of a group of eight 5 MHz

fundamental mode plano-convex SC-cut resonators were reprocessed repeatedly with increasing contours. The nominal angles of cut of were $\phi = 21^{\circ}56'$ and $\theta = 33^{\circ}38'$. The spread in angles was such that some of the blanks resulted in resonators that had a positive frequency vs. temperature slope at the inflection temperature, and some resulted in resonators that had a negative slope. Each blank was reprocessed a minimum of five times, up to a maximum curvature of 3.0 diopters.

The results for one of the repeatedly reprocessed resonators is shown in Figure 5, in which the normalized frequency change is plotted vs. temperature change for six different contours. Both axes in figure 5 are referenced to the inflection point.

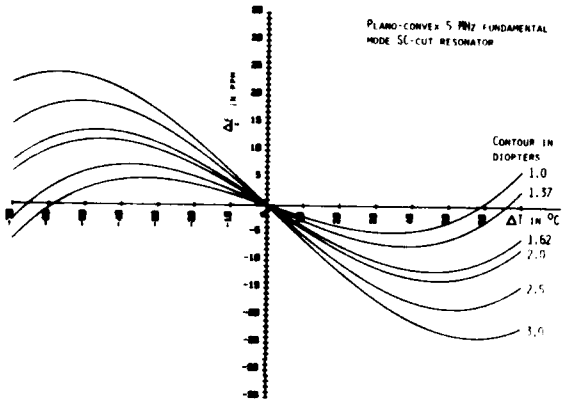


Fig. 5 - Frequency-Temperature vs. Contour

The slope of the frequency vs. temperature curve at the inflection point decreased approximately linearly with increasing blank curvature. For the eight repeatedly reprocessed resonators, the average change in slope was 21.6×10^{-8} per diopter. (The standard deviation was 1.1×10^{-8} per diopter.)

Third Overtone 5 MHz and 10 MHz SC-Cut Resonators

The experiment was repeated with 5 MHz 3rd overtone and 10 MHz third overtone resonators. The nominal angles of cut were $\phi = 21^{\circ}56'$ and $\theta = 34^{\circ}01'$.

The frequency vs. temperature characteristics of six 5 MHz 3rd overtone, 4.0 diopter plano-convex resonators were measured. The blanks were recontoured in 1.0 diopter steps. Three of the blanks were recontoured to 5.0 diopters, then to 6.0 diopters, then to 7.0 diopters, then down to 3.0 diopters. The other three blanks were recontoured to 3.0 diopters, then to 2.0 diopters, then up to 5.0 diopters, then to 6.0 diopters. As before, the angles of cut of the flat sides were not changed during reprocessing.

For these six blanks, the average change in slope at the inflection temperature was 3.33×10^{-8} per diopter. (The standard deviation was 0.46×10^{-8} per diopter.) The slope decreased with increasing blank curvature, as it did for the fundamental mode resonators.

A group of eight 10 MHz 3rd overtone resonators was first measured when the blanks had a contour of 1.0 diopter. Two of the resonators had a small positive slope at the inflection point, the other six had small negative slopes. Each of the blanks was recontoured in 0.5 diopter steps, up to 3.0 diopters. The average decrease in slope at the infection temperature was 2.68×10^{-8} per diopter. (The standard deviation was 0.79×10^{-8} per diopter.)

"Angle Correction" by Contouring

Since the slope of the frequency vs. temperature curve at the inflection point appears to vary linearly with blank contour, when a blank is recontoured, the new slope can be predicted from the following expression:

$$S_N = S_I + K\,(D_N - D_I) \qquad (1)$$

where S_N = slope at the inflection temperature with the new contour
S_I = slope at the inflection temperature with the initial contour
D_N = new contour, in diopters
D_I = initial contour, in diopters

and K =
- -21.6×10^{-8} per $^{\circ}C$ per diopter for 5 MHz fundamental mode resonators
- -3.3×10^{-8} per $^{\circ}C$ per diopter for 5 MHz 3rd overtone resonators
- -2.7×10^{-8} per $^{\circ}C$ per diopter for 10 MHz 3rd overtone resonators

Thus, for example, for 5 MHz fundamental mode resonators, for each 1/4 diopter increase in contour, the slope at the inflection decreases by 5.4×10^{-8} per $^{\circ}C$. To place this into perspective, if the slope at inflection is changed from zero to -5.4×10^{-8} per $^{\circ}C$, the lower turnover temperature is lowered from the inflection point by $17^{\circ}C$, which, according to Figure 3, corresponds to an equivalent θ-angle change of about 1.5 minutes.

The slope, S, at the inflection temperature, T_i, is related to the lower turnover temperature, T_L, by

$$S = -3\,c_i\,(T_i - T_L)^2$$

where c_i is the third order temperature coefficient when the reference temperature is T_i. From the best fit to the data from a selection of 82 frequency vs. temperature curves, ranging from $T_i - T_L = 3^{\circ}C$ to $T_i - T_L = 58^{\circ}C$, the empirical relationship between S and T_L is

$$S = -1.959 \times 10^{-10}\,(T_i - T_L)^2 \qquad (2)$$

This result agrees, to a close approximation, with a curve generated by Ward[8].

For AT-cut crystals, the contour is a critical part of the design. For a given overtone and frequency, there is usually a "best" contour for maximum Q and minimum incidence of activity dips. Even small deviations from this contour can significantly degrade performance.[9] For SC-cut resonators, it has been found that acceptable performance can be achieved over a wide range of contours. For example, for 5 MHz fundamental mode resonators, in the range of 1.25 diopters to 3.0 diopters, neither Q degradation nor activity dip problems have been observed. As the contour is increased from 1.25 diopters to 3.0 diopters, the spacing between the c-mode and the closest inharmonic mode increases by about 50%, while the resistance of the inharmonic modes between the c-mode and b-mode remain much higher than the c-mode resistance. The capacitance ratio also increases by 50%, from about 1200 at 1.25 diopters to about 1800 at 3.0 diopters.[10] No significant change in acceleration sensitivity has been observed between resonators made with 1.37 diopters and those with 2.5 diopters.[11]

Since no adverse performance as a consequence of contour variations has yet been discovered, the contour dependence of the frequency vs. temperature characteristics can be used to "angle correct", i.e., to intentionally change the frequency vs. temperature characteristics, by recontouring the blank.

To determine the new contour required in order to produce a desired slope at the inflection temperature, one can rewrite equation (1) as follows:

$$D_N = D_I - \frac{1}{K}(S_I - S_N) \qquad (3)$$

Equation (3) has successfully been used to produce fundamental mode 5 MHz SC-cut resonators that closely approximate the "ideal" curve shown in Figure 1. Initially, the blanks are ordered to have a low θ-angle such that when resonators are made with a 1.25 diopter contour, most of the resonators have near zero to a small positive slope at the inflection point. The slopes at the inflection point are measured, and the blanks are recontoured in accordance with equation (3).

For example, the process for making blanks for 5.115 MHz fundamental mode resonators consists of final-lapping the blanks to 4.660 MHz with 1 μm aluminum oxide abrasive, using an electrometer for marking the fast-etch sides[12] (i.e., the negative-on-compression sides) with water-insoluble ink, contouring the marked side with 1.25 diopters using 3 μm or 5 μm abrasive, making solder-sealed resonators, measuring the frequency vs. temperature characteristics, stripping the blanks, recontouring in accordance with equation (3), and chemical polishing[13] up to the final blank frequency.

Figure 6 shows the frequency vs. temperture characteristics of one of the first resonators on which the method was tried. (The breaks in this frequency vs. temperature curve are due to the fact that only the three least significant digits of the counter output are plotted. After Δf = 99 Hz, the plotter returns to zero.) The initial contour was 1.37 diopters, which produced a slope at the inflection point of 12.4×10^{-8} per °C. To shift this slope to zero, equation (3) yields $D_N = 1.94$ diopters. The blank was recontoured to 2.0 diopters, the closest available diopter cup. Figure 7 shows the result. This "angle correction" method has also been successfully applied by Ward.[14]

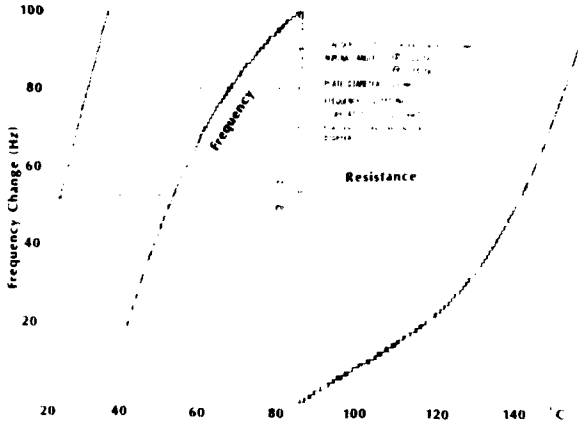


Fig. 6 - SC-cut Frequency vs. Temperature Characteristics Before Correction

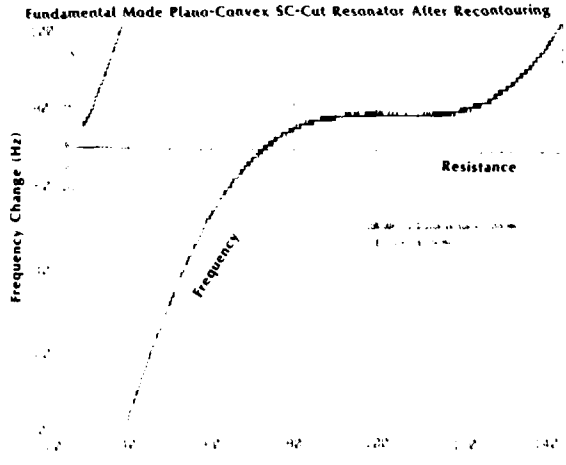


Fig. 7 - SC-cut Frequency vs. Temperature Characteristics After Correction

The Effects of Electrode Dimensions

The effects of electrode diameter and of plateback have been investigated for 5 MHz fundamental mode SC-cut resonators. The electrode material was pure gold for all resonators. In order to maximize the sensitivity to apparent angle-of-cut changes, resonators with near zero slopes at the inflection point were selected.

The Effects of Electrode Diameter

Seven resonators were plated $1f^2$ (i.e. 0.5% mass loading) using a 7 mm diameter circular electrode mask. The blank contours were 1.37 diopters. After the frequency vs. temperature characteristics had been measured, the electrodes were removed, the resonators were replated with the same electrode thickness using a 4 mm diameter electrode mask, and the frequency vs. temperature characteristics were remeasured.

The smaller electrodes shifted the apparent angles of all resonators to higher values. The slopes at the inflection point shifted by an average of -2.6×10^{-8} per ^{o}C. (The standard deviation from the mean was 1.1×10^{-8}.)

According to equation (2), a shift in slope from zero to -2.6×10^{-8} per ^{o}C lowers the lower turnover temperature to 11^{o} below the inflection temperature, which, according to Figure 3, corresponds to an apparent θ-angle change of about 40 seconds.

The Effects of Plateback

Four 1.0 diopter resonators were plated $1f^2$, then $2\frac{1}{2}f^2$, then $4f^2$. To minimize the effects due to electrode dimension variations, each plating step was performed in one step, starting with unplated blanks, rather than overplating the previously deposited electrodes.

The average shift in the slopes at the inflection point varied linearly with plateback, and was -2.5×10^{-8} per f^2 plateback. (The standard deviation was 0.97×10^{-8} per f^2.) This shift corresponds to an apparent θ-angle increase of about 40 seconds per f^2, or to about 1 minute 20 seconds per 1% mass loading.[15]

Conclusions

The frequency vs. temperature characteristics of SC-cut resonators are highly sensitive to blank contour. This sensitivity has two important consequences: 1. the effect can be used to adjust the frequency vs. temperature characteristics, and 2. even when one has "perfect" angle-of-cut control, the contour must be controlled carefully in order to produce resonators of predictable frequency vs. temperature characteristics.

SC-cut crystals' temperature coefficients are also sensitive to electrode dimensions. Reducing the electrode area and increasing the plateback both result in an increase in the apparent θ-angle. Since no adverse consequences, such as activity dips, were observed during the experiments, adjusting the plateback and the electrode area also may be useful for "angle correcting" over a small range.

SC-cut crystals have been found to be remarkably free of activity dips over a wide range of parameters. Of the numerous designs evaluated, only one marginal design has so far produced activity dips. (A small activity dip was repeatedly observed near $20^{o}C$ in fundamental mode 5 MHz resonators with a 1.12 diopter blank contour, however, 1.12 diopter is near the contour below which Q-degradation begins to occur.)

Until economical methods of X-ray orienting and angle correcting SC-cut crystals become available, the use of the SC-cut's contour sensitivity can serve as a means for producing "good" SC-cut resonators. To avoid the need for double processing, an economical method for X-ray orienting is necessary. Once such an orientation method becomes available, unless some adverse consequences of varying the contour are uncovered later, varying the contour may be a competitor to other angle correction techniques. For example, rather than sorting blanks according to the amounts and directions of angles-of-cut shifts necessary, blanks can be sorted according to the contours that will produce the desired frequency vs. temperature characteristics. Whichever angle correction technique is eventually adopted, accurate control of blank contour will still be essential.

Only 14 mm blank diameters were used in these experiments. Due to the relative insensitivity of the designs to blank contour (with respect to Q and activity dips), it is probable that, especially for the 5 MHz fundamental mode resonators, acceptable performance can be obtained with smaller blank diameters.

Acknowledgements

The authors gratefully acknowledge the assistance of John Kosinski and Ed Simon with the crystals' evaluation. The authors also wish to thank Jack Kusters for permission to reproduce Figures 2 and 3.

References

1. J.A. Kusters and C.A. Adams, "Production Statistics of SC (or TTC) Crystals," Proc. 34th ASFC*, pp. 167-174, 1980.

2. R. Burgoon and R. Wilson, "Design Aspects of an Oscillator Using the SC-cut Crystal," Proc. 33rd ASFC*, pp. 411-416, 1979.

3. Quartz Resonator Handbook - Manufacturing Guide for AT-Type Units, edited by R.E. Bennett, prepared for the Department of the Army under contract No. DA36-039-SC-71061 by Union Thermoelectric Div., 1960, pp. 89-103, AD 251289.

4. L.A. Tyler, "The Design of Fundamental Mode Thickness Shear Quartz Resonators (Military Types: CR-18/U and CR-19/U)" prepared for the Department of the Army under contract No. DA36-039-SC-71061 by Union Thermoelectric Div., 1960, pp. 35-41, AD 274031.

5. Jerry Leach, "5 MHz BT-cut Resonators," Proc. 24th ASFC*, pp. 117-125, 1970, AD 746210.

6. Universal Shellac & Supply Co., 495 West John Street, Hicksville, NY 11801.

7. Ref 3, p. 59.

8. Roger Ward, "The SC-cut Crystal - A Review," April 1980, reprints available from Colorado Crystal Corp., 2303 West 8th Street, Loveland, CO 80537.

9. Ref 4, pp. 41-53.

10. R.L. Filler and J.R. Vig, "Fundamental Mode SC-cut Resonators," Proc. 34th ASFC*, pp. 187-193, 1980.

11. R.L. Filler and J.R. Vig, "The Acceleration and Warmup Characteristics of Four-Point-Mounted SC and AT-cut Resonators," elsewhere in these Proceedings.

12. John G. Gualtieri, "A Simple Method for Location of the Mounting Positions for Low Acceleration Sensitivity SC-cut Resonators," R&D Technical Report DELET-TR-81-5, US Army Electronics Research and Development Command, Fort Monmouth, NJ 07703, February 1981, AD A097001.

13. J.R. Vig, R.J. Brandmayr and R.L. Filler, "Etching Studies on Singly and Doubly Rotated Quartz Plates," Proc. 33rd ASFC*, pp. 351-358, 1979.

14. Roger W. Ward, "Design of High Performance SC Resonators," elsewhere in these Proceedings.

15. A. Ballato and T. Lukaszek, "Higher Order Temperature Coefficients of Frequency of Mass Loaded Piezoelectric Crystal Plates," Proc. 29th ASFC*, pp. 10-25, 1975.

* ASFC - Annual Symposium on Frequency Control, US Army Electronics R&D Command, Fort Monmouth, NJ 07703. Copies of the Proceedings are available either from the Electronics Industries Association, 2001 Eye Street NW, Washington, DC 20006, or the National Technical Information Service, Sills Building, 5285 Port Royal Road, Springfield, VA 22161.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

THE ACCELERATION AND WARMUP CHARACTERISTICS OF FOUR-POINT-MOUNTED SC AND AT-CUT RESONATORS

Raymond L. Filler and John R. Vig

US Army Electronics Technology & Devices Laboratory (ERADCOM)
Fort Monmouth, NJ 07703

Abstract

The acceleration sensitivities and warmup characteristics of four-point-mounted SC and AT-cut resonators have been investigated. The acceleration sensitivities were determined from the vibration induced sidebands. For each resonator, measurements were made as functions of vibration direction and vibration frequency. The resonator parameters studied included the angles of cut (AT vs. SC), the mounting orientation, overtone number, blank geometry (plano-plano, plano-convex or biconvex), mounting clip stiffness, and blank material (natural quartz, cultured quartz or swept cultured quartz).

The warmup characteristics were measured by monitoring the resonator frequency with a CI-meter, immersing the resonator in ice water, and after thermal equilibrium had been reached, rapidly immersing the resonator into boiling water. The warmup times of various designs, fundamental-mode and 3rd overtone, AT and SC-cuts are compared. It is shown that four-point-mounted ceramic flatpack enclosed SC-cut resonators are capable of warmup from $0^{o}C$ to within 2×10^{-7} of the $100^{o}C$ frequency in 20 seconds.

Key Words: Quartz, quartz crystals, quartz resonators, SC-cut, AT-cut, acceleration, warmup, frequency control.

Introduction

This is a progress report on characterizing the warmup and acceleration characteristics of ceramic flatpack enclosed quartz crystal resonators of various designs. Fast warmup is desirable in many applications. The effect of rapid temperature change was investigated to determine the parameters that are important for minimizing warmup time. Acceleration sensitivity[1] is becoming a limiting factor in system performance as other noise sources are being diminished. This is true for more than just mobile applications. All precision frequency sources are subject to accelerations, often in the form of building vibrations and shocks transmitted through the floor and amplified by resonances in benches and equipment racks.

Fabrication Details

Resonator Enclosures

The resonators were packaged in the ceramic flatpack enclosures described previously.[2-9] The packages are currently fabricated in two sizes. Both are made from high alumina ceramic, with tungsten feedthroughs. All resonators were sealed at an elevated temperature, in a high vacuum (10^{-7} torr), after a $350^{o}C$ vacuum bake. The larger package, figure 1, is 19mm across the flats and is 5.5mm thick. The blank diameters of the 5MHz and 10MHz resonators enclosed in this package are both 14mm. The smaller package is square, 1cm x 1cm, and is 3.7mm thick. It houses the 6.4mm diameter, 22MHz fundamental mode high-shock-resistant resonators described previously.[8]

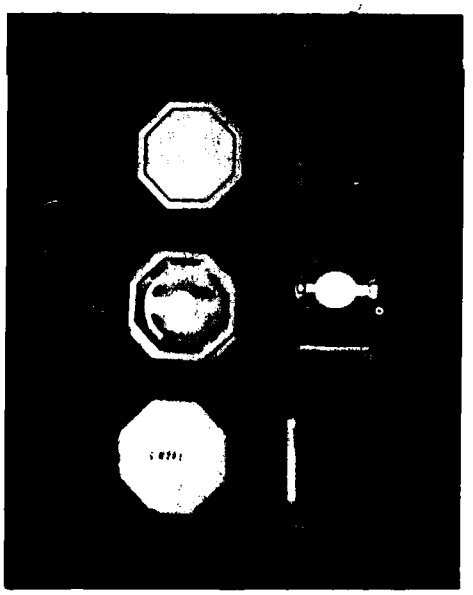


Fig. 1 - Ceramic Flatpack Resonator (Left)

Mounting Structure

The details of the mounting structure are important in both the acceleration (especially

vibration) response, and the warmup behavior. The mounting structure has been described previously.[8] The mounting clip, called the "L" clip, is depicted in Figure 2. The material is 25 μm thick molybdenum ribbon bent into an inverted "L" shape. The clips are brazed to the ceramic enclosure, and the quartz plate is bonded to the clips using a silver filled polyimide adhesive.[8]

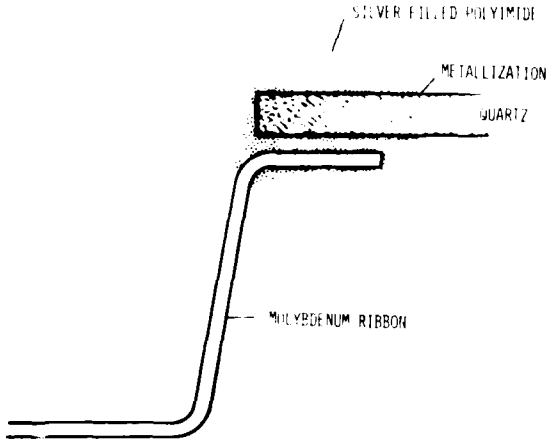


Fig. 2 - "L" Clip

The plano-convex plates are mounted onto the clips so that the plano sides are resting on the clips. Since the two faces of an SC-cut plate are not equivalent, it was decided in the early stages of these experiments to control the SC-cut blank faces to be contoured. Most of the plano-convex SC-cut plates were contoured on the fast-etch sides[10] (i.e., the negative-on-compression sides).

The width of the clip was 0.75 μm for the 22MHz units and 1.5 μm for the 5MHz and 10MHz units. The height of the clip was 0.70mm for the 22MHz units and also for the 5 and 10MHz units with the designation "short clips". The height of the clips designated "long clips" was 1.1mm.

The orientation of the clips with respect to each other and with respect to the crystallographic axes of the quartz plate was varied. The 22MHz units had 4 clips located either ±45° from the ZZ' axis, (at ψ = 45°, 135°, 225°, 315°), called the 90° mount, or located ±30° from the ZZ'axis (ψ = 60°, 120°, 240°, 300°), and called the 60° mount. Three different mounting configurations were used in the larger package. In one, the 4 clips were spaced 90° apart and located on the XX' and ZZ' axes (ψ = 0°, 90°, 180°, 270°). The second configuration also used 4 clips spaced 90° apart, but the location was rotated by 14° from the previous configuration, (ψ = 14°, 104°, 194°, 284°). The third configuration consisted of three clips located at ψ = 14°, 194° and 284°.

Acceleration Sensitivity

Measurement Procedure

The acceleration sensitivities were determined from the vibration induced sidebands, as is described elsewhere in these Proceedings.[1] In this report, the sensitivities are described by $|\Gamma|$, which is the magnitude of the acceleration sensitivity coefficient in the most sensitive direction.

Parameters Investigated

The effects of several fabrication parameters on the acceleration sensitivity coefficient were investigated. The first was the angle of cut: AT-cut vs. SC-cut of the same frequency and overtone. The second parameter was the mounting configuration: three-point vs. four-point mounting, and the clip orientations. The third parameter was the geometry of the quartz plate: plano-plano vs. plano-convex vs. bi-convex, and the thickness of the quartz plate. The fourth parameter was the type of quartz material used: natural quartz vs. cultured quartz vs. swept cultured quartz.

Experimental Results

The average values for the magnitude of the acceleration sensitivity vector, $|\Gamma|$, for all of the resonators measured is given in Table I. The right-most column gives the maximum and minimum value for each group. Figure 3, is a graphical summary of Table I. The vertical axis is the logarithm of the average value of $|\Gamma|$, with the limits of the bars being the maximum and minimum of the group. The numbers correspond to the line in

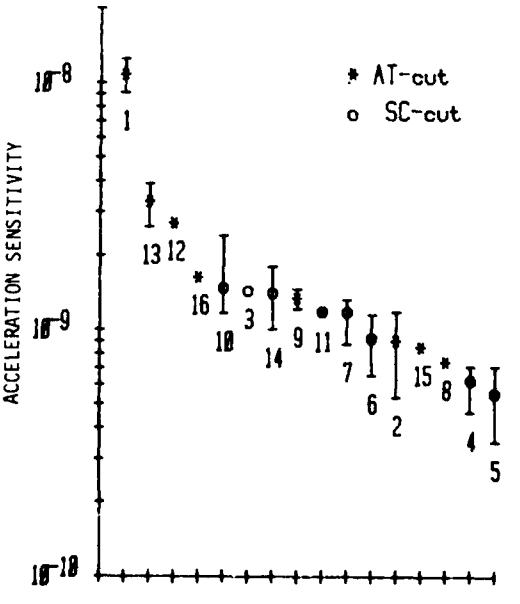


Fig. 3 - Graphical Summary of Table I (See Text)

TABLE I

	FREQ. (MHz)	CUT	O/T	GEOM.	MOUNT	QUARTZ MATERIAL	$\|\Gamma\|$ ($\times 10^{-10}$/g)	(MAX/MIN) ($\times 10^{-10}$/g)
1	5.115	AT	1	Plano-Convex	4pt. (X-Z) Short	Natural	108.	(125/91)
2	"	"	"	Bi-Convex	"	"	9.0	(11.8/53)
3	"	SC	"	Plano-Convex	3pt. (14°) Short	"	14.3	*
4	"	"	"	"	4pt. (14°) Short	"	6.2	(7.1/4.6)
5	"	"	"	"	4pt. (X-Z) Short	"	5.5	(7.1/3.5)
6	"	"	"	"	4pt. (X-Z) Long	"	9.2	(11.5/6.5)
7	"	"	"	"	"	Swept Cultured	11.7	(13.2/8.7)
8	"	AT	3	"	4pt. (X-Z) Short	Natural	7.4	*
9	"	"	"	"	4pt. (X-Z) Long	"	13.4	(14.6/12.1)
10	"	SC	"	"	"	Unswept Cultured	14.7	(24.0/10.5)
11	"	"	"	"	"	Swept Cultured	11.8	(12.6/11.6)
12	10.0	AT	3	"	4pt. (X-Z) Short	Natural	27.	*
13	"	"	"	"	4pt. (X-Z) Long	"	33.	(39/26)
14	"	SC	"	"	"	Unswept Cultured	14.	(18/10)
15	21.9	AT	1	Plano-Plano	4pt. (90°) Short	Natural	8.5	*
16	"	"	"	"	4pt. (60°) Short	"	16.3	*

*Only one unit of this type was evaluated to date.

Table I for each group of resonators. The groups were arranged in order of acceleration sensitivity for clarity. There is no physical significance to the horizontal axis.

It is interesting to note that the acceleration sensitivity of 5MHz fundamental mode AT-cut resonators decreases by a factor of 12, to 9×10^{-10} per g, by changing the geometry from plano-convex (line 1) to the more symmetrical bi-convex shape (line 2). The 22MHz units (line 15) are also more symmetrical, being plano-plano. The acceleration sensitivity for the 22MHz units with 90° mounting is 8.5×10^{-10}/g. The sensitivity for 60° mounting (line 16) is about twice that of the 90° mounting. The units of lines 1, 2, 15 and 16 were fabricated with natural quartz and were mounted on "short" clips.

Comparison of the 3rd overtone plano-convex 5 MHz AT-cut units (line 8) to the fundamental mode plano-convex 5MHz AT-cut units (line 1) shows a decrease in sensitivity by nearly a factor of 15. These overtone units are virtually identical in performance to the fundamental mode bi-convex AT-cut units (line 2). The 5MHz 3rd overtone AT-cut units with long clips (line 9) are about a factor of 2 worse than the similar units with short clips (line 8).

The 10 MHz third overtone AT-cut units, both with short (line 12) and long (line 13) clips, showed disappointing performance. These units had four times poorer sensitivity than the 5MHz 3rd overtone AT-cut units with short clips. The contour on the 5MHz 3rd overtone units was about twice (4.0 diopters) that on the 10MHz 3rd overtone units (1.62 or 2.0 diopters).

The lowest sensitivity was obtained with fundamental mode 5MHz plano-convex SC-cut resonators (line 5). Rotation of the mounting orientation by 14° (line 4) from the ψ = 0°, 90°, 180°, 270° orientation did not significantly change performance. Removal of one clip from the 14° orientation in an attempt to produce the configuration predicted[11] for optimum performance (line 3) resulted in a rather significant increase in the acceleration sensitivity. Units similar to those of line 5, but with long clips (line 6), also showed an increase in acceleration sensitivity. Substitution of swept cultured quartz (line 7) produced almost no change in performance from that of the natural quartz units (line 6).

The performance of the third overtone 5 MHz SC-cut units was also disappointing. Neither the units fabricated from swept cultured quartz (line 11) nor those fabricated from unswept cultured quartz (line 10) had acceleration sensitivities lower than any of the four-point mounted SC-cut fundamental mode units (line 4-7). It does not appear that there is a difference between units made with swept or unswept cultured quartz, or between units made with cultured and natural quartz. The range of results in each group is larger than any differences between the groups.

Third overtone 10MHz SC-cut units were fabricated from unswept cultured quartz only (line 14). These units had approximately the same sensitivity as the 5MHz 3rd overtone SC-cut units (lines 10 and 11), and were better by a factor of two when compared to 10MHz 3rd overtone AT-cut units (lines 12 and 13).

The conclusions one can draw from the results in Table I are as follows:

1. The best SC-cut designs evaluated to date have failed to produce a substantial improvement over the best AT-cut designs.

2. For AT-cut resonators, the acceleration sensitivity along the direction normal to the blank is significantly higher than along the worst direction in the plane of the blank. For SC-cut resonators, the sensitivity along the normal to the blank is comparable to the sensitivity along the worst direction in the plane of the blank i.e. [illegible] is nearly along the normal to the blank for AT-cut resonators, but not near the normal nor near the plane of the blank for SC-cut resonators.

3. For fundamental mode AT-cut resonators, symmetrical (i.e. plano-plano and biconvex) blank geometries produce a significant improvement in acceleration sensitivity over plano-convex geometries. Although only plano-convex SC-cut blank geometries have been evaluated in these studies to date, no systematic variation in acceleration sensitivities was observed for fundamental mode SC-cut resonators over a 1.0 diopter to 2.5 diopter contour range. The importance of symmetry for AT-cut resonators agrees with the results obtained previously by others.[12,13,14] Since SC-cut crystal plates are inherently asymmetrical in the thickness direction, the meaning and significance of a "symmetrical" mounting configuration for SC-cut crystals is less clear-cut.

4. For plano-convex AT-cut resonators, the third overtone units have substantially lower acceleration sensitivities than the fundamental mode units. On the other hand, among the plano-convex SC-cut resonators, the 3rd overtone units have so far not shown an improvement over the fundamental mode units.

5. Based on these experiments and on experiments reported by others perviously,[15,16] finding a design that can consistently produce a few parts in 10^{11} per g acceleration sensitivity has been elusive to date. In applications where less than 1×10^{-10} vibration sensitivity is required, pairs of resonators[9,17] or active compensation[18,19,20] methods need to be used.

Warmup Characteristics

The warmup time of a crystal resonator is the time required for the frequency to arrive to within some specified offset from the equilibrium frequency after a step change in temperature. The warmup time, in addition to depending on the temperature step and the frequency offset chosen, depends on the thermal properties of the enclosure with its mounting clips, and on the response of the resonator to thermal transients. To minimize warmup time, the resonator should be thermal transient compensated, that is, the frequency vs. temperature characteristic during warmup should be the same as the static frequency vs. temperature characteristic.

Parameters Investigated

Most of the resonators used in this warmup study had a frequency of 5MHz, were enclosed in the ceramic flatpack shown in Figure 1, and were supported on four mounting clips located 90° apart. The parameters investigated were: 1. blank thickness, (i.e. the overtone number), 2. angles of cut, (i.e. AT vs SC), and 3. clip length. The warmup characteristics of 22MHz fundamental mode AT-cut resonators, enclosed in the smaller flatpack, were also investigated.

Experimental Procedure

The warmup characteristics were measured by monitoring the resonator frequency with a CI-meter, immersing the resonator in ice water, and after thermal equilibrium had been reached, rapidly immersing the resonator into boiling water. The curve in Figure 4 illustrates the results obtained for an AT-cut fundamental mode resonator. The vertical axis is frequency; the horizontal axis is time.

The details of the curve require some explanation. The dynamic range of the digital-to-analog converter used to interface the frequency counter to the recorder was 1000Hz; the resolution was 1Hz. The graph, therefore, only represents the three least significant digits of the 5MHz frequency. Every time the frequency passes through a multiple of 1000Hz the plotter pen returns to zero. For example, in figure 4, the 0°C frequency is 5,115,294Hz. The line labeled 0°C is at 294Hz. The 100°C line is at 710Hz, which appears to be higher, but due to the "roll-over" it is actually lower, being at 5,114,710Hz.

Results

In figure 4, the large undershoot that is characteristic of AT-cut resonators takes place during the initial 5 seconds. At 5 seconds, the frequency **vs. time** curve's slope is reversed. If the frequency had tracked the static frequency vs. temperature curve, the frequency vs. time curve would have approached the 100°C frequency from above. The frequency is within 0.2 ppm (1Hz) of the final frequency after 47 seconds. For comparison, it took an HC-36 enclosed, 2 point-mounted AT-cut resonator about 22 seconds just to turn around from the undershoot. The 22MHz fundamental mode AT-cut crystals in the smaller flatpacks turned around from the undershoot in 1.5 seconds, after experiencing an undershoot of 1.5 kHz.

Figure 5 shows the results for the same experiment performed on an SC-cut fundamental mode resonator. Note the absence of any significant

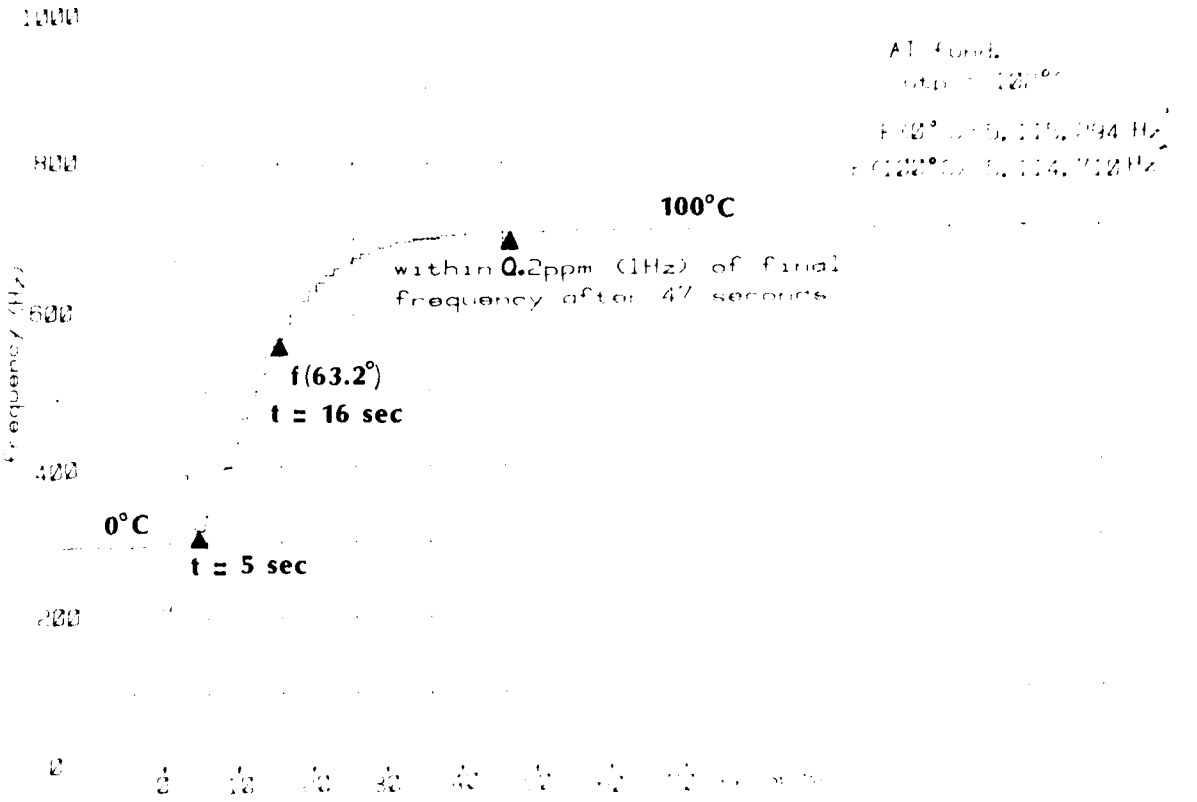


Fig. 4 -Warmup of 5MHz AT-cut Fundamental Mode Resonators

overshoot. The frequency tracks the static frequency vs. temperature curve nearly perfectly. This resonator was within 0.2 ppm of the final frequency after 24 seconds. If there had been no overshoot, rather than the 0.4 ppm observed for this particular unit, the time to warmup to within 0.2 ppm would have been less than 20 seconds.

Table II is a summary of the results for several types of 5 MHz resonators. It can be seen that the SC-cut units warm up substantially faster than the AT-cut units. The "time constant" column, due to the minimal overshoot of the SC-cut resonators, is the thermal time constant of the ceramic flatpack enclosed resonators. It is interesting to note that the time constant for the long-clip fundamental mode unit is nearly identical to that of the long-clip third overtone unit (that has a quartz plate which is three times more massive).

It has been previously stated[21] that the warmup time for a third overtone unit is about three times that of a fundamental mode unit. A simple model will show why that is not necessarily so. The thermal equivalent circuit of a resonator is shown in Figure 6. The two thermal resistances are the clips (R_c) and the quartz plate (R_p). The quartz plate's heat capacity, represented by a capacitor (C_p), is much greater than the heat capacity of the clips. The thermal time constant of the clip-quartz plate system is therefore given by

$$\tau = (R_c + R_p)\, C_p \qquad (1)$$

The thermal resistance of a fundamental mode quartz plate is three times that of a third overtone plate of the same frequency because it has 1/3 the thickness, i.e.

$$R_p\,(\text{fund}) = 3\, R_p\,(\text{3rd overtone}) \qquad (2)$$

Also, the heat capacity of a fundamental mode plate is three times smaller than that of a third overtone plate because it has 1/3 the mass, i.e.

$$C_p(\text{fund}) = 1/3\, C_p\,(\text{3rd overtone}) \qquad (3)$$

Therefore, the term R_pC_p is independent of overtone. If R_c is much greater than R_p, then from equation (1),

$$\tau = R_cC_p \qquad (R_c >> R_p) \qquad (4)$$

and therefore

$$\tau(\text{3rd O/T}) = 3\,\tau(\text{fund}); \qquad (5)$$

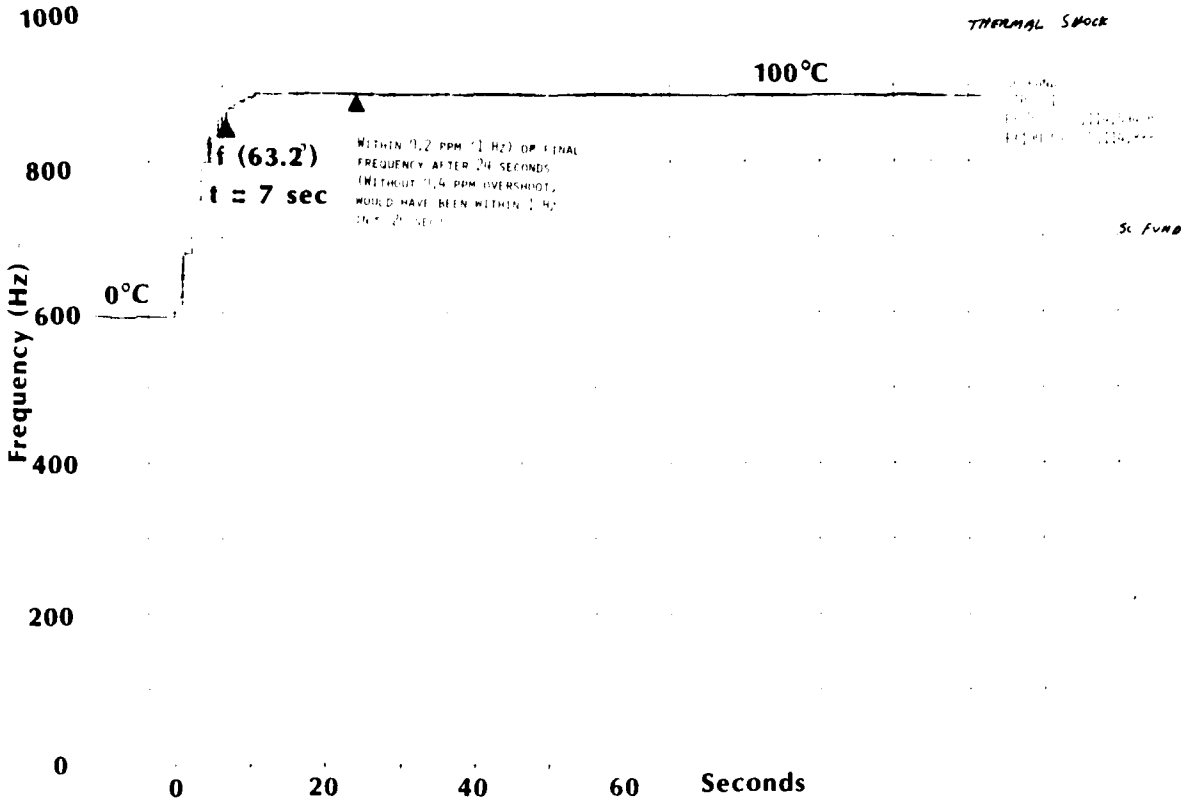


Fig. 5 -Warmup of 5MHz SC-cut Fundamental Mode Resonators

TABLE II

FREQUENCY/ OVERTONE/ CUT	CLIP LENGTH*	TIME CONSTANT (SECONDS)	FREQUENCY OVERSHOOT (PPM)	WARMUP TO 2×10^{-7} of F (100°C)
5 MHz/1/SC	SHORT	7	0.5	24 SEC
5 MHz/1/SC	LONG	10	0.2	26 SEC
5 MHz/3/SC	LONG	12	0.8	----
5 MHz/1/AT	SHORT	--	77	47 SEC
5 MHz/3/AT	SHORT	--	54	84 SEC (UTP= 92°C)
5 MHz/3/AT	SHORT	--	44	----

*SHORT = 0.76 mm
LONG = 1.5 x SHORT

The situation is different if the thermal resistance of the clips is much less than that of the quartz plate. In that case

$$\tau = R_p C_p \qquad (R_c << R_p) \qquad (6)$$

which is independent of overtone.

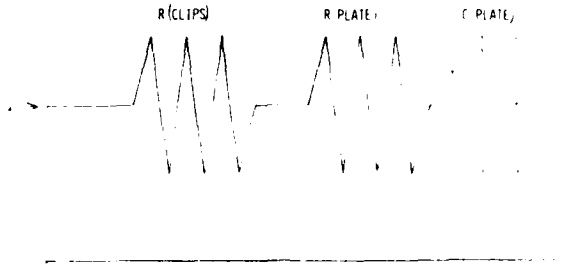


Fig. 6 - Thermal Equivalent Circuit of a Ceramic Flatpack Resonator

In the ceramic flatpack, the geometry is such that the ratio of the thermal resistance of the quartz plate to the thermal resistance of the clips is approximately equal to the ratio of the thermal conductivity of molybdenum to quartz. Since the thermal conductivity of molybdenum is about ten times that of quartz, for the ceramic flatpack enclosure, R_c is much less then R_p, and equation (6) is a close approximation.

It has been shown, therefore, that the thermal time constant of ceramic flatpack enclosed resonators is approximately independent of blank thickness. Properly designed ceramic flatpack enclosed 5MHz SC-cut resonators can warm up to within 2 x 10^{-7} of the equilibrium frequency in less than 20 seconds after a 100°C step change in temperature. The potential therefore exists for a warmup time of substantially less than one minute by using fast warmup oscillators with ceramic flatpack resonators.

Acknowledgements

The authors wish to thank J. Keres, T. Snowden and the resonator development group at GEND for their assistance in fabricating the resonators used in these experiments. Thanks are also due to E. Simon and D. Boyce of ERADCOM for their assistance in evaluating the resonators, and to W. Washington, R. Brandmayr, J. Gualtieri and H. Spaights of ERADCOM for their assistance in preparing the resonators blanks.

References

1. R.L. Filler, "The Effects of Vibration on Frequency Standards and Clocks," These proceedings.

2. J.R. Vig and E. Hafner, "Packaging Precision Quartz Crystal Resonators," Technical Report ECOM-4134, US Army Electronics Command, Fort Monmouth, NJ, July 1973, AD 763215*.

3. P.D. Wilcox, G.S. Snow, E. Hafner and J.R. Vig, "A New Ceramic Flatpack for Quartz Resonators," Proc. 29th ASFC**, pp. 202-219 (1975), AD AU17466*.

4. R.D. Peters, "Ceramic Flatpack Enclosures for Precision Quartz Crystal Units," Proc. 30th ASFC**, pp. 224-231 (1976), AD A046089*.

5. J.R. Vig, J.W. LeBus and R.L. Filler, "Chemically Polished Quartz," Proc. 31st ASFC**, pp. 131-143 (1977), AD A088221*.

6. J.R. Vig, R.J. Brandmayr and R.L. Filler, "Etching Studies on Singly and Doubly Rotated Quartz Plates," Proc. 33rd ASFC*, pp. 351-358 (1979).

7. J.R. Vig, "UV/Ozone Cleaning of Surfaces: A Review," In Surface Contamination: Genesis, Detection and Control, K.L. Mittal ed., Vol. 1, pp. 235-254, Plenum Press, New York, 1979.

8. R.L. Filler, J.M. Frank, R.D. Peters and J.R. Vig, "Polyimide Bonded Resonators," Proc. 32nd ASFC**, pp. 290-298 (1978).

9. R.L. Filler, L.J. Keres, T.M. Snowden and J.R. Vig, "Ceramic Flatpack Enclosed SC and AT-cut Resonators," 1980 Ultrasonics Symposium Proceedings, IEEE, (1980).

10. J.G. Gualtieri, "A Simple Method for Location of the Mounting Positions for Low Acceleration Sensitivity SC-cut Resonators," R&D Technical Report DELET-TR-81-5, US Army Electronics Research and Development Command, Fort Monmouth, NJ 07703, February 1981, AD A097001*.

11. P.C.Y. Lee and K.M. Wu, "Nonlinear Effect of Initial Stresses in Doubly-Rotated Crystal Resonator Plates," Proc. 34th AFCS**.

12. A.W. Warner and W.L. Smith, "Highly Stable Crystal Oscillators for Missile Applications," Proc. 13th AFCS**, pp. 191-206, (1959), AD 298325*.

13. D.L. Hammond, "Precision Quartz Resonators," Proc. 15th AFCS**, pp. 125-138, (1961), AD 265455*.

14. R. Besson, J.J. Gagnepain, D. Janiand, M. Valdois, "Design of a Bulk Wave Quartz Resonator Insensitive to Acceleration," Proc 33rd AFCS**, pp. 337-343, (1979).

15. J.A. Kuster and C.A. Adams, "Production Statistics of SC (or TTC) Crystals," Proc. 34th AFCS**, (1980), pp. 167-174.

16. B. Goldfrank, J. Ho and A. Warner, "Update of SC-cut Crystal Resonator Technology," These proceedings.

17. A. Ballato, "Resonator Compensated for Acceleration Fields," Proc. 33rd AFCS**, (1979), pp. 322-336.

18. J.M. Przyjemski, "Improvement in System Performance Using a Crystal Oscillator Compensated for Acceleration Sensitivity," Proc. 32nd AFCS, (1978), pp. 426-431.

19. D.A. Emmons, "Acceleration Sensitivity Compensation in High Performance Crystal Oscillators," Proc. 10th PTTI, NASA Tech. Memorandum 80250, Greenbelt, MD (1978), pp. 55-82.

20. V.J. Rosati and R.L. Filler, These Proceedings.

21. W.H. Hicklin and A.L. Bennett, "Dynamic Temperature Behavior of Quartz Crystal Units," R&D Technical Report, ECOM-0440-F, US Army Electronics Command, Fort Monmouth, NJ, 1970, NTIS Accession Nr. AD 878448*.

*Copies available from National Technical Information Service (NTIS), Sills Building, 5285 Port Royal Road, Springfield, VA 22161.

**Annual Symposium on Frequency Control, US Army Electronics R&D Command, Fort Monmouth, NJ 07703. Copies of the Proceedings are available either from NTIS or the Electronics Industries Association, 2001 Eye Street, NW, Washington, D.C. 20006.

Proc. 35th Ann. Freq. Control Symposium, USERADCOM, Ft. Monmouth, NJ 07703, May 1981

REDUCTION OF THE EFFECTS OF VIBRATION ON SC-CUT QUARTZ CRYSTAL OSCILLATORS

VINCENT R. ROSATI and RAYMOND L. FILLER

US Army Electronics Technology & Devices Laboratory (ERADCOM)
Fort Monmouth, NJ 07703

Abstract

To provide the necessary spectral purity for modern communications, navigation, and radar systems, reduction in the acceleration sensitivity (γ) of quartz resonators, e.g., from $\gamma = 10^{-10}/g$ to $\gamma = 10^{-12}/g$, is required. Several passive (i.e., involving the resonator only) methods to accomplish this reduction are being pursued. Two examples of the passive approach are the proper design of single resonators, and dual resonators. An active method (involving electronic circuitry) has also been investigated and shown to be capable of an additional reduction by a factor of about 50. However, the nonlinearity of the varactors used for implementation of this method limit the ultimate amount of reduction possible.

The work presented here is an alternative active method which exploits the linear voltage/frequency characteristic of SC-cut crystals. The method consists of applying a properly phased electrical replica of the vibration directly to the crystal's electrodes. A series of experiments, using several different SC-cut crystals, was conducted to determine the efficacy of the approach.

A crystal oscillator was modified to permit the application of the compensation voltage without affecting the oscillator circuitry.

The range over which the compensation was tested were: acceleration - 1 to 14 g; vibration frequency - 50 to 500 Hz; magnitude of compensating voltage - 20 to 20 V P-P, depending on the initial value of γ.

Early results show that γ can be reduced by almost three orders of magnitude at discrete frequencies, and by one order of magnitude over the 50 to 500 Hz range.

KEY WORDS

Quartz, quartz crystals, quartz resonator, quartz oscillator, crystal oscillator, SC-cut, acceleration, vibration, phase noise, acceleration compensation.

Introduction

Many modern communications, navigation, and radar systems must operate from vibrating platforms such as aircraft, tanks, trucks, etc. To realize the systems' full potential it is necessary that the effects of vibration on frequency and phase

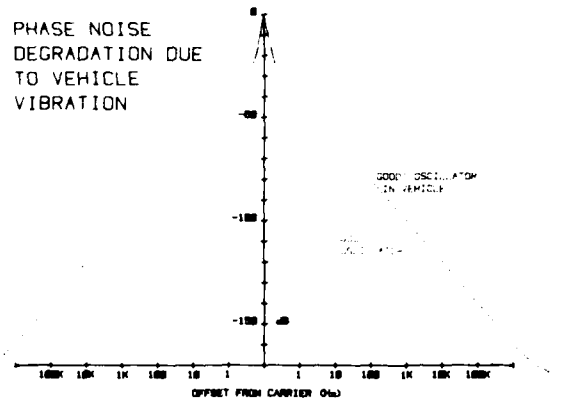


FIGURE 1: Phase noise degradation of a high quality oscillator subjected to 1 g random vibrations.

noise be greatly reduced. Figure 1 shows the vibration-induced phase noise degradation of a "good" oscillator. The lower curve is a plot of the phase noise specifications of a typical high quality oscillator. This curve represents the performance to be expected when the oscillator is operated in a vibration-free environment. The upper curve is a plot of the calculated phase noise when the same oscillator is subjected to uniform, broad-band vibration. The calculation was based on an acceleration of 1 g and a vibration sensitivity of 2×10^{-9} per g. It is readily seen that when low phase noise is required, improvement of the vibration resistance of crystal oscillators is necessary, e.g., from parts in 10^9 per g to parts in 10^{12} per g if "laboratory" performance is to be maintained while the oscillator is under vibration. (Hereafter, the vibration sensitivity will be referred to by the use of "γ" when referring to the scalar quantity and "$\vec{\Gamma}$" when referring to the vector quantify.[1])

Acceleration Sensitivity Minimization

- Isolation — Passive Methods
- SC-Cut
- Enantiomorphic Pairs

- Accelerometer Feedback — Active Methods
  - Varactor
  - Direct-to-Crystal

TABLE 1

Various approaches to minimizing the effects of vibration on quartz crystal oscillators.

Table 1 summarizes the methods which are available to reduce the vibration sensitivity of crystal oscillators. First among these is vibration isolation. That is, the oscillator is suspended by a support structure designed to minimize the transmission of the vehicle's vibrations to the oscillator. However, inspection of Figure 2 leads one to conclude that isolation is one of the least attractive alternatives. In an undamped system, there is always a region of amplification, exacerbating the problem. If the vibrations that must be suppressed are low frequency (common in some vehicle-borne applications), a very "soft" isolation system is needed, implying a structure inconsistent with the trend toward compactness. On the other hand, a damped isolation system, while having less amplification, also has less isolation.

Another approach is to use a resonator which is, itself, less sensitive to vibration. Much has been written about the SC-cut crystals' resistance to vibration [2,3,4] and no further treatment will be given here except to say that sensitivities of the order of parts in 10^{10} per g have been reported.

Still another technique is the use of paired resonators, [5,6] the axes of which are aligned anti-parallel, in order to minimize vibration sensitivity.

The foregoing methods are termed "passive", i.e., there is no electronic sensing, amplification, or feedback involved. The following methods are termed "active", i.e., the vibratory motion is sensed, amplified, and fed back to the oscillator (or crystal) by electronic circuitry.

The first active method, reported by Pryzjemski[7], uses an accelerometer-derived signal which is fed back to the frequency-trim varactor in the oscillator circuit, phased so that the varactor drives the oscillator frequency in the direction opposite to the acceleration-induced frequency shift. This method has the advantage that it can be used with any type of crystal cut, but suffers from the nonlinear behavior of the varactor, as will be shown.

A second active approach has been investigated and is reported here. It exploits the linear voltage-frequency property of SC-cut crystals [8,9].

The Voltage-Frequency Effect

What we have termed the "voltage-frequency effect" is also known as the Polarizing Effect. The bulk of the investigation of this phenomenon has been conducted by Hruska, with important contributions by Kusters. (The reader is directed to references 8 and 9, which give more complete bibliographies.)

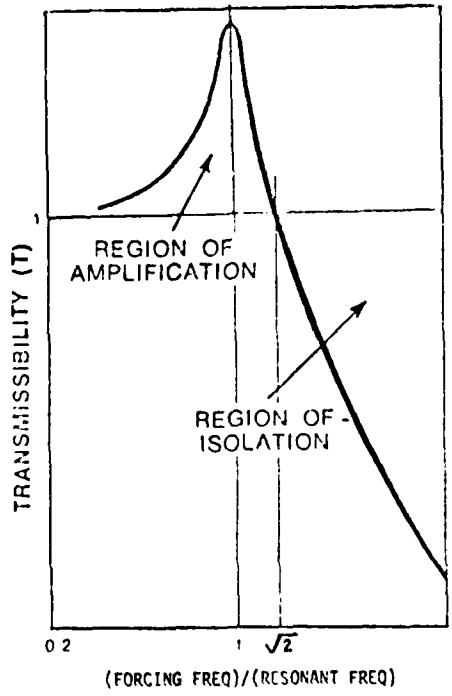


FIGURE 2: Generalized curve of the transmissibility vs. frequency for a spring-type vibration isolation system. Note the presence of a region of amplification.

Simply stated, the voltage-frequency effect is the change in resonant frequency of piezoelectric resonators due to an impressed electric field. The frequency change is a highly linear function of the applied voltage. Laboratory measurements show that the linearity is at least as good as 1 part per 10^5 up to + 40 volts. The measured voltage sensitivity[10] for 5 MHz, fundamental mode, SC-cut crystals is

$$\frac{\Delta F}{F} \times \frac{1}{V} \quad 7 \times 10^{-9}/\text{volt},$$

in agreement with Hruska and Kusters. AT-cut crystals theoretically have zero voltage sensitivity, but values up to 5×10^{-10}/volt have been observed.[11] However, this low value renders the direct-to-crystal vibration suppression method impractical for AT-cut crystals. For SC-cut crystals, vibrations of 10 g can be highly suppressed by the application of about 20 volts.

Direct-to-Crystal Vibration Compensation

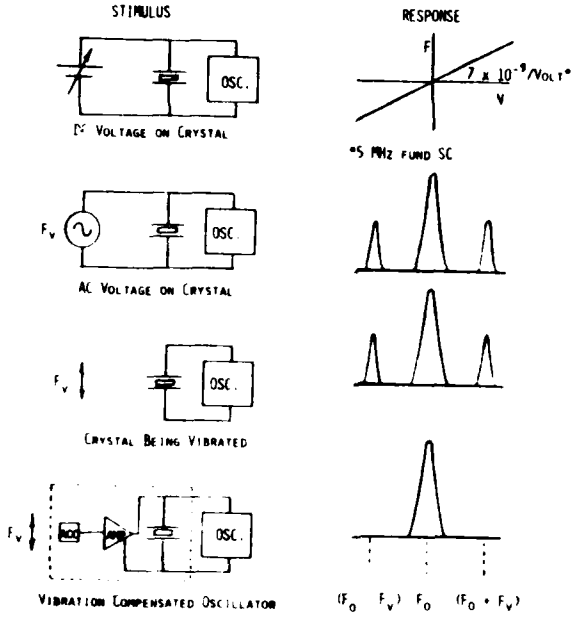


FIGURE 3: Development of the direct-to-crystal method for suppressing vibration-induced sidebands.

Reference to Figure 3 will expose the rationale of the direct-to-crystal approach. The uppermost set of drawings illustrates the linear nature of the SC-cut voltage-frequency effect. Next, if the stimulus is changed from DC to AC, observation of the oscillator output on a spectrum analyzer will show sidebands occuring above and below the carrier frequency (F_O), separated from it by the excitation frequency (F_V). The next lower set of drawings shows a spectrum analyzer display of a crystal oscillator vibrated at a frequency F_V. Notice that the two spectra are identical. The bottom set of drawings shows the result of applying an electrical replica of the vibration frequency directly to the crystal, with amplitude and phase adjusted to cause suppression of the vibration-induced sidebands.

To compare the varactor and direct-to-crystal approaches, an experiment was performed wherein the suppression voltage was derived from an audio oscillator which was phase-locked to the driving frequency of the shake table upon which commercially available test oscillators were mounted.

The phase-locked oscillator was used to facilitate adjustment of the magnitude and phase of the suppression voltage. Figure 4 shows the results of the experiment. The upper curve is the sideband level of the uncompensated oscillator. The lower curve, marked with "X", is the result of applying the suppression voltage to the oscillator's frequency trim varactor. The curve marked with "0" is the result of applying the suppression directly to the crystal. (The adjustment of the amplitude and phase to achieve greater than 50dB suppression was very tedious. It is the authors' opinion that both curves would be congruent had not our patience run out.)

The two active methods appear to offer equivalent suppression. However, another experiment was performed and results show that they do not.

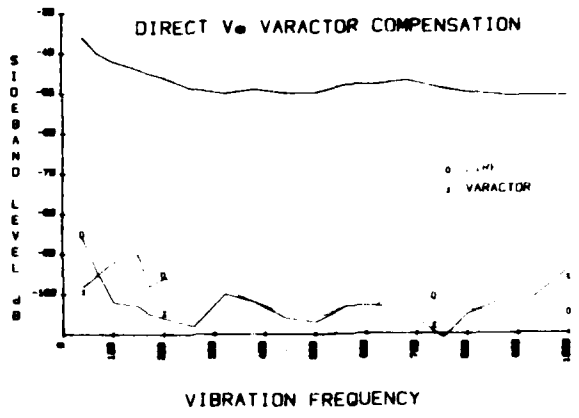


FIGURE 4: Sideband suppression obtainable at discrete frequencies using varactor and direct-to-crystal schemes.

Effect of Varactor Nonlinearity

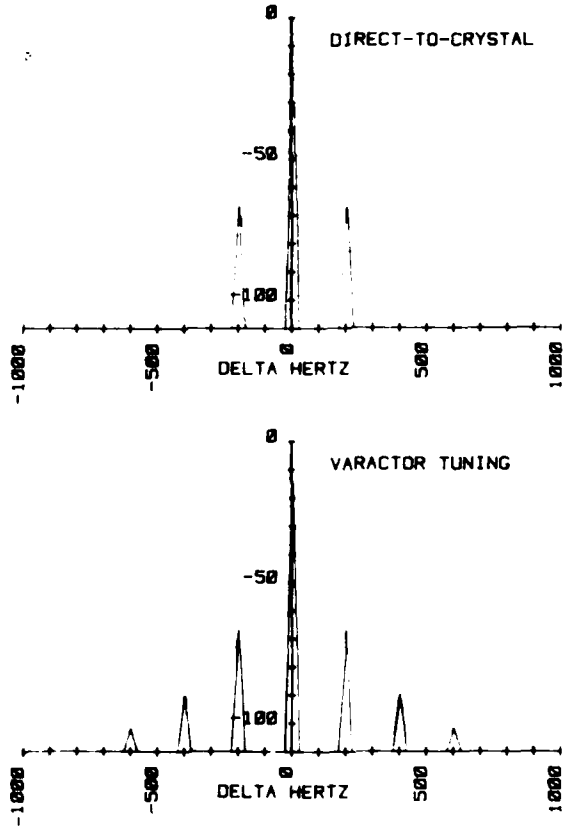


FIGURE 5: "Artificial" sidebands generated by applying a 200 Hz voltage directly to the crystal electrodes (upper spectrum) and to the frequency trim varactor (lower spectrum).

Shown in Figure 5 are plots of data taken from the spectrum analyzer for the following conditions: The upper spectrum is the result of applying a 200 Hz voltage directly to the crystal. Sidebands are present at + 200 Hz, as expected. The lower spectrum is the result of applying the same voltage to the varactor. Note the presence of the second and third harmonics, representing a total harmonic distortion of about 8 percent. These harmonics are due to the nonlinearity of the varactor. The varactor built into the oscillator circuit was intended only for the adjustment of frequency by the application of a DC voltage so linearity was not important. Lower harmonic content would be obtained by substituting a more linear varactor, but unless a perfectly linear voltage/capacitance characteristic becomes available, some harmonics will still be generated. Therefore, if an active suppression scheme is implemented by the varactor method, suppression of one frequency gives rise to harmonics. If broadband (i.e., random) vibrations are considered, the overall effect will be that the degree of possible suppression is substantially reduced. Clearly, a linear suppression scheme is preferred.

Broadband Behavior of the Direct-to-Crystal Scheme

A crystal oscillator was mounted on a shake table and subjected to vibrations with a magnitude of 8 g. The frequency was swept from 50 to 1000 Hz. The acceleration sensitivity, γ, was calculated, making use of the sideband amplitudes measured on a spectrum analyzer. Figure 6 is a plot of γ versus frequency, arrived at by the formula

$$\gamma = \frac{2Fv}{aF_o} \; 10^{\alpha/20},$$

where F_v is the vibration frequency, a is the acceleration, F_o is the center frequency of the oscillator, and α is the sideband amplitude in dBc.

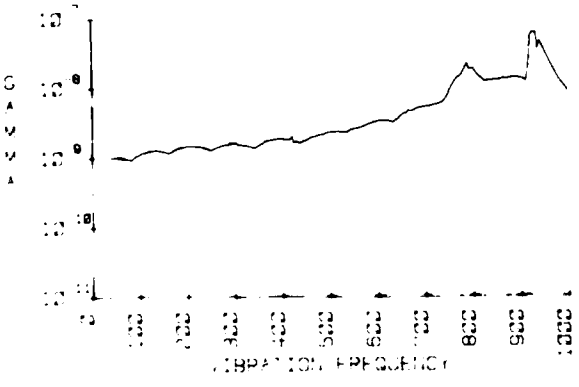


FIGURE 6: Acceleration sensitivity (γ) of a crystal oscillator vs. vibration frequency. Resonant peaks are evident at 800 and 950 Hz.

For an ideal crystal/oscillator combination the curve of γ vs. vibration frequency should be flat. The positive slope of the curve is due to the resonances occuring at about 800 and 950 Hz. These resonances are most likely due to the circuit board and the crystal mount. (Several oscillators of the same design showed similar resonances.) It should be pointed out that great care should be taken in the mechanical design of crystals and oscillators to avoid any resonances, since more and more applications are surfacing which require precision oscillators to be operated in vibratory environments.

The test oscillator was modified by adding 100 pF capacitors in series with each side of the crystal, and by connecting 2.5 megohm resistors to each crystal terminal. The compensation signal was fed to the crystal through the resistors. The signal itself was generated by an accelerometer mounted to the shake table.

For these experiments, the accelerometer position was not adjusted to be paralled to the crystal's acceleration sensitivity vector, $\vec{\Gamma}$, (which is not, in general, normal to the crystal plane), since the vibration was unidirectional. In actual use, however, the accelerometer should be carefully positioned to be parallel to $\vec{\Gamma}$. The accelerometer was connected to a single operational amplifier (type 741) the output of which was fed to the crystal through the isolation resistors. The voltage (gain) was adjusted to obtain as deep a null in sideband amplitude as possible. The oscillator was again vibrated at 8 g from 50 to 500 Hz. Figure 7 shows the result. The upper curve represents

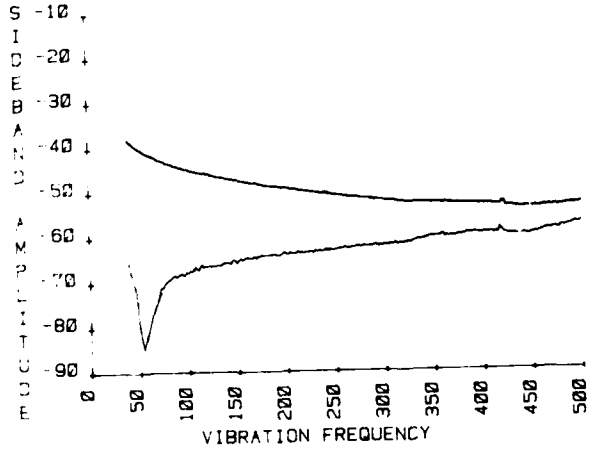


FIGURE 7: Direct-to-crystal suppression of sidebands, adjusted for maximum suppression of 50 Hz.

the sideband level with the suppression circuit turned off. (This curve would be logarithmic with a slope of -20 dB per decade if there were no resonances in the crystal/oscillator combination). The lower curve represents the suppressed sideband level. At 50 Hz (where the amplifier gain was adjusted) a suppression of 45 dB was achieved. The gradual loss of suppression is due to two factors: First, the resonances of the oscillator board and crystal cause greater accelerations inside the oscillator enclosure than were sensed by the externally mounted accelerometer, so the circuit, in its present configuration, could not fully suppress those peaks. (The accelerometer would have to be attached to the crystal.) Second, the amplifier which was used does not have a constant phase shift over the frequency band studied. (More suitable amplifiers did not arrive in time to be used for this report.)

It should be noted that for two sine waves to completely cancel, their amplitudes must be exactly equal and their relative phase must be exactly 180 degrees. Figure 8 shows the results obtained when the suppression circuit is "tuned" at 200 Hz. Maximum suppression in this case is only 15 dB but is effective over a much broader bandwidth.

Conclusion

We have shown that vibration-induced sidebands can be suppressed by 60 dB, at a single vibration frequency, using active methods. Due to harmonics produced by the nonlinearity of the varactor, the direct-to-crystal method appears superior (unless a superlinear varactor becomes available in the future.)

Further, the potential for deep suppression of broad-band vibrations has been established. In order to achieve full potential, the following conditions must be satisfied: 1. the crystal resonator and oscillator structures must not have mechanical resonances, 2. the (uniaxial) accelerometer must be mounted so that its sensitive direction is aligned parallel to $\vec{\Gamma}$ (Therefore, any vibration direction, relative to $\vec{\Gamma}$, is automatically adjusted for amplitude by the accelerometer.) and 3. a high quality, large gain-bandwidth amplifier must be used so that the phase shift through the amplifier is constant over the frequencies of interest.

Static accelerations, due to gravity or a tight turn by an aircraft, for example, can be accomodated by the use of accelerometers that have DC response (piezoresistive or capacitive types.)

The effects of constant low voltage DC impressed on SC-cut crystals is being investigated to determine if any long-term frequency drift is caused by "sweeping" impurities from the crystal.

It should also be pointed out that the various active and passive methods can be combined to yield greater overall suppression. For example, a pair of SC-cut resonators, either or both of which is (are) vibration suppressed by the application of voltage to the electrodes, can be built into an oscillator which is in turn mounted on a simple vibration isolation platform. Evidently, there are several possible alternatives from which to choose for the most economical implementation of suppression of the effects of vibration.

References

1. R.L. Filler, "The Effects of Vibration on Frequency Standards and Clocks", These Proceedings.

2. A. Warner, et al, "Low 'g' Sensitivity Crystal Units and Their Testing," Proc. 33rd Annual Frequency Control Symposium, (1979).

3. M. Valdois, J.J. Gagnepain, R.J. Besson, "Influence of Environmental Conditions on a Quartz Resonator," Proc. 28th Annual Frequency Control Symposium (1974).

4. R.L. Filler, "The Effect of Vibration on Quartz Crystal Resonators," Technical Report DELET-TR-80-10, ERADCOM, Fort Monmouth, NJ, Available from NTIS, AD AO 85642.

5. A. Ballato, "Resonators Compensated for Acceleration Fields, "Proc. 33rd Annual Frequency Control Symposium, (1979).

6. J.J. Gagnepain and F.L. Walls, "Quartz Crystal Oscillators with Low Acceleration Sensitivity," Technical Report NBSIR 71-855, National Bureau of Standards, Washington, DC 20234, March 1977.

7. J.M. Przyjemski, "Improvement in System Performance Using a Crystal Oscillator Compensated for Acceleration Sensitivity," Proc. 32nd Annual Frequency Control Symposium, (1978).

8. C.K. Hruska, "Zero Polarizing Effect with Doubly Rotated Quartz Plates Vibrating in Thickness," IEEE Trans. Sonics and Ultrasonics, Vol. SU-27, March 1980.

9. J. Kusters, "The Effect of Static Electric Fields on the Elastic Constants of Alpha Quartz", Proc. 24th Annual Frequency Control Symposium, (1970).

10. R.L. Filler and J.R. Vig, "Fundamental Mode SC-cut Resonators". Proc. 34th Annual Frequency Control Symposium, (1980) pp 187-193.

11. C.K. Hruska, "On the Linear Polarizing Effect with α-Quartz AT Plates", IEEE Trans. Sonics and Ultrasonics, Vol. SU-28, March 1981, pp 107-110.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

INCIDENT AMPLITUDE FREQUENCY EFFECT IN RESONATORS WORKING ON TWO FREQUENCIES

J.J. Gagnepain, Ch. Guérin, P.J. Besson

Ecole Nationale Supérieure de Mécanique et des Microtechniques
La Bouloie - Route de Gray - 25030 BESANCON CEDEX - FRANCE -

SUMMARY

Using a dual mode crystal, several resonant frequencies can be selected at the same time, applying even small electrical voltages. Several effects can cause the operation to fail. However, when amplitude dependent signal instabilities or linear effects appear.

The relationship between various vibrations is studied. Results :
- Evidence of coupling between oscillators.
- Coupling due to the vibrating material (not the usual linear effect).

In this paper we will consider coupling of two resonant frequencies and experimental data using AT-cut resonators.

To avoid the electrical coupling between oscillators, the two crystals are always tested using two different frequency synthesizers.

The variation in frequency of the principal vibration will be presented in terms of the amplitude of the other mode. For clarity this effect will be called incident amplitude frequency effect (or simply amplitude frequency effect).

Experimental curves are presented for several AT cut 5 MHz crystals using third and fifth overtone and B mode. Experimental results are presented and some applications are given.

INTRODUCTION

Resonant devices have been built to operate at one working frequency and even different working frequencies.

Usually, it is sought to determine the temperature of the resonator [1].

Internal heating and thermal regulation [2] have allowed the use of a vibrational mode different to the usual oscillating frequency control.

Some authors [3] have used the 3rd overtone to maintain stability over a wide temperature range [4].

Several different overtones or modes can be excited at the same time in the same resonator material [5].

Under those conditions, the coupling between two working frequencies has to be investigated with precision.

EXPERIMENTAL SET-UP DESCRIPTION

Two kinds of experimental set-ups are possible.

1/ Active set-up : 2 different oscillators O_1 and O_2 are using the same crystal X (see Fig 1). In this case, coupling due to electronics cannot be avoided. Then, it becomes difficult to separate electronics and resonator contributions. In addition, instabilities can appear as shown previously.[6] As a consequence, careful A.G.C. is needed.

2/ Passive set-up : 2 different frequency sources S_1 and S_2 are used independently to drive the crystal (see Fig 2). S_1 and S_2 are isolated by adequate isolation amplifiers I_1 and I_2. P_1 and P_2 are power amplifiers. A band-pass filter centered on ω' allows measurement of output ω' signal amplitude and two band elimination filters allow measurement of phase and power related to ω". Passive set-up has generally been prefered yielding drive-levels up to 20 milliwatts.

PRINCIPAL EXPERIMENTAL RESULTS

The evolution of the main frequency (i.e. which is used in frequency control applications) has been recorded as an other mode was suddenly driven. AT cut 5th overtone and SC cut 3rd overtone 5 MHz units have been used for experiments. The second considered modes were respectively 5 MHz 3rd overtone and 8.3 MHz 5th overtone. Fig 3 and Fig 4 exhibit the main frequency evolution versus time. An immediate frequency offset Δf_o followed by an exponential variation is noticed. When the second signal is suppressed another frequency offset occurs and the resonator returns to its original frequency.

A typical pattern for SC cut is shown on Fig 5. The two frequency offsets are equally 30 Hz. Thermal exchange time constants are clearly visible on Fig 3, Fig 4 and Fig 5. Actually, material temperature increases as previously pointed out [7]. In the case of Fig 5 experiments were performed at 25°C and temperature elevation was approximately 1°C.

Experimental results yielding frequency offset versus second mode drive level are represented on Fig. 6, Fig. 7 and Fig. 8 which exhibit a linear dependence. For clarity this effect will be called indirect amplitude frequency effect (I.A.F. effect).

Since regular amplitude frequency effect [8] (A.F. effect) is not well known for very high drive levels, measurements have been performed (Fig. 9, Fig. 10 and Fig. 11). It turns out that I.A.F. effect is roughly 50% of A.F. effect for a given drive level. In the case of AT cuts the two effects are of opposite signs.

MODEL FOR COUPLING BETWEEN MODES

We assume an elastic coupling K coefficient depending on drive levels ($K \ll 1$). Principle of experiments is shown Fig. 12 and governing equations are :

$$\frac{1}{\omega_1^2}\ddot{\theta}_1 + \frac{1}{Q_1\omega_1}\dot{\theta} + \theta_1 = K\theta_2 + A_1 \cos \Omega_1 t \qquad (1)$$

$$\frac{1}{\omega_2^2}\ddot{\theta}_2 + \frac{1}{Q_2\omega_2}\dot{\theta} + \theta_2 = K\theta_1 + A_2 \cos \Omega_2 t \qquad (2)$$

where $K = f(A_1, A_2)$

Q_1 and Q_2 are the quality factors.

Solutions are :

$$\theta_1 = X_1 \cos(\Omega_1 t + \Phi_1) + x_1 \cos(\Omega_2 t + \Psi_1)$$
$$\theta_2 = X_2 \cos(\Omega_2 t + \Phi_2) + x_2 \cos(\Omega_1 t + \Psi_2)$$

Initial conditions yielding X_1, X_2, x_1, x_2, Φ_1, Φ_2, Ψ_1, Ψ_2. In particular :

$$X_1 = A_1 \left[1 + \frac{2K^2}{\left(1 - \frac{\Omega_1^2}{\omega_1^2}\right)\left(1 - \frac{\Omega_2^2}{\omega_2^2}\right) - K^2} \right]^{1/2}$$

$$\times \left[\left(1 - \frac{\Omega_2^2}{\omega_2^2}\right)^2 \left(1 - \frac{\Omega_1^2}{\omega_1^2}\right)^2 - K^4 \right] \Big/ \left[\left(1 - \frac{\Omega_1^2}{\omega_1^2}\right)\left(1 - \frac{\Omega_2^2}{\omega_2^2}\right) - K^2 \right]$$

For case of equal [illegible] one gets frequency for equation (3) :

$$\left(1 - \frac{\omega'^2}{\omega_1^2}\right)\left(1 - \frac{\omega'^2}{\omega_2^2}\right) - K^2 = 0 \qquad (3)$$

[illegible]

$$\omega' = \omega_1 \left(1 - \frac{\omega_2^2}{(\omega_1^2 - \omega_2^2)} \frac{K^2}{2}\right) \qquad (4)$$

and

$$\omega'' = \omega_2 \left(1 + \frac{\omega_1^2}{(\omega_1^2 - \omega_2^2)} \frac{K^2}{2}\right) \qquad (5)$$

where $\omega_2 > \omega_1$.

Similar equations are obtained for the second mode (second vibrator of model) solutions are identical to ω' and ω'' of Eq (4) and Eq (5). If resonator is driven both by ω' and ω'' vibrational amplitudes are maximal and given by :

$$\theta_{1,max} = X_1(\omega') \cos\left[\omega' t + \Phi_1(\omega')\right] + \frac{K X_2(\omega'')}{\left(1 - \frac{\omega''^2}{\omega_1^2}\right)} \cos\left[\omega'' t + \Psi_1(\omega'')\right] \qquad (6)$$

$$\theta_{2,max} = X_2(\omega'') \cos\left[\omega'' t + \Phi_2(\omega'')\right] + \frac{K X_1(\omega')}{\left(1 - \frac{\omega'^2}{\omega_2^2}\right)} \cos\left[\omega' t + \Psi_2(\omega')\right] \qquad (7)$$

Since K is very small interesting experiments correspond to the first term of Eq (6) and Eq (7). It must be pointed out that this model is only valid for passive set-ups.

It turns out that it is impossible to explain IAF effect only through AF effect on each mode. Also capacitive coupling cannot give correct account of results. In other words, in Eq (4) and Eq (5) K must be assumed to depend on drive levels. This dependence mainly causes IAF effect.
Two cases are to be considered.

a/ Main frequency f' is the lowest :

When the second mode is not driven ($P_2 = 0$) f' is f'_o and the coupling coefficient is K'_o. From Eq (4) :

$$\Delta f' = f' - f'_o = \frac{f_1 f_2^2}{2(f_2^2 - f_1^2)} (K'^2_o - K^2)$$

From experiments $\Delta f'$ is linear versus P_2 and practically does not depend on P_1 (P_1 and P_2 being drive levels for the two modes). Then :

$$K'^2_o - K^2 = i_{21} P_2$$

From experiment (see Fig 6 and Fig 7) in the case of SC cut third overtone ($f_1 = 5$ MHz, $f_2 = 5.$ MHz)

$$i_{21} \simeq 0 \times 10^{-4} \times W^{-1}$$

b/ Main frequency f'' is the largest :

If f'' = f''_o and K = K''_o for $P_1 = 0$, $\Delta f''$ is obtained from Eq (5) :

$$\Delta f'' = - \frac{f_2 f_1^2}{2(f_2^2 - f_1^2)} (K''^2_o - K^2)$$

with $K''^2_o - K^2 = i_{12} P_1$

From experiment (see Fig 8) in the case of AT cut fifth overtone ($f_1 = 3$ MHz, $f_2 = 5$ MHz)

$$i_{12} \simeq 92.5 \times 10^{-4} \times W^{-1}$$

FURTHER EXPERIMENTAL RESULTS

In this section, two problems will be considered leading to some results :

1°/ Reverse influence of the main mode :

Previous experiments were dealing with influence of a given mode on the main metrological mode. For applications purposes, the reverse influence is also of interest. Experiments have been achieved yielding influence of 5 MHz 3^{rd} overtone on 5 MHz 7^{th} overtone of an SC cut resonator (see Fig 13). The effect is important and corresponds to :

$$\frac{\Delta f_{21}}{f_{21}} = -13.5 \times 10^{-6} \times W^{-1}$$

2°/ Consequences for P.S.R. consideration :

In the case of doubly rotated resonators, the frequency yields the "true" temperature of working crystal. Measurement of the 1st indicate a power transfer from B mode drive level to the main frequency to be measured. This result is given on Fig 14 and corresponds to a slope of 10^{-7} Hz/μW. Correcting this effect yields (f_1 = 5 MHz, f_2 = 5.5 MHz B = 30) :

$$\frac{\Delta f_{21}}{f_{21}} = 2.5 \times 10^{-14} \times W^{-1}$$

Recently, internal heating has been developed in the laboratory for this purpose, using an SC cut resonator. The frequency was driven hard. The B mode was used as temperature sensor. In this case, influence of 5.5 MHz drive level on B mode frequency and temperature indications had to be measured. Fig 15 yields experimental approximative data (due to difficulty of experiment). The correction is typically $2.5 \times 10^{-13} \times Hz/\mu W$.

CONCLUSION

In this paper, IAI effects importance has been pointed out and experimental results have been given together with a phenomenological model. The coupling factor turns out to be dependent on the drive level of the coupled mode. The previous results are useful for not only for frequency adjustment, but warm-up and thermal regulation applications [9, 10].

REFERENCES

1 - J.A. Kusters - M.F. Hlavac and D.L. Beach, "Quartz mode perturbations temperature and stress compensated crystal" Proc. 32nd Annual Symposium Frequency Control - 1978.

2 - J.J. Gagnepain - P. Besson, "Résonateur à quartz à régulation directe interne en température" C.R. Acad. Sc. Paris, t.289 Série B - 1979.

3 - M. Mac, will be currently done to appear in 1981. Thesis. Laboratoire de Chronométrie Electronique et Piézoélectricité - ENSMM - Besançon.

4 - A. Thirard, "Etudes topographiques des vibrations de cisaillement d'épaisseur de résonateurs à quartz" Thèse de Docteur-Ingénieur - ENSMM - Besançon 1980.

5 - J.H. Balbi, J.A. Duffaud and R.J. Besson, "A new non linear analysis method and its application to quartz crystal resonator problems" Proc. 32nd Annual Symposium Frequency Control. 1978.

6 - B. Van der Pol, "The non linear theory of electric oscillators" P.I.R.E. Vol. 22, N°9, Sept. 1934.

7 - J.P. Valentin, "Internal heating and thermal regulation of bulk quartz resonators" Proc. 34th Annual Frequency Control Symposium - 1980.

8 - J.J. Gagnepain, "Non linear phenomena in bulk and surface acoustic wave resonators" 5th Conference on Piezoelectronics - Warsovie - 1980.

9 - P. Besson et J.P. Valentin, "Oscillateur haute fréquence sur thermostat" French Patent 79-18-553 - 1979.

10 - P.J. Besson and L.A. Emmons, "Initial results on 5 MHz quartz oscillators equipped with R.V.A. resonators" F.T.T.I. Washington - Nov. 1979.

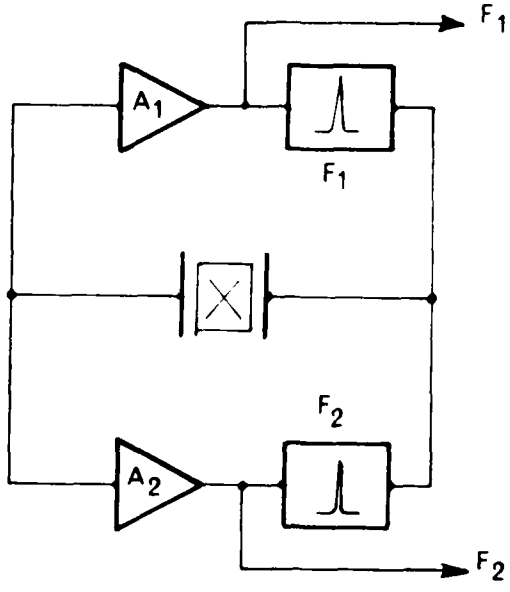


Figure 1 : Active set-up

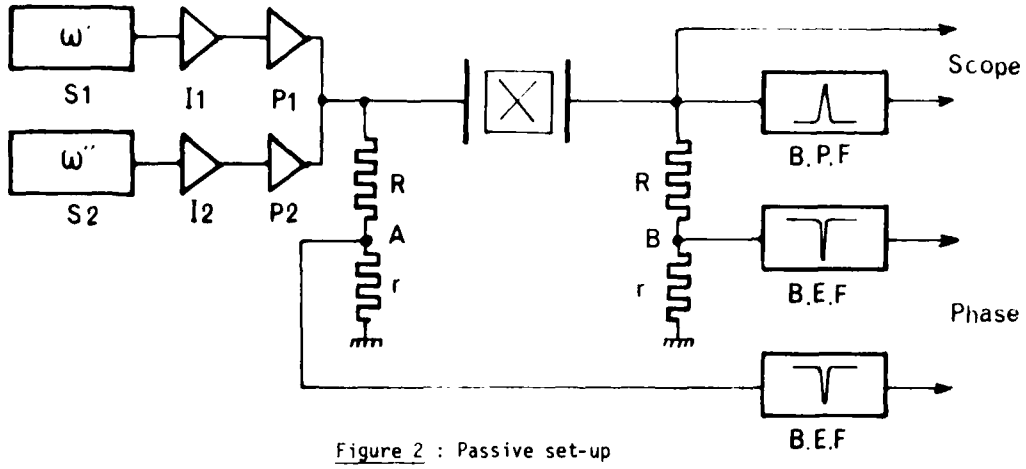


Figure 2 : Passive set-up

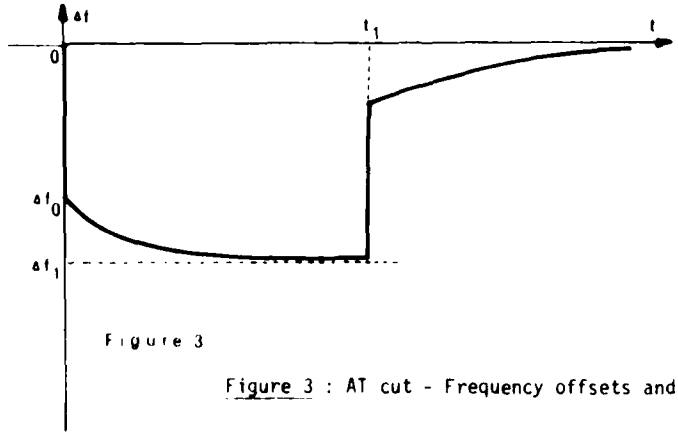


Figure 3 : AT cut - Frequency offsets and exponential evolutions.

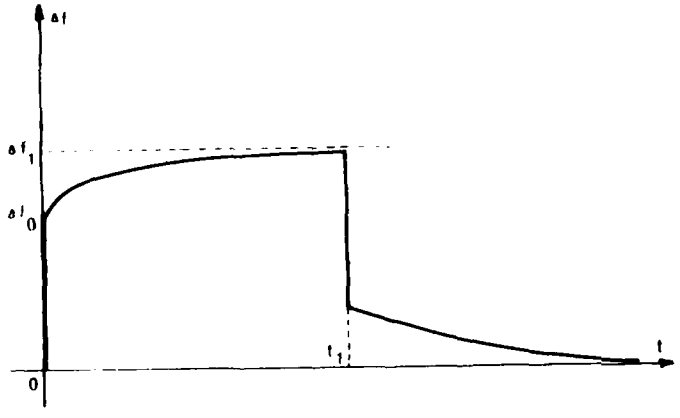


Figure 4 : SC cut - Frequency offsets and exponential evolutions.

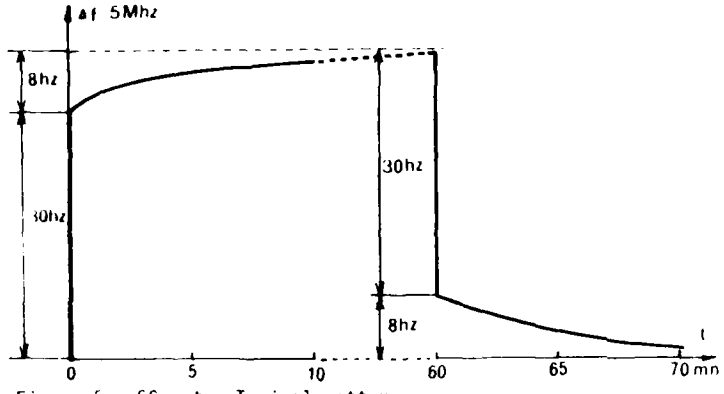


Figure 5 : SC cut - Typical pattern
Power of 5 MHz = 2 µw - Power of 8.3 MHz = 13.4 mW.

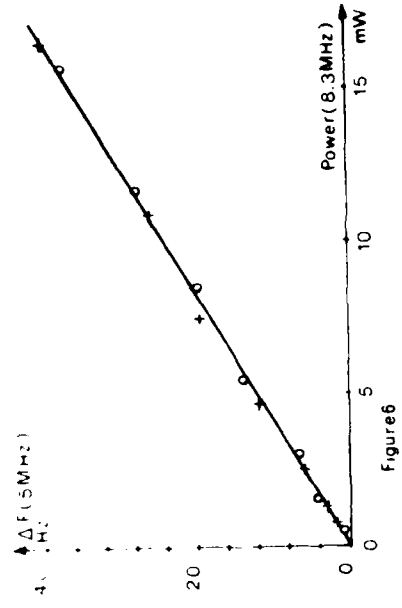


Figure 6 : IAF effect - SC cut
+ Power of 5 MHz = 1 µw - o Power of 5 MHz = 11 µw.

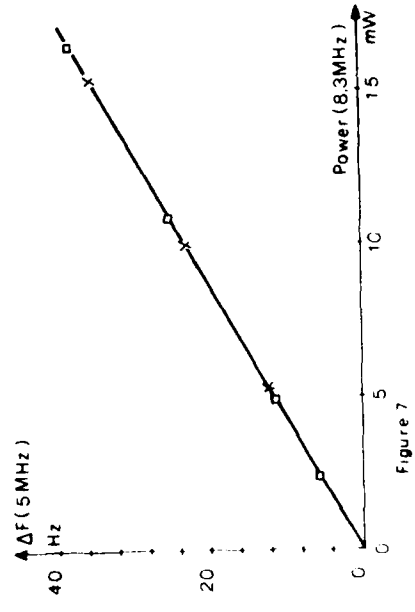


Figure 7 : IAF effect - SC cut
□ Power of 5 MHz = 0,5 mW - xPower of 5 MHz = 1 mW

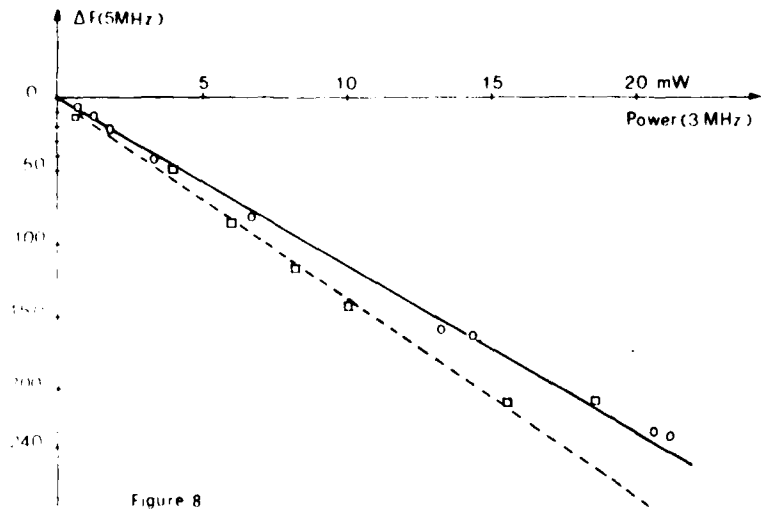


Figure 8 : IAF effect - AT cut

□ Power of 5 MHz = 2 µw - o Power of 5 MHz = 100 µw

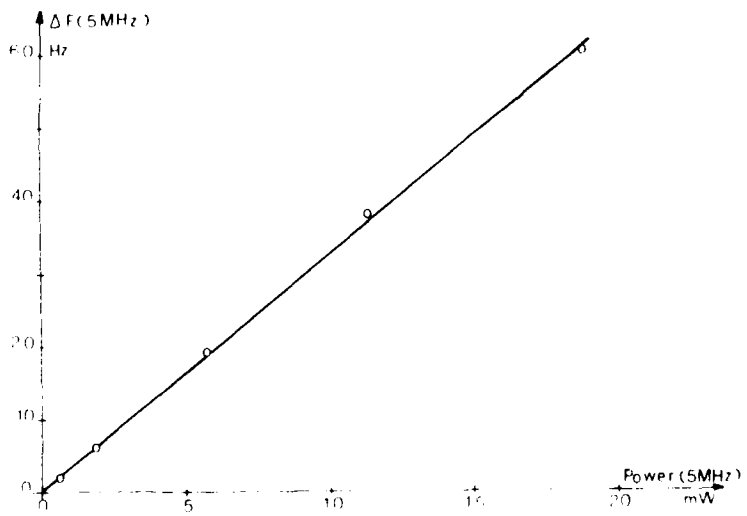


Figure 9 : AF effect - SC cut, 5 MHz 3rd overtone.

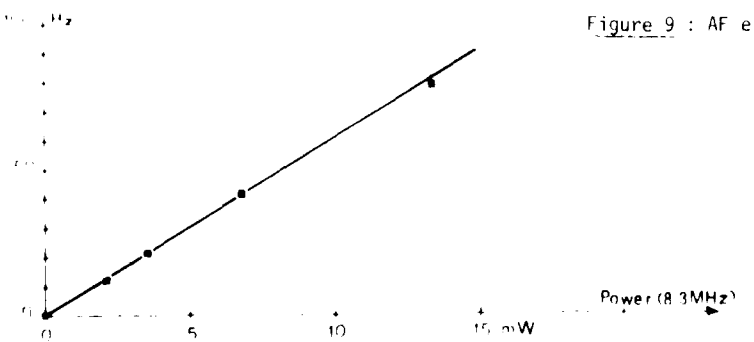


Figure 10 : AF effect - SC cut, 8.3 MHz 5th overtone.

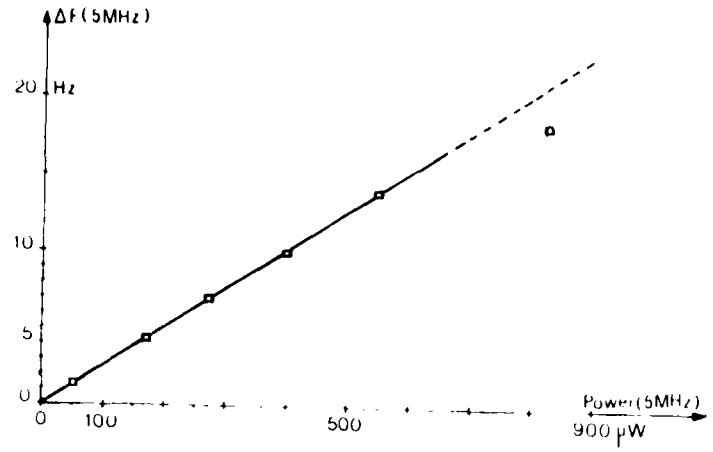


Figure 11 : AF effect - AT cut, 5 MHz 5th overtone.

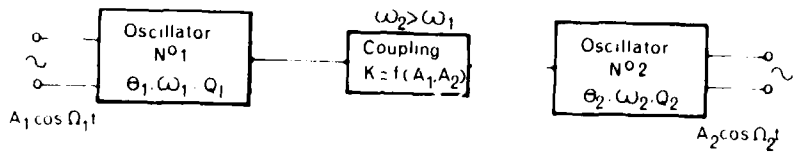


Figure 12 : Principle of experiments.

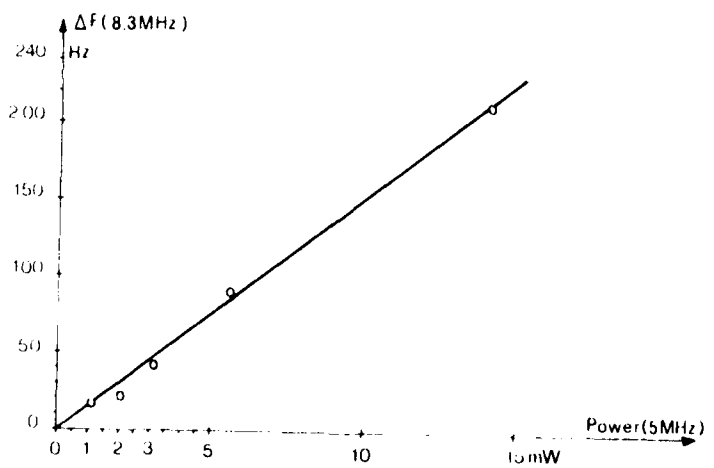


Figure 13 : Reverse IAF effect - SC cut.

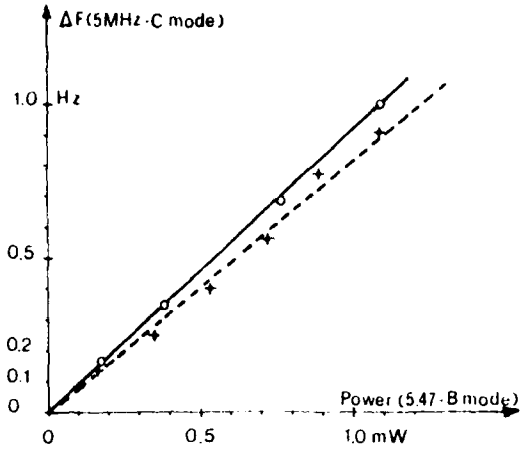


Figure 14 : IAF effect of B mode on C mode (main frequency) - SC cut
o Power of 5 MHz = 2 µw - + Power of 5 MHz = 5 µw

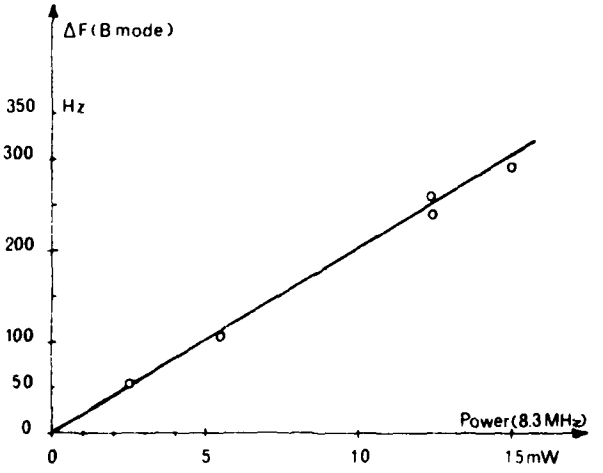


Figure 15 : IAF effect for B mode - SC cut.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

QUARTZ TUNING FORK CRYSTAL USING OVERTONE FLEXURE MODES

CHIH S. CHUANG

AGU R&D CORPORATION
Orange, CA

Summary

We have developed high frequency quartz tuning fork crystals using different overtone flexure modes over the frequency range from 150KHz to 1MHz. These quartz tuning fork crystals were produced by using photolithographic, chemical etching and laser frequency adjustment manufacturing techniques, and packaged in a relatively small ceramic package. With its ultraminiature size and ability to withstand high mechanical shock, these crystals have important applications in this frequency range.

In this paper we shall first describe the motions of overtone flexure modes of a tuning fork and their frequencies related to the fundamental flexure mode. One of the problems in obtaining a high quality factor of the quartz crystal in the overtone flexure modes tuning fork designs, is the energy loss due to the crystal mounting in the package. The motions of two crystal tines tend to couple to the displacement of a portion of the base of the tuning fork in longitudinal and lateral directions which cause a great energy loss of the flexure motion of two crystal tines. We shall show the effect on the crystal Q factor and frequency shift in different mountings and modification of tuning fork shape. With proper crystal tuning fork shape and mounting, we have achieved Q of better than 1.6×10^6.

The turning point temperature of overtone flexure modes is much higher than fundamental flexure mode. With the change in crystal orientation, we can design a quartz crystal tuning fork which has a turning point temperature over a very wide range.

The crystal parameters including the motional capacitance, quality factor Q, and the temperature-frequency characteristic of the quartz crystal tuning fork in different overtone flexure modes are analyzed and discussed. The selection of which flexure mode should be used in this frequency range depends upon the crystal parameters and the tuning fork as will also be discussed.

Introduction

In the past ten years the tuning fork quartz crystal in flexure mode design has evolved to become one of the dominate designs in the low frequency spectrum. The large volume application for the tuning fork quartz resonator operating at 32768 Hz is for the wrist watch.

The reasons that the tuning fork flexure mode design is so unique are as follows:

a) small size.
b) high mechanical shock.
c) better aging characteristic.
d) low manufacturing cost.

The use of photolithographic, chemical etching and laser frequency adjustment techniques[1] in the manufacture of the tuning fork quartz resonator provide accuracy, repeatability in dimensions, and in addition, further improvements in resonator size and cost are possible.

With the advantage of this tuning fork flexure mode design, one would try to extend the use of this design to a wide frequency spectral range. The tuning fork quartz crystal resonator design has been mainly used in its fundamental flexure mode because of the simplicity of this mode, and also it has been investigated for many years[3,4]. When the fundamental flexure mode is used in extending toward the higher frequency by simply reducing the length of the tuning fork, some of the crystal parameters become unacceptable in practical applications, and it also becomes difficult to manufacture. Figure 1 shows an example of the Q and the motional capacitance of a quartz tuning fork fundamental flexure mode resonator as a function of the frequency. The Q of the crystal increases as the frequency is increased, but it reaches a maximum and then decreases. The motional capacitance decreases very fast when the frequency is higher than 125KHz. Beyond this frequency range the use of the overtone flexure modes of the tuning fork provides better crystal parameters and also has advantages in the manufacturing processes.

Tuning Fork Flexure Modes

Flexure Motion of Tuning Fork

To obtain a complete solution of the flexure motion of a tuning fork as shown in Figure 2, is not an easy problem. In practice, the result of the analysis on the flexure motion of a bar can give an important indication of the motion of the tine and the frequency equation in the first approximation. The following flexure motion

equation is based on this simplified analysis[4]. The relative displacement of the tine as a function of the position of the length direction is approximately:

$$u(y) = P\{\cos\ (m\ \frac{y}{\ell}) - \cosh\ (m\ \frac{y}{\ell})\} + \sinh(m\ \frac{y}{\ell}) - \sin\ (m\ \frac{y}{\ell}) \qquad (1)$$

where

$$P = \frac{\sinh m \sin m}{\cosh m \sin m + \sinh m \cos m}$$

and m is a constant which is determined by the equation.

$$\cosh m \cos m + 1 = 0$$

For different flexure modes m has the following values:

flexure mode	m
fundamental	1.8751
first overtone	4.6941
second overtone	7.8547
third overtone	10.9955

Figure 3 shows the lateral displacement of the tine for the fundamental and the overtone flexure modes.

The frequency of the flexure modes including rotary inertia is approximately:

$$f = \frac{m^2 w}{\ell^2}\sqrt{\frac{E}{12\rho}}\left\{1 - \frac{m}{24}\left(\frac{w}{\ell}\right)^2\left[\frac{\mu'}{\mu} + m\left(\frac{\mu'}{\mu}\right)^2\right]_\ell\right\} \qquad (2)$$

where w: tine width
ℓ: tine length
ρ: mass density
E: Young's modulus
μ': the derivative of tine displacement μ with respective to the tine length direction.

When the ratio of tine width w to the tine length ℓ is very small, the frequency ratios of the overtone modes with respect to the fundamental flexure mode are 6.23, 17.28, 33.4 for first, second and third overtone flexure mode respectively.

Electrode Design

The electrodes are placed on the piezoelectric resonator to obtain the maximum electrical mechanical coupling. Since there are more than one nodal points of the overtone flexure motion of the tines, polarity changes of the electrodes along the tines are necessary. The positions where the polarity of the electrodes change can be determined by finding the minimum stress points along the tines.

The stress due to the flexure motion is approximately proportional to the second derivative of the lateral displacement of the tines with respect to the direction along the length. Figure 4 shows the relative stress as function of the position along the tine length for the fundamental and overtone flexure modes. The nodal points shown in the figure will be the approximate locations in a practical tuning fork. The actual locations will depend on the detail geometry of the tuning fork. However, the locations shown in Figure 4 do provide the initial positions for the electrode design.

Figure 5 and Figure 6 illustrate examples of the electrode designs of the first and second overtone flexure modes of a tuning fork. In these designs we have considered the optimization of the motional capacitance and also suppression of unwanted fundamental and other modes. Types A and B are provided for different designs. The selection of type A or B electrode designs depend on the crystal orientation and the thickness of the crystal.

Experimental Results

Quartz Crystal Wafer

The quartz tuning fork crystals were produced using photolithographic, chemical etching, manufacturing techniques. The tuning fork crystals were made from a quartz wafer oriented[5] in two possible ranges:

(a) (ZYwℓ) -25° to $+25^\circ$/-35° to $+35^\circ$

(b) (XYtℓ) -25° to $+25^\circ$/-55° to $+55^\circ$

depending on the wafer thickness and the application of the crystal. In the following, experiments were done with case (a) using a type A electrode design. Figure 7 shows an example of a finished wafer which has 78 crystals within an approximately one square inch wafer. The frequency of the crystal is approximately 700KHz and operates in the second overtone flexure mode. Individual crystal parameters can be probed and measured while it is still attached to the wafer and the frequency can be adjusted by laser tuning, in situ.

Mounting and Packaging

When a tuning fork, quartz crystal is mounted in a package, depending on the clamped position from the crotch of the tuning fork, the frequency and Q of the crystal resonator may be different. As part of the experimental program, the influence of the mounting location on the resonator Q and frequency was investigated. Several different crystals that have different tine widths and tine lengths were used. In the experiment, the frequency of the crystal was measured while the crystal was still attached to the wafer. The crystal was then removed from the wafer and mounted in a small ceramic package. Conductive and nonconductive epoxies were used to glue the crystal to the pedestal in the package and also to provide the electrical connection from the crystal to the package electrical terminals. After the package and crystal went through the sealing processes where the crystal is placed in a vacuum environ-

ment, the motional resistance R_1 and the frequency of the crystal were measured.

Figure 8 shows an example of the experimental results. In the figure the frequency shift ΔF is the frequency difference between the measured frequency after the package was hermetically sealed and the measured frequency when the crystal was in the wafer. The experimental results indicate a drastic increase in R_1 and an abrupt shift in the frequency at certain discrete base lengths. In a certain discrete base region, the crystal resonator R_1 is low and the crystal frequency is relatively stable. A similar result was observed with steel tuning forks[6]. These discrete base lengths for the first overtone flexure mode quartz tuning fork depend on the frequency of the resonator and are approximately inversely proportional to the frequency.

These experimental results indicate that there is a strong coupling between the lateral motions of two tines and the portion of the mounting base of the tuning fork when the tuning fork is vibrating in its first or second overtone flexure mode. This mounting problem for the fundamental flexure mode of a tuning fork has been investigated in the past. Michales[7] first suggested to modify the tuning fork tine shape to reduce the coupling between the motion of tines and the mounting base of the tuning fork which is operated at the fundamental flexure mode. We have tried similar modifications on the tuning fork operated in first and second overtone flexure modes.

Figure 9 shows experimental results for different tine shapes on the Q of the second overtone flexure mode. These results indicate that an improvement of Q can be obtained by shaping a notch in the outside edges of the tines of the tuning fork similar to the result obtained for the fundamental flexure mode. When the notches were placed inside the tines the Q of the quartz crystal decreases.

A similar experiment was performed in the first overtone flexure mode, the results show that the modification of tines along the inside of the tines will improve the Q of the first overtone flexure mode. These results give good agreement with the recent calculations based on a finite element analysis technique[8].

The optimum tuning fork shape modification also depends on the tuning fork crotch dimension and the mounting base geometry.

Crystal Parameters

Crystals using: 1) a first overtone flexure mode design in the frequency range from 150KHz to 600KHz, and 2) a second overtone flexure mode design in the frequency from 600KHz to 1MHz range; were made and packaged in the small ceramic packages as shown in Figure 10. To demonstrate the crystal parameters, a crystal design with tine width equal to 16 mils was chosen. The crystal thickness was 5 mils. By changing the tine length different crystal frequencies were obtained.

Figure 11 shows the frequency ratio of the first overtone flexure mode to the fundamental flexure mode. The frequency ratio obtained by the experiment is much smaller and changes more rapidly as function of the tine width to tine length ratio in comparison with calculated values based on equation 2. The coupling effects between two tines shear and torsion modes and the geometry deviation of the tuning fork compared to the ideal case due to the anisotropic quartz etching may all contribute to this difference. Similar results but with greater deviation from the calculated value for the second overtone mode was obtained and shown in Figure 12.

The electrode designs of these crystals are the type A designs as shown in Figure 5 and Figure 6. The crystals were mounted at the proper positions as discussed above. The packages were hermetically sealed with approximately 1 torr of gas in the package.

The motional capacitance and the Q of the crystals as a function of the frequency of the crystal were measured and are shown in Figure 13. The width dimension and the gap between electrodes have been kept the same for different crystal frequency designs. The motional capacitance in principle should be approximately inversely proportional to the square root of the frequency, since the motional capacitance is approximately proportional to the tine length. The deviation from theory in the higher frequency range is due to the fact that the electrode length has been shortened more than the tine length because of manufacturing considerations. The variation of Q for different crystal frequencies is due to the fact that the design and process has not been optimized for individual frequencies. At certain frequency the Q of the crystal resonator is more than 2.5×10^5.

The motional capacitance of the second overtone flexure mode designs is higher than the first overtone flexure mode. This is due to the reason that the tine length of the second overtone flexure mode is longer than the first overtone flexure mode. The motional capacitance of the first overtone flexure mode is 1.5 fF at 170KHz and drops to 0.44 fF at 600 KHz. For the second overtone flexure mode the motional capacitance is 0.8 fF at 600KHz and drops to 0.25 fF near 1MHz.

Frequency Temperature Characteristic

The relation between the crystal frequency and the environmental temperature for the overtone flexure modes is generally very similar to the fundamental flexure mode, i.e., a parabolic relation. However, there are two basic differences between F-T characteristic for the overtone flexure modes and the fundamental flexure mode of the quartz tuning fork. The first difference is that for a given dimension of the tuning fork, the turning point temperature of the second overtone

flexure mode is higher than the first overtone flexure mode, and the turning point temperature of the first overtone mode is much higher than the fundamental flexure mode. An example is given in Table 1. In this example the difference is approximately 140°C. The K constant for the overtone flexure mode is slightly less than the fundamental flexure mode and is in the vicinity of 0.03 ppm/°C².

The second difference is the effect of the tine width to the tine length ratio on the turning point temperature. The turning point temperature of the overtone flexure modes increases as the ratio of the tine width to the tine length is increased, which is exactly in the opposite direction for the fundamental flexure mode of the quartz tuning fork. The theoretical analysis which was done by Nakazawa et al[9], qualitatively is similar to the experimental result obtained in the overtone flexure modes. The difference in the F-T characteristic between fundamental and overtone flexure modes may be due to the effect of the coupling between the flexure modes and the torsional modes.

Figure 14 shows experimental results for the turning point of the fundamental and first overtone flexure mode tuning fork quartz crystals as function of frequency (or tine width to tine length ratio). The tine width, the thickness, and the crystal orientation of these crystals are identical. Figure 15 gives the comparison of the turning point temperature of the first and second overtone flexure mode of quartz tuning fork. The crystal orientation of the crystal shown in Figure 15 is different from the ones shown in Figure 14.

With the high turning point temperature of the overtone flexure modes of the quartz tuning fork, one can design a quartz crystal tuning fork with a combination of different orientations, which will provide a turning point temperature over a very wide range.

Selection of Flexure Modes

With the combination of the fundamental and the overtone flexure mode designs, we can produce a quartz tuning fork crystal resonator packaged in a small ceramic package to cover the frequency spectrum from 10KHz to 1MHz. There are certain ranges where the fundamental and the first overtone flexure modes, or the first overtone and the second overtone flexure modes can be chosen. The determining factor for which mode should be used has to be established by the crystal parameters required for the individual applications.

Figure 16 shows an example of the proposed frequency ranges where the fundamental, first overtone and second overtone flexure modes of the quartz tuning fork are useful. Below 125KHz the fundamental flexure mode design has a preference because of the advantage of small overall dimensions and ease of manufacture. In some applications, the small motional capacitance and lower turning point temperature may not be desirable. The first overtone flexure mode design may be used to improve these two parameters. Beyond 125KHz the Q of the fundamental flexure mode of the tuning fork decreases as the frequency of the crystal is increased. Also, the dimension of the tine length becomes too small to manufacture easily. The only advantage of using the fundamental flexure mode design is to achieve very high mechanical shock tolerance.

The first overtone flexure mode design may be used in the range from 100KHz to 600KHz. When the frequency is higher than this range, the motional capacitance is too small and also there is difficulty in manufacturing.

The second overtone flexure mode may be used in the range from 450KHz to 800KHz. It is possible to extend to 1MHz, but the small motional capacitance may limit its application. The third overtone flexure mode may have to be considered beyond this frequency.

Conclusion

In this paper it has been shown that with the combination of the fundamental and the overtone flexure modes of quartz tuning forks, we can produce a quartz tuning fork in the frequency range from 10KHz to 1MHz. The quartz tuning fork crystal can be produced by using photolithographic chemical etching and laser frequency adjustment manufacturing techniques and packaged in a relatively small ceramic package. With it's ultraminiature size and high mechanical shock characteristic, the tuning fork quartz crystal will have important new applications in this frequency range.

Acknowledgment

The author would like to thank Terry Macicl for her assistance in mask modification work and Sandy Reynes in assistance in the quartz crystal parameters measurements.

Reference:

1. J.H. Staudte, Proceedings of the 27th Symp. on Frequency Control, 1973, pp. 50-54.

2. I. Koga, J. Inst. of Elec. Engrs., (Japan), Vol.48, 1928, pp. 100-101.

3. W.P. Mason, Physical Acoustics, D. Van Nostrand Co. New York, 1958, p. 54.

4. J.W.S. Rayleigh, The Theory of Sound, Vol.1, New York: Dover, Second Edition, 1945, pp. 255-305.

5. IEEE Standard on Piezoelectricity, IEEE Std. 176-1978.

6. J.A. Kusters, C.A. Adams and H.E. Karrer and R.W. Ward, Proc. 30th Ann. Frequency Control Symp., 1976, pp. 179-183.

7. S.E. Michales, U.S. Patent No. 2247960, July 1, 1941.

8. Yoshiro Tomikawa, Kimio Sato, Masashi Konno, IEEE Trans. Sonics, Ultrason, Vol.SO-27, No.5. September, 1980, pp. 253-257.

9. M. Nakazawa, Y. Nakamura, S. Miyashita, IEEE Trans. Soncis, Ultrason., Vol.SO-26, No.5. September, 1979, pp. 369-376.

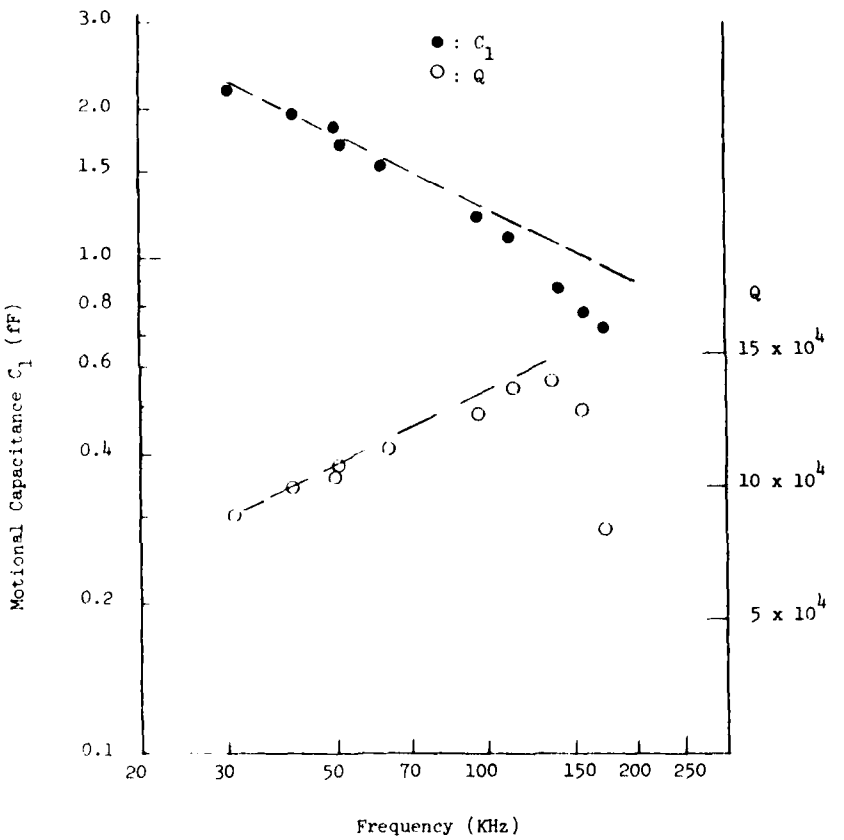
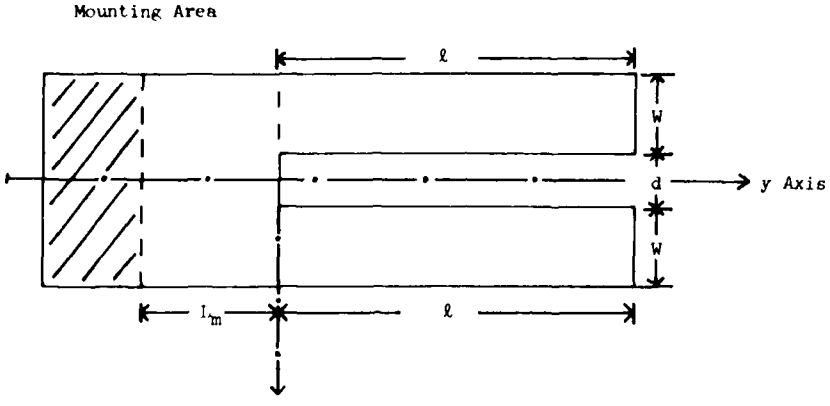
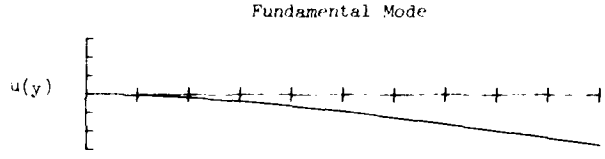


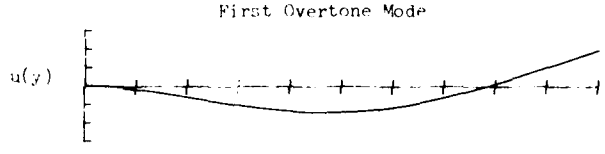
Figure 1. The Q and the motional capacitance of the quartz tuning fork fundamental flexure mode resonator as function of the frequency.

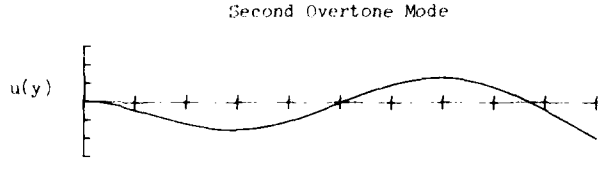


ℓ : Tine length
W : Tine length
L_m: Base length
t : Thickness

Figure 2: A typical tuning fork.







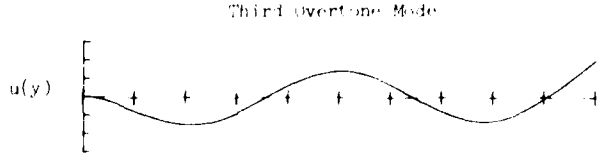


Figure 3. Lateral displacement of the tines of the flexure modes of a tuning fork.

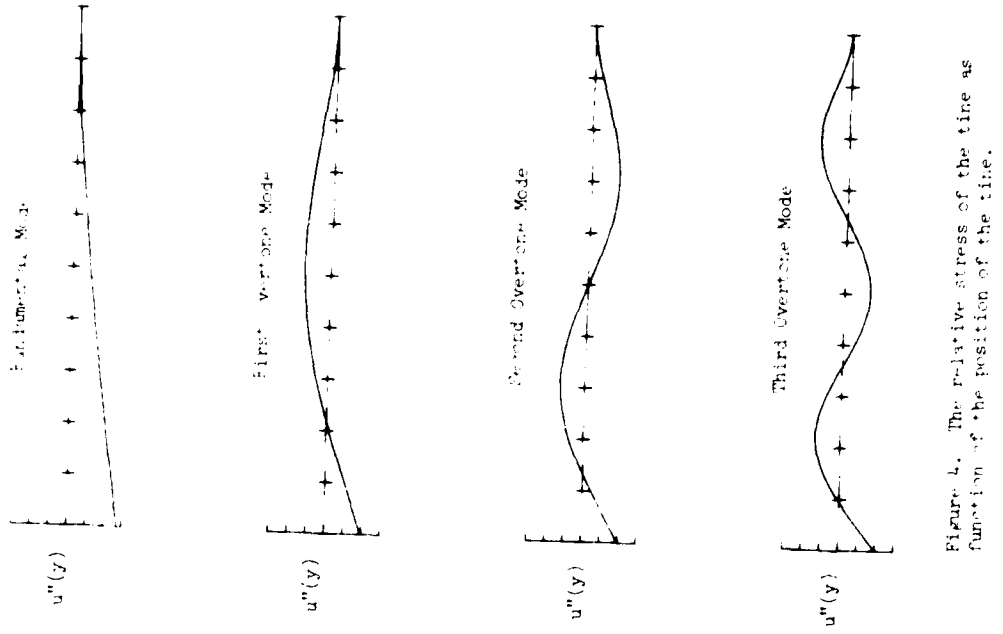
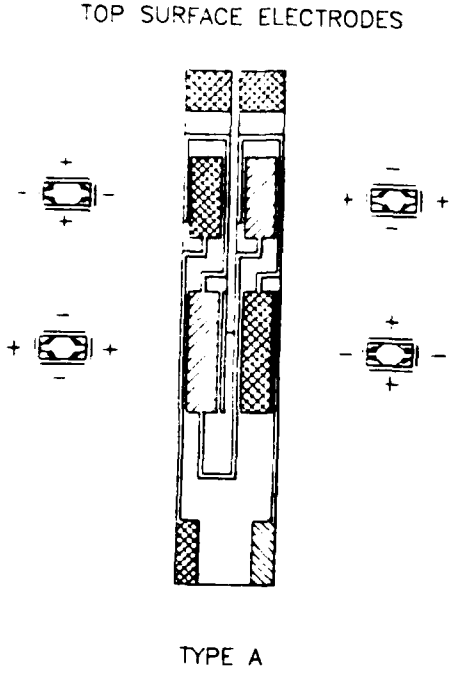


Figure 4. The relative stress of the tine as function of the position of the tine.



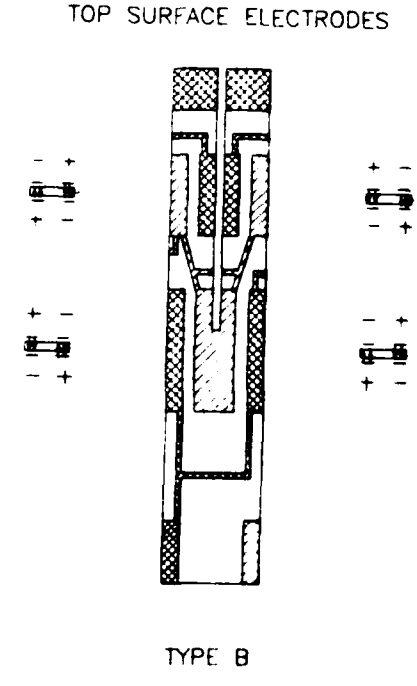


Figure 5. Electrode designs of the first overtone flexure mode of a tuning fork.

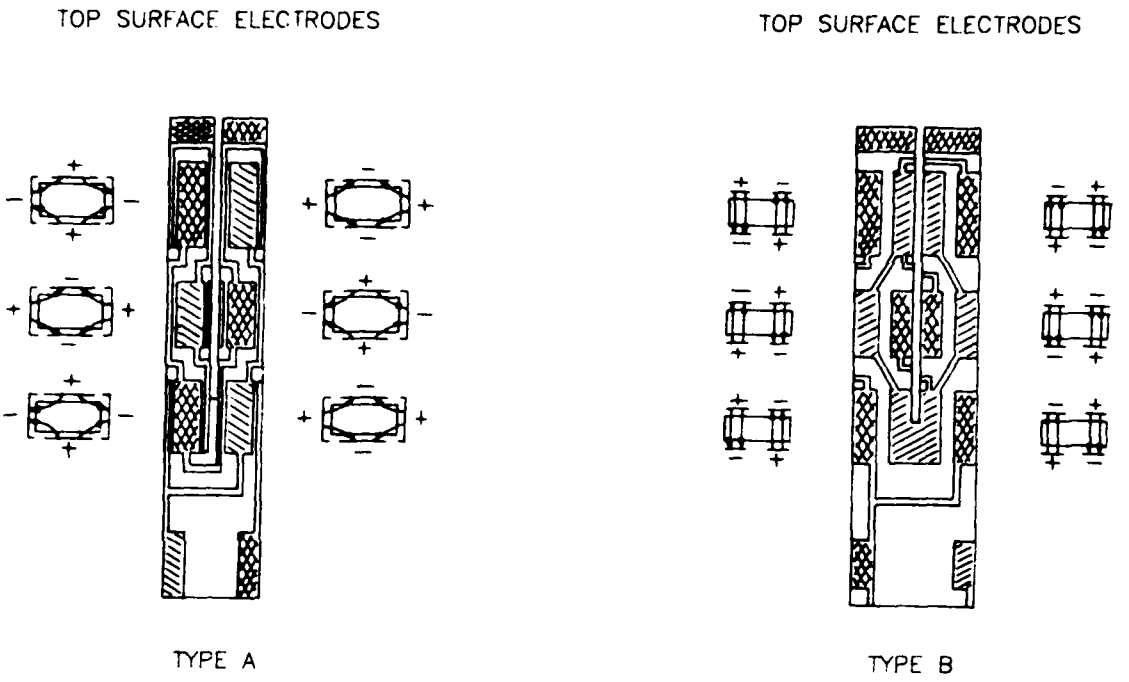


Figure 6. Electrode designs of the second overtone flexure modes of a tuning fork.

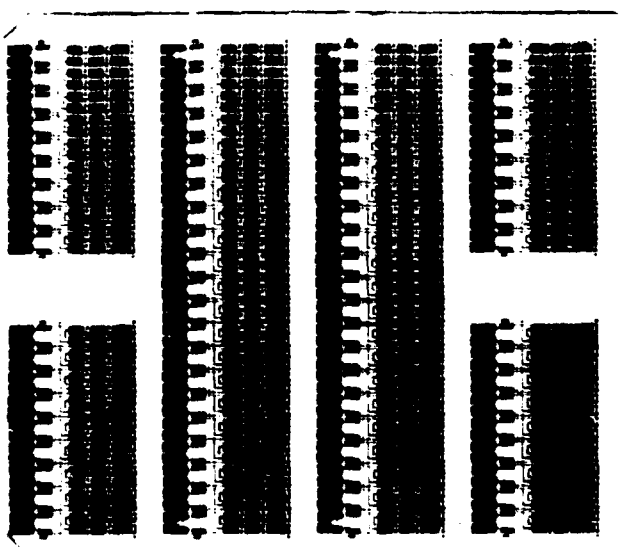
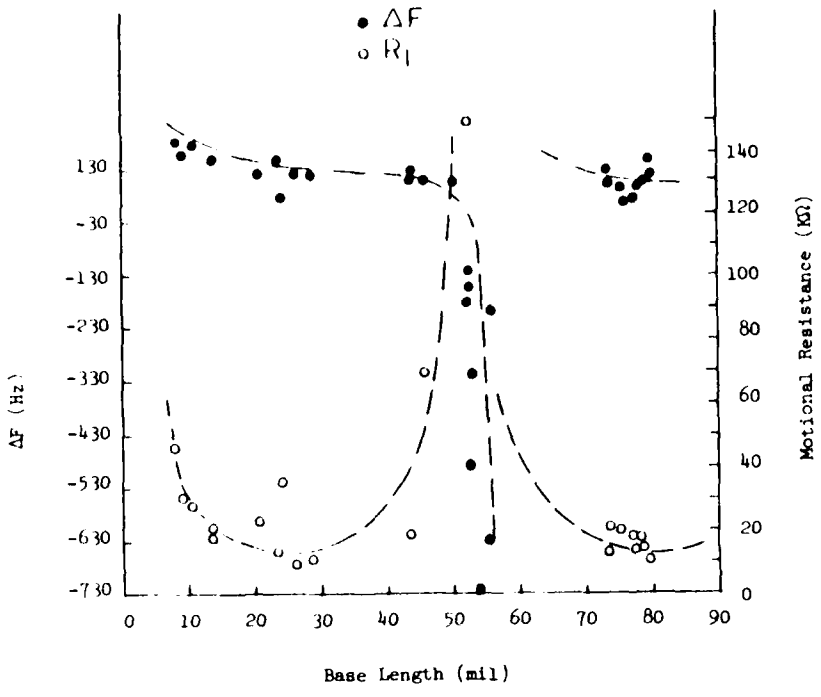
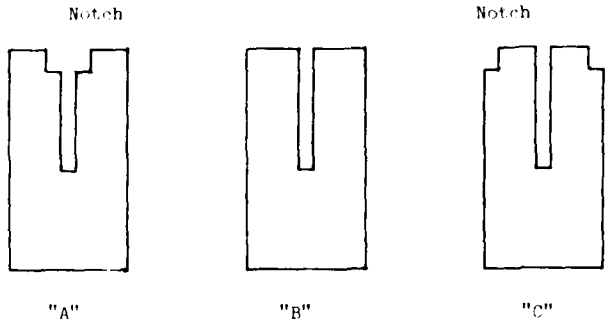


Figure 7. An example of a quartz wafer containing 78 tuning fork crystals in second overtone flexure mode design and produced by the photolithographic, chemical etching techniques.



Crystal Base Width: 35 mil
Tine Width: 16 mil
Crystal Frequency : 307 KHz

Figure 8. The frequency shift and the R_1 of the overtone flexure mode of the quartz tuning fork as function of the base length.



Notch dimension: 4 mil x 7 mil

Crystal dimensions: W : 16 mil
ℓ :100 mil
t : 5 mil

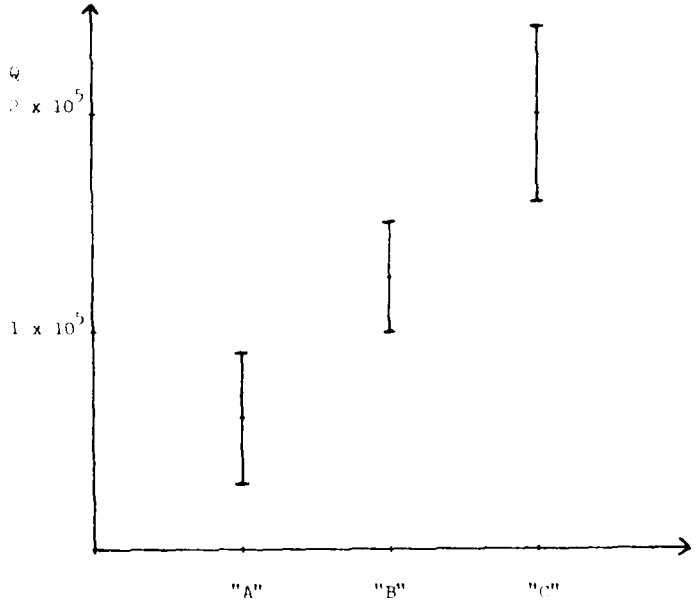


Figure 9. The effect of modifications on the tine shapes on the Q of the second overtone flexure mode of the quartz tuning fork.



Figure 10. Example of a tuning fork quartz crystal packaged in a small ceramic package.

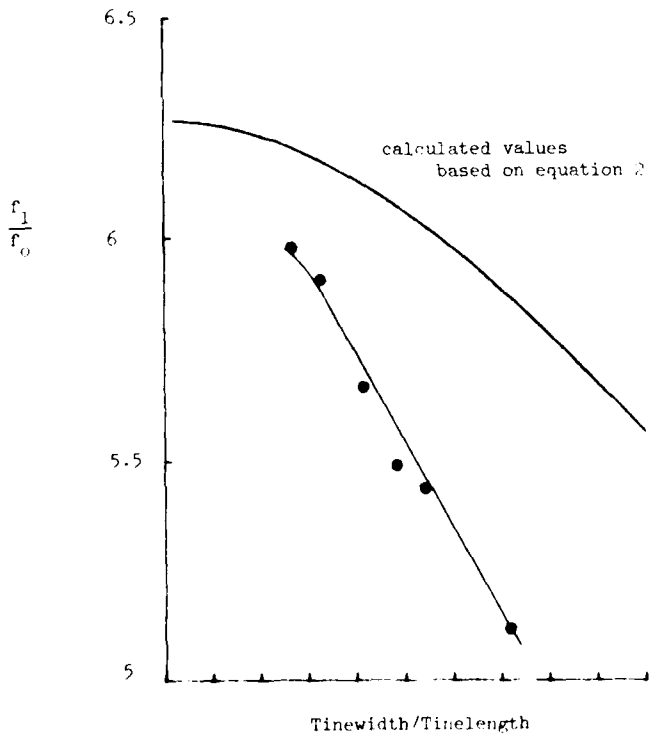


Figure 11. The frequency ratio of the first overtone flexure mode to the fundamental flexure mode.

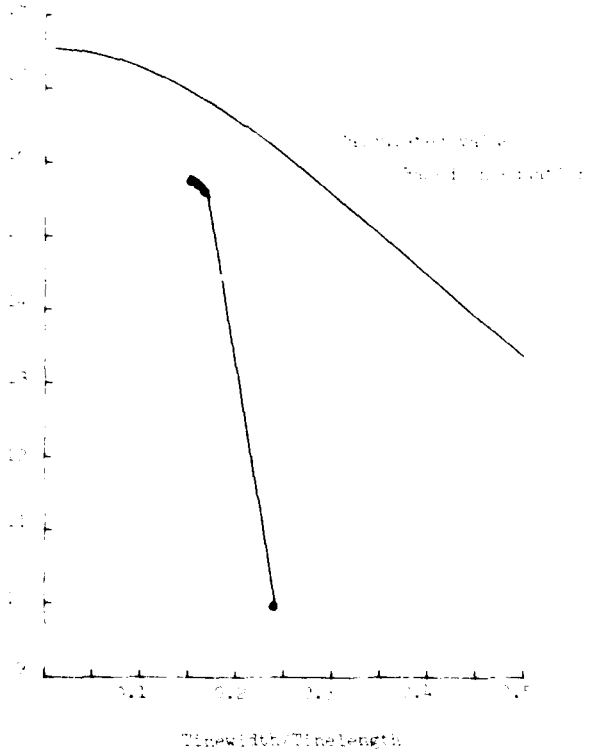


Figure 12. The frequency ratio of the first overtone flexure mode to the fundamental flexure mode.

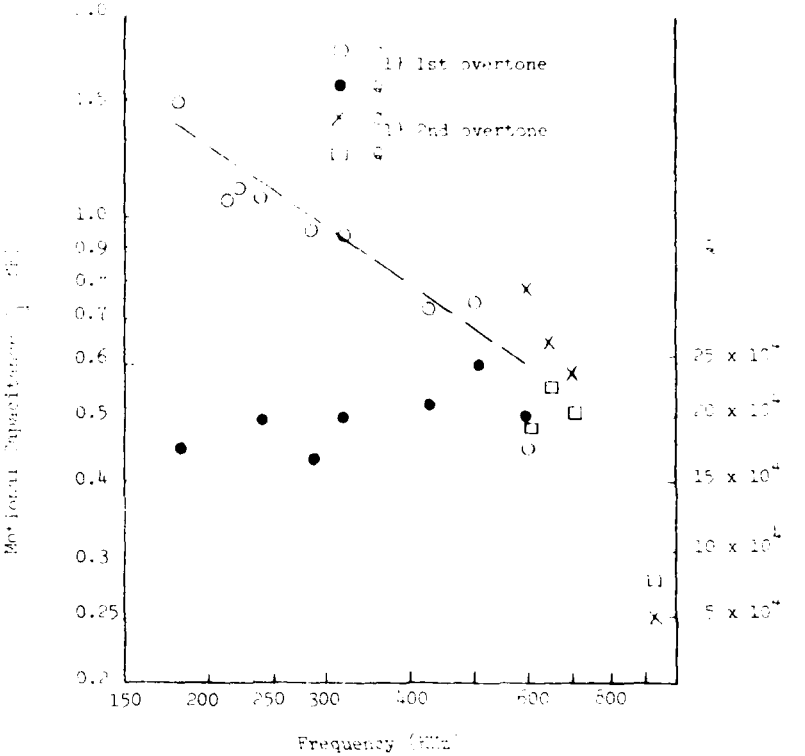


Figure 13. The Q and the motional capacitance of the fundamental, first overtone and second overtone flexure modes of the quartz tuning fork.

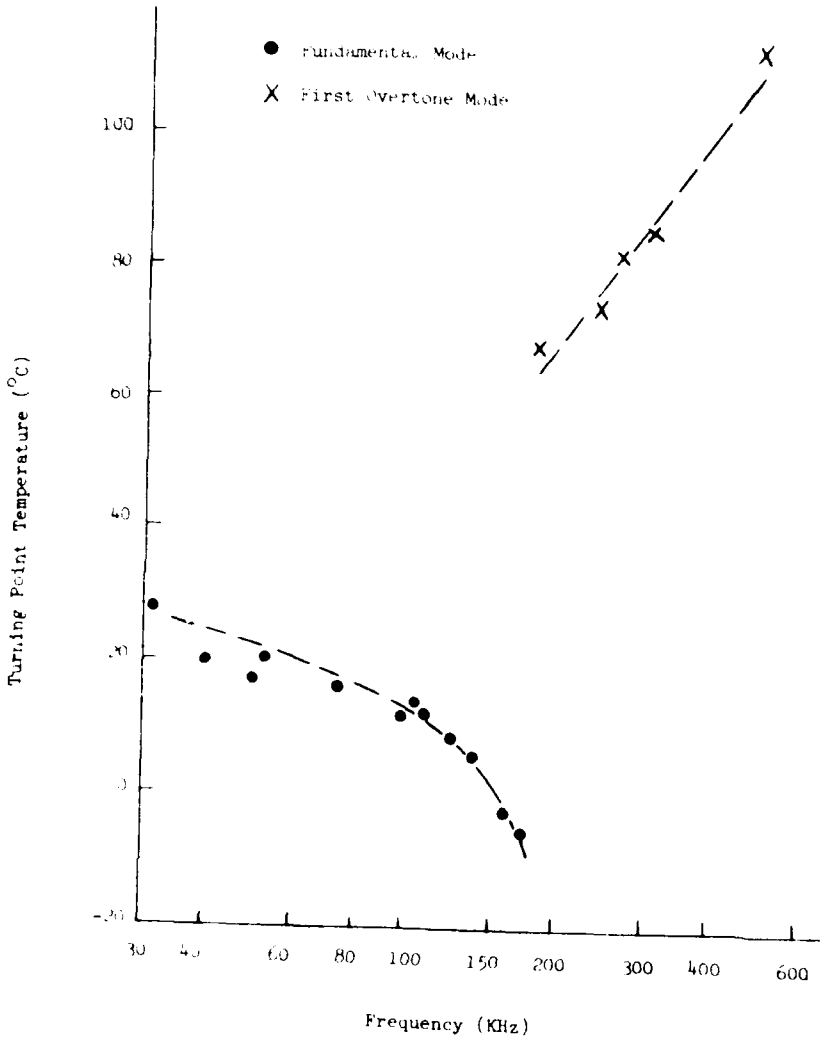


Figure 14. The turning point temperature of the fundamental and first overtone flexure modes of the quartz tuning fork as function of the resonator frequency.

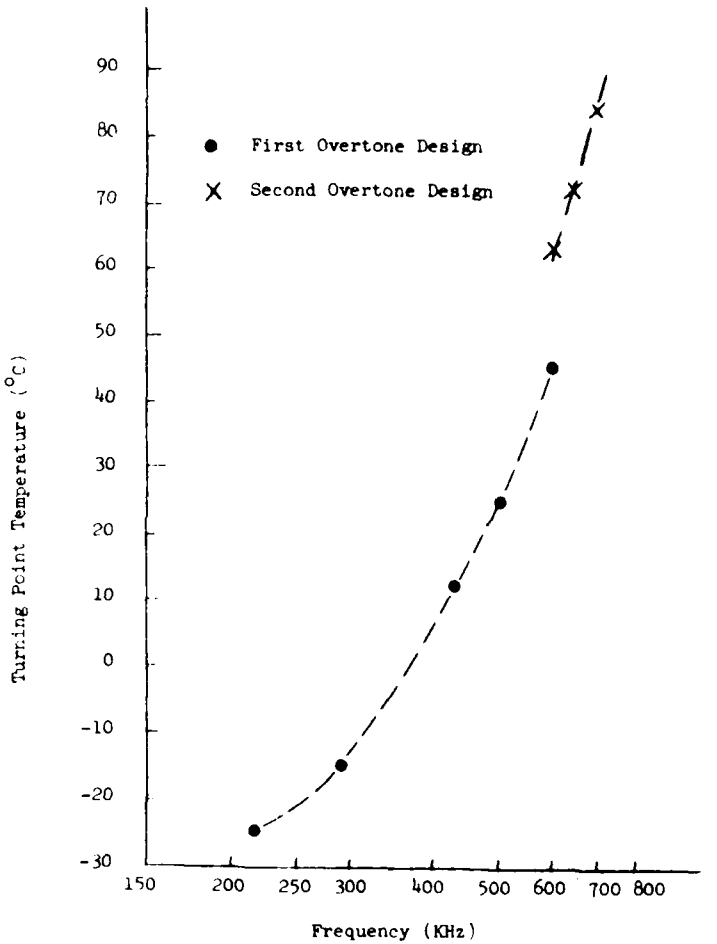


Figure 15. The comparison of the turning point temperature of the quartz tuning fork in first and second overtone flexure mode designs as function of the resonator frequency.

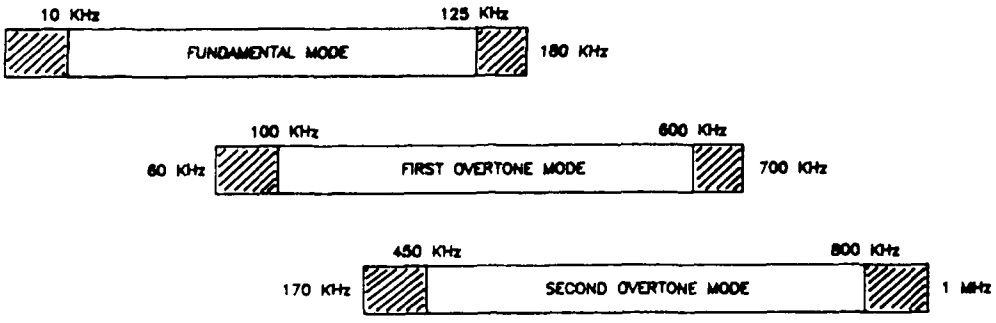


Figure 16. An example of proposed frequency ranges where the fundamental, first overtone and second overtone modes of the quartz tuning fork crystal resonator may be adopted.

(The shaded areas are the ranges where individual mode may be extended with certain limitations.)

MODE	FREQUENCY (KHz)	TURNING POINT TEMPERATURE (°C)
FUNDAMENTAL	41.1	20
FIRST OVERTONE	246	74
SECOND OVERTONE	650	160

TUNING FORK DIMENSIONS AND ORIENTATION

TINE WIDTH = 16 MILS

TINE LENGTH = 104 MILS

THICKNESS = 5 MILS

CROTCH WIDTH = 3 MILS

ORIENTATION = (ZYW)2°

Table 1. The comparison of the turning point temperature of the quartz tuning fork operating in different flexure modes.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

A MINIATURE QUARTZ RESONATOR VIBRATING AT 1 MHz

R.J. Dinger

ASULAB SA - Neuchâtel - Switzerland

Abstract

A 1 MHz miniature quartz resonator is presented. The resonator consists of the resonant bar, vibrating in the length extension mode and an integral suspension system. The unit may be manufactured using photolithographic techniques and chemical milling and assembled in standard watch tuning fork packages.

1. Introduction

The growing use of microelectronics for control of machinery, tools, appliances and toys creates an increasing market for frequency control devices. Microprocessor based systems in particular need a clock-signal to synchronize the treatment of the information. While the requirements of systems operating with 4 bits can generally be met by ceramic resonators, 8 and more bit processors need the frequency stability of a quartz resonator or other element with equal performance.

Since the manufacturing costs of integrated circuits increases with higher speed of the processor, the designers of electronic systems are using systems operating at a frequency that is sufficiently high to perform the functions of the system in the required time, but no higher. This has led to a substantial market for 1 MHz quartz crystal resonators.

Even though it is possible to manufacture AT-cut 1 MHz crystals, these units become bulky, expensive and their electrical and mechanical parameters, especially the quality factor Q, are considerably inferior to the parameters of AT-cut units in the 5 to 10 MHz range. This indicates that a 1 MHz AT-cut crystal is beyond the reasonable design limits of thickness shear resonators. On the other hand it is possible to use overtone modes of wrist-watch or similar tuning-fork designs to obtain rugged and cheap crystals that are well suited for the applications mentioned above up to frequencies of several hundred kHz [1]. The frequency region between several hundred kHz and a few MHz is the domain of the torsion-bar, face-shear and length-extension resonators. Torsion bar resonators suffer from a low piezoelectric coupling and face shear devices require expensive mounting techniques that are often shock-sensitive, due to the necessity of clamping the resonators in the nodal axis which goes through the center of the plate.

This paper presents a length-extension resonator, operating at 1 MHz, having an integral tuning fork like suspension system which allows a cheap and rugged mounting similar to that of a wrist-watch tuning fork.

Although the resonant element itself is a simple length-extension bar, the structure of the entire resonator is too complicated for mechanical manufacturing techniques. The device has therefore been designed to be manufactured using photolithographic techniques and chemical milling [2].

2. Concept

The active part of the resonator consists of a "z + 2°" cut length extension bar. It is well known that this device has a nodal axis parallel to the x-axis of the crystal going through the center of the bar. The classical suspension of these resonators consists of two wires in the nodal axis, attached to the resonators surface as shown in fig. 1.

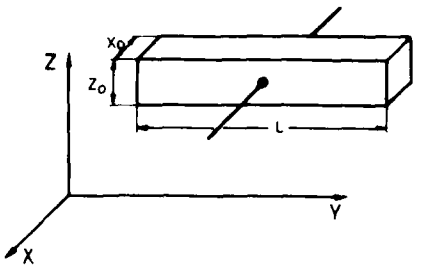


Fig. 1: classical length extension resonator

The frequency is given by one dimension only, the resonator length L [3].

$$f_n = \frac{2n+1}{2} \cdot \frac{1}{\sqrt{S_{22} \cdot \rho}} \cdot \frac{1}{L} \approx 2.72 \cdot \frac{2n+1}{2} \cdot \frac{1}{L} \text{ MHz} \cdot \text{mm}, \qquad (1)$$

n giving the order of the mode (fundamental: n = 0) and S_{22} and ρ being the compliance modulus and the density. Even harmonics also exist but due to the clamping of the bar in the suspension they have a low quality factor. Eq. 1 shows that the length-extension mode is well suited for resonators in the frequency range from about 500 kHz (giving an overall length of 5 mm) to several MHz if the first overtone is used as well. For a 1 MHz resonator the length becomes 2.72 mm.

To overcome the mounting problems of the structure in fig. 1, the resonant bar was completed with a tuning fork type suspension system as shown in fig. 2. Similar suspension systems have been used by J. Michel [4] already in 1975.

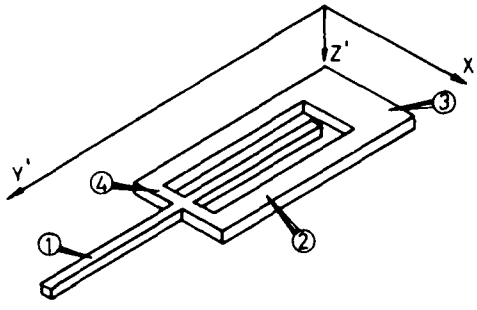


Fig. 2: structure of the 1 MHz length-extension resonator
1 resonant bar, 2 "tuning fork"-like suspension system, 3 mounting area, 4 connection part

This kind of a suspension system cannot be considered as a clamping in the nodal axis according to fig. 1 because the dimensions of the suspension of the resonator are equal or even larger than the width x_o and thickness z_o of the resonant bar itself. For this reason the resonator has to be thought of as an assembly of two resonant bars clamped at one end. To compensate the momentum of each they are vibrating with equal amplitude and opposite phase. This situation is similar to a quartz tuning fork which, due to its heavy and massive base part behaves like two clamped cantilevers.

The length extension mode is coupled to a face-shear mode in the y-z plane (except for a -18.5° cut angle) which excites a torsional movement of the resonant bar around the nodal axis and a torsional stress in the suspension system. In addition the suspension system is stressed in flexure like a tuning fork due to Poisson coupling. Although the amplitudes of these stresses are small these vibrations have to be considered and care has to be taken not to couple the main resonance with a fundamental or overtone resonance in the suspension system.

Three different electrode systems can drive the crystal. Fig. 3 shows the cross section through the resonant bar in all three systems. Fig. 3a) shows an electrode arrangement that uses electrodes on

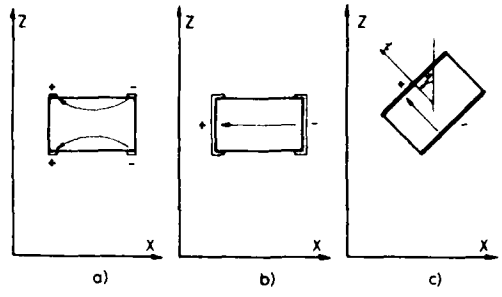


Fig. 3: Cross section of the resonant bar with electrode systems

only the two main surfaces of the resonator. This is an elegant solution from a manufacturing point of view but the piezoelectric coupling is limited to a value for C_1 of about 0.15 fF per mm of electrode length. The coupling becomes considerably larger if the metallization pattern according to fig. 3b) is used. Since a side metallization cannot be performed before the resonator is cut from the wafer, a second metallization step is needed. The piezoelectric coupling may be calculated for this arrangement.

$$C_1 = \frac{1}{125} \cdot \varepsilon_o \cdot \varepsilon_r \cdot \frac{z_o L}{x_o}, \qquad (2)$$

where C_1 is the motional capacitance, ε_o and ε_r the dielectric constants of the vacuum and quartz and z_o, L and x_o the thickness, the electrode length and the width of the resonator respectively.

The largest coupling can be obtained by the electrode arrangement shown in fig. 3c). It should be noted however that this configuration needs a second cut angle φ. The piezoelectric coupling becomes

$$C_1 = \frac{1}{125} \cdot \varepsilon_o \cdot \varepsilon_r \cdot \frac{x_o L}{z_o} \cdot \sin \varphi, \qquad (3)$$

Electrode arrangements b) and c) are compared in fig. 4. The figure has been plotted keeping in mind that the resonator has to be produced using watch tuning fork production methods for the manufacturing of the resonator itself and for its assembly. This restricts the width, thickness and the second cut angle φ to the domain shown in the

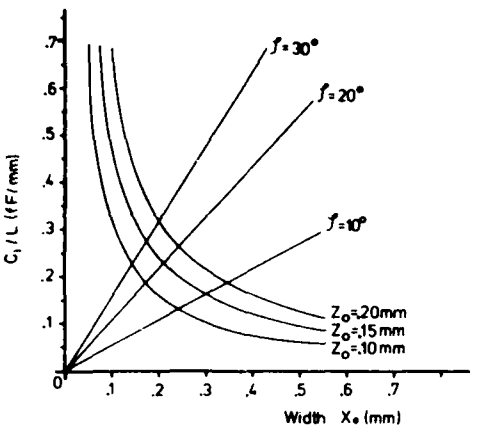


Fig. 4: Piezoelectric coupling of the electrode arrangements according to fig. 3b) and 3c).
z_o: Thickness of the resonator according to fig. 3b).
φ : Second cut angle of the resonator according to fig. 3c). Resonator thickness = 0.1 mm.

figure. Fig. 4 leads to the following conclusions:

1. Within the dimensional limits discussed the two electrode systems give about equal piezoelectric coupling of maximum 0.7 to 0.9 fF per mm of electrode length. This means that for a 1 MHz crystal the motional capacitance will become 1.5 to 2.0 fF, a value that remains considerably below the C_1 value of AT-cut crystals of the same frequency.

2. If the motional capacitance is to be increased further either extreme aspect ratios (which may lead to asymmetry problems) or a high second cut angle will have to be used. In the latter case the dimensional limits also cannot be kept. Either of these two solutions will have severe technological drawbacks.

In the course of the work described in this paper, the electrode arrangements a) and b) were used to avoid the complications due to the second cut angle.

3. Unwanted modes

Two families of unwanted vibrations have to be considered for the design of this resonator: flexural vibrations of the main resonant bar and resonances in the suspension system. Coincidence of any of these unwanted resonances with the main resonance must be avoided.

3.1. Flexural vibrations

The resonant bar may oscillate in the fundamental or an overtone flexural vibration in the x-y plane or the y-z plane. In addition, the free-free mode and the clamped-free mode of one half of the bar have to be considered. For the free-free mode the resonant frequencies are [5]

$$f_n = \frac{\alpha_n^2}{2\pi} \frac{1}{\sqrt{S_{22}\cdot\rho}} \cdot \frac{1}{\sqrt{12}} \cdot \frac{x_o}{L^2} \qquad (4)$$

where α_n is a constant depending on the mode and the other symbols as in eq. 1. The formula holds for x-y flexure, for y-z flexure x_o has to be replaced by z_o. For the clamped-free resonances eq.4 holds as well, but it has to be considered that the length L is only half the length of the bar. The values for α_n are:

	clamped-free	free-free
α_o	1.88	4.73
α_1	4.70	7.86
α_2	7.86	10.99
α_n	$\frac{2n+1}{2}\pi$	$\frac{2n+3}{2}\pi$

Combining eq. 1 and 4 the forbidden dimensional ratios become:

$$\frac{x_o}{L} = \pi \frac{2\sqrt{3}}{\alpha_n^2} \quad \text{and} \quad \frac{x_o}{L} = \pi \frac{8\sqrt{3}}{\alpha_n^2} \qquad (5)$$

The first equation giving the condition for free-free vibration and the second for a clamped-free vibration and L being in either case the total length of the bar.

The same dimensional ratios lead to y-z flexural coupling if the thickness to length z_o/L takes the above values.

Considering electrode arrangements according to fig. 3a) and b) a thickness and width between 50 µm and 200 µm are of **special interest.** For a 1 MHz resonator of total length 2.75 mm the forbidden dimensions are shown in fig. 5.

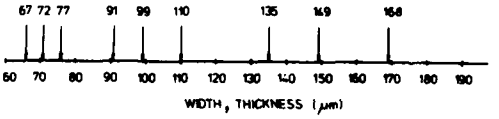


Fig. 5: Forbidden dimensions for a 1 MHz resonator

Fig. 5 shows that there are three regions to be used for the width and the thickness: the first around 85 µm, the second around 125 µm and the third around 185 µm.

3.2. Vibrations of the suspension system

The resonances in the suspension system which may coincide with the main resonance are the x-y and y-z flexural vibrations. These are more difficult

to calculate than for the resonant bar. In fact, numerical methods must be used. However, for the design of the resonator, approximate calculations are often sufficiently precise.

To estimate the flexural resonances of the suspension system, eq. 4 is used, assuming the "tuning fork" like suspension system vibrating as a clamped cantilever. Eq. 4 holds for a clamped cantilever of constant cross section and does not take into account the small section of the suspension system that connects the resonant bar with the "tuning fork" (part No 3 in fig. 1). This latter part can be treated as an added mass to the cantilever and its influence on the frequency of the "tuning fork" becomes

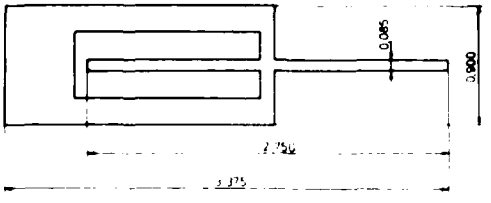
$$f = \frac{n}{M} \qquad (6)$$

where "f" is the frequency shift caused by the connecting part whose mass is M. M is the mass of one tine of the "tuning fork". Eq. 6 is approximate only if M is small compared to unity. It follows from eq. 6 in connection with eq. 4 and 5 that the suspension system should be designed such that the ratio of the length-extensional resonance between the second and third overtone of the suspension system "tuning fork".

[illegible]

[illegible]

[illegible]



[illegible] MHz resonator. All [illegible] mm.

[illegible] of the [illegible] and the measured [illegible] are:

[illegible] fig. 3

	a	b
[illegible]	1'0[illegible] kHz	
[illegible]	[illegible]	
[illegible]	0.1 fF	1.[illegible] fF
[illegible]	[illegible] kΩ typ.	600 Ω typ.

Compared to the available 1 MHz crystals these data look very promising even though they have been measured on prototype series and there may be some changes before the design goes into production.

The resonant frequency of 1'020 kHz confirms the hypothesis that the resonator can be visualized as composed of two clamped-free resonant bars vibrating in antiphase. The effective length of the clamped bar is 1.475 mm which gives, according to eq. 1, a frequency of 1'020 kHz. The mass loading due to the electrodes reduces the frequency by roughly 1% which gives a resonant frequency of 1'010 kHz. If the resonator would vibrate as a free-free bar the frequency would be 980 kHz. The experimental result obviously confirms the model of the two clamped-free resonant bars.

4.3. Frequency distribution

Since the frequency is influenced only by the crystal length, which is the largest dimension of the resonator, a narrow frequency distribution may be expected. This has been verified experimentally and the frequency distribution is well within ±.1% of the center value. This means that for certain applications these resonators could be used without any further tuning.

4.4. Frequency Spectrum

Even though the litterature mentions a strong coupling between length-extension and flexure [6] and there are many flexure modes in the vicinity of the length-extension resonance (cf. fig. 5) the frequency spectrum is very pure. As shown in fig. 7, no other peaks than the main resonance at 1'020 kHz can be seen between 200 kHz and 1.8 MHz and no peak

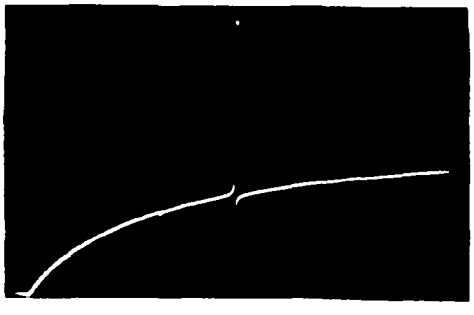


Fig. 7: Frequency Spectrum of the length-extension resonator, vert. scale: 5 dB/div, horiz. scale: 200 kHz/div, marker: 1'020 kHz.

at all appears between 0 and 200 kHz (Photo not shown).

4.4. Temperature behaviour

A typical frequency vs temperature curve is shown in fig. 8.

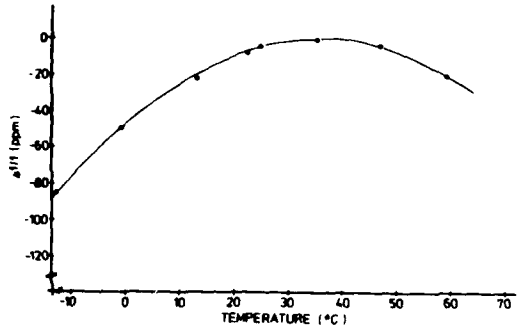


Fig. 8: Temperature behaviour

The curve shows the expected parabolic behaviour. The turning point is at 36°C, a value slightly higher than the 25°C observed with flexural mode resonators (tuning forks) made from the same material and cut angle. The latter was the usual "z + 2°" cut used for watch tuning forks. The agreement between the experimental value and the theoretical predictions using Adams [7] constants is satisfactory and the turning point may be adjusted by changing the cut angle. For the applications this resonator has been designed for it is an advantage to use a turning point above room temperature. The operating temperature range is often specified from -20°C to +80°C or from -20°C to +100°C giving an optimum value for the turning point of 30°C to 40°C.

5. Conclusions

This paper presents a miniature length-extension crystal that can be produced using photolithographic techniques and chemical milling. The present resonator has been designed for frequencies around 1 MHz, but the concept is useful for resonators from several hundred kHz to several MHz.

The mechanical and electrical parameters of this crystal are comparable or better than the AT or face-shear (DT) crystals presently on the market for low-cost applications. Due to its small size (the resonator has been packaged in watch tuning fork cans of 2 mm diameter and 6 mm length and in ceramic packages of 1.5 x 2.4 x 6.7 mm) this resonator is a miniature component on a printed circuit board and can be assembled with an integrated circuit in a standard dual in-line plastic package to form a miniature time base module.

Acknowledgements

The author would like to thank M. Christen and T. Gladden for their assistance in the preparation of the manuscript.

References

[1] S.S. Chuang, preceeding paper in this conference

[2] J.A. Kuster et al, Proc. of the 30th Symposium on Frequency Control (1976), p. 167

[3] R.A. Heising, Quartz Crystals, D. van Nostrand Co. Inc., 1946, p. 45

[4] J. Michel, Swiss Patent No 12124/75

[5] L.D. Landau, E.M. Lifshitz, Elastizitätstheorie, Akademie-Verlag, Berlin, 1966, p. 129

[6] Same as [3], p. 538

[7] C.A. Adams et al, Proc. of the 24th Symposium on Frequency Control (1970), p. 55

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

INVESTIGATION OF SPURIOUS MODES OF CONVEX DT-CUT QUARTZ CRYSTAL RESONATORS

T. Adachi, Y. Tsuzuki
Yokohama National University
Yokohama, Japan 240

and

C. Takeuchi
Fujitsu Ltd.
Kawasaki, Japan 211

Summary

Investigation has been made on the spurious modes of convex DT-cut quartz crystal resonators and a method to avoid these spurious modes has been established.

The convex DT-cut quartz crystal resonator is a thin bar having the same orientation as the usual square DT-cut resonators. Its length is parallel to the X-direction and one of its major surface is slightly curved as a circular shape. The main mode of the resonator is the width-shear mode.

Detailed displacement distributions of the spurious mode have been investigated by means of laser holographic interferometer. It became clear that the spurious modes were the inharmonic thickness-flexure modes of even order along the length direction.

The relation between the resonance frequency and the dimensions of the resonators has been calculated. Satisfactory agreements have been obtained between the calculated results and the measurements, and the proper dimensions to avoid the spurious response troubles have been determined.

INTRODUCTION

Convex DT-cut quartz crystal resonators have been in practical use in the frequency range from 500 kHz to 1 MHz, i.e., the boundary region between the thickness-shear mode resonators and the contour mode resonators. The resonator is a thin bar having the same orientation as the usual square DT-cut resonators. Its length is parallel to the X-direction (the electric axis) and one of its major surface is slightly curved as a circular shape (FIG. 1). The main mode of this resonator is the width-shear mode and is excited by the electrodes evaporated on the surface of the plate. The main mode is confined to the center region of the plate by the convex structure, therefore the plate can be supported tightly at its ends. Recently, it has been found out that a very strong spurious response appears occasionally near the main resonance. Consequently, in order to establish a method to avoid these spurious modes, it has become necessary to investigate the vibration modes of the spurious modes.

EXAMPLE OF SPURIOUS RESPONSE

TABLE 1 shows the plate dimensions of three samples having the spurious response near the main resonance. The admittance characteristics of sample 1 is shown in FIG. 2. There are two strong responses close to each other in this admittance characteristics. FIG. 3 shows the temperature characteristics of two admittance-maximum frequencies. The spurious mode has negative frequency-temperature coefficient and has strong coupling with the main mode.

OBSERVATION OF VIBRATION MODE

The laser holographic interferometer has been applied to the investigation of the spurious mode.[1] By means of the laser holographic interferometer, the observation of the displacement component perpendicular to the surface of a plate can be made independently of the displacement components parallel to the plate. It has been known by experiments that the reduction of the plate thickness increases the frequency of the spurious mode. This suggests that the spurious modes have a displacement component perpendicular to the plate. It has already been confirmed that the main mode has only the displacement component parallel to the plate. Accordingly, the displacement component perpendicular to the plate is that of the spurious mode, even though the resonator vibrates in the main mode at the same time.

FIG. 4 shows the principle of holographic observation method of displacement component perpendicular to the plate. The vibration displacement is observed by the following manner. First, a laser beam is projected onto a mirror and the resonator under study. The reflected beam from the mirror acts as a reference beam and a part of the reflected beam from the resonator becomes a signal beam. The two beams reach a photographic plate and the interference fringes are recorded on the plate. The developed photographic plate is called hologram. Next, the resonator is removed and the hologram is placed in the initial position. The reference beam is diffracted by the hologram and the original signal is reproduced. As a result, the three dimensional reconstructed image

of the resonator can be observed at the initial position through the hologram. This reconstructed image is photographed by a camera. In the case of vibrating object, a pattern of dark lines showing the vibration displacement appears in the reconstructed image.

FIG. 5 shows the reconstructed images of the spurious modes of three samples depicted in TABLE 1. The centers of brightest portions in the images indicate locations of nodes where the amplitude of the displacement component is zero and the dark lines represent a group of equal-displacement curves. In all reconstructed images, there are two nodal lines parallel to the length direction. Hence, the displacement distribution along the width direction is similar to that of the fundamental thickness-flexure vibration of a rectangular plate. In addition, there are even number of loops along the length direction in all images. there are two, four, and six loops respectively. From these results, it can be said that these spurious modes are the inharmonic thickness-flexure modes having even order along the length direction.

The reason, why only even-order inharmonic thickness-flexure modes appear, can be explained in the following manner. When the three displacement components of spurious mode are denoted as U_1 , U_2 , and U_3 with reference to **x**, **y**, and **z** axis as shown in FIG. 6, U_1 , U_2 , and U_3 can be expressed as follows.[2]

$$U_1 = y\ \psi_x\ (x,z)$$

$$U_2 = \eta\ (x,z) \qquad (1)$$

$$U_3 = y\ \psi_z\ (x,z)$$

Then, shearing stress components S_5 and S_6 are expressed by η , ψ_x , and ψ_z in the following relations.

$$S_5 = \frac{\partial U_1}{\partial z} + \frac{\partial U_3}{\partial x} = y\ \left(\frac{\partial \psi_x}{\partial z} + \frac{\partial \psi_z}{\partial x} \right)$$

$$S_6 = \frac{\partial U_1}{\partial y} + \frac{\partial U_2}{\partial x} = \psi_x + \frac{\partial \eta}{\partial x} \qquad (2)$$

The y component of electric displacements for rotated Y-cut plate is given by the following equation.

$$D_2 = e_{25}\ S_5 + e_{26}\ S_6 \qquad (3)$$

where e_{25} , e_{26} are the piezoelectric coefficients defined by IRE Standard. Accordingly, the total charge Q on the electrode can be calculated by the following equation.

$$Q = \int_S \left(\frac{1}{t} \int_{-\frac{t}{2}}^{\frac{t}{2}} D_2\ dy \right) dx\ dz$$

$$= e_{26} \int_S \left(\psi_x + \frac{\partial \eta}{\partial x} \right) dx\ dz \qquad (4)$$

In the above equation, it can be assured that ψ_x is proportional to $\partial \eta / \partial x$. Then, the total charge Q becomes zero when η is even function along x and even function along z, or η exists when η is odd along x and even along z.

In order to discuss the resonance frequency of spurious mode, the displacement distribution has been determined from the holographic patterns. The displacement amplitude U , at the location of the n-th dark line counted from the brightest portion in the reconstructed image, can be calculated from the following equation.[1]

$$U = \frac{(4n-1)\ \lambda}{8\ |\cos \theta_i + \cos \theta_r|} \qquad (5)$$

where λ : wave length of the laser
θi : the angle of the incidence
θr : the angle of the reflection (FIG. 4).
In the experiment, λ = 6328 [A] (He-Ne laser), and $\theta_i = \theta r = 12$ [deg.], then Eq. (5) becomes

$$U = 404\ (4n-1) \qquad [A] \qquad (6)$$

Using this equation, the displacement distributions have been obtained for sample 1 and 2 as shown in FIG. 7. FIG. 7(a) shows the distribution of U_2 along the width direction obtained from FIG. 5(a). This distribution is very close to that of the fundamental thickness-flexure vibration of a rectangular plate. The same distributions have been obtained from FIG. 5(b) and (c). FIG. 7(b) and (c) show the distributions of U_2 along the length direction obtained from FIG. 5(a) and (b) respectively.

CALCULATION OF RESONANCE FREQUENCY OF SPURIOUS MODE

$$f = \frac{1}{2\pi}\left[\frac{\int\int_{ij} S_i S_j \, dxdydz}{\int\int\int U_j \, dxdydz}\right]^{\frac{1}{2}} \qquad (7)$$

$$U_1 = X(x)\,Z_1(z) + Z_2(z)$$

$$= H_{n-1}\left(\frac{x}{a}\right)\exp\left(-\frac{x^2}{4a^2}\right)\left(A_1\cos\gamma_1 z + A_2\cosh\gamma_2 z\right) \qquad (8)$$

$$U_2 = -y\frac{dX(x)}{dx}\left(Z_1(z) + Z_2(z)\right) \qquad (9)$$

$$U_3 = yX(x)\left(\gamma_1 Z_1(z) + \frac{dZ_1(z)}{dz}\right) + \left(\gamma_2 Z_2(z) + \frac{dZ_2(z)}{dz}\right) \qquad (10)$$

where H_{n-1} is the $(n-1)$th order Hermite function, and is given by the following equation.

$$a = K\sqrt[4]{\frac{R\,W^3\,D_0}{16\pi^2\,D_4}}$$

In Eq. (11), K is a constant which is equal to unity, if $n = 1$. The value of the constant K in Eq. (11) is determined experimentally so that the distribution function S_3 determined by Eq. (7) may be the most approximate to the measured distribution. The optimum distribution function S_3 determined by Eq. (7) and shown by the solid curve in Fig. 7 is the case, where the value of K have been set to unity. It can be said that the approximation of distribution is fairly good. The resonant frequencies of spurious modes of three samples have been calculated by using Eq. (7) - (11). The differences between calculations and measurements are less than four percent.

The undesirable dimensions can be easily avoided by setting the resonance frequency of each spurious mode given in Eq. (7) equal to that of the main mode. The resonance frequency of main mode is given by the following equation.

$$f = \frac{1}{2W}\left(\frac{1}{\rho S_{55}}\right)^{\frac{1}{2}}\left(1 + \sqrt{\frac{WS_{55}}{RS_{11}}}\,\frac{1}{\pi}\right)^{\frac{1}{2}} \qquad (12)$$

where S_{ij} : elastic compliance constant.

The calculated results are shown as the solid lines in Fig. 8. In the calculation, the constant K in Eq. (11) is assumed to be unity. Six resonators have been made so as to have the undesirable dimensions shown in Fig. 8. Satisfactory agreements have been obtained between the calculated results and the measured ones with the estimated error less than four percent.

From the results obtained above, a method to avoid the spurious modes can be described as follows. When there is no restriction on the thickness of the plate, i.e., no restriction on the value of equivalent series inductance, the spurious modes can be avoided by making the thickness larger than the values determined by curve 2, i.e., when the thickness should be smaller than the values determined by curve 2, i.e., the equivalent series inductance should be selected to a specified value, the thickness must be determined to be the value apart sufficiently in the middle of two adjacent curves in Fig. 8. In the case of a full electrode, the value of equivalent series inductance

is given in Fig. 4, when the equivalent series inductance should be adjusted to a specified value, the thickness must be not only determined to the value approximately in the middle of two adjacent curves but also must be selected to make the value of inductance smaller than the value determined from Fig. 4. By the use of partial electrode, the value of inductance can be adjusted to any specified value higher than that of full-electrode plate.

CONCLUSION

Holographic observation of vibration modes has been made on the series of spurious modes of convex DT-cut crystal resonators. It has been found that these spurious modes are the anharmonic thickness-flexure modes having even order along the length direction. The resonance frequencies of these spurious modes have been calculated and the undesirable dimensions, where the resonance frequency of spurious mode coincide with that of the main mode, have been determined. From these results, a method to avoid the spurious modes in a convex DT-cut crystal resonator has been established.

REFERENCES

1. T. Ikeda, Y. Hirose, and K. Iijima: "Measurement of vibration modes of piezoelectric resonators by means of holography", Proc. of the 25th Annual Freq. Cont. Symp., pp. 113-117, April 26-28, 1971.
2. R. D. Mindlin: "Thickness-shear and flexural vibrations of crystal plates", J. Appl. Phys., 22, 3, pp. 316 (Mar. 1951).
3. C. J. Wilson: "Vibration modes of AT-cut convex quartz resonators", J. Phys. D, 7, pp. 2449 (July 1974).

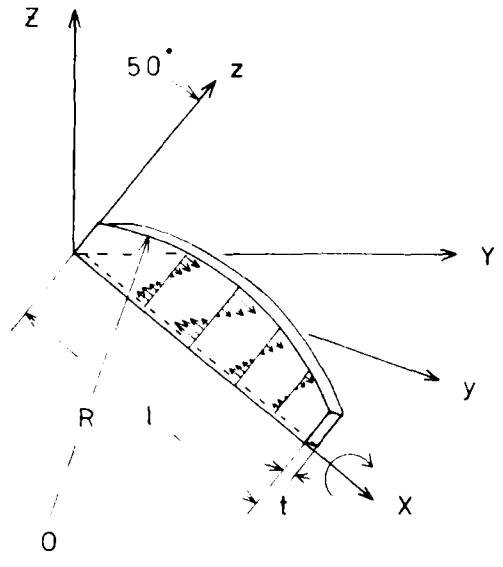


FIG. 1 DIMENSIONS AND VIBRATION MODE OF CONVEX DT-CUT CRYSTAL RESONATOR

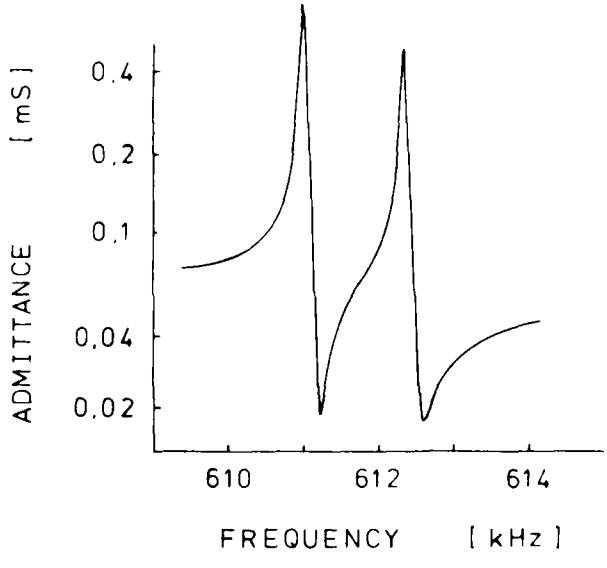


FIG. 2 EXAMPLE OF ADMITTANCE CHARACTERISTICS SHOWING THE SPURIOUS RESPONSE

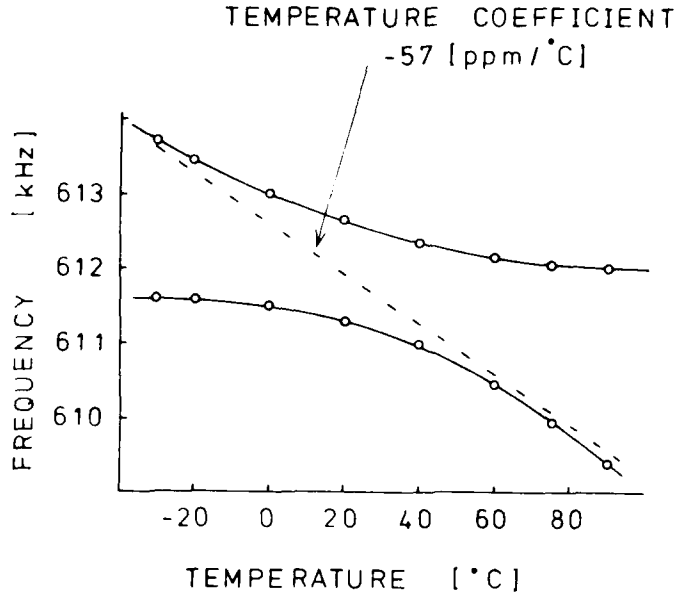


FIG. 3 TEMPERATURE CHARACTERISTICS OF TWO ADMITTANCE MAXIMUM FREQUENCIES

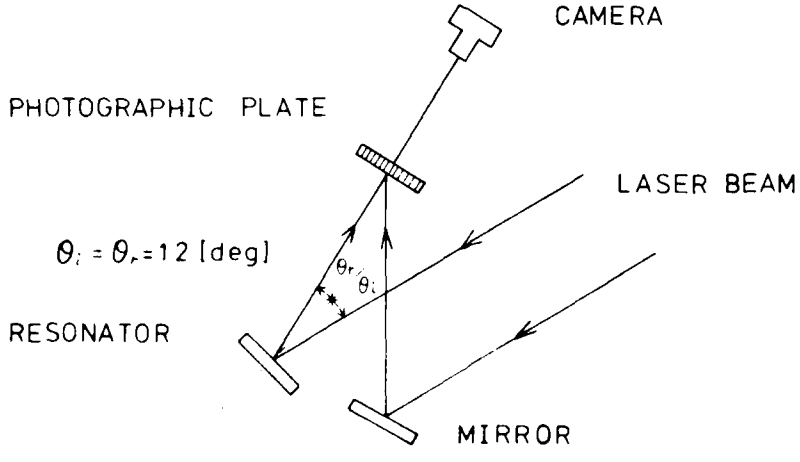
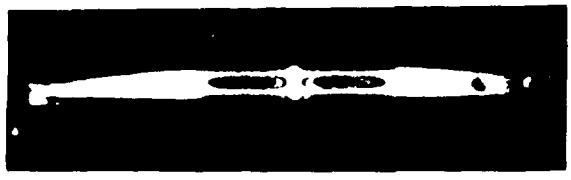
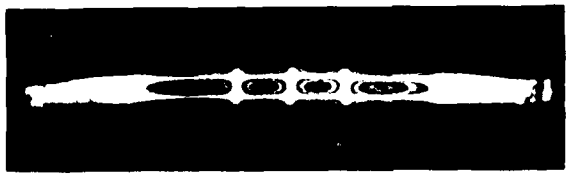


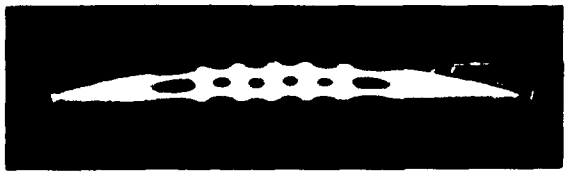
FIG. 4 HOLOGRAPHIC OBSERVATION METHOD OF VERTICAL VIBRATION DISPLACEMENT



(a) SAMPLE 1 (611.4kHz)

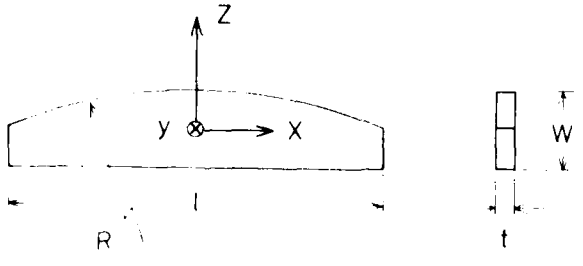


(b) SAMPLE 2 (526.3kHz)



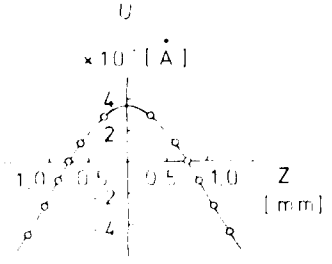
(c) SAMPLE 3 (571.4kHz)

FIG. 5 RECONSTRUCTED IMAGES OF SPURIOUS MODES OF THREE SAMPLES



0

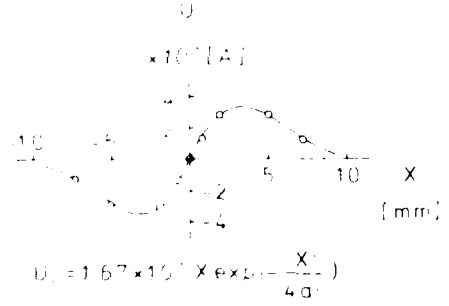
FIG. 6 COORDINATES OF DT-CUT RESONATOR



$U_z = 4.36\times10^{-7}[\cosh(1.73Z)-0.17\cosh(1.26Z)]$

(a) DISTRIBUTION ALONG WIDTH DIRECTION

(SAMPLE 1)

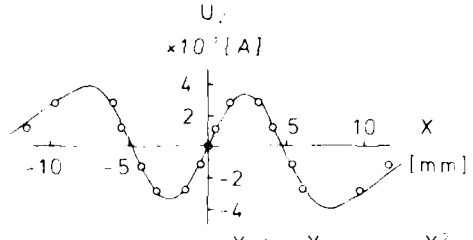


$U_x = 1.67\times10^{-7} X \exp(-\frac{X^2}{4a^2})$

(a = 2.42 mm)

(b) DISTRIBUTION ALONG LENGTH DIRECTION

(SAMPLE 1)



$U_x = 2.07\times10^{-7}[(\frac{X}{a})^3 - 3(\frac{X}{a})]\exp(-\frac{X^2}{4a^2})$

(a = 2.75 mm)

DISTRIBUTION ALONG LENGTH DIRECTION

(SAMPLE 2)

FIG. 2 DISTRIBUTION OF DISPLACEMENT COMPONENT U

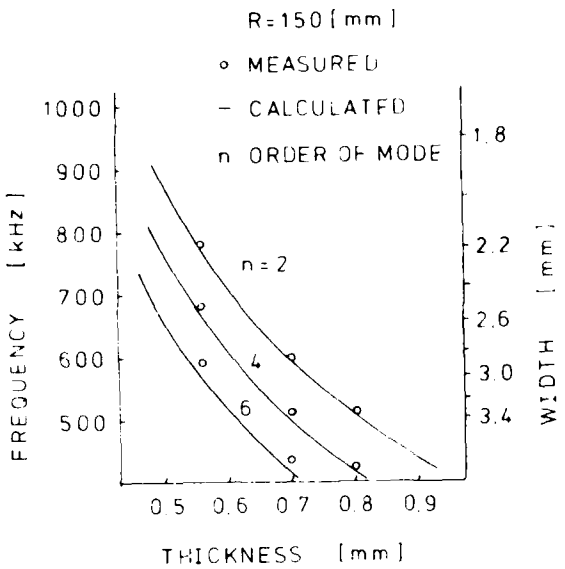


FIG. 8 PLATE DIMENSIONS WHERE RESONANCE FREQUENCY OF SPURIOUS MODE COINCIDE WITH THAT OF MAIN MODE

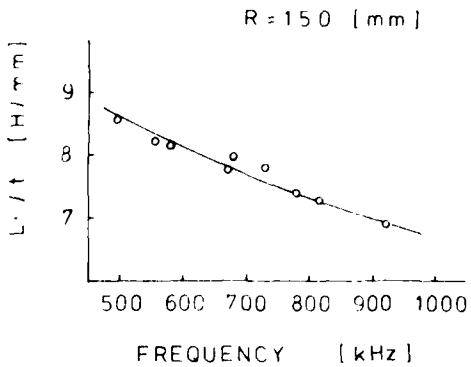


FIG. 9 RELATION BETWEEN EQUIVALENT SERIES INDUCTANCE PER UNIT THICKNESS AND RESONANCE FREQUENCY (MAIN MODE)

TABLE 1 PLATE DIMENSIONS OF THREE SAMPLES

SAMPLE NO.	WIDTH W [mm]	THICKNESS t [mm]	LENGTH l [mm]	CURVATURE R [mm]
1	2.80	0.80	40.0	150
2	3.28	0.80	40.0	150
3	3.04	0.59	39.5	70

Proc. 35th Ann. Freq Control Symposium, ECAERADCOM, Ft. Monmouth, NJ 07703, May 1981

THE EDGE MODE RESONATOR

Daniel C.L. Vanyhelnwe* and E. Dennis Fletcher**

* N.V. Philips Gloeilampenfabrieken, Eindhoven, Nederland.
** Philips Research Laboratories, Redhill, England.

Summary

A new concept of a small resonator has been studied with the finite element method. The resonator vibrates in an edge mode, which is a combined motion of symmetric shear and dilatation, confined to an edge. For quartz, temperature compensated cuts are found theoretically. The mounting is at the side opposite to the active edge.

It is shown theoretically with the finite element method, that the quality factor of the resonator, which is sensitive to asymmetry, can be improved by balancing.

Experiments confirm the temperature compensated cuts, found theoretically. The electrodes also excite other modes which are damped by the mounting, whereas the response of the edge mode is relatively unaffected by the mounting. By adding a small mass the quality factor of the resonaters is improved. After balancing a quality factor of 50000 has been measured for these resonators.

Introduction

Following the latest developments in the miniaturisation of quartz resonators for watches and clocks, we have been working on a new type of resonator, which appears to be similar to the tuning fork resonator in many respects. These are, the temperature coefficient of frequency, the balancing and the mounting on one side. Especially, the second order behaviour of the temperature coefficient is comparable to the 5°X cut tuning fork. The differences with the tuning fork resonator are: its frequency of application, which is well above 100 kHz up into the MHz range, and the motion, which is shear and extension instead of flexure and torsion.

The mode appearing in the new design was first noticed by E. Shaw[1] in a circular disk of $BaTiO_3$, poled in the thickness direction. Because it is confined to the edge of the disk it has been called edge mode. It was subsequently reported by Onoe[2] on a disk of ADP and by Ikegami[3] on a disk of $PbTiO_3$. In 1960, the edge mode was explained theoretically by Gazis and Mindlin[4] from a branch with complex wavenumber of the dispersion equation of a plate. The branch starts at the cut off frequency of the thickness extensional mode and extends to lower frequencies with a complex wavenumber. According the Mindlin and Gazis, the edge mode cannot exist on its own, but is always associated with an extensional (longitudinal) mode. The complex wavenumber, however, suggests that the edge mode can be trapped, with the distribution of the vibration amplitude away from the edge in the form of a damped oscillation. Because of this oscillating longitudinal motion, the total momentum can be conserved, making confinement a possibility.

With its particularly large attenuation coefficient (3x better than the maximum obtainable for facetted crystals), the edge mode branch is interesting for the design of small resonators. For this reason, a study on this mode with the finite element program ASKA[5] was started. The calculation time of the program was reduced by assuming that the edge mode motion is restricted to a plane, leaving the motion in the third direction free (this is equivalent to having a small dimension in that direction). A rectangular resonator was divided into 4 elements along the thickness and 16 elements along the length, with each element being a two dimensional one with 9 nodal points (giving a total of 594 degrees of freedom). Two elements with low density and weak elasticity were used to support the structure.

From ASKA two resonances resulted with an edge mode appearance and a frequency different of only $5\ 10^{-6}$ for a length to thickness ratio of 3.7. From this we concluded that the two resonances are only weakly coupled due to the attenuation. The two resonances arise because the two edge modes on the opposing edges of the rectangular resonator can be either in phase or in anti-phase. This suggests that the edge mode can exist in a pure form by making one edge different from the other. The result of using ASKA to study this can be seen in figure 1. The first picture is the crystal at rest showing the elemental distribution, with the bevel at the right side in order to break the symmetry. The second picture shows the motion at one of the edge mode resonances (of the left edge). The calculations confirm the idea, that the edge mode can exist in a pure and trapped state.

The edge mode can be further understood by analysing its motion. Assume the thickness direction along the edge and the length direction

at right angles to it. The thickness determines the frequency of the edge mode resonator and its motion is attenuated along the length.

Two main components of motion can be seen (figure 1):

1) a shear motion, which changes sign, when going from the upper to the lower half of the resonator. In this way, the bending moment in the upper half is exactly opposed by the bending moment of the lower half. For this reason, the motion of the shear is similar to the motion of the second harmonic of the thickness shear mode.

2) a dilatational motion, which is similar to the motion of the fundamental of the thickness dilatational mode.

The edge mode is attenuated away from the edge, because its frequency is below the cut off frequency of its two main components of motion. At the frequency of the edge mode the two above components of motion are in phase at the edge and thus built up a resonance.

We have also studied the temperature dependence of frequency of the edge mode in quartz, because for this material the possibilities for a rotated cut with temperature compensation are present.

First Order Temperature Coefficient

The existence of a temperature compensated cut in quartz can be suspected by considering the two motions from the introduction. The shear motion stiffness in the xy' plane, c_{66}, has a positive temperature coefficient for a rotation about the x axis and the dilatational motion with a stiffness c_{22} has a negative temperature coefficient for the same rotation. Because the edge mode is composed of these 2 motions its temperature coefficient (1st order) of frequency can be expected to cross the zero line for a rotation about the x axis. For this reason, such rotations of quartz were studied, using ASKA, in order to find a temperature compensated cut for the edge mode resonator.

We used the division into elements of figure 1. Because the stiffness c_{66} must be involved, the xy' plane is the plane of motion for the ASKA calculations. In order to avoid spurious modes, the z' direction thickness is assumed to be small compared with the other dimensions of the resonator. The z' dependent motions are then below their cut off frequencies and strongly attenuated. The small z' dimension can be simulated by introducing into the ASKA program slightly reduced stiffness constants for the xy' plane, according to the following equation:

$$c'_{ij} = c_{ij} - c_{i3}c_{3j}/c_{33} - c_{i4}c_{4j}/c_{44} - c_{i5}c_{5j}/c_{55} \quad i,j=1,2,6$$

The stiffness constants and their temperature dependence up to the 3rd order as given by Bechman[6] have been used in the ASKA calculations. A further rotation about the z axis is considered to be impractical, because it introduces non zero stiffness constants c_{16},c_{26}. Because of this, the motion in the xy' plane becomes asymmetrical and confinement to only one edge becomes impossible. Two possibilities with full symmetry of motion in the xy' plane remain: a) with the x axis along the edge, b) with the x axis perpendicular to the edge. In case a) and in case b) the temperature compensated cuts are found with the ASKA program. Figure 2 gives the first order temperature coefficient of frequency against the angle of rotation about the x axis for case b). Zero's are apparent at angles of rotation of about -11° and 25°. For case a) similar results are obtained with zero's at about -37° and 26°, as can be seen in figure 4. For case b), the 2nd order coefficient, as obtained from the frequencies at 3 different temperatures is -40 and -64 $10^{-9}/(^\circ C)^2$ for angles of rotation of -11° and 25° respectively.

We define the frequency constant of the edge mode as the frequency times the thickness (being the dimension along the edge). For case a) this is the dimension in the x direction, for case b) it is the dimension in the y' direction.

Figures 3 and 5 give the frequency constant against the angle of rotation about the x axis for cases b) and a) respectively.

Theoretical Quality Factor

As mentioned in the introduction, the motion of the edge mode oscillates as a function of the distance to the edge, with the amplitude decaying. The oscillation is necessary in order to conserve the total momentum of the longitudinal motion (this is the motion perpendicular to the edge). Figure 6 gives the two components of motion of the edge mode from ASKA as a function of the distance from the edge for a rotation about the x axis of respectively 25° and -10°, with the y' axis parallel to the edge. The xy' shear is represented in figure 6 by the maximum strain S_6 (at d/4 from the centre) and the dilatation in y' direction by the strain S_2 at the centre. The difference in S_6 for the two angles of cut is apparent. The decay factor for the 25° cut resonator is the largest one. We have found, that the decay of the amplitude of both motions can be described accurately by an exponential function.

The loss due to the mounting of the resonator is a critical factor in the design of small crystals. For this reason we estimated the quality factor Q due to the mounting loss for the edge mode. Assuming, that the loss per period in the mounting is cu_c^2 and the energy of the resonator is cu_o^2, the quality factor is given by: $Q = 2\pi(u_o/u_c)^2$. u_c is the amplitude at the mounting distance and u_o is the average amplitude over the active region. c is an arbitrary

constant with the dimension of elasticity, if the energy is taken per unit of length in the z direction. The ratio u_0/u_e can be derived from the ASKA results. The following table gives the results:

Table 1

Theoretical length to thickness ratio (ℓ'/d) for a $Q=10^5$ due to mounting loss

Angle of Cut	Length/Thickness for $Q=10^5$	Attenuation at distance d (dB)
-10°, y'// edge	5.1	-13.8
$+25^\circ$, y'// edge	3.2	-22.6
$+25^\circ$, x// edge	4.9	-14.7

In the table, the mounting distance in length/thickness is given for $Q = 10^5$. As can be seen from the table, the relative mounting distance ℓ/d is between 3.2 and 5 for the angles of interest. This is small compared to ℓ/d = 15 for the AT type resonator with the shape of a bar, proposed by Onoe[7] and the design with the same shape, proposed by Zumsteg[8].

Design

The design with the y' axis parallel to the edge has been developed further because at the temperature compensated angles of cut, the coupling factor to the xy' shear is about at a maximum[9]. A typical edge mode resonator is shown in figure 7. The electric field, which drives the resonator, gives the maximum coupling when it corresponds to the field coupled to the resonance. This field pattern can be inferred from the motion at resonance, neglecting the piezoelectric effect. We may therefore use the ASKA result of figure 6 to design the electrodes on the edge mode resonator. It can be seen from the figure, that the absolute value of strain S_6 is a maximum at distances ± d/4 from the centre along the y' axis. This motion is excited by an electric field in the +y direction for y = d/4 and in the -y direction for y = -d/4. As can be seen in figure 7, the 3 electrode pairs perform this function approximately. Because the shear decays considerably over a distance equal to the thickness, the electrodes do not need to extend beyond that distance from the edge. The electric contact is made at the other end in x direction, where the crystal is mounted rigidly. As we will discuss later on, the rigid mounting is necessary for the suppression of spurious resonances.

Balancing Considerations

As was mentioned in the introduction, the frequency of the edge mode is below the cut off frequency of its two components of motion, one of which is the xy' shear, which is symmetric along the thickness. For the anti-symmetric xy' shear however, the cut off frequency is below the frequency of the edge mode (being the fundamental frequency for thickness shear). As a consequence of this, any asymmetry in the shear motion of an edge mode resonator will result in the propagation of this anti-symmetric motion and add to the loss at the mounting. Any imperfection, which disturbs the symmetry can cause this. Suppose for instance, that the edge is oblique to the faces, as shown in figure 9. The (twisting) moment created by the xy' shear in the triangle AOC of figure 9 cannot be compensated for by an opposing moment and anti-symmetric shear and flexure components will arise. A first estimate shows that the corners at the edge must be rectangular to within 1' for a quality factor of 10^5.

Because this tolerance is considered too small, the possibility of balancing by adding a thin layer near one of the corners, was studied. Adding mass to the corner A' in figure 9 will compensate for the redundant mass contained in the triangle AOC and restore to a first order the balance of moments at the edge. This could be a solution to improve the Q of imperfect resonators. The balancing has been simulated by ASKA by displacing 3 nodal points at the corner A' by amounts d/800 and d/400 respectively for a 15° cut resonator, with the edge oblique at 17' to the x axis. The resulting displacement is given in the two lower pictures of figure 8. The first and the second pictures of figure 8 are respectively the crystal at rest and the displacement at the edge mode resonance without the balancing. The last picture in figure 8 indicates, that balancing can reduce the spurious modes caused by asymmetric imperfections of the crystal.

Experimental Work

We have experimented with edge mode resonators designed as in figure 7 with electrodes of gold on chromium, defined by photolithography and dimensions of respectively 3.2 x 0.8 x 15.2 and 2.0 x 0.5 x 9.5 mm. The first number is the dimension parallel to the edge in the y' direction, the second in z', and the third in x. The first number determines the frequency of the edge mode and the frequencies for the above resonators are respectively 712.6 and 1180 kHz for a -12° cut angle. The crystals have been made at various angles of cut for a rotation about the x axis of quartz.

Figure 2 gives the experimental results for the first order temperature of frequency. The results follow the theoretical curve, although on average they appear to be somewhat lower in value. The temperature compensation point is rather -8° than the theoretical value -11°.

The experimental results of the frequency constant against the rotation angle are close to the prediction as can be seen in figure 3. The table below shows that the second order temperature

coefficient of frequency also agrees with the experiment.

Table 2

The second order temperature coefficient of frequency

Angle of cut (y' // edge)	Theory $(10^{-9}/(^{o}C)^{2})$	Experiment $(10^{-9}/(^{o}C)^{2})$
25°	-64	-56
-12°	-40	-39

As we can see from the table, the 2nd order coefficient of the -12° cut resonator is comparable to a value of $-35\ 10^{-9}/(^{o}C)^{2}$ for the 5°X cut tuning fork resonator.

Other modes are excited as well by the electrodes of figure 7 as can be seen in figure 11 for a lightly damped resonator. The group of closely spaced responses at 1.6 MHz corresponds to a group of 4 resonances from ASKA. The displacement of these resonances is shown in figure 10 and the responses can be classified as thickness extensional modes. Figure 12 for a resonator glued solidly to a base, shows that all responses, except the edge mode are heavily damped. This is a definite proof that the edge mode exists in this type of crystal.

Without balancing, the Q value of the edge mode resonator is smaller than 10^{4}. In order to improve the Q value, some resonators have been balanced with polystyrene cement on the edge. Figure 13 shows the increase of the response due to the balancing. The balancing is given in terms of the frequency shift it causes. Resonators at -12° show an improved quality factor (in vacuo) after the balancing. Similar results have also been obtained by electroplating one of the electrode pairs at the outside of the resonator. As a result of the first balancing experiments, we have obtained the following properties for the -12° cut resonator:

Table 3

Properties after the first balancing experiment

Size (y'xz'xX) (mm)	Freq. (kHz)	C_0* (pF)	C_1 (L_1) (fF)	Q (R)
3.2x0.8x15.2	712.6	0.13	0.13 (390 H)	$4\ 10^{4}$ (40 kΩ)
2.0x0.5x9.5	1180	0.080	0.086 (212H)	$4\ 10^{4}$

* Calculated

Conclusion

The edge mode resonator as predicted with ASKA has been realized experimentally. The experiments confirm the predictions about the temperature dependence of frequency for a rotation in quartz about the x axis and temperature compensated cuts do exist. Also the idea of improving the quality factor by balancing has been confirmed.

The edge mode resonator and the tuning fork resonator have several aspects in common. These are: the second order temperature coefficient, of frequency, the balancing and the mounting (on one side). They differ in the frequency range of application. Considering the size, the edge mode may find an application in the range of 100 kHz to 10 MHz. In this range, the edge mode resonator is small compared with other designs and can be produced by the same large scale production methods as the tuning fork crystal.

The small motional capacitance of the crystal, however, requires a specially designed oscillator.

Acknowledgement

We are pleased to acknowledge valuable discussions with Dr R. Milsom and also the contributions made by Dr J.E. Curran, our Technology Group and C.C. Li in manufacturing and measuring the experimental devices.

References

1. E.A.G. Shaw, "On the Resonant Vibrations of Thick Barium Titanate Disks", J. Acoust. Soc. Am. 28, 38 (1956).

2. M. Onoe and Y.H. Pao, "Edge Mode of Thin Rectangular Plate of Barium Titanate", J. Acoust. Soc. Am. 33, 1628 (1961).

3. S. Ikegami, I. Ueda and S. Koboyashi, "Frequency Spectra of Resonant Vibration in Disk Plates of $PbTiO_3$ Piezoelectric Ceramic", J. Acoust. Soc. Am. 55, 339 (1974).

4. D.C. Gazis and R.D. Mindlin, "Extensional Vibrations and Waves in Circular Disk and Semi-infinite Plate", J. Appl. Mech. 27, 541 (1960).

5. D.C.L. Vangheluwe, "Finite Element Analysis of AT-cut Crystals", Proc. 32nd Annual Symposium on Frequency Control, pp 134-141 (1978).

6. R. Bechman, A.D. Ballato and T.J. Lukaszek, "Higher-order Temperature Coefficients of the Elastic Stiffnesses and Compliances of Alpha-quartz", Proc. I.R.E. 50, 1812 (1962).

7. M. Onoe and M. Okazaki, "Miniature AT-cut Strip Resonators with Tilted Edges", Proc. 29th Annual Symposium on Frequency Control, pp 42-48 (1975).

8. A.E. Zumsteg and P. Suda, "Properties of a 4 MHz Miniature Flat Rectangular Quartz Resonator Vibrating in a Coupled Mode",

Proc. 30th Annual Symposium on Frequency Control, pp 196-201 (1976).

9. A. Ballato, Physical Acoustics XIII, p 136, Academic Press (1977).

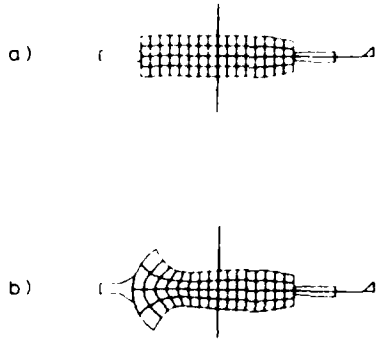


Fig 1 Finite element calculation of the edge mode
a) Elemental distribution in XY' plane
b) Displacement at resonance

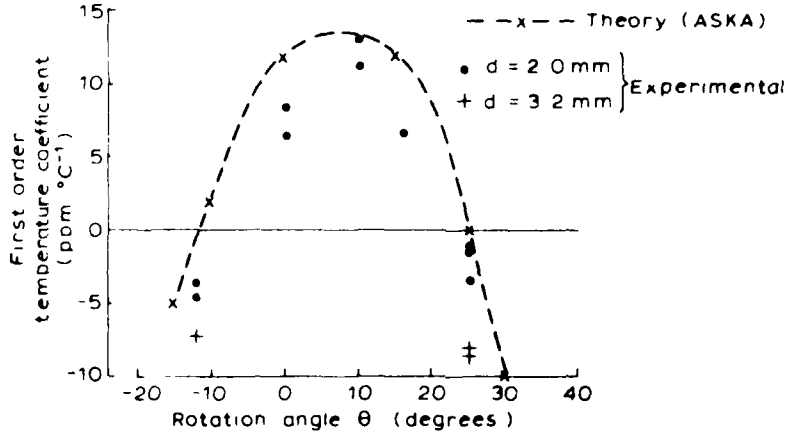


Fig 2 First order temperature coefficient as a function of rotation angle θ (width in direction Y')

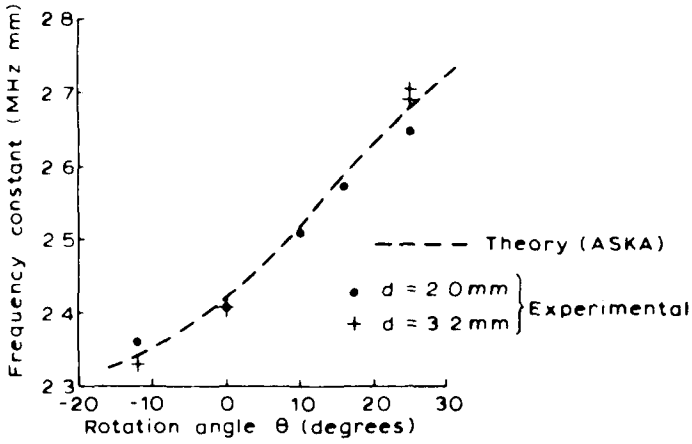


Fig 3 Frequency constant as a function of rotation angle θ (width in direction Y')

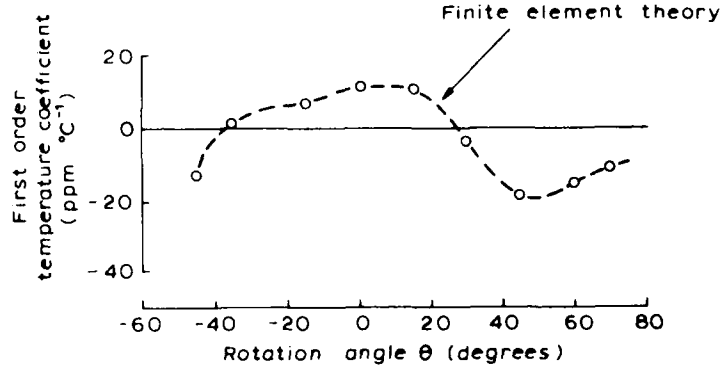


Fig 4 First order temperature coefficient as a function of rotation angle θ (width in direction X)-Theoretical

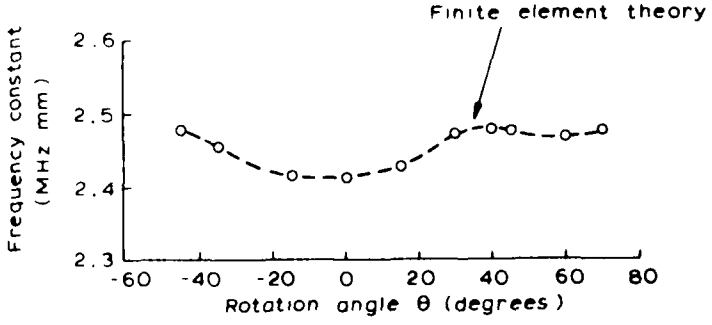


Fig. 5 Frequency constant as a function of rotation angle θ (width in direction X)- Theoretical

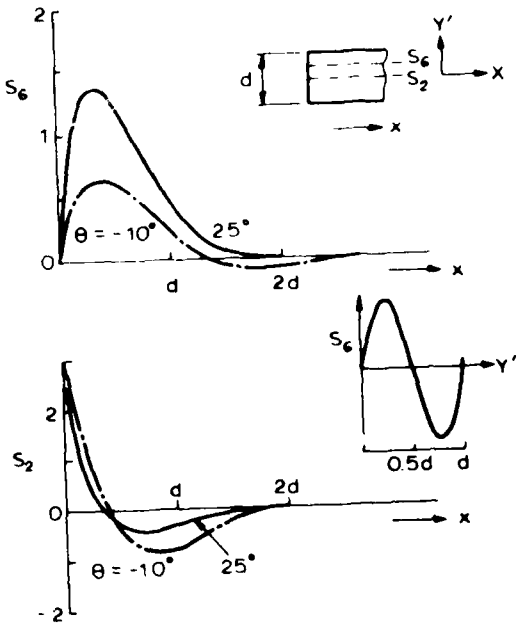


Fig 6 Strain amplitude distributions for the edge mode

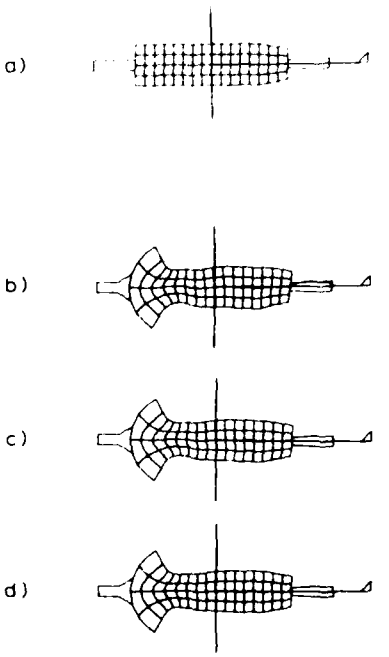


Fig 8 Finite element calculations of balancing considerations

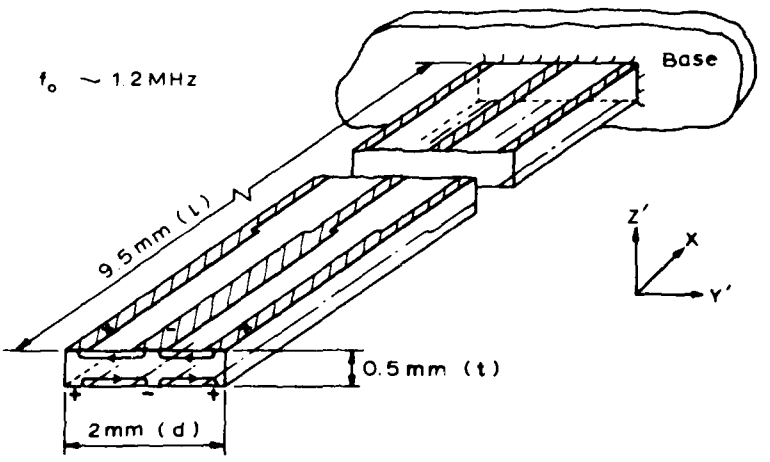


Fig 7 A typical edge mode resonator

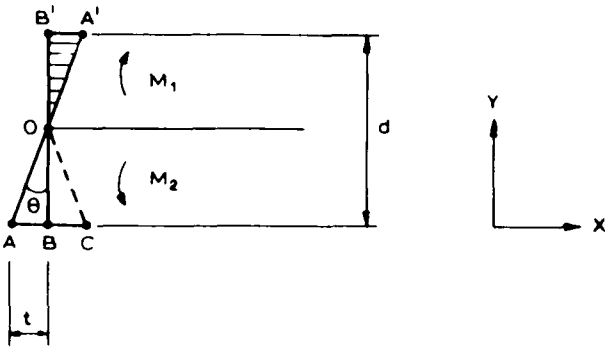


Fig 9 The definition of an oblique edge for balancing considerations

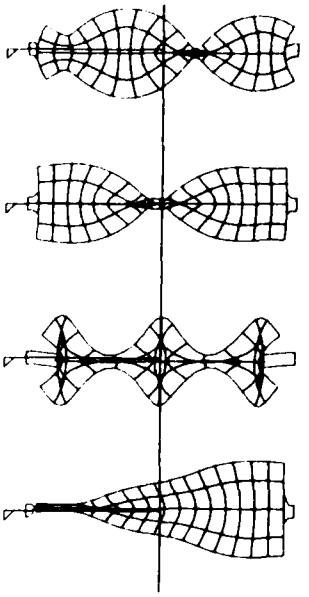


Fig 10 Some thickness-extensional modes by finite element analysis

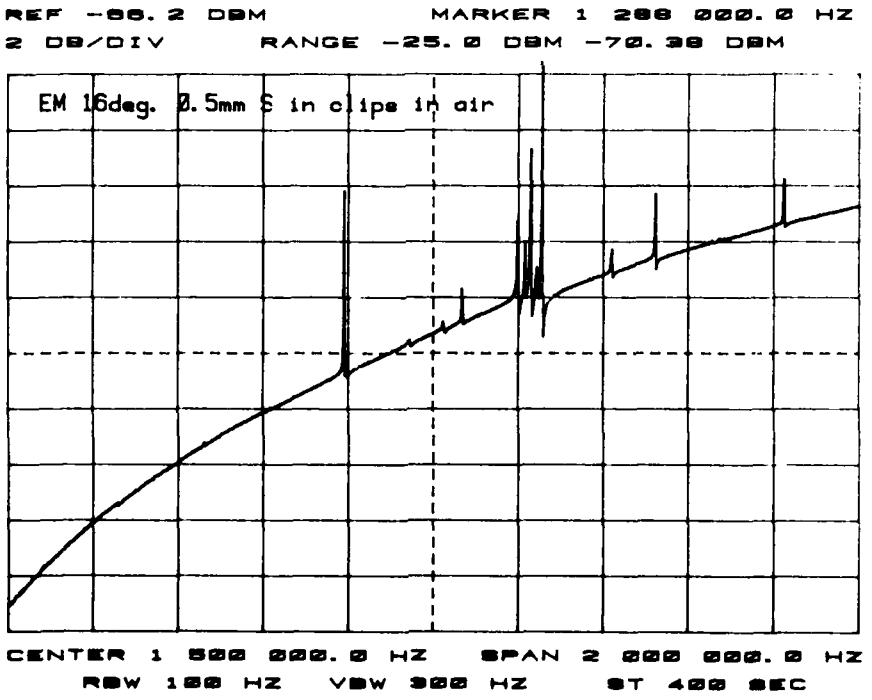


Fig 11 Mode spectrum of a lightly damped resonator

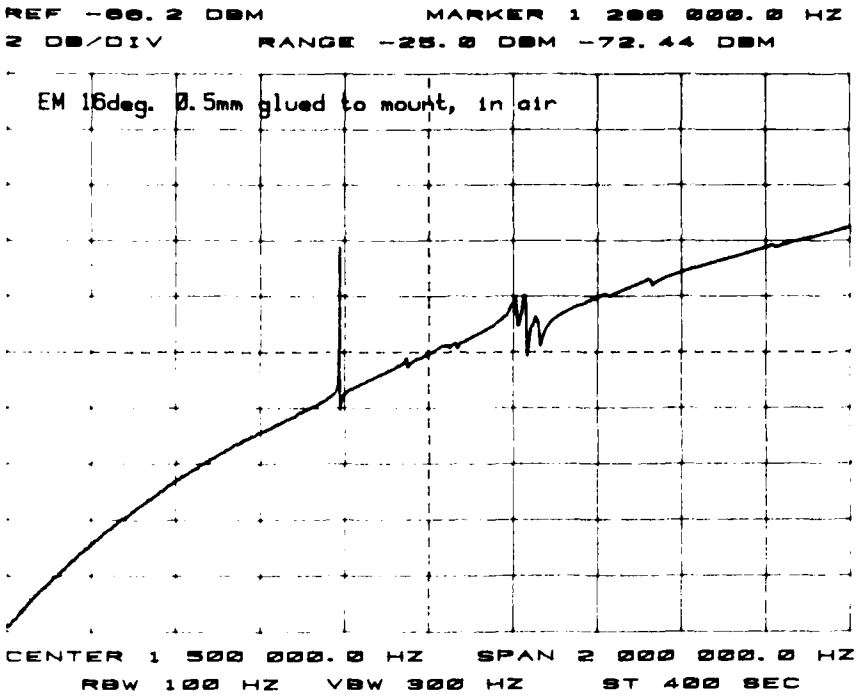


Fig 12 Mode spectrum of a resonator glued to a base

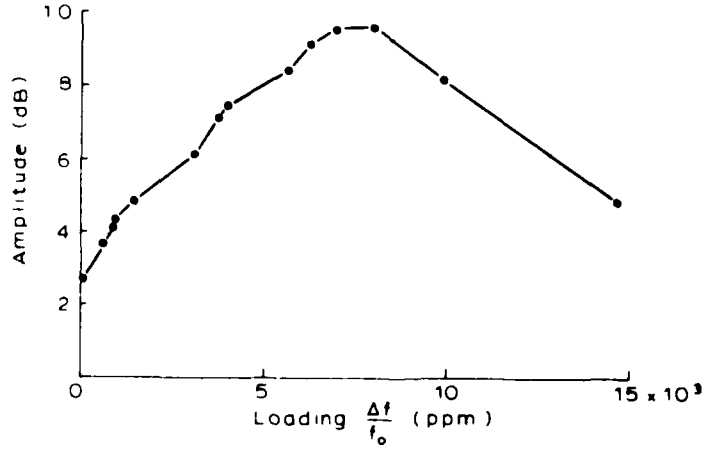


Fig 13 Balancing an edge mode resonator (with polystyrene cement)

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

4.19 MHz Cylindrical AT-Cut Miniature Resonator

S. Okano, T. Kudama, K. Yamazaki and H. Kotake

Toyo Communication Equipment Co., Ltd.
Kanagawa, Japan

Summary

This paper describes a 4.19 MHz AT-cut crystal resonator for wristwatches. Generally, if an AT-cut crystal resonator is made small enough to meet the mounting conditions of a wristwatch, the coupling of unwanted vibrations with the thickness shear vibrations and the affects of the mounting system cannot be ignored. We succeeded in producing at AT-cut crystal resonator for wristwatches featuring superior frequency-temperature characteristics, low equivalent series resistance, and outside dimensions of 2(T) x 3(W) x 8(L) mm by:

(1) Miniaturization through the use of a rectangular resonator elongated in the X-direction.
(2) Increasing the energy trapping effect and eliminating the affect of the X-direction unwanted vibrations by the adoption of a cylindrical shape in the X-direction.
(3) Plotting and studying in detail, a Z' direction unwanted vibrations mode chart. As a result, we decided that the ideal dimensions are R/t=40-60, l/t=12-17, and w/t=3.7-4.13. The actual dimensions at 4.19 MHz are l=6.0 mm, w=1.6 mm, t=0.4 mm, and R=20 mm.
(4) weakening the coupling between the Z' axis direction unwanted vibrations and thickness shear vibration by tilting the edge planes approximately 5 .
(5) Attaching electrodes to the edge plane and effectively using the charge thus generated.

The main features of this resonator are:

1. Average equivalent series resistance of 50 Ohms and mean Q of 200 x 10^3.
2. Frequency-temperature characteristic of approximately +1.5 x 10^{-6} from 10 C to 35 C.
3. Frequency aging characteristic of +0.5 x 10^{-6} / year.

Key words (for information retrieval)
Crystal, Miniature crystal resonator, Edge plane tilting, Rectangular AT-cut, Thickness shear, Unwanted vibration.

Introduction

Quartz wristwatches having a much higher precision than conventional mechanical wristwatches have become a reality, and greater precision is expected. The most widely used crystal resonator for wristwatches is the 32.768 kHz (2^{15} Hz, +5 X-cut tuning fork type, but its frequency-temperature characteristic is a parabolic curve and high precision over a wide temperature range cannot be expected. It is well known that this defect can be substantially overcome by using an AT-cut crystal resonator that shows a third-order curve frequency-temperature characteristic. The AT-cut resonator has an excellent frequency aging characteristic and comparatively large ones have been widely employed as the easiest resonator to manufacture for communications equipment, stable signal sources, and consumer products.

In 1970, Switzerland announced the world's first so-called mega-quartz wristwatch using an AT-cut crystal resonator. Shortly thereafter, in 1975, Japan also announced the same kind of watch. These wristwatches have a guaranteed precision of several seconds / year and utilize the superior characteristics of the AT-cut crystal resonator. However, low-priced, popular models are still not available because of:

1) The higher power consumption of LSI caused by use at a high frequency.
2) Miniaturization of the AT-cut crystal element is difficult.

With the amazing progress made in IC technology in recent years, the appearance of a practical IC equal to the present tuning fork crystal resonator from the standpoints of power consumption and price cannot be far off. The power consumption of present 4.2 MHz CMOS LSI is 3-5 μA at 1.5 Volt.

In the case of the latter, the AT-cut crystal resonator has been miniaturized, but, generally, if the outline dimensions of an AT-cut crystal resonator are made small enough to satisfy the mounting conditions of a wristwatch, coupling between the thickness shear vibration and the unwanted contour vibrations is difficult to avoid. If coupling with the unwanted vibrations occurs, the frequency-temperature characteristic worsens, the equivalent series resistance increases, other undesirable phenomena occur and the superior characteristics of the AT-cut crystal resonator are obstructed.

The results of experiments on these problems have been reported by J. J. Royer[1]. Besides, if the crystal is made smaller, the affect of the mounting system on the resonating system cannot be ignored and a stable frequency is not obtained.

A large amount of research on solving these problems has also been reported. One is the MRI plate proposed by Fukuyo[2] who has produced a resonator having an extremely high Q. Another is the proposal by Onoe[3] and others who have improved the frequency-temperature characteristic by introducing the edge plane tilt principle into a beveled

rectangular resonator elongated in the X-direction. A. E. Zumsteg and P. Suda[4] have proposed a resonator with a simplified mounting system using a rectangular resonator elongated in the Z'-direction. S. Yamashita[5] has also studied the unwanted vibrations of a beveled rectangular resonator elongated in the X-direction in detail.

However, the optimum value has been pursued with these resonators to avoid the many unwanted vibrations that are propagated in the X and Z' directions and efficient mass production is difficult.

We have increased the energy trapping effect and substantially weakened the coupling with the X-direction unwanted vibrations by using a convex shaped rectangular resonator elongated in the X-direction. We also tilted the edge plane on the tail approximately 5° and pursued the pure thickness shear mode by weakening the coupling with the Z' direction unwanted vibrations.

As a result, a miniature AT-cut resonator having outside dimensions of 2(T) x 3(W) x 8(L) mm and the superior characteristics of AT-cut crystal was produced. The developement process and experimental results are discussed below.

Design objectives

The performances demanded of AT-cut resonators by wristwatch manufacturers are extremely severe, a frequency-temperature characteristic of $\pm 1 \times 10^{-6}$ or less at 10°C to 35°C and a frequency aging characteristics of $\pm 3 \times 10^{-7}$ / years or less being typical. Since satisfying these performances was, of course, deemed to be extremely difficult in realizing a wristwatch having a precision of several seconds per year, we proceeded with developement on the following design objectives:

1. F=4.194304 MHz (2^{22} Hz)
2. 100 Ohms or less equivalent series resistance at -20°C to +70°C.
3. Frequency-temperature characteristic of $+2 \times 10^{-6}$ or less at 10°C to 35°C and $+5 \times 10^{-6}$ or less at 0°C to 50°C.
4. Smallest possible outside dimensions.
5. Superior shock-resistance and aging characteristics.
6. High reproducibility and production efficiency design.

The 4.19 MHz of item 1. is considered to be the optimum frequency from the standpoints of LSI power consumption and resonator miniaturization.

Deciding the resonator shape

1. X-dimension (l) selection and adoption of cylindrical shape

From the standpoints of improved equivalent series resistance and frequency-temperature characteristic, a rectangular resonator elongated in the X-direction was adopted. Figure 1 shows its axis directions. To trap the thickness shear mode energy and thus to eliminate the affects of mounting, the bevel shape generally used was adopted in the X-direction. Figure 2 (a) shows the beveled blank. We studied the width and angle of this bevel shape in detail and produced many test units stably produced, and high bevel shape machining precision is required to control the frequency of the unwanted vibrations.

Numerous papers have been announced on research on the bevel shape. However, the exact effect of beveling is not clear and weakening the coupling with the flexure vibrations that are propagated in the X-direction of a small rectangular resonator is extremely difficult.

Generally, the unwanted vibrations propagated in the X-direction of an AT-cut resonator are the high order flexure vibration given by the equation[6]

$$f(\text{flexure}) = \frac{p}{2l}\sqrt{23\,\frac{\tanh(t/l\,P)}{\rho}} \qquad (1)$$

Where, p:degree, normal odd number, l:X dimension length, t:thickness, ρ:crystal density.
The l dimension is designed to adequately separate the flexure vibrations of Eq. (1) from the thickness shear vibrations and to make their coupling with the thickness shear vibration small, and to prevent degradation of the equivalent resistance by mounting. Actually, a fairly large number of X-direction unwanted vibrations having a various vibration modes are also observed.

Therefore we abandoned the bevel shape and studied the convex shape often used with comparatively low frequency miniature AT-cut circular resonators. Figure 2(b) shows the cylindrically shaped resonator. This shape is a cylinder that follows a circle only in the X-direction. Actually, adoption of this shape adequately reduces the X-direction energy leakage and unwanted vibrations coupling so that the l dimension could be virtually ignored. Since precision lapping is possible, the w dimension can be simply selected to avoid unwanted vibrations.

Figure 3(a) and Figure 3(b) are the typical frequency response of a beveled resonator and cylindrical resonator. These measurements are for l/t=15. As can be seen in the figures, the cylindrical shape reduces both the number and magnitude of the unwanted vibrations.

The cylindrical radius R and length l dimensions were selected at R/t=40-60 and l/t=12-17 from the results of experiments taking into account the degree of energy trapping and the degree of equivalent resistance degradation caused by mounting.

Since a large amount of energy cannot be trapped completely at about R/t=50, the equivalent resistance deteriorates somewhat when the resonator is mounted. To prevent this, the rear of the cylindrical surface was beveled about c/l=0.17.

2. Z' dimension (w) selection

Next we studied the Z' direction unwanted vibrations in detail to select the w dimension. Figure 4 is a mode chart showing the frequency response of the R/t=50, l/t=15, c/l=0.17, w/t=3.0-4.4 test pieces. This figure is for comparatively large responses.

Because the resonator is cylindrical, it is an extremely simple type in which X-direction unwanted vibrations are not observed. Straight line t of Ft=1/30 is the thickness shear vibration. Strong unwanted vibrations are observed at w/t=3.5 and w/t=4.2. These are given by equation below and are considered to be the 2nd order longitudinal vibrations and 3rd order face shear vibrations.[6]

$$f_{(\text{longitudinal})} = \frac{n}{2w}\sqrt{\frac{C'_{33}}{\rho}} \qquad (2)$$

$$f_{(\text{face-shear})} = \frac{n}{2w}\sqrt{\frac{C'_{55}}{\rho}} \qquad (3)$$

where, n:degree, w:width dimension, C'_{33}, C'_{55}: stiffness constant.

Actually, the w dimension should be selected with this figure to avoid the unwanted vibrations. However, since the frequency-temperature characteristic of the unwanted vibrations has a large temperature coefficient of about -1×10^{-5} / °C to -2×10^{-4} / °C, to ensure decoupling with the thickness shear vibration from -20°C to +70°C at 4.19 MHz, the thickness shear vibration and unwanted vibrations must be at least ±20 kHz apart.

Figure 5 was used to measure the frequency response. Moreover, to increase the unwanted vibration response sensitivity, a measurement method in which the crystal is touched with a 50 μ fine gold wire probe was employed. The detection sensitivity is approximately 80 dB.

5. Tilting of edge planes

Weakening the coupling between the thickness shear vibration and Z'-axis face shear vibrations by tilting the edge planes approximately 5° was proposed by R. D. Mindlin[7], Onoe[3], and Mochizuki[8]. According to this theory, since the stiffness constant C'_{56} is not zero, the edge plane traction-free boundary conditions are not satisfied by the thickness shear vibration alone.

Eq. (4) gives the edge plane tilt angle that satisfied this condition.

$$\varepsilon = \arctan\left(C'_{56}/C'_{66}\right) \qquad (4)$$

where ε:edge plane tilt angle, C'_{56}, C'_{66}: stiffness constant. ε is approximately 5° for an AT-cut resonator. Figure 6 shows this edge plane tilt.

Figures 7 and 8 are comparisons of the frequency-temperature characteristic when the edge planes are tilted. ε=0° and ε=5° and the w/t is varied. Figure 9 is plotted so that this relationship can be more easily understood and shows the temperature coefficient at 25 C. The solid line is for ε=5° and the broken line is for ε=0°. The crystal cut angle is +35°06'15"±15".
In other words, tilting the edge plane 5° effectively weakens the coupling between the thickness shear vibration and Z'-axis direction unwanted vibrations.

The realization of pure thickness shear vibration was also confirmed from the temperature coefficient of Figure 9. Coupling with w/t=3.5 longitudinal vibrations is also weak. The reason for this was not theoretically analyzed but is assumed to be due to mode conversion between the longitudinal mode and face shear mode.

From the results of these experiments, w/t=3.7-4.13 was selected and the final value of the w dimension was decided by studying the equivalent resistance temperature characteristic.

The dimensions of a 4.19 MHz resonator blank selected by the above technique are l=6.0 mm, w=1.6 mm, t=0.4 mm, R=20 mm, and c=1.0 mm. Figure 10 is a photograph of this blank.

Electrode shape design

Generally, electrodes are attached only to the top and bottom of the X-Z' plane of an AT-cut crystal resonator. However, with a rectangular resonator of about Z/t=4, the charge induced in the X-Y' plane (edge plane) shown in Figure 11 can not be ignored. To effectively use this charge, the equivalent resistance was made substantially lower than that of a resonator without edge plane electrodes by making the electrodes the shape shown in Figure 12. The electrodes are plated with gold to a thickness of approximately 1500 Angstrom.

Outline of production process

Figure 13 outlines the production process. This process is basically the same as that of the conventional AT-cut resonator, except that batch machining is possible. Actually, the cutting, lapping, and cylindrical machining processes are more efficient than those of conventional circular resonators. The usage efficiency of synthetic quartz is also about 15 times better than that of 8 mm diameter circular resonators.

Mounting method and holder

The shock and vibration resistance is improved by mounting the resonator with conductive adhesive at two points at the end of the X-direction. To prevent the thermal expansion difference of the holder and resonator blank from adversely affecting the frequency-temperature characteristics, the mounting section is made flexible. Figure 14 shows the mounting structure.

The holder is a cold welded vacuum sealed type to improve the equivalent series resistance and guarantee the frequency aging characteristic. The outline dimensions are shown in Figure 19. The presented type has a volume 1/18 that of the HC-43/U.

Resonator characteristics

The results of measurement of the various characteristics of 200 resonators designed and manufactured as previously described are given. They showed very good characteristics as initial developement products.

1. Equivalent circuit parameters

Table 1 shows the typical data of the resonator equivalent circuit parameters measured by the IEC-PUB-444 method.

Figures 15(a) thru 15(e) are the histograms.

F 4.194304(MHz)	
R1 53.9(Ohm)	Figure 15(a)
Q 223 x 10^3	Figure 15(b)
L1 91.0(mH)	Figure 15(c)
C0 1.35(pF)	Figure 15(d)
C1 .041	Figure 15(e)

Table 1

The distribution of the equivalent series resistance R1 is somewhat poor, but is considered to be capable of improvement at the next stage of development.

2. Frequency-temperature characteristic and equivalent resistance-temperature characteristic

Figure 16 shows the frequency and equivalent resistance-temperature characteristics. A frequency change of F of $\pm 2.0 \times 10^{-6}$ is adequately satisfied between 15 C and 35 C. The frequency-temperature characteristics are a 3rd order curve, and the equivalent resistance is almost flat over the -10 C to +60 C temperature range, indicating ideal thickness shear vibration. The point of inflection temperature is approximately 31 C to 32 C.

3. Frequency aging characteristics

Figure 17 shows the frequency aging characteristic when the resonator was stored unloaded at 25 C. Results of $+0.5 \times 10^{-6}$ / 3500 H, or about one year were obtained.

The equivalent series resistance was simultaneously measured, but showed virtually no change.

Results of +5 seconds in about one year were obtained from tests performed by a watch manufacturer with the resonator mounted in an actual watch and its superior frequency-aging characteristic even in continuous loaded operation was confirmed.

4. Shock resistance and other environmental performances

These are the most stringent of all the performances demanded of a resonator for wristwatches. Figure 18 shows the change in the frequency and equivalent resistance when the resonator was dropped onto a hard wooden board from a height of 75 cm. The average frequency change is about -0.3×10^{-6} and shows that the resonator is adequately shock resistant. Adequate results were also obtained in environment testing, vibration testing, thermal shock, heat cycle, and other tests.

Conclusion

It was confirmed that the effects of unwanted vibrations can be easily avoided even with a small rectangular resonator by using a cylindrical shape in the X-direction and tilting the edge planes about 5°.

As a result, a miniature AT-cut crystal resonator with characteristics comparable to those of a large resonator was realized.

We intend to continue to direct our efforts toward greater miniaturization and improvement of the equivalent resistance and frequency-temperature characteristics of the resonator in the future.

Acknowledgement

The authors wish to thank Prof. Fukuyo of Chiba University, Prof. Onoe of Tokyo University, and Assistant Prof. Mochizuki of Shizuoka University for their technical advice.

References

1) J. J. Boyer, "Rectangular AT-Cut Resonators", Proc. 27th AFCS, 1973.
2) H. Fukuyo, "Mass-Productive Miniature AT-Cut Crystal Resonator for Watch", Proc. CIC, 1974.
3) M. Onoe and M. Okazaki, "Miniature AT-Cut Strip Resonators with Tilted Edges", Proc. 29th AFCS, 1975.
4) A. E. Zumsteg and P. Suda, "Properties of 4 MHz Miniature Flat Rectangular Quartz Resonator Vibrating in a Coupled Mode", Proc. 30th AFCS, 1976.
5) S. Yamashita et al, "A 4.19 MHz Beveled Miniature Rectangular AT-Cut Quartz Resonator", Proc. 32nd AFCS, 1978.
6) H. Fukuyo, "Researches in Modes of Vibrations of Quartz Crystal Resonators by Means of The Probe Method", Bull. of Tokyo Inst. Tech., Ser. A, No.1, 1955 (in Japanese).
7) R. D. Mindlin, "Thickness-Twist Vibrations of a Quartz Strip", Proc. 24th AFCS, 1970.
8) Y. Mochizuki, "Simple Exact Solutions for Thickness Shear Mode of Vibration of a Crystal Strip", Proc. 29th AFCS, 1975.
9) M. Onoe, "Analysis and Detection of Unwanted Modes in Quartz Crystal Units for Oscillators", J. Inst. Commun. Eng. of Japan, Vol. 47, No.1, Jan. 1964 (in Japanese).

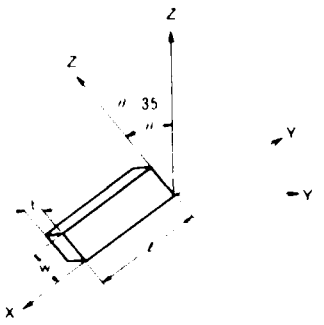


Fig.1 Rectangular AT-Cut Quartz Plate.

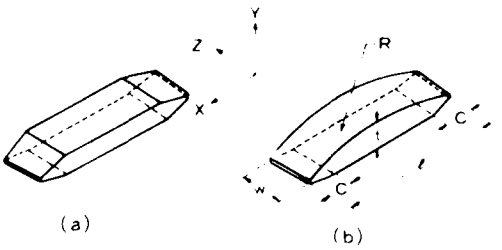


Fig.2 (a) Beveled Blank.
(b) Cylindrical Blank.

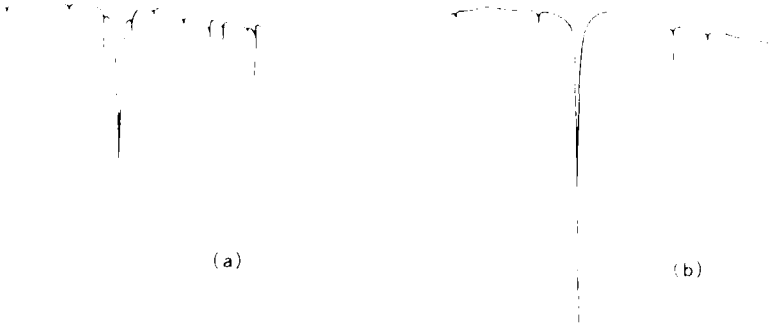


Fig.3 Typical Frequency Response of Beveled Resonator (a) and Cylindrical Resonator (b).

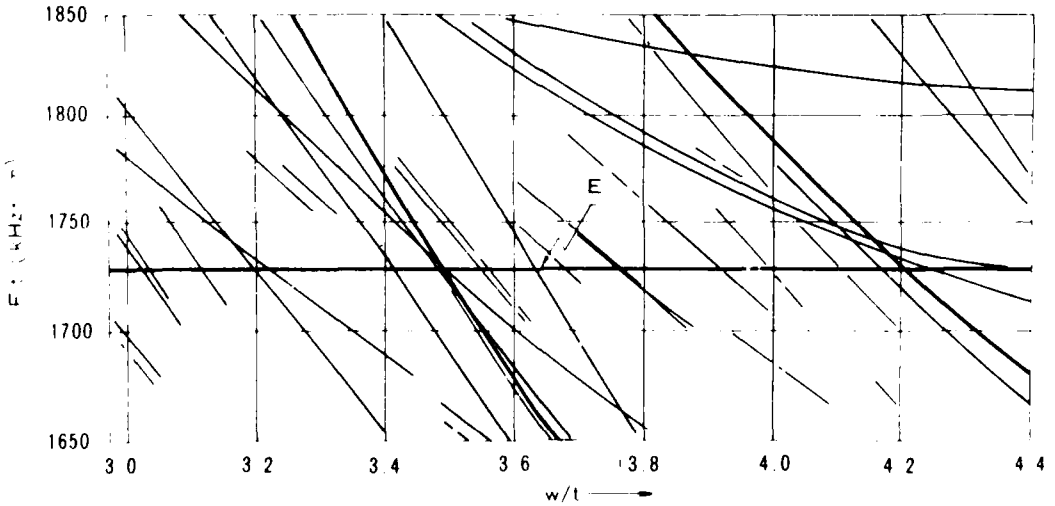


Fig.4 Mode Chart.

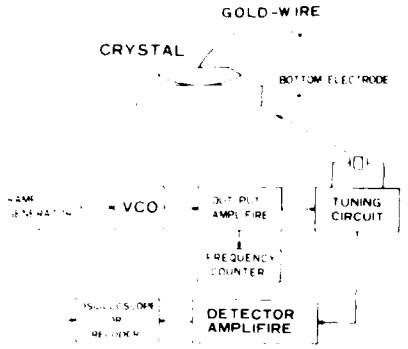


Fig.5 Block Diagram of Frequency Response Measurement.

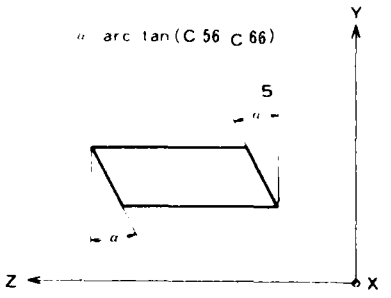


Fig.6 Edge Plane Tilt.

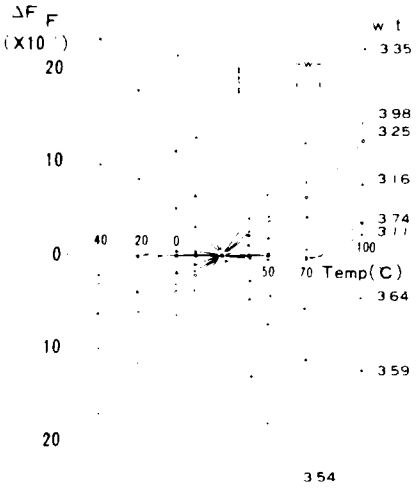


Fig.7 F-T Characteristics (Edge Plane Tilt Angle α=0 and w t is a Parameter).

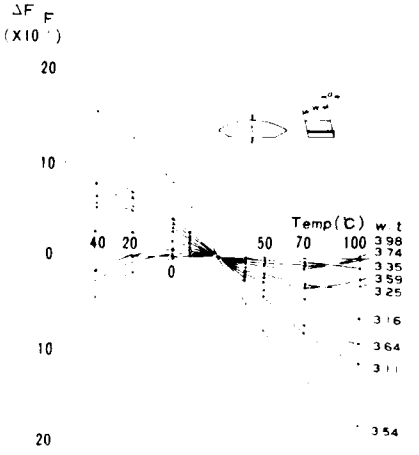


Fig.8 F-T Characteristics (Edge Plane Tilt Angle α 5 and w t is a Parameter).

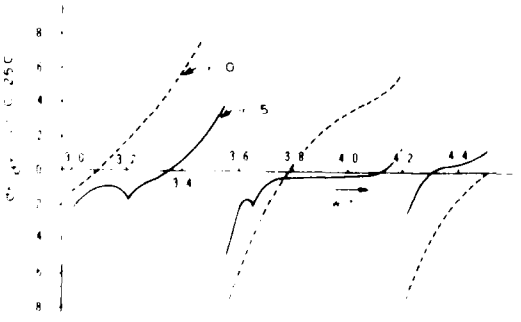


Fig 9 Temperature Coefficient at 25C for α 5 (Solid Line) and α=0 (Broken Line)

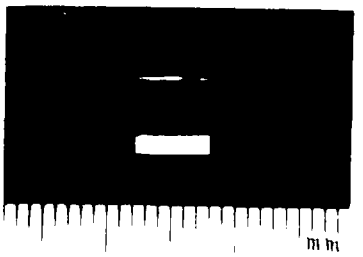


Fig.10 Photograph of Resonator Blank.

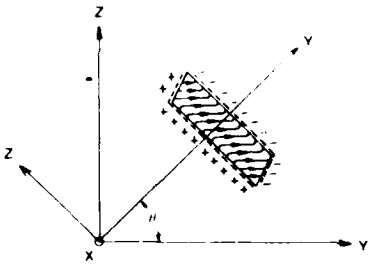


Fig.11 Induced Charge at Edge Planes.

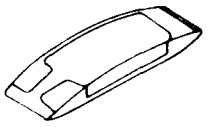


Fig.12 Electrode Shape.

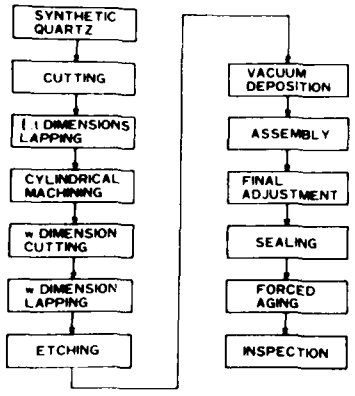


Fig.13 Production Process.

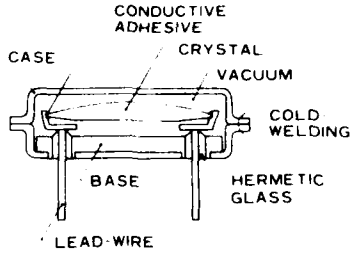


Fig.14 Mounting Structure and Its Holder.

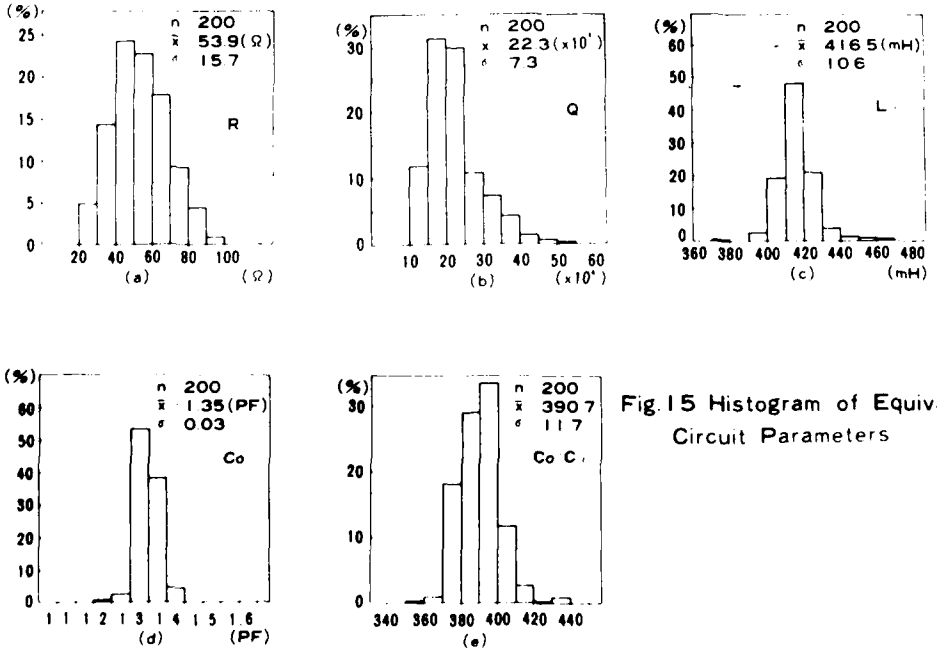


Fig.15 Histogram of Equivalent Circuit Parameters

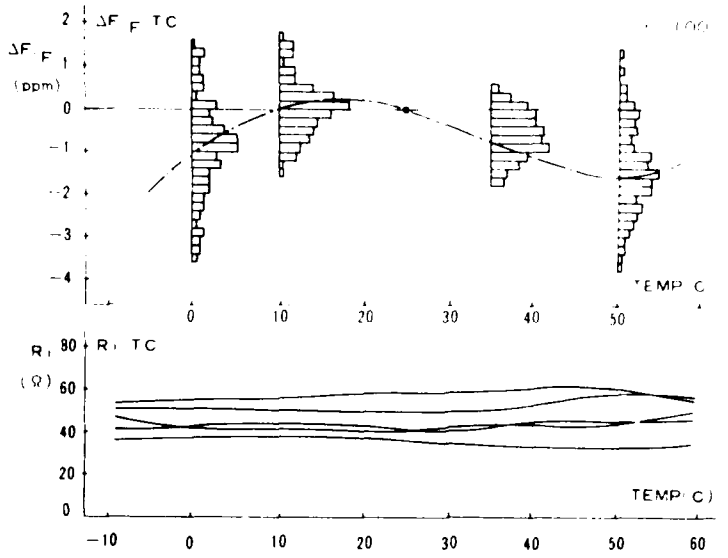


Fig.16 Frequency and Equivalent Series Resistance versus Temperature Characteristics.

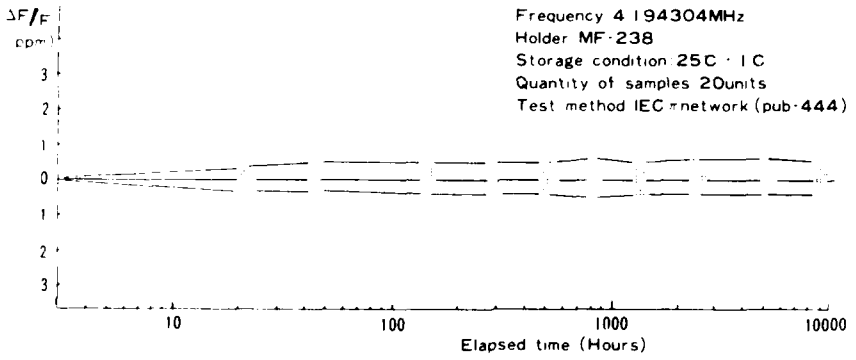


Fig.17 Frequency Aging Characteristics.

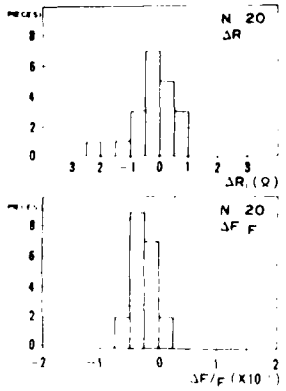


Fig.18 Shock Resistance.

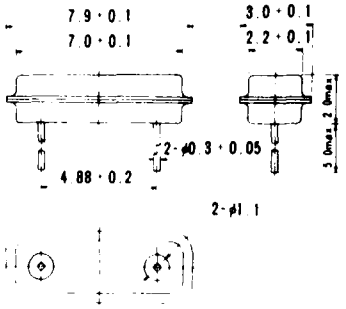


Fig.19 Outline Dimensions.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

THREE-DIMENSIONAL MODE-MATCHING THEORY OF RECTANGULAR BAR RESONATORS USING COMPLEX WAVE-NUMBERS

* R.F. Milsom
** D.T. Elliott
** M. Redwood

* Philips Research Laboratories, Redhill, Surrey, England.
** Department of Electrical and Electronic Engineering, Queen Mary College, University of London, England.

Summary

The miniature rectangular bar resonator introduced by Zumsteg in 1976 was shown to have excellent electrical performance as well as offering the advantages of small size and simple shape. Earlier attempts to correctly model all the parameters of this resonator have been unsuccessful, either because over-simplification has led to considerable inaccuracy or because over-complication has led to excessive demands on computing resources. Here we describe a new three-dimensional model which is comprehensive, but nevertheless straightforward in concept. The solution is expressed as the linear sum of partial modes, each having a wave-number whose components in all three coordinate directions are in general complex. Mode-matching theory is used to find the relative amplitudes of the partial modes. Newton's equations of motion and the quasi-static approximation to Maxwell's equations are satisfied exactly, taking full account of piezoelectric coupling. The boundary conditions are satisfied exactly on the major faces and approximately on the minor faces of the bar. The mass-loading and electrical shorting effects of the electrodes are both included.

The fundamental mode of the bar and its inharmonic overtones are shown to be coupled modes which are linear combinations of the TS1, TS3 and F1 modes of a plate. The dispersion of these coupled modes is found to vary strongly with changes in the cross-sectional aspect ratio of the bar. This implies that energy-trapping also varies with this ratio. In some instances the spatial attenuation of amplitude towards the ends of the bar is in the form of a damped oscillation rather than the more usual simple exponential decay. Solutions are presented for frequency, motional capacitance, quality-factor and frequency-temperature behaviour as functions of aspect ratio, angle of orientation, bar length, electrode length and mass-loading. Solutions are also presented for the frequencies and quality factors of the inharmonic overtone modes. Excellent agreement is shown with published experimental data.

Introduction

Several novel designs of miniature quartz crystal resonator have been reported in recent years. One of the most interesting of these is the design proposed by Zumsteg in 1976[1]. This combines the advantages of a simple rectangular shape, considerably smaller overall size than a conventional AT-cut disc resonator and exceptional electrical performance. Unfortunately this performance has not been fully understood. Early analysis employing two-dimensional plate equations gave good agreement for frequency but only moderate agreement for frequency-temperature performance and very poor agreement for energy-trapping[2]. Subsequently we developed a three-dimensional variational model which gave better agreement for energy-trapping, but which involved a very large amount of computation to obtain even a single solution[3]. This model was therefore not a useful design tool. Much the same problem was found with the finite element program ASKA[4] when applied to this type of resonator. Here we describe a new three-dimensional model which uses mode-matching theory giving the solution in terms of a linear sum of partial modes with wave-numbers which in general are complex for all three rectangular co-ordinate directions.

The Zumsteg resonator is shown in Figure 1. It consists of a rectangular bar taken from a rotated Y-cut quartz plate with the width dimension parallel to the X-axis of the crystal. Angles of rotation θ of greatest interest are close to 35°, which is the AT-cut. The electrodes are rectangular and occupy the full width of the central region of each face. The principal mode of vibration contains contributions from both flexural and thickness-shear plate modes which are coupled together at the faces normal to X. This will be discussed more fully later. The mode is sufficiently complicated to suggest that a fully three-dimensional analysis is required to adequately describe its behaviour. The two-dimensional plate equations are based on the assumption that one dimension of the resonator is significantly smaller than the other two and that is not the case in the devices considered here. It can therefore be expected that a two-dimensional analysis based on these equations[2], though efficient, might lead to some inaccuracy. The solution described here satisfies the full set of Newton's equations and the quasi-static approximation to Maxwell's equations, taking account of the piezoelectric coupling between these two. This self-consistent approach to the coupling between electrical and mechanical fields is not usually necessary in the case of quartz which is piezoelectrically weak. Nevertheless more strongly

piezoelectric materials may be required for some miniature resonator applications in order to achieve adequate motional capacitance. It is mainly with this possibility in mind that piezoelectricity is treated as a first-order effect. The boundary conditions are satisfied exactly on the major faces (i.e. those normal to Y') within the limitations of the mass-loading and zero normal flux approximations described later. The boundary conditions on the minor faces (i.e. those normal to X and Z') and at the interfaces between the plated and unplated regions of the bar are satisfied approximately. The analysis treats the plated and unplated regions as sections of acoustic waveguide, and the central problem is that of finding the dispersion characteristics of the guided modes. Firstly, however, it is necessary to consider the modes of a piezoelectric continuum, since a linear combination of these describes each guided mode.

Differential Equations and Modes of a Continuum

The differential equations governing the behaviour of a piezoelectric continuum are Newton's laws of motion and the quasi-static approximation to Maxwell's equation. (This latter approximation is valid because acoustic waves are typically five orders of magnitude slower than electromagnetic waves). Thus:

$$T_{ij,j} - \rho \ddot{u}_i = 0 \qquad \text{(Newton's Law)} \tag{1}$$

and

$$D_{i,i} = 0 \qquad \text{(Gauss' Law)} \tag{2}$$

where T, ρ, u and D are respectively stress, density, particle displacement and electric flux density. Normal tensor notation is used so that comma denotes differentiation with respect to spatial coordinates and dot denotes differentiation with respect to time. These equations are coupled by the piezoelectric equations of state,

$$T_{ij} = c^E_{ijkl} S_{kl} - e_{kij} E_k \tag{3}$$

$$D_i = e_{ikl} S_{kl} + \varepsilon^S_{ik} E_k \tag{4}$$

where S, E, c^E, e, ε^S are respectively strain, electric field, and the stiffness, piezoelectric and permittivity tensors of the crystal, referred to the rotated coordinate axes, x_1, x_2, x_3, shown in Figure 1. In addition,

$$S_{ij} = \tfrac{1}{2}(u_{i,j} + u_{j,i}) \tag{5}$$

$$E_i = -\phi_{,i} \tag{6}$$

where ϕ is electric potential. Substituting equations (5) and (6) into (3) and (4) and hence into (1) and (2),

$$\begin{aligned} -\rho \ddot{u}_j + c^E_{ijkl} u_{k,li} + e_{kij} \phi_{,ki} &= 0 \\ e_{ikl} u_{k,li} - \varepsilon^S_{ik} \phi_{,ki} &= 0 \end{aligned} \tag{7}$$

Now consider a solution to equation (7) of the form:

$$\begin{aligned} u_j &= A_j \exp\left[-jk_0(\xi x_1 + \alpha x_2 + \zeta x_3) + j\omega t\right] \quad j=(1,2,3) \\ \phi &= A_4 \exp\left[-jk_0(\xi x_1 + \alpha x_2 + \zeta x_3) + j\omega t\right] \end{aligned} \tag{8}$$

where k_0 is a real constant, ξ, α and ζ are complex normalised wave-numbers, ω is angular frequency, and t is time. Substituting this trial solution into equation (7) gives

$$H_{ij} A_j = 0 \qquad (i=1,2,3,4) \tag{9}$$

where the elements of the matrix H are functions of the material constants, the trial wave-number components and frequency. Non-trivial solutions are given by those combinations of wave-number and frequency for which

$$|H| = 0 \tag{10}$$

Each such solution represents a mode of the continuum. Resubstituting these values of frequency and wave-number into any three of equation (9) and solving the linear equations yields the ratio $A_1:A_2:A_3:A_4$ appropriate to the mode.

Boundary Conditions and Modes of the Acoustic Waveguide

Consider the rectangular acoustic waveguides shown in Figure 2. In each case the bar is of infinite length, width ℓ_1 and thickness ℓ_2. Guided modes which either propagate along the bar or alternatively are evanescent along the bar (i.e. cut-off modes) must satisfy the electrical and mechanical boundary conditions on the faces normal to both x_1 and x_2. The two cases shown must be studied. In the first case the faces normal to x_2 are electrically and mechanically free, and in the second they are each loaded with a thin conducting film density ρ_m and thickness h. For short-circuit modes the two conducting films are externally shorted together. In both cases the surfaces normal to x_1 are free. The boundary conditions for the plated and unplated guides are then,

Unplated guide:

$$T_{1j} = 0 \quad \text{at} \quad x_1 = \pm \ell_1/2 \quad (i=1,2,3) \tag{11}$$

$$D_1 = 0 \quad \text{at} \quad x_1 = \pm \ell_1/2 \tag{12}$$

$$T_{2i} = 0 \quad \text{at} \quad x_2 = \pm \ell_2/2 \tag{13}$$

$$D_2 = 0 \quad \text{at} \quad x_2 = \pm \ell_2/2 \tag{14}$$

Plated guide:

As for the unplated guide, equations (11) and (12)

$$T_{2i} = \mp \rho_m h \ddot{u}_i \quad \text{at} \quad x_2 = \pm \ell_2/2 \tag{15}$$

$$\phi = 0 \quad \text{at} \quad x_2 = \pm \ell_2/2 \tag{16}$$

Two approximations are made here. In equations (12) and (14) the normal component of flux should strictly be continuous across the boundary rather than zero. However for guides with either high dielectric constant (as with lithium niobate) or weak piezoelectric coupling (as with quartz) the approximateion is good. Secondly, since $h \ll \ell_2$, it can be assumed that the mechanical effect of the metal film is to simply mass-load the surface yielding equation (15). These boundary conditions, along with the ditterential equations (7) and the trial solutions (8) are now used to find the guided modes.

It will be appreciated that fixing any three of frequency , and the normalised wave-numbers, ζ, and ξ restricts possible values of the fourth to solutions of equation (10). If ω, ξ and ζ are fixed then equation (10) is an eighth order polynomial in α, with complex coefficients. The zeros of this polynomial α_n yield eight partial modes of the form:

$$u_i^{(n)} = B_n a_{in} \exp\left[-jk_o(\zeta x_1 + \alpha_n x_2 + \xi x_3) + j\omega t\right]$$
$$(i=1,2,3,4, n=1.....8) \qquad (17)$$

where B_n is an unknown amplitude constant, $a_{in} = A_i$, $A_4 = 1$, and α is rewritten as α_n for convenience. k_o is arbitrarily defined by $k_o = \pi/\ell_2$ so that all values of ζ, α and ξ will be of unit order of magnitude.

In order to satisfy the boundary conditions at $x_2 = \pm \ell_2/2$ a linear combination of these eight modes is required.

Thus,

$$u_i = \sum_{n=1}^{8} B_n a_{in} \exp\left[-jk_o \alpha_n x_2\right] \exp\left[-jk_o(\zeta x_1 + \xi x_3) + j\omega t\right] \qquad (18)$$

Substituting this into equations (13) and (14) for the unplated bar or equations (15) and (16) for the plated bar yields linear equations of the form:

$$\sum_{n=1}^{8} K_{jn} B_n \qquad (j=1.....8) \qquad (19)$$

where the elements of the matrix K are functions of ζ, ξ and ω as well as constants of the material. Possible non-trivial solutions are given by those combinations of ζ, ξ and ω for which

$$|K| = 0 \qquad (20)$$

Resubstituting these wave-numbers and frequency into any seven of equation (19) and solving yields the ratios of the B_n. Each such solution is a Lamb mode of an infinite plate of thickness ℓ_2 and may be written,

$$u_i^{(m)} = C_m \left(\sum_{n=1}^{8} b_{mn} a_{imn} \exp\left[-jk_o \alpha_{mn} x_2\right] \right) \exp\left[-jk_o(\zeta_m x_1 + \xi x_3) + j\omega t\right] \qquad (21)$$

where ζ_m is one solution of equation (20) for a given combination of ξ and ω. C_m is an unknown constant, $b_{mn} = B_n$, $B_8 = 1$, and the number of subscripts for a and α has been increased to indicate both the plate mode number m and the continuum mode number n. An infinite number of plate modes exist for each choice of ξ and ω, and indeed an infinite number would be required to satisfy the boundary conditions of equations (11) and (12) exactly at every point on the faces normal to x_1. This is clearly impracticable. However, it is found that a small number of judiciously chosen modes gives an accurate approximation to the boundary conditions. Earlier analysis[1],[2] has shown that the predominant plate modes are TS1 (thickness-shear across the bar) and F1 (flexure across the bar). These modes are shown schematically in Figure 3. For certain width-to-thickness ratios there is also coupling to the face shear mode FS1 (not shown). However such ratios should be avoided in practice because of the close proximity of another bar mode which is predominantly face shear[1].

We now consider to what extent the stress boundary conditions of equation (11) can be satisfied by the TS1 and F1 modes, and what additional mode if any might be required to satisfy these conditions more accurately. The distribution of stress $T_{ij}^{(m)}$ across the thickness associated with each plate mode m may be found for any given choice of ω and ξ by substituting equation (13) into equations (5), (6) and (3). If $\xi = 0$, that is if the mode is uniform along the bar, there is a large odd component of T_{11} and a large even component of T_{12} associated with each mode (odd and even in the sense of the spatial symmetry about $x_2 = 0$). Other components of stress are very small in comparison. If $\xi \neq 0$, which is the situation that arises with energy-trapping, there is also a large odd component of T_{13} present in each mode. This suggests that for the energy-trapped bar mode there will be strong coupling to the TS3 mode (thickness-shear along the bar, see Figure 3) whose predominant stress components for $\xi \neq 0$ also include the odd components of T_{11} and T_{13} and even component of T_{12}. Since we cannot satisfy equation (11) exactly, we modify the boundary conditions such that only the fundamental Fourier component of appropriate symmetry is zero at $x_1 = \pm \ell_1/2$ for the three stresses T_{i1} $(i = 1,2,3)$. Thus:

$$\left.\begin{aligned} \int_{-\ell_2/2}^{\ell_2/2} T_{11} \sin\left(k_o x_2\right) dx_2 &= 0 \\ \int_{-\ell_2/2}^{\ell_2/2} T_{12} \cos\left(k_o x_2\right) dx_2 &= 0 \\ \int_{-\ell_2/2}^{\ell_2/2} T_{13} \sin\left(k_o x_2\right) dx_2 &= 0 \end{aligned}\right\} \text{ at } x_1 = \pm \ell_1/2 \qquad (22)$$

The TS1, F1 and TS3 modes in fact constitute six plate modes because if ζ_m is a solution of equation (20) then $-\zeta_m$ is also a solution (representing the same mode propagating or evanescent in the opposite x_1-direction). A linear combination of these six modes can be made to satisfy equations (22). The coupled mode solution then has the form,

$$u_i = \sum_{m=1}^{6} \sum_{n=1}^{8} C_m b_{mn} a_{imn} \exp\left[-jk_o\left(\zeta_m x_1 + \alpha_{mn} x_2 + \xi x_3\right) + j\omega t\right] \quad (23)$$

Substituting equation (23) into equation (22) yields linear equations of the form,

$$\sum_{m=1}^{6} G_{jm} C_m \qquad (j = 1.....6) \quad (24)$$

where the elements of the matrix G are functions of ξ and ω. Non-trivial solutions to equation (24) are given by combinations of ξ and ω for which,

$$|G| = 0 \quad (25)$$

These solutions give the dispersion of the guided modes of the bar. Resubstituting the values of ζ and α for such a mode into any five of equation (24) yields the ratios of the C_m for that mode. The solution for the mode pattern is then given by

$$u_i^{(p)} = D_p \sum_{m=1}^{6} \sum_{n=1}^{8} c_{pm} b_{pmn} a_{ipmn} \exp\left[-jk_o\left(\zeta_{pm} x_1 + \alpha_{pmn} x_2 + \xi_p x_3\right) + j\omega t\right] \quad (26)$$

where ξ_p is one solution of equation (25) for given ω, D_p is an unknown amplitude constant, $c_{pm} = C_p$, $C_1 = 1$, and the number of subscripts for b, a, α and ζ has been increased to indicate the bar mode number p, the plate mode number m, and the continuum mode number n.

The accuracy of the approximation to the boundary conditions on the width faces implied by equation (22) could be increased by taking additional modes to satisfy the conditions for the higher harmonic Fourier components. However it is reasonable to assume that at frequencies of interest such modes would be very strongly attenuated away from the width faces and therefore have negligible influence on the solution. Similarly we may assume that the electrical boundary condition of equation (12) can be ignored.

Guided Modes in AT-Z' Quartz Bars

We now apply the analysis of the proceeding sections to the specific case of a bar cut from an AT plate with the length dimension parallel to Z' (i.e. the x_3 coordinate axis). Initially we find the frequencies of modes which are uniform in x_3, that is the untrapped modes. These frequencies, which are essentially the cut-off frequencies of the guided modes, are found by putting ξ equal to zero, including only the TS1 and F1 plate modes and using only the first two of equation (22), the third being satisfied identically. In this special case G is reduced to a 4 x 4 matrix and the frequencies are given by the roots of equation (25) using this reduced form. Typical solutions for an unplated bar are shown in Figure 4 as functions of the width-to-thickness ratio of the bar L (= ℓ_1/ℓ_2). All frequencies Ω_o shown are normalised to the cut-off frequency of the TS1 mode of the infinite plate which is the lowest root of equation (20) with $\zeta = \xi = 0$. The quartz constants used are those of Ref. 5 and the angle of the plate $\theta = 34.8^o$. N depicts the number of half-cycles of the flexural mode F1 in the coupled mode. The solutions shown agree well with other methods of analysis[1,6]. Those curves marked with an asterisk indicate modes which are known to couple to face-shear modes[1]. For miniature resonators narrow bars are desirable, so the modes of greatest interest are the 4th and 8th harmonic flexural modes (N = 4 and N = 8). Here we shall concentrate on 4th harmonic bars, which have an aspect ratio L of approximately 3 for frequencies close to that of the infinite plate.

The dispersion characteristic of the Nth harmonic guided mode is found by solving equation (25) for ξ at normalised frequencies Ω above and below its cut-off frequency Ω_o. Typical dispersion curves for four different values of L are shown in Figure 5. The important point to note it that in all cases there are two branches, indicating the existance of two guided modes of the bar, both of which are a linear combination of the TS1, TS3 and F1 plate modes. We shall call these the Y and Z modes. The lengthwise component of wave-number of both modes is either real or imaginary above the normalised frequency Ω_c. Below Ω_c the two modes are complex conjugates, and the two curves shown in this frequency range are the real and imaginary components of the wave-number of one of these. The negatives of the wave-numbers shown are also solutions, so that there is a total of four possible guided modes at each frequency. The main cause of the strong dependence of the dispersion on aspect ratio is the presence of the TS3 mode in the solution. Frequencies of interest are substantially below the infinite plate cut-off frequency of this mode, which means that its vibrational energy decays quite sharply in the x_1-direction away from the width faces. However, the bar is so narrow that the TS3 mode can still have an influence on both faces simultaneously in a manner that depends strongly on the aspect ratio. We shall consider the significance of this later. The solutions shown in Figure 5 are for unplated bars. No significant change in these curves is found when we solve for the plated bar, other than the expected overall downward shift in frequency. This shift is found to depend linearly on mass-loading μ for practical film thickness, mass-loading being defined by,

$$\mu = \frac{\rho_m}{\rho} \frac{2h}{\ell_2} \quad (27)$$

The electrical shorting effect of the electrodes considered on its own produces a frequency lowering equivalent to a mass-loading of about 0.002. It follows that this electrical effect should not be ignored, since relatively large changes in wavenumber can occur for the corresponding small change in frequency.

Trapped Modes of the Resonator

The guided modes of an infinite bar are now used to solve for the trapped modes of the resonator using mode-matching theory, analogous to that for the trapped thickness-shear mode of a plate[7]. Firstly we define three regions, the E region, S^- region and S^+ region as shown in Figure 6. The solution will be described by a linear summation of the guided modes defined by equation (26). These solutions already satisfy the boundary conditions on faces normal to x_1 and x_2. It therefore remains to satisfy the boundary conditions on faces normal to x_3, including the interfaces between the three regions. In normal energy-trapping theory modes are "trapped" under the electrodes because the resonant frequency is above the cut-off frequency of the guided mode in the E region and below the cut-off frequency in the S regions. Here the situation is more complicated because additional modes are present. If the bar is rigidly mounted at both ends the boundary conditions are:

$$\left.\begin{aligned} u_i &= 0 && x_3 = \pm L/2 \\ u_i &\text{ is continuous at} && x_3 = \pm W_1/2 \\ \partial u_i/\partial x_1 &\text{ is continuous at} && x_3 = \pm W_1/2 \end{aligned}\right\} \begin{matrix} (i=1,2,3) \\ (j=1,2,3) \end{matrix} \qquad (28)$$

The electrical boundary conditions on faces normal to x_3 will be ignored. To satisfy equations (28) for all x_1 and x_2 at each face would require an infinite number of modes. However it is found that the modes already described give a sufficiently accurate solution. This simplification is possible firstly because the spatial distribution of the vibration patterns in the x_1-x_2 plane of the three plate modes are nearly constant over the frequencies of interest, and secondly it is found that the ratio of amplitudes of the TS1 and F1 modes is also nearly constant over this range. However the ratio of TS3 to the other two modes varies considerably. For the first two modes u_1 and u_2 are the dominant components of motion, while for the TS3 mode u_3 dominates. Using this information the boundary conditions may be simplified by defining appropriate Fourier components of vibration amplitude. Thus, taking account of the symmetry of the guided modes,

$$U_1 = \int_{-W_2/2}^{W_2/2}\int_{-W_1/2}^{W_1/2} u_1 \cos\{\pi x_1/W_1\}\sin\{\pi x_2/W_2\}\,dx_1 dx_2$$

$$U_3 = \int_{-W_2/2}^{W_2/2}\int_{-W_1/2}^{W_1/2} u_3 \sin\{\pi x_1/W_1\}\sin\{\pi x_2/W_2\}\,dx_1 dx_2 \qquad (29)$$

Equations (28) then simplify to,

$$U_1 = U_3 = 0 \quad \text{at} \quad x_3 = \pm L/2 \qquad (30)$$

U_1, U_3 are continuous at $x_3 = \pm W_2/2$ (31)

$\dfrac{\partial U_1}{\partial x_3}$, $\dfrac{\partial U_3}{\partial x_3}$ are continuous at $x_3 = \pm W_2/2$ (32)

Equation (30) is satisfied approximately by including only those bar modes which decay towards the ends of the bar in the two S regions. The approximation is valid if the length of each end region is significantly greater than the thickness. Equations (31) and (32) may be satisfied by including all four bar modes for the E region. The general solution may then be written as the sum of modes defined in equation (26). Thus,

$$u_i^{(E)} = \sum_{p=1}^{8}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} D_p c_{pm}^{(E)} b_{pmn}^{(E)} a_{ipmn}^{(E)} \exp\left[-jk_o\left\{\xi_{pm}^{(E)} x_1 + \zeta_{pmn}^{(E)} x_2 + \xi_p^{(E)} x_3\right\} + j\omega t\right]$$

$$u_i^{(S^-)} = \sum_{p=1}^{8}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} D_p c_{pm}^{(S^-)} b_{pmn}^{(S^-)} a_{ipmn}^{(S^-)} \exp\left[-jk_o\left\{\xi_{pm}^{(S^-)} x_1 + \zeta_{pmn}^{(S^-)} x_2 + \xi_p^{(S^-)} x_3\right\} + j\omega t\right]$$

$$u_i^{(S^+)} = \sum_{p=1}^{8}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} D_p c_{pm}^{(S^+)} b_{pmn}^{(S^+)} a_{ipmn}^{(S^+)} \exp\left[-jk_o\left\{\xi_{pm}^{(S^+)} x_1 + \zeta_{pmn}^{(S^+)} x_2 + \xi_p^{(S^+)} x_3\right\} + j\omega t\right] \qquad (33)$$

where the superscripts refer to the appropriate region. Substituting equation (33) into equations (31) and (32) yields linear equations of the form

$$\sum_{p=1}^{8} F_{jp} D_p = 0 \qquad (j = 1,\ldots,8) \qquad (34)$$

where the elements of the matrix F are functions of ω. Solutions for the short-circuit resonant frequencies of the bar are then found by solving,

$$|F| = 0 \qquad (35)$$

The ratios of D_p are found by resubstituting for ω in any 7 of equation (34). This gives the complete solution for the particle motion associated with each mode of the resonator (except for an unknown overall amplitude constant). Typical solutions for U_1 and U_3 for the first three modes with L = 2.8 and W = 20 are shown in Figure 7. R is the inharmonic overtone mode number, i.e. the number of half-wavelengths of U_1 trapped under the electrode. Note that U_1 and U_3 have opposite symmetry. We would not expect the even-numbered inharmonics to be excited because the charge associated with U_1 cancels due to symmetry, and the charge associated with U_3 is very small due to weak piezoelectric coupling to this component. The damped oscillation

in the S regions is a fairly unfamiliar form of energy-trapping, although a similar solution has been obtained recently from a two-dimensional analysis of the width-extensional-mode ceramic resonator[5]. The lengthwise amplitude distribution of T_1 measured by Zumsteg[6] with a laser-probe does not show this damped oscillation. However this is because the resonator geometry used produces near critical damping. The predicted and measured distributions agree well as shown in Figure 8. A typical spectrum of inharmonic overtone modes is shown for different values of mass-loading and fixed aspect ratio L in Figure 9. Frequencies shown are normalised to the cut-off frequency in the S regions. To ensure that only the fundamental mode (k = 1) is excited it is clear that the ratio of electrode length to crystal thickness must be less than a certain critical value. This value is a function of both L and σ. Spurious inharmonic overtone modes are easier to control in this type of resonator than the conventional AT-cut disc, since in the latter it is possible to trap more than one half-cycle of vibration in any direction in the plane of the electrodes. Here we need only be concerned with trapping in the Z' direction. [illegible] control of spurii partly accounts for the very clean spectrum observed[1]. Figures [illegible] show the predicted vibration patterns for the fundamental mode of a typical resonator in planes normal to Z' and X respectively. In particular the two central diagrams in Figure 10 show the vibration predicted in two nearly co-[illegible] planes close to the electrode edge, one [illegible] just inside the E region and the other just inside the S region. The smooth transition between the solutions in the two regions is clearly demonstrated both here and in all the diagrams in Figure 11.

Quality Factor

A number of different phenomena affect the quality factor Q of a resonator. In the device considered here the energy-trapping and air-loading effects are the most significant. The damping due to air-loading is a strong effect because there are large components of vibration normal to the surfaces (see Figures 10 and 11). Typically evacuation of the enclosure increases Q by an order of magnitude from 50,000 to 500,000. However, achieving such a high value also requires efficient trapping of the mode under the electrode. Figure 12 shows schematically how an estimate of Q (considering energy-trapping alone) can be made. The kinetic energy stored in a resonator of infinite length is found by integration over the range $-\infty < x < \infty$, and this is subdivided into the energy stored in the actual resonator, and the remainder which may be regarded as the energy absorbed at the mountings. Q is the ratio of energy stored to energy lost. It should be understood that this approach merely provides a rough estimate of Q since in practice not all energy at the physical ends of the resonator is absorbed. Nevertheless, such a calculation provides a useful method of comparing the energy-trapping in different designs.

We now refer back to the dispersion curves shown in Figure 5. For a trapped mode the end sections of the bar are excited at a frequency below ω_{c1}, such that both Y and Z modes decay. The overall rate of decay of the trapped mode is proportional to the smaller of the imaginary parts of ξ. For a given lowering of the frequency, the rate of decay is shown to vary strongly with width-to-thickness ratio L. This implies that Q can be maximised by appropriate choice of L, this choice being influenced by both electrode length and mass-loading which together determine the relationship between resonant frequency and cut-off frequency.

We now consider how the Q's of the fundamental and inharmonic overtone modes vary with electrode length and crystal length. Computed values are shown for a particular choice of L and σ in Figure 13. Clearly there are optimum values of electrode length/crystal thickness W and these generally occur below the cut-off electrode length for the inharmonic modes (see Figure 9).

The optimum for a given bar length arises because a short electrode implies an operating frequency close to cut-off and therefore a small attenuation coefficient, and a long electrode leaves insufficient space between the electrodes and the ends of the bar to achieve adequate decay of the amplitude. Figure 14 shows similar curves for a slightly different value of L. As can be seen by comparing Figures 13 and 14 the Q has increased considerably, illustrating the significant effect of the change in dispersion shown in Figure 5. The double peak in Q for the longer bar results from the interaction of the Y and Z modes. The order of magnitude of predicted Q shown in Figure 14 agrees well with typical measured values[7].

Frequency-Temperature Performance

Figure 15 shows typical theoretical and experimental values of first-order temperature coefficient α as a function of L. Measured values are taken from Ref. 1. Solid curves show predicted values for the untrapped mode of an unplated bar using Mason's constants[4]. The overall shape of each curve is parabolic and displaced by about 0.35 ppm/°C for each 0.1° change in φ. The effect of plating the bar to a mass-loading of 1% is to produce a similar shift equivalent to about 4 minutes of arc. The effect of trapping on these curves is not considered here, but nevertheless good agreement with experiment is shown. Agreement using Bechmann's constants[4] (dashed curve) is less good, the difference between the two sets of data being equivalent to some 23 minutes of arc.

Similar agreement between experiment and theory using Mason's constants has also been obtained for the 8th harmonic flexural mode. The important conclusion from Figure 15 is that angle φ and nominal aspect ratio L can be optimised to give only second order errors in temperature coefficient for first order errors in dimensional

ratio. Manufacturing tolerance can then be relaxed. A typical set of computed frequency-temperature curves for near-optimum conditions is shown in Figure 16. These confirm that the second-order variation of frequency-stability with first-order error in L applies over a wide temperature range, and also that it is possible to achieve very high frequency stability at the hotter end of the range. Total variation of no more than 3 ppm between 25° and 100° is possible, comparing favourably with the temperature performance of the SC-cut resonator[12]. These predictions are broadly confirmed by experiment[11] as shown in Figure 17.

Equivalent Circuit Parameters

We now consider the description of the resonator in terms of its equivalent circuit parameters. The normal form of circuit is shown in Figure 18. C_s is the capacitance at infinite frequency, and $C_{N,R}$ $L_{N,R}$ and $R_{N,R}$ are the motional capacitance, motional inductance and equivalent series resistance of the Rth inharmonic overtone of the Nth flexural mode. The static or D.C. capacitance is given by

$$C_o = C_s + \sum_N \sum_R C_{N,R} + \sum_q C_q \qquad (36)$$

where the last term represents the additional branches corresponding to other types of spurious mode.

Neglecting the fringing effect, an approximate value for static capacitance is given by

$$C_o = \varepsilon_{22} \frac{S}{t} W \qquad (37)$$

The motional capacitance of the N,Rth mode is given by a generalised form of an expression derived by Reilly[7] from the theory of Lloyd[12]. Thus,

$$C_{N,R} = \frac{1}{\omega_{N,R}^2} \frac{\int_{-t_1/2}^{t_1/2} \int_{-W_2/2}^{W_2/2} \int_{-t_3/2}^{t_3/2} \left[D_2 f(x_1 x_3) \right]_{x_2 = t_2/2} dx_1 dx_3}{\int_{-t_1/2}^{t_1/2} \int_{-t_2/2}^{t_2/2} \int_{-t_3/2}^{t_3/2} \left(u_1^2 + u_2^2 + u_3^2 \right) dx_1 dx_2 dx_3} \qquad (38)$$

where $\omega_{N,R}$ is the angular frequency. The other parameters, if required, may be found from

$$L_{N,R} = \frac{1}{\omega_{N,R}^2 C_{N,R}} \qquad (39)$$

$$R_{N,R} = \frac{\omega_{N,R} L_{N,R}}{Q_{N,R}} \qquad (40)$$

where $Q_{N,R}$ is the quality factor of the N,Rth resonance. It is found convenient to define normalised capacitance constants ε_o and $\varepsilon_{N,R}$ such that

$$C_o \text{ (in pF)} = \varepsilon_o A \text{ (in cm}^2\text{)} . t \text{ (in MHz)} \qquad (41)$$

$$C_{N,R} \text{ (in fF)} = \varepsilon_{N,R} A \text{ (in cm}^2\text{)} . t \text{ (in MHz)} \qquad (42)$$

where A and f are electrode area and resonant frequency. Typical values of ε_o and $\varepsilon_{4,1}$ versus normalised electrode length W are shown in Figure 19 for two values of a and fixed L. For this aspect ratio it can be seen that the important capacitance ratio $C_o/C_{4,1}$ has a minimum value of about 253. This is somewhat higher than the ratio for AT-cut thickness-shear mode resonators.

Figure 20 shows a comparison of the predicted and measured[1] motional capacitances of 2nd, 4th, 6th and 8th harmonic flexural mode resonators as functions of aspect ratio. Agreement is generally good, and the small discrepancy could be accounted for by the fact that the resonator blanks were etched after measuring the dimensions, thus slightly changing the ratio of width to thickness. Maximum motional capacitance occurs for values of L at which the resonant frequency is close to that of the infinite plate.

Conclusions

A new three-dimensional theory of rectangular bar resonators has been developed. This has been shown to give good agreement with experiment for frequency, motional capacitance, frequency-temperature behaviour and energy-trapping. It has also been shown that the theory can be used to design the physical dimensions of the resonator and its electrodes such that one or more of its electrical parameters is optimised. In particular the values of width-to-thickness of the bar for maximum energy-trapping, minimum capacitance ratio $C_o/C_{4,1}$ and minimum sensitivity of the frequency-stability to dimensional errors are approximately equal. A compromise design with near-optimum performance of many of its parameters is therefore possible using the type of simple rectangular resonator described.

References

1. ZUMSTEG, A.E. and SUDA, P., "Properties of a 4 MHz miniature flat rectangular quartz resonator vibrating in a coupled mode". Proc. 30th AFCS, pp.196-201, 1976.

2. ZUMSTEG, A.E., SUDA, P. and ZINGG, W., "Energy-trapping of coupled modes in rectangular AT-cut resonators". Proc. 32nd AFCS, pp.260-266, 1978.

3. MILSOM, R.F., "Three-dimensional variational analysis of small crystal resonators". Proc. 33rd AFCS, pp.263-270, 1979.

4. VANGHELUWE, D.C.L., "Finite-element analysis of AT-cut crystals". Proc. 32nd AFCS pp.134-141, 1978.

1. BECHMANN, R., BALLATO, A.D. and LUKASZEK, T.J. "Higher temperature coefficients of the elastic stiffness and compliances of alpha-quartz". Proc. IRE, 50, pp.1812-1822, 1962.

2. MINDLIN, R.D., "Thickness-shear and flexural vibrations of crystal plates". J. Appl. Phys. 22, pp.316-323, 1951.

3. SMITH, W.R., and SHOCKLEY, M., "Wave propagation analysis of the monolithic-crystal filter". Proc. IRE, 116, pp.1555-1564, 1964.

4. WATANABE, H., NAKAMURA, K. and SHIMIZU, H., "A new type of contra-trapping caused by contributions from the complex branches of dispersion curves". Proc. 1980 Ultrasonics Symposium, p.752-755.

5. MASON, W.P., "Zero temperature coefficient quartz crystals for very high temperatures". Bell Sys. Tech. J., 30, pp.360-380, 1951.

6. STEVENS, A., "Doubly rotated thickness mode plate vibrators". In "Physical Acoustics", (Mason, W.P. and Thurston, R.N., ed), pp.11-18, Academic Press, New York, 1977.

7. SMITH, A.E., (Private Communication).

8. TIERSTEN, H., "Equations governing the electrical behavior of an arbitrary piezoelectric resonator having N electrodes". Bell Sys. Tech. J., 56, pp.1881-1900, 1967.

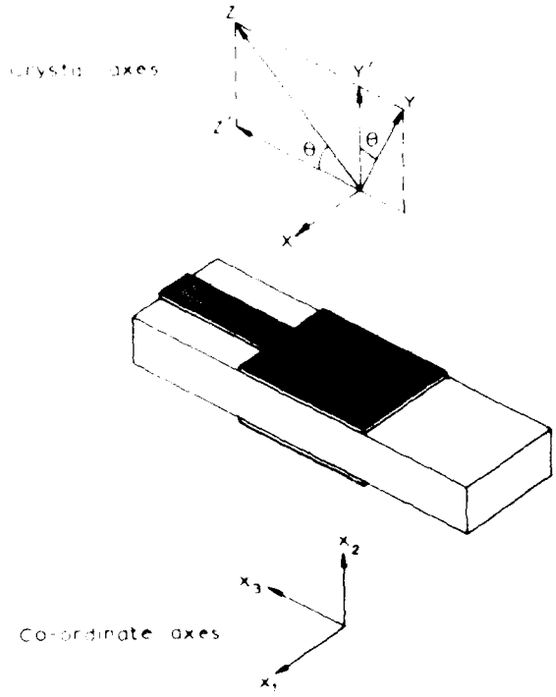


Fig 1 Zumsteg bar resonator

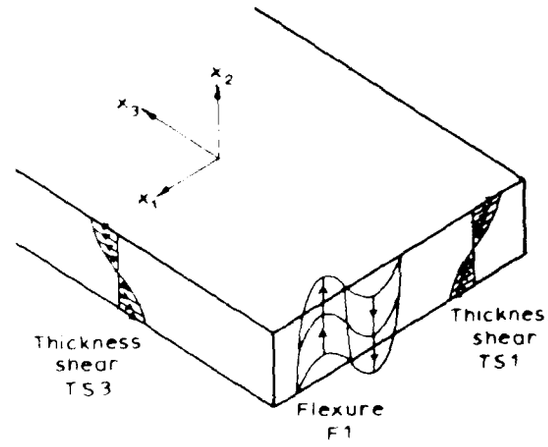
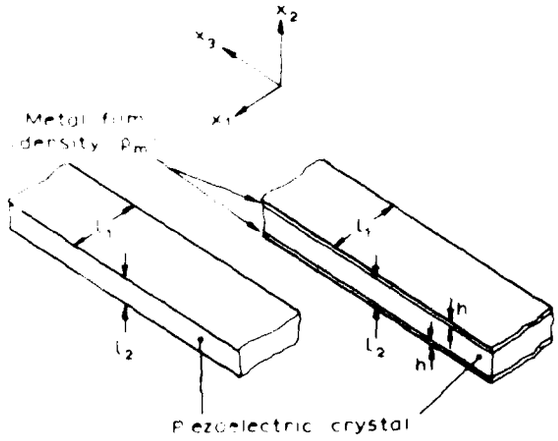


Fig 3 Predominant plate modes in Zumsteg bar resonator



Unplated guide Plated guide

Fig 2 Infinite rectangular acoustic waveguide

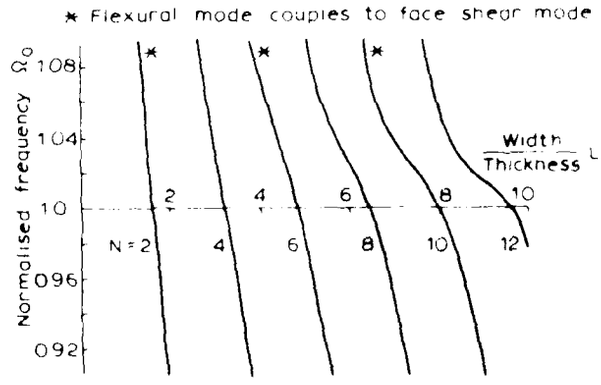


Fig 4 Mode spectrum for infinitely long bars, with $\xi = 0$, $\theta = 34.8°$

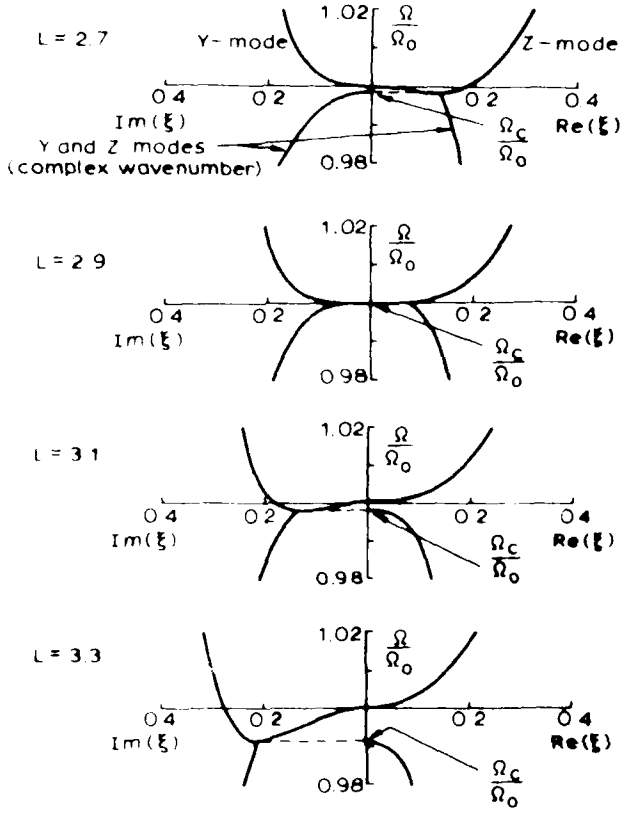


Fig 5 Dispersion of Y and Z bar modes for 4 values of width / thickness L

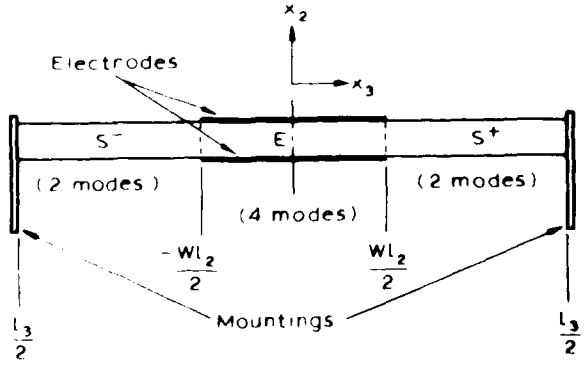


Fig 6 Definition of S^-, E and S^+ waveguide regions and number of guided modes used in solution in each region

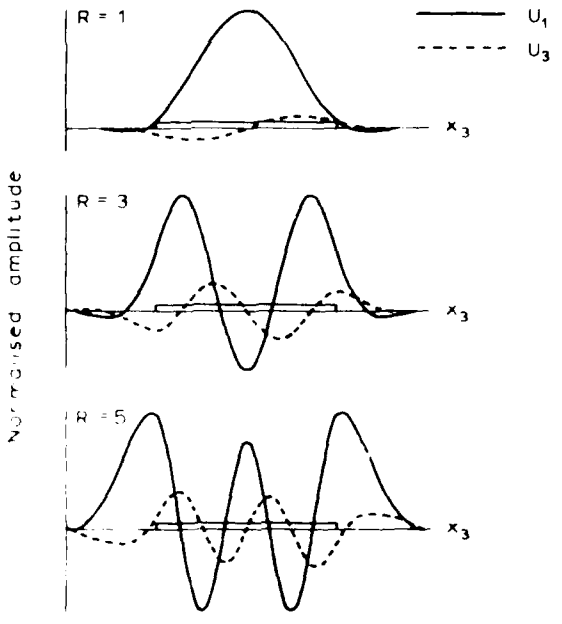


Fig 7 Lengthwise amplitude distribution of fundamental and inharmonic overtone modes

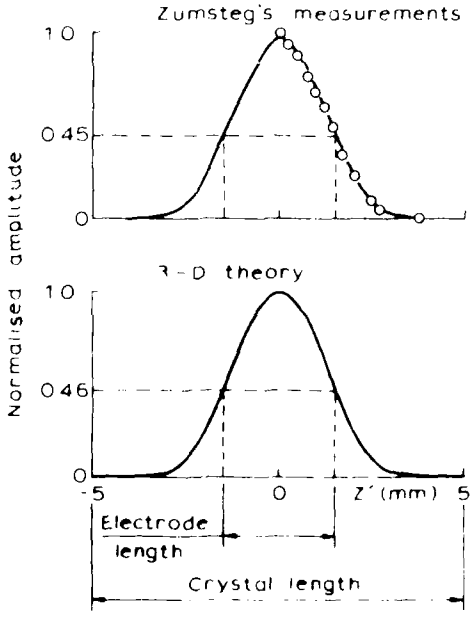
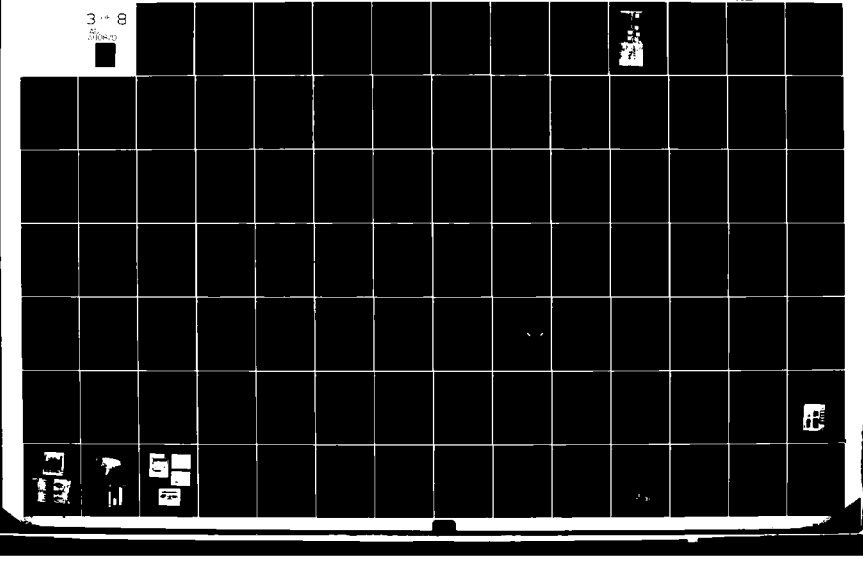


Fig 8 Predicted and measured lengthwise spatial distribution of component of vibration normal to electroded surface (L = 3.125, W = 7.5, μ = 0.008)

AD-A110 870 ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND AD--ETC F/G 9/5
PROCEEDINGS OF THE 35TH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL, --ETC(U)
1981

UNCLASSIFIED

NL



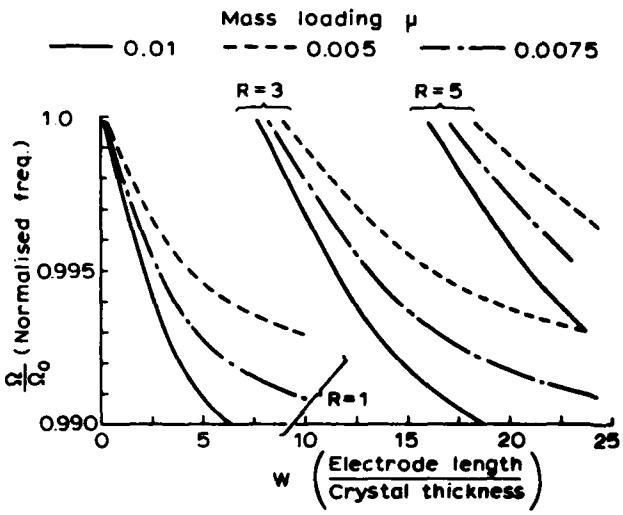


Fig.9 Inharmonic overtone mode spectrum for bar with L = 2.85

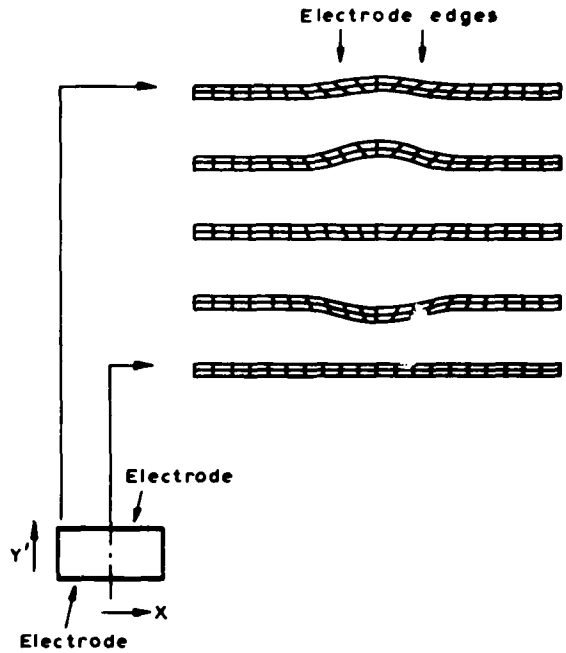


Fig. 11 Vibration in equally-spaced planes normal to X.

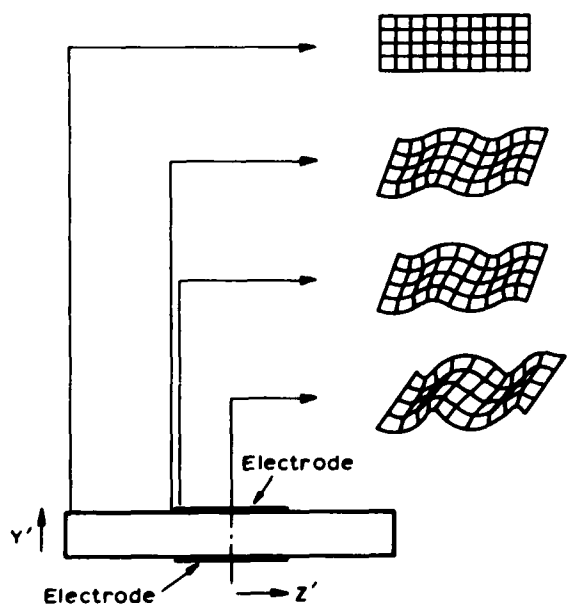
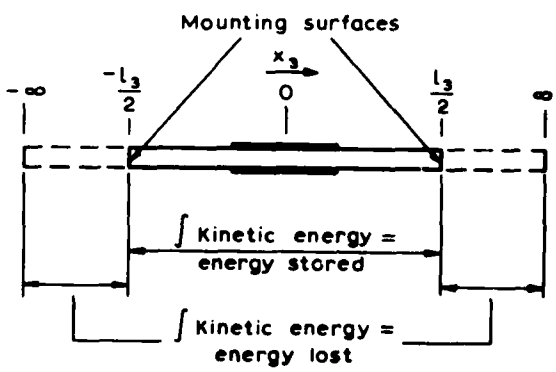


Fig. 10 Vibration in planes normal to Z'.



Q = Energy stored / Energy lost

Fig. 12 Scheme for estimating Q-factor.

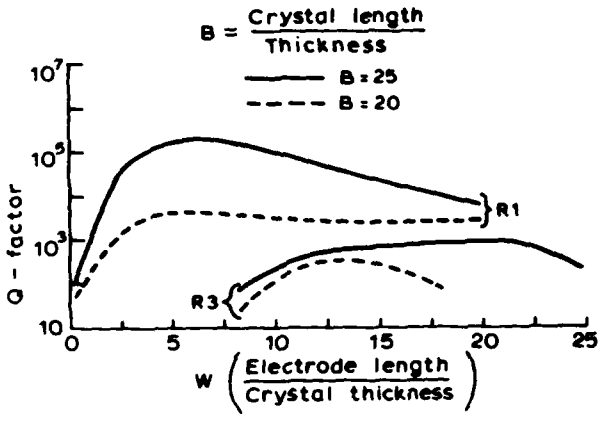


Fig. 13 Quality factors of fundamental and first spurious mode ($L = 2.85$, $\mu = 0.0075$)

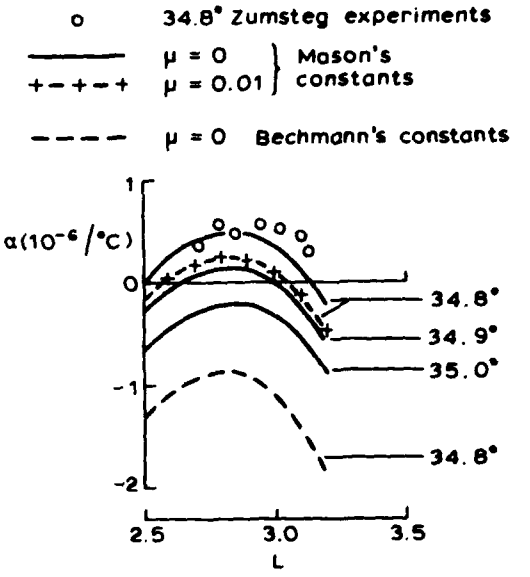


Fig. 15 First-order temperature coefficient α at 25°C.

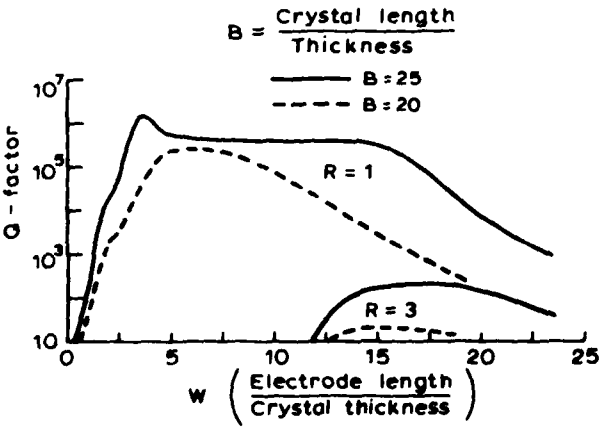


Fig. 14 Quality factors of fundamental and first spurious mode ($L = 3.125$, $\mu = 0.0075$)

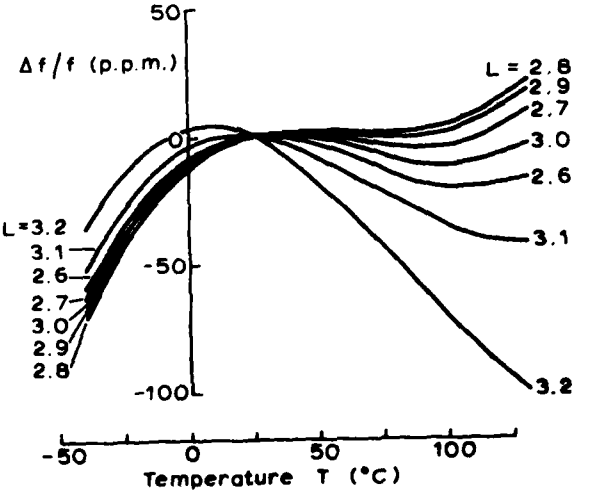


Fig. 16 Frequency - temperature characteristics - $\theta = 34.8°$, $\mu = 0.01$.

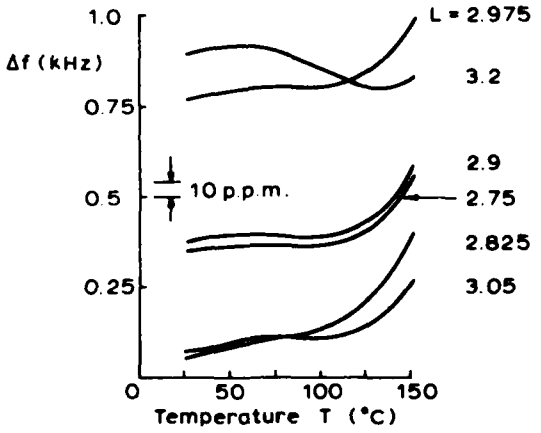


Fig.17 Measured frequency-temperature characteristics – θ = 34.8°.

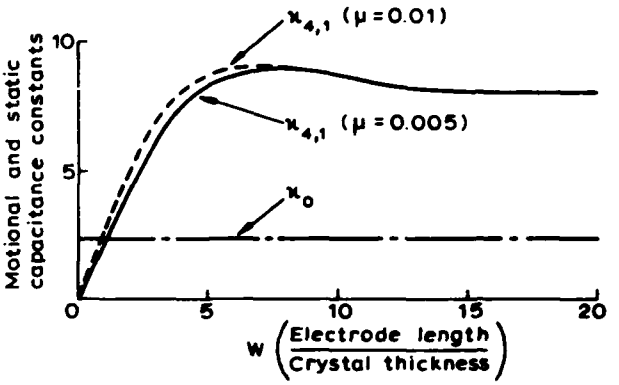


Fig. 19 Capacitance constants for L = 2.85

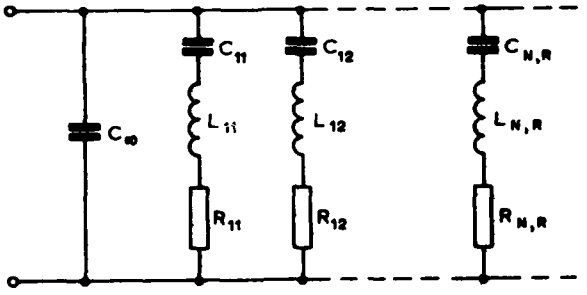


Fig 18 Electrical equivalent circuit of bar resonator.

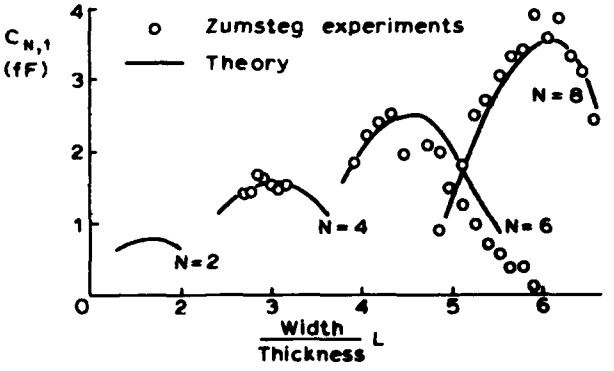


Fig. 20 Comparison of predicted and measured motional capacitance for 4MHz resonators.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

SIMPLE MODEL FOR AN AT CUT RECTANGULAR QUARTZ PLATE

J.H. BALBI AND M.I. DULMET

Laboratoire de Mécanique Appliquée, associé au C.N.R.S.
Faculté des Sciences de Besançon. La Bouloie. Route de Gray
25030 BESANCON CEDEX (FRANCE)

SUMMARY

A very simple theoretical model is proposed for an AT Cut rectangular plate and nevertheless gives a very correct account of experimental results. Avantages of the model are obvious. In particular, the model yields resonant frequencies, vibration modes, quality factors. Also it is very suitable for further studies of particular non linear phenomena. Model pattern habe been experimentally obtained by X-ray topography and show a good agreement with theory.

INTRODUCTION

A very simple model for the description of thin rectangular quartz plate's vibrations[1] is given. The proposed solution is reliable and simple. This model yields analytical expressions for frequencies and modal patterns of eigenmodes. It will be used for further linear and non-linear investigations. Under some assumptions, we obtain a two equations system which gives the components u_1 and u_2 of the mechanical displacement as fonctions of the three spatial coordinates.

ASSUMPTIONS

Some usual assumptions are made :

1/ The Piezoelectric coupling, small in quartz will be neglected : resonator vibrations are essentially elastic. Piezoelectricity appears as an excitation or detection mean.

2/ This thickness shear resonator is a rectangular AT Cut plate, the thickness of which (2b) is small compared to the length (2a) and the width (2c).

3/ Under these circunstances[2] the dominant deformation, i.e displacement gradient, is $u_{1,2}$ because :

i) the wawe-numbers along the plate are much smaller then the thickness number,

ii) the other mechanical displacements are much smaller than u_1 displacement.

$$u_1 >> u_2,\ u_3$$

$$u_{i,2} >> u_{i,1},\ u_{i,3} \quad , \quad i = 1,2,3$$

A comma followed by an index denotes differentiation with respect to the reference coordinates.

4/ Though the lapping has been performed in such a way to get a plate as plane as possible, some convexity will still subsist ; therefore, the energy will be trapped. This volume phenomenon will be taken into account by boundary conditions ; i.e instead of writting that the plate is slightly convex while the boundary is free, the choosen phenomenological model will be a plane plate with a zero displacement u_1 at the boundary :

$$u_1\ (\pm a) = 0$$

$$u_1\ (\pm c) = 0$$

5/ The resonator is clamped on the faces $x_3 = \pm c$ such that $u_2\ (\pm c) = 0$.

FREE UNDAMPED MODEL

With these assumptions we can write the stresses :

$$T_{11} = c_{11}\, u_{1,1} + c_{12}\, u_{2,2} \qquad (1)$$

$$T_{22} = c_{12}\, u_{1,1} + c_{22}\, u_{2,2}$$

$$T_{13} = c_{55}\, u_{1,3}$$

$$T_{12} = c_{66}\, (u_{1,2} + u_{2,1})$$

And constitutive equations :

$$T_{11,1} + T_{12,2} + T_{13,3} = \rho \ddot{u}_1 \qquad (2)$$

$$T_{12,1} + T_{22,2} = \rho\, \ddot{u}_2$$

or

$$c_{11}u_{1,11}+c_{66}u_{1,22}+c_{55}u_{1,33}+(c_{12}+c_{66})u_{2,12} = \rho\,\ddot{u}_1$$
$$(c_{66}+c_{12})u_{1,12}+c_{66}u_{2,11}+c_{22}u_{2,22} = \rho\,\ddot{u}_2 \qquad (3)$$

these equations admit the following particular solutions :

$$u_1 = A\cos\zeta x_1 \sin\eta x_2 \cos\xi x_3 \cos\omega t$$
$$u_2 = B\sin\zeta x_1 \cos\eta x_2 \cos\xi x_3 \cos\omega t \qquad (4)$$

Substituting the u_1, u_2 expressions in equations (3) we obtain the dispersion relation :

$$(c_{11}\zeta^2+c_{66}\eta^2+c_{55}\xi^2-\rho\omega^2)(c_{66}\zeta^2+c_{22}\eta^2-\rho\omega^2) - (c_{12}+c_{66})^2\eta^2\zeta^2 = 0 \qquad (5)$$

Boundary conditions :

The two second order differential equations system implies two boundary conditions on each face of the quartz.

1) faces $x_1 = \pm a$

$u_1 = 0$ (from H_4)

and $T_{12} = c_{66}(u_{1,2} + u_{2,1}) = 0$

then $\zeta = m\frac{\pi}{2a}$ m odd (6)

2) faces $x_3 = \pm c$

From H_4 and H_5 it follows that

$u_1(\pm c) = 0$

$u_2(\pm c) = 0$

then $\xi = p\frac{\pi}{2c}$ p odd (7)

3) faces $x_2 = \pm b$

they are free so :

$$T_{12} = c_{66}(u_{1,2} + u_{2,1}) = 0$$
$$T_{22} = c_{12}u_{1,2} + c_{22}u_{2,1} = 0 \qquad (8)$$

from relations (8), it appears that two wave numbers[2] η_1, η_2 are needed in order to satisfy the boundary conditions. Solutions of the differential equations (3) are now in the form :

$$u_1 = (A_1\sin\eta_1x_2+A_2\sin\eta_2x_2)\cos\zeta x_1\cos\xi x_3\cos\omega t$$
$$u_2 = (B_1\cos\eta_1x_2+B_2\cos\eta_2x_2)\sin\zeta x_1\cos\xi x_3\cos\omega t \qquad (9)$$

Since ω is in the vicinity of thickness shear frequency and since the wave numbers along the plate ζ and ξ are small, equation (5) gives the approximated values :

$$\eta_1^2 = \frac{\rho\omega^2}{c_{66}}$$
$$\eta_2^2 = \frac{\rho\omega^2}{c_{22}} \qquad (10)$$

from relation (3) we have :

$$B_1 = \beta\frac{\zeta}{\eta_1}A_1$$
$$B_2 = -\frac{\eta_2}{\zeta\beta}A_2 \qquad (11)$$

where : $\beta = -\frac{(c_{12}+c_{66})}{(c_{22}-c_{66})}$ (12)

equations (11) and (9) yield :

$$c_{66}(\eta_1+\beta\frac{\zeta^2}{\eta_1})A_1\cos\eta_1b+\eta_2c_{66}(1-\frac{1}{\beta})A_2\cos\eta_2b=0$$
$$(c_{12}+c_{22}\beta\zeta A_1\sin\eta_1b+(c_{12}\zeta-\frac{\eta_2^2}{\beta\zeta}c_{22})A_2\sin\eta_2b = 0 \qquad (13)$$

from (10) we note that $\eta_2 = \eta_1\frac{c_{66}}{c_{22}}$ (14)

Since we are in the vicinity of a thickness shear mode we take :

$$\eta_1 = \frac{n\pi}{2b}+\frac{\epsilon}{b} \qquad n \text{ odd} \qquad (15)$$

with some approximations equations (13), (14), (15) give :

$$\epsilon = -\frac{\zeta^2}{\eta_1^2}\cdot\frac{(c_{12}+c_{22}\beta)(1-\beta)}{\sqrt{c_{66}c_{22}}}\cdot\cot g\frac{n\pi}{2}\frac{c_{66}}{c_{22}} \qquad (16)$$

then :

$$\frac{A_2}{A_1} = (-1)^{\frac{n+1}{2}}\frac{(c_{12}-\beta c_{22})\zeta}{\sqrt{c_{66}c_{22}}\,\eta_1\sin\eta_1b\frac{c_{66}}{c_{22}}} \qquad (17)$$

from equations (17), (3), we obtain :

$$\rho\omega^2 = c_{66}\eta_1^2+c_{55}\xi^2+\left|c_{11}+\frac{(-1)^{\frac{n+1}{2}}(c_{12}-\beta c_{22})}{(c_{66}+c_{12})^{1/2}\sin\frac{c_{66}}{c_{22}}\eta_1b}\right|\xi^2 \qquad (18)$$

If we only consider thickness shear vibrations, we can neglect the compression term and then :

$$u_1 = A_1\cos\zeta x_1\sin\eta_1x_2\cos\xi x_3\cos\omega t$$
$$u_2 = B_1\sin\zeta x_1\cos\eta_1x_2\cos\zeta x_3\cos\omega t \qquad (19)$$

N.B.:

These solutions are particular solutions, for example $u_1=A_1\sin\zeta_1x_1\sin\eta_1x_2\sin\xi x_3\cos\omega t$ is an other solution, consequently in the expression of ω the integer m and p can be taken odd or even.

FREE DAMPED MODEL

We assume[3] that the boundary conditions are not perturbated by damping and we only keep the part of the solutions which describe thickness shear vibrations then we look for damped solutions of the following type :

$$u_1 = (A_1'\cos\Omega t+A_1^2\sin\Omega t)e^{-\delta t}\sin\eta_1x_2\cos\zeta x_1\cos\xi x_3$$
$$u_2 = (B_1'\cos\Omega t+B_1^2\sin\Omega t)e^{-\delta t}\cos\eta_1x_2\sin\zeta x_1\cos\xi x_3 \qquad (20)$$

where, as in an unidimensionnal free damped oscillator Ω is obtained under the form :

$$\Omega^2 = \omega^2 - \delta^2 \qquad (21)$$

ω is an eigenfrequencie of the undamped model and δ is the damping factor. Expanding the solutions (20) in the new constitutive equations :

$$c_{11}u_{1,11}+c_{66}u_{1,22}+r_{66}\dot{u}_{1,22}+c_{55}u_{1,33}+(c_{12}+c_{66})u_{2,12}+ +r_{66}\dot{u}_{1,12} = \rho\ddot{u}_1 \qquad (22)$$

$$(c_{12}+c_{66})u_{1,12}+r_{66}\dot{u}_{1,12}+c_{66}u_{2,11}+c_{22}u_{2,22}+r_{66}\dot{u}_{2,11}= \rho\ddot{u}_2$$

we obtain the algebric system[3] :

$$\begin{vmatrix} K_1 & R_1 & P_1 & P_2 & A_1^1 \\ -R_1 & K_1 & -P_2 & P_1 & A_1^2 \\ P_1 & P_2 & K_2 & R_2 & B_1^1 \\ -P_2 & P_1 & -R_2 & K_2 & B_1^2 \end{vmatrix} = 0 \qquad (23)$$

In order to avoid the trivial solution δ must be a zero of the matrix-determinant. Taking into account the smalness of δ, this polynomial equation may be simplified :

$$N\delta^2 - 2r_{66}\delta P + M = 0 \qquad (24)$$

δ is obtained as a function of the elastic coefficients of r_{66} and of the three integers n,m,p. So we have an indication of theoretical value of the Q-factor.

EXPERIMENTAL RESULTS

Some experiments have been made in the Laboratoire de Chronométrie et Piézoélectricité de Besançon, with an AT Cut rectangular quartz plate (2a = 14,5 mm ; 2b = 0,6278 mm ; 2c = 14 mm). This resonator has quasi symmetrical electrodes ; we put this plate in an X-ray beam. Natural mechanical and electric vibrations frequencies are measured experimentally using the devices[4] illustrated in figure I . Topographs are made in order to obtain the main modes. Figure II shows overtone 9 spectrum with various topographs. Measured frequencies show a good agreement with theoretical calculated frequencies (figure III). We note that the mechanical spectrum exhibits resonance modes with m odd or even, p odd or even. We also note that some nodal lines are distorded. We are going to explain this by a coupling between two close resonance frequencies.

For example :
Take (0,9,1) and (1,9,0) modes, they are described by :

$$u_1^1 = A^1 \sin\frac{9\pi}{2b}x_2\cos\frac{\pi}{2a}x_1\sin\frac{2\pi}{2c}x_3\cos\omega t$$

$$u_1^2 = A^2 \sin\frac{9\pi}{2b}\sin\frac{2\pi}{2a}x_1\cos\frac{\pi}{2c}x_3\cos\omega t$$

u_1^1 et u_1^2 are eigenmodes.

(we have a #c) then consider :

$$u_1^1+u_1^2 \sim \cos\frac{\pi}{2a}x_1\sin\frac{2\pi}{2a}x_3+\sin\frac{2\pi}{2a}x_1\cos\frac{\pi}{2a}x_3$$

$$u_1^1+u_1^2 \sim 2\sin\frac{\pi}{4a}(x_1+x_3)\cos\frac{\pi}{4a}(x_1-x_3)$$

We obtain nodal lines for $\frac{\pi}{4a}(x_1+x_3)= k\pi$
if $k = 0$, $x_1 = -x_3$.

So consider :

$$u_1^1-u_1^2 \sim \cos\frac{\pi}{2a}x_1\sin\frac{2\pi}{2a}x_3-\sin\frac{2\pi}{2a}x_1\cos\frac{\pi}{2a}x_3$$

$$\sim 2\sin\frac{\pi}{4a}(x_3-x_1)\cos\frac{\pi}{4a}(x_1+x_3)$$

Here nodal line is $x_3 = x_1$.

The theoretical nodal lines were $x_3 = 0$ in the first case and $x_1 = 0$ in the second case.

Then the interaction between (0,9,1) and (1,9,0) modes yields a rotation of the nodal lines.

Second example :

Mode (0,9,2) and mode (1,9,1).
If we superpose these modes :

$$u_1^1 + u_1^2 \sim \cos\frac{\pi}{2a}x_1\left|\cos\frac{3\pi}{2a}x_3+2\sin\frac{\pi}{2a}x_1\sin\frac{2\pi}{2a}x_3\right|$$

Theoretical nodal lines passes by A $(0,\frac{a}{3})$.

At the vicinity of A $\quad x_1 = y_1$

$$x_3 = \frac{a}{3} + y_3$$

$$u_1^1 + u_1^2 \sim \left|\sqrt{3}\sin\frac{\pi}{2a}y_1 - \sin\frac{3\pi}{2a}y_3\right|$$

Then nodal line is in the neighbrough of the straight line :

$$3y_3 - \sqrt{3}\,y_1 = 0$$

the same study at the point $(0,-\frac{a}{3})$ gives :

$$-3y_3 + \sqrt{3}\,y_1 = 0$$

the influence of (1,9,1) mode on (0,9,2) mode implies the rotation of real nodal lines versus theoretical nodal lines.

CONCLUSION

The analysis presented in this paper makes possible the calculations of resonant frequencies and displacements of thickness modes in a rectangular plate. Calculated frequencies are compared with experimental frequencies. Topographs of displacement show a rotation of nodal lines which was explained by a coupling between natural frequencies.

The obtained accuracy is good enough for a practical use in a future non linear study.

ACKNOWLEDGMENT

The authors wish to thank R.J. BESSON and his staff for their scientific and technic help.

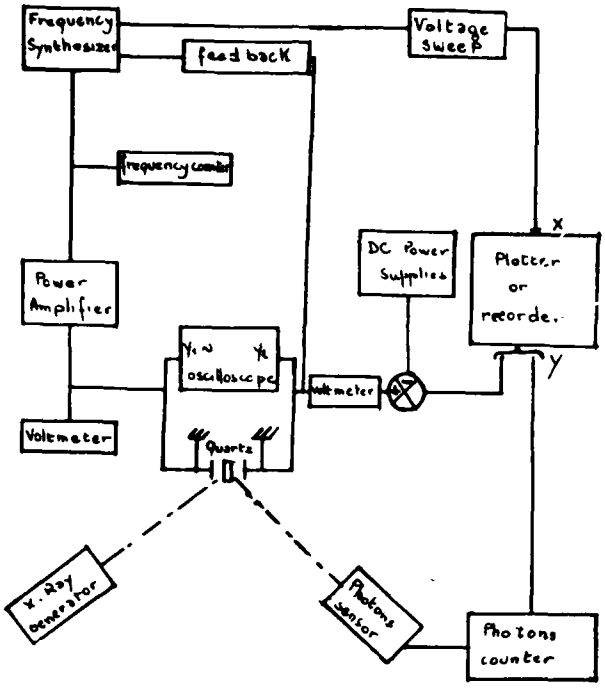
REFERENCES

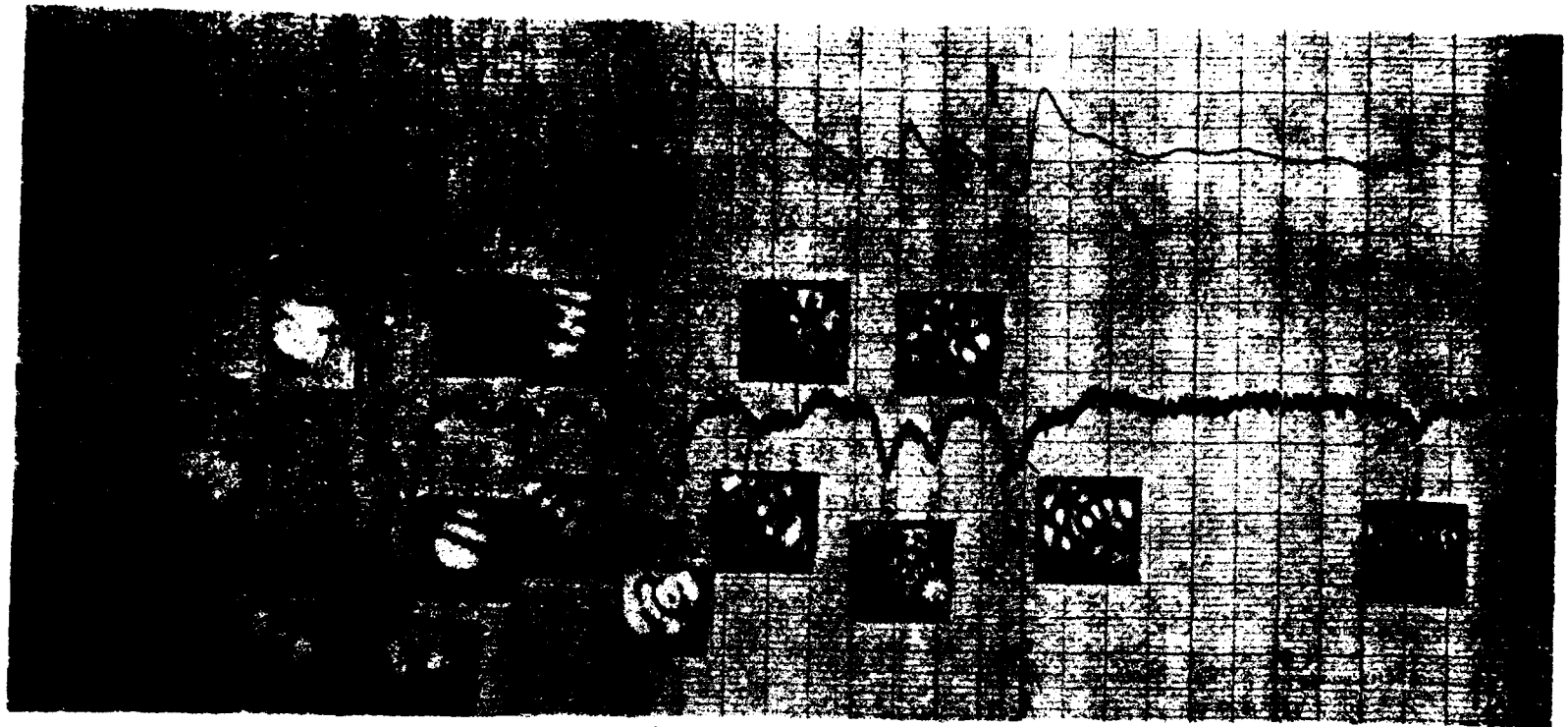
1- R.D. Mindlin, W.J. Spencer,
"Anharmonic, Thickness Twist overtone of thickness shear and Flexural vibrations of rectangular AT Cut quart plates"" J. Acoust. Soc. Am. 42 - 1268 - 1277 (1967)

2- H. F. Tiersten
"Analysis of intermodulation in thickness shear and trapped energy resonator", J. Acoust. Soc. Am. 57 - 667 - 681 (1975)

3- M. Dulmet,
"Quelques phénomènes linéaires et non linéaires dans les résonateurs à quartz", Thèse de Docteur-Ingénieur, Université de Besançon 1980.

4- A. Thirard,
"Etudes topographiques des vibrations de cisaillement d'épaisseur des résonateurs à quartz", Thèse de Docteur-Ingénieur. Université de Besançon 1980.


Frequency Synthesizer
feed back
Voltage sweep
frequency counter
DC Power Supplies
Plotter or recorde.
x
y
Power Amplifier
Yt
Xt
oscilloscope
Voltmeter
Voltmeter
Quartz
X-Ray generator
Photons sensor
Photons counter



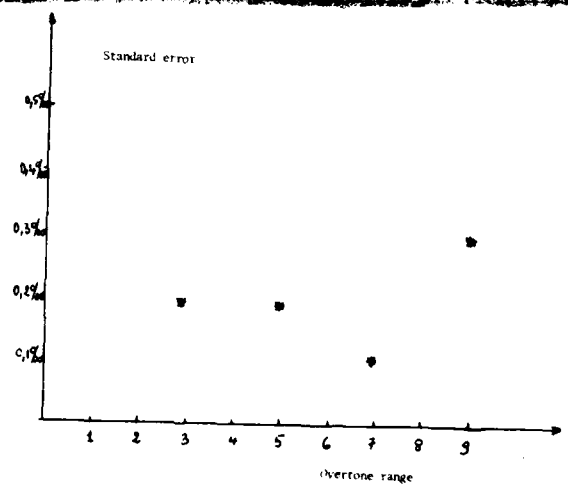


FIGURE III

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

STRESSES IN RECTANGULAR CANTILEVER CRYSTAL PLATES UNDER TRANSVERSE LOADING

P.C.Y. Lee and C.S. Lam
Department of Civil Engineering
Princeton University
Princeton, NJ 08544

Summary

A second-order theory governing approximately the symmetric flexural vibrations (or deflection) in crystal plates is derived from the Cauchy two-dimensional flexural theory of elastic, anisotropic plates.

Close form solutions of the second-order equations of flexure are obtained for rectangular cantilever plates subject to: (a) a uniformly distributed shear load along the edge parallel to the clamped edge, and (B) a uniformly distributed load over the entire plate.

For load case (A), two-dimensional displacement and stresses are computed and plotted along the x_1 and x_3 axes of AT-cut plates of quartz for various values of azimuth angle ψ. As for load case (B), the solution is employed to compute the displacement and stresses in isotropic, elastic plates so that the present results are compared, in tabulated form, with existing results by other analytic and numerical methods.

Introduction

Bending of rectangular cantilever plates of isotropic elastic materials under transverse loading still remains as one of the very important and very difficult problems in the theory of elasticity. Many approximate methods have been employed to obtain two-dimensional distributions of stresses and displacement as the solutions of the Lagrange-Germain equation of flexure of elastic plates. Nash had analyzed the problem with three different approximate methods, i.e. finite-difference and two collocation methods.[1] Chang, by the concept of generalized simply supported edge and method of superposition, solved the problem in terms of a series of infinite simultaneous equations, and had included results by L.T. Wu from finite element method for comparison.[2] Many related works to this problem can be found from the references of these two papers.[1-2]

In contrast, solutions of similar problems for anisotropic plates are scarce, and they are needed for studying the effect of transverse loading and acceleration on the resonant frequencies of crystal plates.[3]

In order to check the accuracy of the present solution, displacement and bending stress are computed for isotropic plates and are compared with those obtained by Nash,[1] Wu and Chang.[2] Results listed in Tables I-IV show that predictions by several different methods all agree with each other within a small range. However, we would like to note that our present approach leads to analytical solutions in close form.

Cauchy's Two-Dimensional Flexural Theory

Let a rectangular plate, which has a length $2a$, thickness $2b$, and width $2c$, be referred to as the rectangular coordinate system (x_i) and its middle plane coincides with the x_1x_3 plane. The azimuth angle ψ, shown in Fig. 1, is the angle between the x_1 axis of the plate and the x_1' axis of the crystal. In a rotated Y-cut plate, we may choose the diagonal axis of quartz as the x_1' axis. Referred to the x_i system, Cauchy's stress equation of flexural vibrations of crystal plates can be written as

$$T^{(1)}_{1,11} + 2\,T^{(1)}_{5,13} + T^{(1)}_{3,33} + q = 2b\rho\,\ddot{u}^{(0)}_2 \quad , \quad (1)$$

where $T^{(1)}_1$, $T^{(1)}_3$ are the bending couples, $T^{(1)}_5$ twisting couple, $u^{(0)}_2$ the deflection or displacement of the plate in the x_2 direction, q the appliec loading intensity (force per unit surface area) in x_2 direction, and ρ the mass density of the crystal.

The stress couples are related to the transverse shear stress resultants $T^{(0)}_6$ and $T^{(0)}_4$ by

$$\begin{aligned} T^{(0)}_6 &= T^{(1)}_{1,1} + T^{(1)}_{5,3} \, , \\ T^{(0)}_4 &= T^{(1)}_{5,1} + T^{(1)}_{3,3} \, . \end{aligned} \quad (2)$$

For anisotropic materials, the couple-displacement relations are

$$T_1^{(1)} = -\frac{2b^3}{3}(\gamma_{11}u_{2,11}^{(0)} + \gamma_{13}u_{2,33}^{(0)} + 2\gamma_{15}u_{2,13}^{(0)})$$

$$T_3^{(1)} = -\frac{2b^3}{3}(\gamma_{31}u_{2,11}^{(0)} + \gamma_{33}u_{2,33}^{(0)} + 2\gamma_{35}u_{2,13}^{(0)}) \quad (3)$$

$$T_5^{(1)} = -\frac{2b^3}{3}(\gamma_{51}u_{2,11} + \gamma_{53}u_{2,33} + 2\gamma_{55}u_{2,13})$$

where γ_{ij} are the elastic material constants. Inserting (3) into (1), we obtain

$$\gamma_{11}u_{2,1111}^{(0)} + 4\gamma_{15}u_{2,1113}^{(0)} + 2(\gamma_{13} + \gamma_{55})\,u_{2,1133}^{(0)} + 4\gamma_{35}u_{2,1333}^{(0)} + \gamma_{31}u_{2,3333}^{(0)} + \frac{3\rho}{b^2}u_2^{(0)} = \frac{3}{2b^2}q\,, \quad (4)$$

which is the displacement equation of flexural vibration of crystal plates.

For isotropic, elastic materials, γ_{ij} can be expressed in terms of the Young's modulus E and the Poisson's ratio ν by the relations

$$\gamma_{11} = \gamma_{33} = \frac{E}{1-\nu^2}\,,\quad \gamma_{13} = \frac{\nu}{1-\nu^2}E\,,\quad \gamma_{55} = \frac{E}{1+\nu^2}\,,\quad \gamma_{13} = \gamma_{35} = 0\,. \quad (5)$$

Hence, (3) are reduced to

$$T_1^{(1)} = -D(u_{2,11}^{(0)} + u_{2,33}^{(0)})$$
$$T_3^{(1)} = -D(u_{2,33}^{(0)} + u_{2,11}^{(0)}) \quad (3')$$
$$T_5^{(0)} = -D(1-\nu)\,u_{2,13}^{(0)}$$

and the displacement equation of motion (4) becomes

$$D\nabla^4 u_2^{(0)} + 2b\rho\,\ddot{u}_2^{(0)} = q \quad (4')$$

for isotropic plates, where $D = \frac{2b^3}{3}\frac{E}{1-\nu^2}$ and

$$\nabla^4 = \frac{\partial^4}{\partial x_1^4} + 2\frac{\partial^4}{\partial x_1^2\,\partial x_3^2} + \frac{\partial^4}{\partial x_3^4}\,.$$

On the edges of the plate, we require that:

(1) $T_1^{(1)}$ or $u_{2,1}^{(0)}$ and $T_6^{(0)} + T_{5,3}^{(1)}$ or $u_2^{(0)}$ be specified at $x_1 = \pm a$.

(2) $T_3^{(1)}$ or $u_{2,3}^{(0)}$ and $T_4^{(0)} + T_{5,1}^{(1)}$ or $u_2^{(0)}$ be specified at $x_3 = \pm c$.

One-Dimensional Equations for Flexure

In order to replace approximately the two-dimensional equations (1) - (4) by a system of one-dimensional equations, we expand displacement in a series of characteristic functions ϕ_n

$$u_2^{(0)}(x_1,x_3,t) = \sum_{n=0}^{\infty} v_2^{(n)}(x_1,t)\,\phi_n(x_3)\,. \quad (6)$$

where

$$\phi_0 = 1\,,\quad \phi_1 = \sqrt{3\eta}\,,\quad (\eta = x_3/c)$$
$$\phi_n = \frac{\cos(\frac{n\pi}{2} + k_n\eta)}{\cos(\frac{n\pi}{2} + k_n)} + \frac{\cosh(\frac{n\pi}{2} + i\,k_n\eta)}{\cosh(\frac{n\pi}{2} + ik_n)}\,, \quad (7)$$

and k_n satisfies the trancendental equations

$$\tan k_n = \pm \tanh k_n\,. \quad (8)$$

The characteristic functions $\phi_n(x_3)$ are the one-dimensional modes of (4) or (4') when the displacement $u_2^{(0)}$ is dependent on x_3 and t only; the transcendental equation (8) is resulted from the traction-free conditions imposed on the edges $x_3 = \pm c$, i.e. $T_0^{(1)} = T_4^{(0)} = 0$ at $x_3 = \pm c$ (or $\eta = \pm 1$). ϕ_n also satisfies the orthogonality condition

$$\int_{-1}^{1} \phi_m\phi_n\,d\eta = 2\delta_{mn}\,. \quad (9)$$

The variational stress equation of motion corresponding to (1) is

$$\int_{t_0}^{T_1}\int_A (T^{(1)}_{1,11} + 2\,T^{(1)}_{5,13} + T^{(1)}_{3,33} + q - 2b\rho\,\ddot{u}^{(0)}_2)\delta u^{(0)}_2\,dA\,dt = 0\,. \tag{10}$$

By substituting (6) into (10), replacing dA by $c\,dx_1\,d\eta$, integrating by parts over $\eta = -1$ to $\eta = +1$, and requiring that the final integral be independent of any arbitrary variation $\delta v^{(n)}_2$, we obtain the n^{th}-order stress equation of motion

$$M^{(n)}_{1,11} - \frac{2}{c}M^{(n)}_{5,1} + \frac{1}{c^2}M^{(n)}_3 + q^{(n)} + V^{(n)} = 4b\rho\,\ddot{v}^{(n)}_2, \tag{11}$$

where

$$\begin{aligned} M^{(n)}_1 &= \int_{-1}^{1} T^{(1)}_1\,\phi_n\,d\eta\,, \\ M^{(n)}_5 &= \int_{-1}^{1} T^{(n)}_5\,\phi'_n\,d\eta\,, \\ M^{(n)}_3 &= \int_{-1}^{1} T^{(n)}_3\,\phi''_n\,d\eta\,, \\ q^{(n)} &= \int_{-1}^{1} q\,\phi_n\,d\eta\,, \\ V^{(n)} &= \left[\frac{1}{c}(T^{(0)}_4 + T^{(1)}_{5,1})\phi_n - \frac{1}{c^2}T^{(1)}_3\phi'_n\right]^{1}_{-1}, \end{aligned} \tag{12}$$

and $\phi'_n = \partial\phi/\partial\eta$, $n=0,1,2,\ldots$. We note that $V^{(n)}$ is determined by the applied shear $T^{(0)}_4 + T^{(1)}_{5,1}$ and the applied bending couple $T^{(1)}_3$ on the edges $x_3 = \pm c$. Therefore, for traction-free edges at $x_3 = \pm c$, $V^{(n)} = 0$.

The one-dimensional couple-displacement relations are obtained by inserting (6) into (3) and then, in turn, into (12). Thus

$$M^{(m)}_1 = -\frac{2b^3}{3}\left(2\gamma_{11}v^{(m)}_{2,11} + \frac{2}{c}\gamma_{55}\sum^{o} E_{mn'}v^{(n)}_{2,1} + \frac{1}{c^2}\gamma_{13}\sum^{e} E_{mn''}v^{(n)}_2\right),$$

$$M^{(m)}_5 = -\frac{2b^3}{3}\left(\gamma_{51}\sum^{o} E_{m'n}v^{(n)}_{2,11} + \frac{2}{c}\gamma_{55}\sum^{e} E_{m'n'}v^{(n)}_{2,1} + \frac{1}{c^2}\gamma_{53}\sum^{o} E_{m'n''}v^{(n)}_2\right),$$

$$M^{(m)}_3 = -\frac{2b^3}{3}\left(\frac{2}{c^2}\gamma_{33}k^4_m v^{(m)}_2 + \gamma_{31}\sum^{e} E_{m''n}v^{(n)}_{2,11} + \frac{2}{c}\gamma_{35}\sum^{o} E_{m''n'}v^{(n)}_{2,1}\right), \tag{13}$$

where

$$\begin{aligned} E_{mn'} &= \int_{-1}^{1}\phi_m\phi'_n\,d\eta\,, \\ E_{mn''} &= \int_{-1}^{1}\phi_m\phi''_n\,d\eta\,, \\ E_{m'n'} &= \int_{-1}^{1}\phi'_m\phi'_n\,d\eta\,, \quad \text{etc.} \\ \sum^{o} &= \sum_{m+n=\text{odd}}\,, \qquad \sum^{e} = \sum_{m+n=\text{even}}\,. \end{aligned} \tag{14}$$

We note that $M^{(0)}_5 = 0$, $M^{(0)}_3 = M^{(1)}_3 = 0$ due to $\phi'_0 = 0$ and $\phi''_0 = \phi''_1 = 0$. The values of the integration constants $E_{mn'}$ etc. in (14) and the values of k_n, roots of (8), have been studied in detail and numerical values are computed and tabulated.[4] These values are independent of material properties of the plate.

Upon substituting (13) into (1), we obtain the one-dimensional displacement equations of flexural motion

$$\begin{aligned} &\gamma_{11}v^{(m)}_{2,1111} + \frac{1}{c}\gamma_{15}\sum^{o}(E_{mn'} - E_{m'n})v^{(n)}_{2,111} \\ &+ \frac{1}{2c^2}\sum^{e}[\gamma_{13}(E_{mn''} + E_{m''n}) - 4\gamma_{55}E_{m'n'}]v^{(n)}_{2,11} \\ &+ \frac{1}{2c^3}\,35\sum^{o}(E_{m''n'} - E_{m'n''})v^{(n)}_{2,1} + \frac{1}{c^4}\gamma_{33}k^4_m v^{(m)}_2 \\ &+ \frac{3}{b^2}v^{(m)}_2 = \frac{3}{4b^3}(q^{(m)} + V^{(m)})\,; \qquad m=0,1,2,\ldots \end{aligned} \tag{15}$$

Equations (11), (13) and (15) form an infinite system of one-dimentional equations to formally replace the two-dimensional equations (1) - (4).

Second-Order Equations for Flexure

In (15), the symmetric displacement $v^{(n)}_2$, $n=0,2,4\ldots$ and the anti-symmetric dispalcement $v^{(n)}_2$, $n=1,3,5\ldots$ are coupled through the material constants γ_{15} and γ_{35}. In order to

extract the first two equations from (15) which govern the symmetric flexural motion (or deformation) of the plate, we adopt the truncation procedure:

(1) Set $\gamma_{15} = \gamma_{35} = 0$.

(2) Retain $v_2^{(0)}$ and $v_2^{(2)}$;
Let $v_2^{(n)} = 0$ $n > 2$ and $n =$ even.

(3) Set $M_1^{(n)} = M_5^{(n)} = M_3^{(n)} = 0$
for $n > 2$ and $n =$ even.

Thus we have the approximate, one-dimensional, second-order equations for flexural motion of crystal plates as follows.

Displacement equations of Motion

$$\gamma_{11} v^{(0)}_{2,1111} + \frac{1}{c^2}\gamma_{13} E_{02"} v^{(2)}_{2,11} + \frac{3\rho}{b^2}\ddot{v}_2^{(0)} = \frac{3}{4b^3} q^{(0)},$$

$$\gamma_{11} v^{(2)}_{2,1111} + \frac{1}{2c^2} E_{02"}\gamma_{13} v^{(0)}_{2,11} + \frac{1}{c^2}(E_{2"2}\gamma_{13} - 2E_{2'2'}\gamma_{55}) v^{(2)}_{2,11} + \frac{1}{c^4}\gamma_{33} k_2^4 v_2^{(2)} + \frac{3\rho}{b^2}\ddot{v}_2^{(2)} = \frac{3}{4b^3} q^{(2)} . \quad (16)$$

Couple-Displacement Relations

$$M_1^{(0)} = -\frac{2b^3}{3}\left(2\gamma_{11} v^{(0)}_{2,11} + \frac{1}{c^2}\gamma_{13} E_{02"} v_2^{(2)}\right),$$

$$M_1^{(2)} = -\frac{2b^3}{3}\left(2\gamma_{11} v^{(2)}_{2,11} + \frac{1}{c^2}\gamma_{13} E_{22"} v_2^{(0)}\right),$$

$$M_5^{(2)} = -\frac{2b^3}{3}\frac{2}{c}\gamma_{55} E_{2'2'} v^{(2)}_{2,1}, \quad (17)$$

$$M_3^{(2)} = -\frac{2b^3}{3}\left(\gamma_{13} E_{2"0} v^{(0)}_{2,11} + \frac{2}{c^2}\gamma_{33} k_2^4 v_2^{(2)}\right).$$

In (16), traction-free edges at $x_3 = \pm c$ have been taken into account by setting $V^{(0)} = V^{(2)} = 0$, and in (17), we recall that $M_5^{(0)} = M_3^{(0)} = 0$. The values of k_2 and integration constants defined in (14) are

$$k_2 = 2.36502, \quad E_{02"} = E_{2"0} = 9.29455,$$
$$E_{22"} = E_{2"2} = -6.15131, \quad E_{2'2'} = 24.7404 . \quad (18)$$

At the edges $x_1 = \pm a$ of the plate, we require the specification of

(1) either $M_1^{(0)}$ or $v^{(0)}_{2,1}$,
(2) either $M_1^{(2)}$ or $v^{(2)}_{2,1}$,
(3) either $M^{(0)}_{1,1} - \frac{2}{c} M_5^{(0)}$ or $v_2^{(0)}$, (19)
(4) either $M^{(2)}_{1,1} - \frac{2}{c} M_5^{(2)}$ or $v_2^{(2)}$.

Once we have solutions $v_2^{(0)}$ and $v_2^{(2)}$ which satisfy the one-dimensional equations (16) and boundary conditions specified according to (19), we are able to obtain the two-dimensional displacement field approximately from (6)

$$u_2^{(0)}(x_1, x_3, t) = v_2^{(0)}(x_1, t) + v_2^{(2)}(x_1, t)\phi_2(x_3) \quad (20)$$

and, by substituting (20) into (2) and (3), the two-dimensional stress couples and stress resultants.

Cantilever Rectangular Plates Subject to Transverse Loadings

Close form solutions of (16) are obtained for cantilever crystal plates subject to two kinds of transverse loading:

(A) Transverse shear loading distributed uniformly over the plate edge parallel to the support.

(B) Uniformly distributed transverse loading over the entire plate.

Since both problems (A) and (B) are static, $\ddot{v}_2^{(0)} = \ddot{v}_2^{(2)} = 0$, and (16) may be written

$$c^4 v^{(0)}_{2,1111} + c^2 Q_2 v^{(2)}_{2,11} = F_0 ,$$

$$c^4 v^{(2)}_{2,1111} + c^2 Q_1 v^{(0)}_{2,11} + c^2 Q_2 v^{(2)}_{2,11} + Q_3 v_2^{(2)} = F_2 , \quad (21)$$

where

$$Q_1 = \bar{\gamma}_{13}E_{22}'' - 2\bar{\gamma}_{55}E_{2'2'} ,$$

$$Q_2 = \frac{1}{2}\bar{\gamma}_{13}E_{02}'' , \qquad (22)$$

$$Q_3 = \bar{\gamma}_{33}k_2^4 , \quad \bar{\gamma}_{ij} = \gamma_{ij}/\gamma_{11} ,$$

$$F_0 = \frac{3c^4 q^{(0)}}{4b^3 \gamma_{11}} , \quad F_2 = \frac{3c^4 q^{(2)}}{4b^3 \gamma_{11}} .$$

Problem (A)

Since there is no applied load over the plate, q must be zero. Then from the third equation of (12), we have $q^{(0)} = q^{(2)} = 0$ and from (22), $F_0 = F_2 = 0$ in (21).

Let the uniform shear loading intensity at edge $x_1 = +a$ be q^*, which has dimension of force per unit length. Then for the cantilever plate with edge shear, we require, according to (19),

$$v_2^{(0)} = v_2^{(2)} = v_{2,1}^{(0)} = v_{2,1}^{(2)} = 0 \quad \text{at} \quad x_1 = -a$$

$$M_1^{(0)} = M_1^{(2)} = 0, \quad M_{1,1}^{(0)} = -2q^* , \qquad (23)$$

$$M_{1,1}^{(2)} - \frac{2}{c}M_5^{(2)} = 0 , \quad \text{at} \quad x_1 = +a$$

We assume the solution form

$$u_2^{(0)} = A e^{\lambda x_1/c} , \quad v_2^{(2)} = B e^{\lambda x_1/c} . \qquad (24)$$

Equations (16) are satisfied by (24), provided

$$\begin{bmatrix} \lambda^4 & Q_2\lambda^2 \\ Q_2\lambda^2 & \lambda^4 + Q_1\lambda^2 + Q_3 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = 0 \qquad (25)$$

The vanishing of the determinant of the coefficient matrix of (25) leads to four repeated zero roots of λ and four non-zero roots λ_p, p=1,2,3,4 which satisfy

$$\lambda_p^4 + Q_1\lambda_p^2 + (Q_3 - Q_2^2) = 0; \quad p=1,2,3,4 . \qquad (26)$$

Therefore, we may express the general solution by

$$v_2^{(0)} = \sum_{p=1}^{4} A_p e^{\lambda_p r\bar{x}_1} + A_5\bar{x}_1^3 + A_6\bar{x}_1^2 + A_7\bar{x}_1 + A_8 ,$$

$$v_2^{(2)} = \sum_{p=1}^{4} \alpha_p A_p e^{\lambda_p r\bar{x}_1} + \alpha_5 A_5\bar{x}_1 + \alpha_5 A_6 , \qquad (27)$$

where

$$\alpha_p \equiv \frac{B_p}{A_p} = -\frac{\lambda_p^2}{Q_2} ; \qquad p=1,2,3,4$$

$$\alpha_5 = -\frac{6Q_2}{Q_3} r^{-2} , \quad \alpha_6 = -\frac{2Q_2}{Q_3} r^{-2} , \quad r=a/c . \qquad (28)$$

Substitution of (27) into (23), with the aid of (17), leads to a system of eight nonhomogeneous equations for the eight amplitudes A_j, j=1,2,...,8, which may be written in the matrix notation

$$[M_{ij}][A_j] = [B_i]; \qquad i,j=1,2,3,\ldots,8, \qquad (29)$$

in which the non-zero elements are given below.

$$M_{1q} = e^{-\lambda_q r} , \quad M_{2q} = \alpha_q e^{-\lambda_q r} , \quad M_{3q} = \lambda_q e^{-\lambda_q r} ,$$

$$M_{4q} = \alpha_q\lambda_q e^{-\lambda_q r} , \quad M_{5q} = (2\lambda_q^2 + \bar{\gamma}_{13}E_{02}''\alpha_q)e^{\lambda_q r} ,$$

$$M_{6q} = (2\alpha_q\lambda_q^2 + \bar{\gamma}_{13}E_{22}''\alpha_q) e^{\lambda_q r} ,$$

$$M_{7q} = (2\lambda_q^3 + \bar{\gamma}_{13}E_{02}''\alpha_q\lambda_q) e^{\lambda_q r} ,$$

$$M_{8q} = [2\alpha_q\lambda_q^3 + (\bar{\gamma}_{13}E_{22}'' - 4\bar{\gamma}_{55}E_{2'2'})\alpha_q\lambda_q] e^{\lambda_q r} ,$$

$$M_{15} = -1 , \quad M_{16} = 1 , \quad M_{17} = -1 , \quad M_{18} = 1 ,$$

$$M_{25} = -\alpha_5 , \quad M_{26} = \alpha_6 , \quad M_{35} = 3r^{-1}, \quad M_{36} = -2r^{-1},$$

$$M_{37} = r^{-1} , \quad M_{45} = \alpha_5 r^{-1} , \quad M_{55} = (12r^{-2} + \bar{\gamma}_{13}E_{02}''\alpha_5),$$

$$M_{56} = (4r^{-1} + \bar{\gamma}_{13}E_{02}''\alpha_6), \quad M_{65} = \bar{\gamma}_{13}E_{22}''\alpha_5 ,$$

$$M_{66} = \bar{\gamma}_{13}E_{22}''\alpha_6 , \quad M_{75} = (12r^{-3} + \bar{\gamma}_{13}E_{02}''\alpha_5 r^{-1}) ,$$

$$M_{85} = (\bar{\gamma}_{13}E_{22}'' - 4\bar{\gamma}_{55}E_{2'2'})\alpha_5 r^{-1} ,$$

$$B_7 = 3q^*c^3/\gamma_{11}b^3 ; \qquad q=1,2,3,4 . \qquad (30)$$

A rectangular AT-cut quartz plae with $r = a/c = 1.50$, shown in Fig. 1, is chosen for the study. Let x_1' axis be the digonal axis of quartz. The material constants γ_{ij} referred to the plate axes x_i must be functions of the azimuth angle ψ according to the rule of tensor transformation between c_{ij} and c_{ij}' of the elastic stiffness coefficients of quartz.

For a given edge shear intensity q^*, amplitudes A_j can be calculated from (29), and $v_2^{(0)}$ and $v_2^{(2)}$ from (27). Then the two-dimensional displacement $u_2^{(0)}$ is obtained from (20), and stress couples and stress resultants from (2) and (3), respectively, with ψ regarded as a parameter.

For the visualization of their two dimensional variations in the plate, the displacement and stresses are computed and plotted along the x_1 and x_3 axes in dimensionless quantities and for various values of ψ. Displacement $u_2^{(0)}$ are shown in Figs. 2 and 3, stress couple $T_1^{(1)}$ in Figs. 4 and 5, $T_3^{(1)}$ in Figs. 6 and 7, $T_5^{(1)}$ in Fig. 8, shear stress resultant $T_6^{(0)}$ in Figs. 9 and 10, and $T_4^{(0)}$ in Fig. 11. We note that both $T_5^{(1)}$ and $T_4^{(0)}$ vanish along the x_1 axis.

Problem (B)

When the plate is subject to a uniform load over the entire plate, then the loading intensity q becomes a constant. Thus, from (12) we get

$$q^{(0)} = 2q, \qquad q^{(2)} = 0, \tag{31}$$

for $\int_{-1}^{1} \phi_2 d\eta = 0$. Therefore in (21) and (22), we have

$$F_0 = \frac{3c^4 q}{2b^3 \gamma_{11}}, \qquad F_2 = 0. \tag{32}$$

Since there is no applied edge force, the boundary conditions (23) reduce to

$$v_2^{(0)} = v_2^{(2)} = v_{2,1}^{(0)} = v_{2,1}^{(2)} = 0 \quad \text{at} \quad x_1 = -a$$

$$M_1^{(0)} = M_1^{(2)} = M_{1,1}^{(0)} = M_{1,1}^{(2)} - \frac{2}{c} M_5^{(2)} = 0 \tag{33}$$

$$\text{at} \quad x_1 = +a.$$

We may regard that the solution of (21) in this case as the sum of the homogeneous solution (when $F_0 = F_2 = 0$) and a particular solution for F_0, F_2 given in (32). The homogeneous solution is identical to (27). The particular solution is obtained in a polynomial form

$$\bar{v}_2^{(0)} = F_0 P_1 r^4 \bar{x}_1^4, \quad \bar{v}_2^{(2)} = F_0 (P_2 r^2 \bar{x}_1^2 + P_3), \tag{34}$$

where

$$P_1 = -\frac{Q_3}{24(Q_2^2 - Q_3)},$$

$$P_2 = \frac{Q_2}{Q_2^2 - Q_3}, \tag{35}$$

$$P_3 = \frac{Q_1 Q_2}{Q_3(Q_2^2 - Q_3)}.$$

We note that the particular solution is completely determined by the applied load F_0 and the coefficients of the governing equation Q_1, Q_2, and Q_3 defined in (22).

By adding (34) to (27), we have the complete solution. Thus

$$v_2^{(0)} = F_0 \Big(\sum_{p=1}^{4} A_p e^{\lambda_p r \bar{x}_1} + A_5 \bar{x}_1^3 + A_6 \bar{x}_1^2 + A_7 \bar{x}_1 + A_8 + P_1 r^4 \bar{x}_1^4 \Big),$$

$$v_2^{(2)} = F_0 \Big(\sum_{p=1}^{4} \alpha_p e^{\lambda_p r \bar{x}_1} + \alpha_5 A_5 \bar{x}_1 + \alpha_6 A_6 + P_2 r^2 \bar{x}_1^2 + P_3 \Big). \tag{36}$$

Again, the requirement of boundary conditions (33) results in a system of eight equations on eight unknowns A_j, $j=1,2,3,\ldots,8$

$$[M_{ij}][A_j] = [B_j'], \tag{37}$$

where the elements M_{ij} are identical to those defined in (30), while the elements B_j' are given as follows:

$$
\begin{aligned}
B_1' &= -P_1 r^4 , \qquad B_2' = -P_2 r^2 , \\
B_3' &= 4 P_1 r^3 , \qquad B_4' = 2 P_2 r , \\
B_5' &= -24 P_1 r^2 - \bar{\gamma}_{13} E_{02''} (P_2 r^2 + P_3) , \\
B_6' &= -4 P_2 - \bar{\gamma}_{13} E_{22''} (P_2 r^2 + P_3) , \\
B_7' &= -48 P_1 r - 2 \bar{\gamma}_{13} E_{02''} P_2 r , \\
B_8' &= (4 \bar{\gamma}_{55} E_{2'2'} - \bar{\gamma}_{13} E_{22''}) P_2 r .
\end{aligned} \tag{38}
$$

We see that (24) and (37) have identical form; they differ only in B_j and B_j'.

Once A_j are calculated from (37), $v_2^{(0)}$ and $v_2^{(2)}$ can be obtained from (36). Then the two-dimensional displacement $u_2^{(0)}$ can be calculated from (20), and stresses from (2) and (3).

To check the accuracy of our solution, we have computed displacement and stresses for isotropic elastic plates by converting the material constants through (5), since many existing solutions by various different methods of approximation are available only for isotropic plates.

First, a square plate ($r = a/c = 1$) with Poisson's ratio $\nu = 0.3$ is chosen. The displacement $u_2^{(0)}$ along the free edge $x_1 = a$ and the bending stress $T_1^{(1)}$ along the clamped edge $x_1 = -a$ are computed as functions of x_3/c, and they are listed in Tables I and II, respectively, together with the results by Chang [2] and by Wu (F.E.M.).

A second comparison is made for a rectangular plate ($r = a/c = 0.5$) and $\nu = 0.3$. In a similar manner, the present results are compared with those by Nash[1] and Chang[2] as shown in Tables III and IV.

It can be seen that the agreement among the results by various different methods is close.

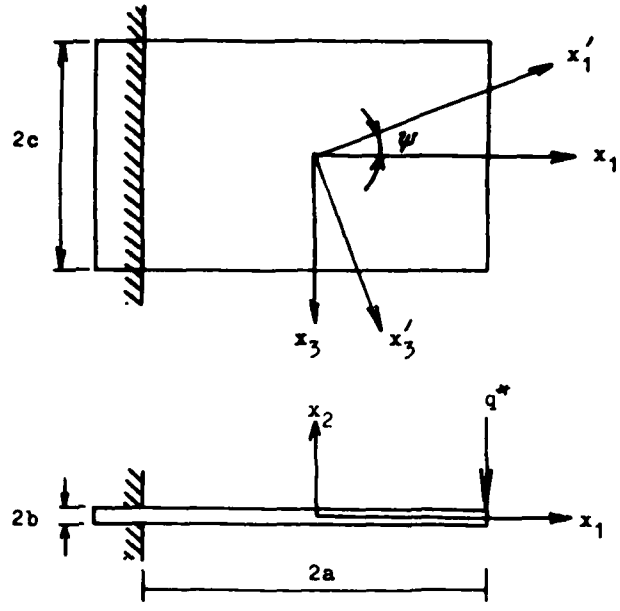


Fig. 1 A Rectangular Cantilever Plate Referred to x_i Coordinate System, x_i' Referred to Axes of Crystal Symmetry, and ψ the Azimuth Angle

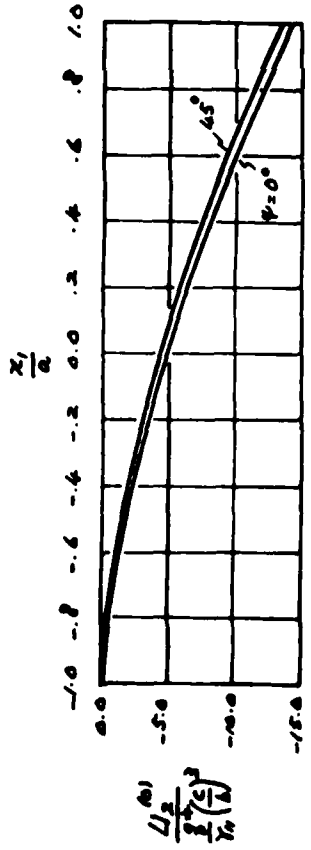


Fig. 2 Displacement $u_2^{(0)}$ Along the x_1 Axis of a Rectangular AT-Cut of Quartz Plate ($r = a/c = 1.50$)

References

1. W.A. Nash, J. Appl. Mech., Vol. 19, pp. 33-36, [1952].
2. Chang Fo-van, Appl. Math. Mech., Vol. 1, No. 3, pp. 371-383 [1981] (English translation of a Chinese Journal in Appl. Math. Mech., Tsing Hua University).
3. P.C.Y. Lee, Y.S. Wang and X. Markenscoff, J. Acoust. Soc. Am., Vol. 59, No. 1, pp. 90-96 [1976].
4. P.C.Y. Lee, "Flexural Vibrations of a Strip," presented at the 15th IUTAM, Toronto, Canada [1980].

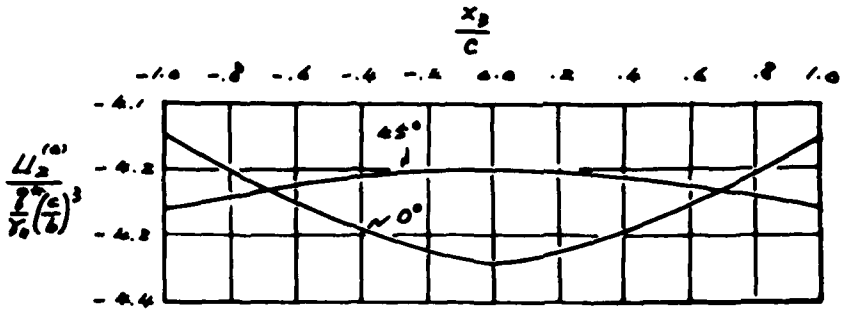


Fig. 3 Displacement $u_2^{(0)}$ Along the x_3 Axis of the Plate

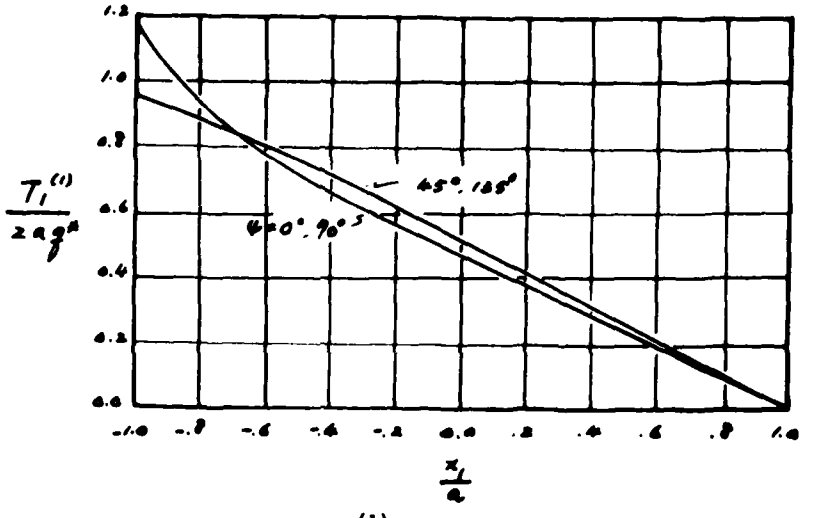


Fig. 4 Bending Couple $T_1^{(1)}$ Along the x_1 Axis of the Plate

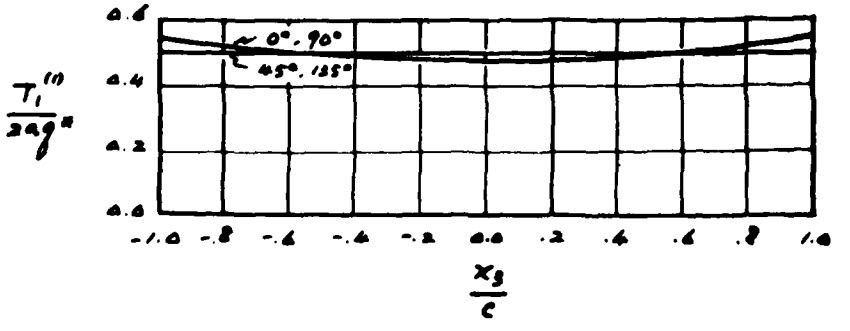


Fig. 5 Bending Couple $T_1^{(1)}$ Along the x_3 Axis of the Plate

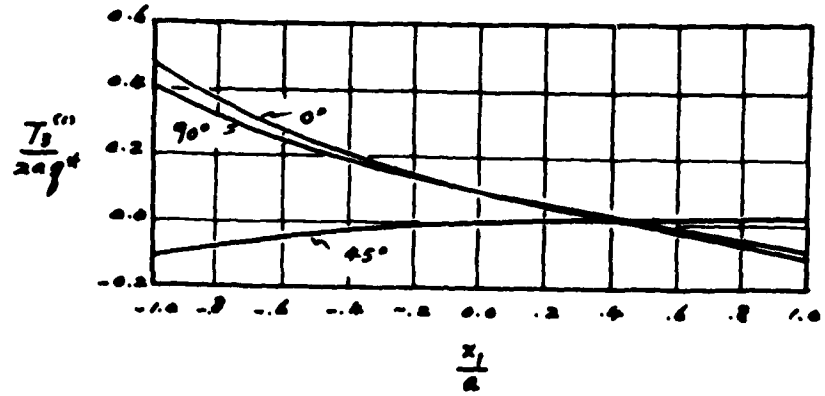


Fig. 6 Bending Couple $T_3^{(1)}$ Along the x_1 Axis of the Plate

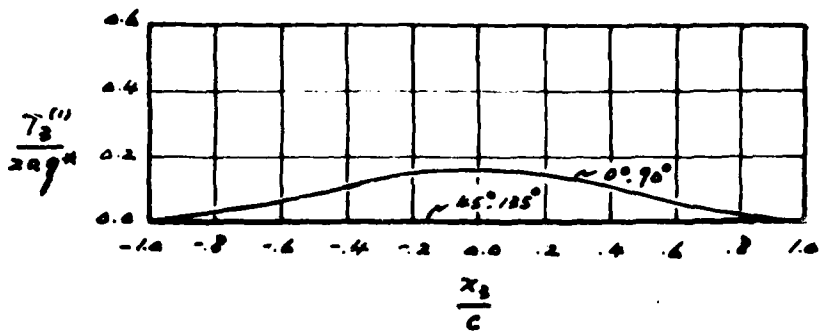


Fig. 7 Bending Couple $T_3^{(1)}$ Along the x_3 Axis of the Plate

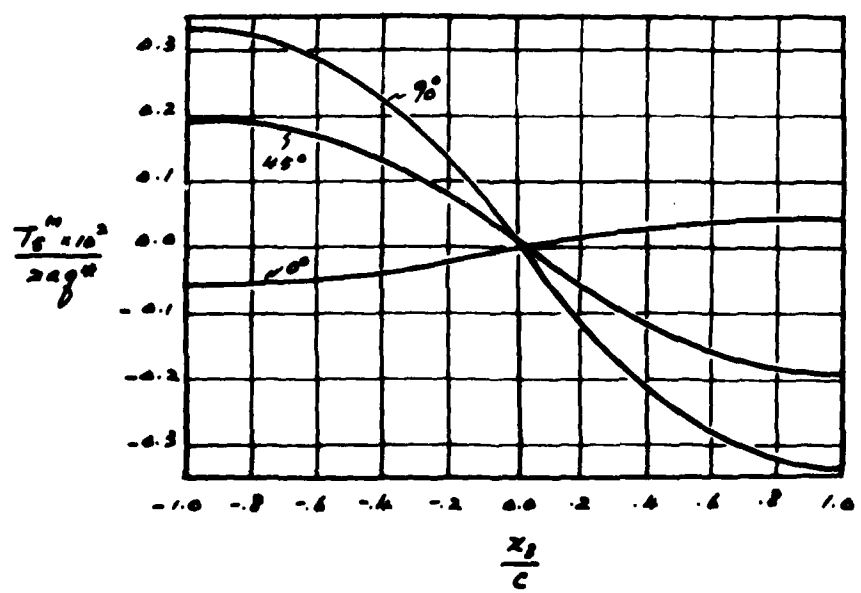


Fig. 8 Twisting Couple $T_5^{(1)}$ Along the x_3 Axis of the Plate

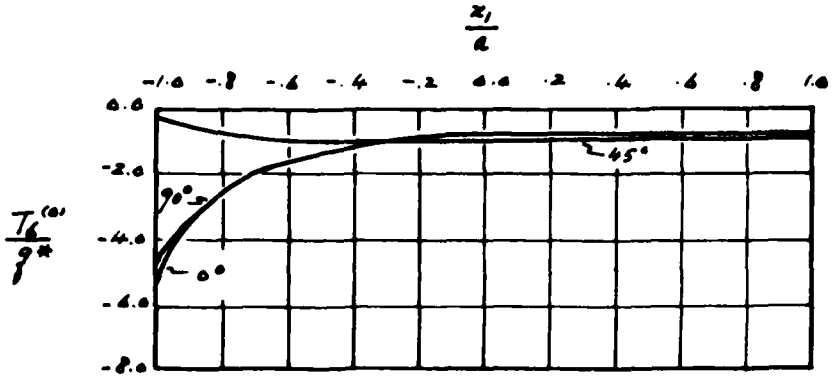


Fig. 9 Transverse Shear $T_6^{(0)}$ Along the x_1 Axis of the Plate

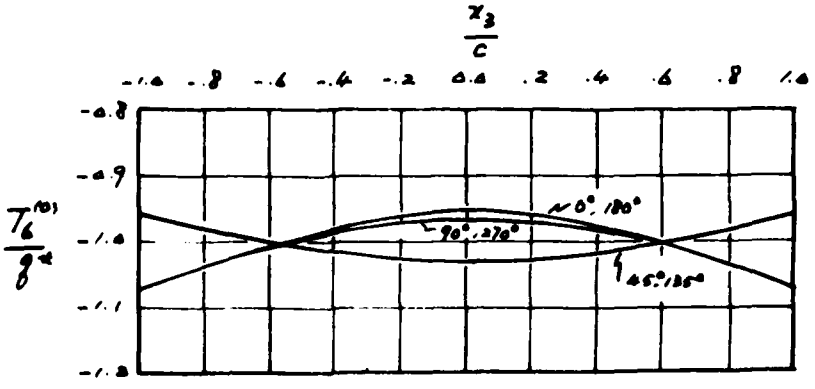


Fig. 10 Transverse Shear $T_6^{(0)}$ Along the x_3 Axis of the Plate

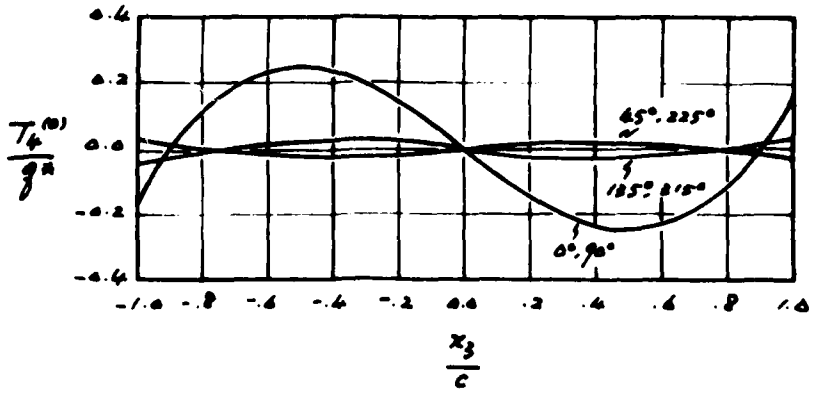


Fig. 11 Transverse Shear $T_4^{(0)}$ Along the x_3 Axis of the Plate

Table I Comparison of displacement $u_2^{(0)}$ along the free edge $x_1 = +a$ for a square plate ($r = a/c = 1$, $\nu = 0.3$).

$x_{3/c}$	0.00	0.25	0.50	0.75	1.00
Chang	$u_2^{(0)} = 0.13102 \frac{q(2a)^4}{D}$	0.13091	0.13056	0.12998	0.12933
F.E.M.	0.12905	0.12892	0.12851	0.12788	0.12708
present	0.1270	0.1268	0.1264	0.1258	0.1252

Table II Comparison of bending couple $T_1^{(1)}$ along the clamped edge $x_1 = -a$ for a square plate ($r = a/c = 1$, $\nu = 0.3$).

$x_{3/c}$	0.00	0.25	0.50	0.75	1.00
Chang	$T_1^{(1)} = -0.53560q(2a)^2$	-0.53550	-0.53353	-0.51270	0.0
F.E.M.	-0.53092	-0.53058	-0.52760	-0.50399	-0.34571
present	-0.5602	-0.5468	-0.5098	-0.4580	-0.4009

Table III Comparison of displacement $u_2^{(0)}$ along the free edge $x_1 = +a$ for a rectangular plate ($r = a/c = 0.5$, $\nu = 0.3$).

$x_{3/c}$	0.00	0.25	0.50	0.75	1.00
Chang	$u_2^{(0)} = 0.12837 \frac{q(2a)^4}{D}$	0.12825	0.12784	0.12691	0.12540
Nash	0.141		0.139		0.135
present	0.1267	0.1264	0.1256	0.1244	0.1232

Table IV Comparison of bending couple $T_1^{(1)}$ along the clamped edge $x_1 = -a$ for a rectangular plate ($r = a/c = 0.50$, $\nu = 0.3$)

$x_{3/c}$	0.00	0.25	0.50	0.75	1.00
Chang	$T_1^{(1)} = -0.51049q(2a)^2$	-0.51451	-0.51386	-0.51074	0.0
Nash	-0.5082		-0.5047		-0.4824
present	-0.5310	-0.5241	-0.5051	-0.4784	-0.4489

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

AN ANALYSIS OF SC-CUT QUARTZ TRAPPED ENERGY RESONATORS WITH RECTANGULAR ELECTRODES

D.S. Stevens and H.F. Tiersten
Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, New York 12181

Abstract

An approximate dispersion equation describing the mode shape along the surface of a doubly-rotated SC-cut quartz plate vibrating in the vicinity of the fundamental and odd overtone essentially thickness-shear frequencies of interest is obtained from the equations of three-dimensional linear piezoelectricity. The influence of piezoelectric stiffening, electrode mass loading and electrical shorting is included in the treatment. The dispersion relations in both the electroded and unelectroded regions of the plate are obtained along with simple approximate edge conditions to be satisfied at a junction between the two regions. The dispersion relations and edge conditions are applied in the analysis of trapped energy resonators with rectangular electrodes on SC-cut quartz plates and a lumped parameter representation of the admittance, which is valid in the vicinity of a resonance, is obtained. In addition, the two-dimensional generalization of Bechmann's number in one dimension is provided.

1. Introduction

Previous analytical work on trapped energy resonators with rectangular electrodes holds only for rotated Y-cuts of quartz[1,2]. In recent years the doubly-rotated SC-cut[3,4] has given every indication of possessing significant technological advantages[5,6] over the AT-cut. On account of the additional anisotropy of the SC-cut, the existing analytical treatment[1] of trapped energy resonators does not hold. In this paper an analysis of SC-cut quartz trapped energy resonators is performed.

The analysis proceeds by decomposing the mechanical displacement vector along[7] the orthogonal eigenvector triad of the pure thickness solution and transforming the independent spatial variables to a Cartesian coordinate system containing the plate normal and the component in the plane of the plate of the large thickness eigen-displacement of interest. The Cartesian coordinate system employed determines the most desirable orientation for rectangular electrodes on SC-cut quartz plates. Since the piezoelectric coupling is small in quartz and we are interested in obtaining the dispersion relation in the vicinity of the thickness-frequencies of interest for small wavenumbers along the plate, only the thickness-dependence of all electrical variables is included in the treatment. A number of relatively small unimportant transformed elastic constants that cause coupling to the third component of thickness eigen-displacement are neglected[8] along with those that cause coupling of the symmetric to the antisymmetric solution for the plate. The resulting system of transformed differential equations and boundary conditions on the major surfaces of the plate are employed in the determination of the asymptotic dispersion relation in the vicinity of each thickness-frequency of interest using a procedure analogous to the one used in the case of rotated Y-cut quartz[1,2]. Both the dispersion relation in the electroded region and slightly different one in the unelectroded region are obtained along with simple approximate edge conditions to be satisfied at a junction between the two regions.

The above-mentioned dispersion relations and edge conditions are applied in the analysis of trapped energy resonators with rectangular electrodes on SC-cut quartz plates and a lumped parameter representation of the admittance, which is valid in the vicinity of a resonance, is obtained. In addition, the two-dimensional generalization of Bechmann's number in one dimension is provided. Some numerical results obtained from the analysis are given.

2. Transformation of Equations

The stress equations of motion and charge equation of electrostatics may be written in the form

$$\hat{T}_{mn,m} = \rho \ddot{\hat{u}}_n, \quad D_{m,m} = 0. \tag{2.1}$$

The linear piezoelectric constitutive equations are given by

$$\hat{T}_{mn} = \hat{c}_{mnrs}\hat{u}_{r,s} + \hat{e}_{rmn}\varphi_{,r}, \tag{2.2}$$

$$D_m = \hat{e}_{mrs}\hat{u}_{r,s} - \epsilon_{mn}\varphi_{,n}. \tag{2.3}$$

The notation is conventional[9,10] and carets have been introduced over the mechanical variables and the elastic and piezoelectric constants because we are going to transform from the Cartesian coordinate system presently being employed to a combination of two very special Cartesian coordinate systems that facilitate the analysis. Moreover, in the original coordinate system we consistently use the tensor indices m, n, r, s. The substitution of (2.2) and (2.3) in (2.1) yields

$$\hat{c}_{mnrs}\hat{u}_{r,sm} + \hat{e}_{rmn}\varphi_{,rm} = \rho\ddot{\hat{u}}_n , \tag{2.4}$$

$$\hat{e}_{mrs}\hat{u}_{r,sm} - \epsilon_{mn}\varphi_{,nm} = 0 , \tag{2.5}$$

which are the differential equations of linear piezoelectricity.

The coordinate axes are oriented with $\hat{x}_2$ normal to the major surfaces of the plate, which are at $\hat{x}_2 = \pm h$, and $\hat{x}_1$ directed along the axis of second rotation of the doubly-rotated quartz plate. Since in the modes of interest the spatial rate of variation of the dependent variables in the plane of the plate is much less than in the thickness direction and the piezoelectric coupling is small, we can ignore the $\hat{x}_w$ dependence of φ and D_v and of any variable in the expression for D_2, and in place of Eqs. (2.3) and (2.5), respectively, write

$$D_2 = \hat{e}_{2r2}\hat{u}_{r,2} - \epsilon_{22}\varphi_{,2} , \tag{2.6}$$

$$\hat{e}_{2r2}\hat{u}_{r,22} - \epsilon_{22}\varphi_{,22} = 0 , \tag{2.7}$$

and we have introduced the convention that the subscripts u, v, w take the values 1 and 3, but skip 2. Moreover, it is advantageous for us to write (2.2) and (2.4), respectively, in the forms

$$\hat{T}_{mn} = \hat{c}_{mnr2}\hat{u}_{r,2} + \hat{e}_{2mn}\varphi_{,2} + \hat{c}_{vrnm}\hat{u}_{r,v} , \tag{2.8}$$

$$\hat{c}_{2nr2}\hat{u}_{r,22} + \hat{e}_{22n}\varphi_{,22} + (\hat{c}_{2rnv} + \hat{c}_{2nrv})\hat{u}_{r,2v} + \hat{c}_{vnrw}\hat{u}_{r,wv} = \rho\ddot{\hat{u}}_n , \tag{2.9}$$

since we are interested in modes in the vicinity of the frequencies of thickness vibration. Substituting from (2.7) into (2.9), we obtain

$$\bar{\hat{c}}_{2nr2}\hat{u}_{r,22} + (\hat{c}_{2rnv} + \hat{c}_{2nrv})\hat{u}_{r,2v} + \hat{c}_{vnrw}\hat{u}_{r,wv} = \rho\ddot{\hat{u}}_n , \tag{2.10}$$

where

$$\bar{\hat{c}}_{2nr2} = \hat{c}_{2nr2} + \frac{\hat{e}_{22n}\hat{e}_{22r}}{\epsilon_{22}} , \tag{2.11}$$

are the usual piezoelectrically stiffened elastic constants.

For thickness vibrations only x_2-dependence is considered and (2.10) takes the form

$$\bar{\hat{c}}_{2nr2}\hat{u}_{r,22} = \rho\ddot{\hat{u}}_n . \tag{2.12}$$

In the case of the unelectroded plate the antisymmetric thickness vibration solution can be written in the form

$$u_r = A_r \sin \eta x_2 \, e^{i\omega t} , \tag{2.13}$$

which satisfies (2.12) provided

$$(\bar{\hat{c}}_{2nr2} - \bar{c}\,\delta_{nr})A_r = 0 , \tag{2.14}$$

where

$$\bar{c} = \rho\omega^2/\eta^2 . \tag{2.15}$$

For a nontrivial solution the determinant of the coefficients of A_r in (2.14) must vanish, i.e.,

$$|\bar{\hat{c}}_{2nr2} - \bar{c}\delta_{nr}| = 0 , \tag{2.16}$$

which yields a cubic equation in $\bar{c}$. Since the $\hat{c}_{2sr2}$ constitute a symmetric, positive-definite matrix[11], Eq. (2.16) yields three real, positive roots[12] $c^{(i)}$ $(i = 1,2,3)$, which yield mutually orthogonal eigenvectors $A_r^{(i)}$ when substituted in (2.14), and we may write

$$A_r^{(i)}A_r^{(j)} = N_{(i)}^2\delta_{ij} , \tag{2.17}$$

where the $N_{(i)}$ are the normalization factors. If we normalize the $A_r^{(i)}$ thus

$$q_r^{(i)} = A_r^{(i)}/N_{(i)} , \tag{2.18}$$

and write the $q_r^{(i)}$ as a 3×3 orthogonal transformation matrix Q_{ir}, from (2.17) we have the orthogonality relations

$$Q_{ir}Q_{jr} = \delta_{ij} , \tag{2.19}$$

and, of course, the other orthogonality relations

$$Q_{ir}Q_{is} = \delta_{rs} , \tag{2.20}$$

hold.

We may now transform the components of the mechanical displacement vector $\hat{u}_r$ in the original coordinate system to the components u_i in the thickness solution eigenvector coordinate system, thus

$$u_i = Q_{ir}\hat{u}_r , \tag{2.21}$$

and, of course, the inverse relation

$$\hat{u}_r = Q_{ir}u_i , \tag{2.22}$$

holds. A schematic diagram showing the original coordinate system and the eigenvector coordinate system is shown in Fig. 1. For a pure thickness solution for an unelectroded plate, one of the thickness eigen-displacements, say u_1, exists and the other two vanish identically. Moreover, for a mode of vibration in the vicinity of the thickness solution for which u_1 exists, although u_2 and u_3 exist they are one or more orders of magnitude smaller than u_1. This is the reason there is a significant advantage in decomposing the mechanical displacement along the eigenvector triad of the pure thickness solution when describing this type of mode.

Clearly, the independent variables x_s must be referred to a Cartesian coordinate system containing the axis $\hat{x}_2$, which is normal to the major surfaces of the plate, and two orthogonal axes in the plane of the plate. Of all the orthogonal coordinate systems in the plane of the plate, one is particularly

well suited for the description of modes of vibration in the vicinity of the pure thickness solution u_1, when, as is the case in this work, it is dominantly a thickness-shear displacement, as shown in Fig.1. That coordinate system contains the axis formed by the projection of the u_1-eigendisplacement direction on the plane of the plate as shown in Fig.2. From (2.22) the angle θ is given by

$$\theta = \tan^{-1}(-Q_{13}/Q_{11}), \tag{2.23}$$

and the planar rotation matrix R_{av} is given by

$$R_{av} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{2.24}$$

where we have introduced the convention that the subscripts a, b, c, d take the values 1 and 3, but skip 2. Clearly, the planar coordinates x_a and $\hat{x}_v$ are related by

$$x_a = R_{av}\hat{x}_v, \quad \hat{x}_v = R_{av}x_a, \tag{2.25}$$

the first of which yields

$$\partial/\partial\hat{x}_w = R_{aw}\partial/\partial x_a. \tag{2.26}$$

Transforming (2.10) with Q_{ja}, substituting from (2.22) for $\hat{u}_r$ and $\hat{u}_a$ and (2.26) for the partial derivatives with respect to $\hat{x}_v$ and $\hat{x}_w$ and employing (2.14) with (2.18) and (2.19), we obtain

$$\bar{c}^{(j)}u_{j,22} + (c_{aij2} + c_{aji2})u_{i,2a} + c_{ajib}u_{i,ab} = \rho\ddot{u}_j, \tag{2.27}$$

where

$$c_{aij2} = Q_{jn}Q_{ir}R_{av}\hat{c}_{vrn2},$$
$$c_{ajib} = R_{av}Q_{jn}Q_{ir}R_{bw}\bar{c}_{vnrw}. \tag{2.28}$$

In the case of the thickness vibrations of the electroded plate with shorted electrodes the three solutions of the differential equations are coupled[13-15] in the transcendental frequency equation. However, since the piezoelectric coupling is small in quartz and the three $\bar{c}^{(j)}$ are widely separated, the transcendental frequency equation approximately uncouples and we effectively have three independent transcendental frequency equations[11], one for each thickness eigen-displacement u_j. Consequently, (2.27) is a useful form for the treatment of modes of vibration in the vicinity of the pure thickness solution u_j for the electroded as well as the unelectroded plate.

For the electroded plate with shorted electrodes the electrical boundary condition is

$$\varphi = 0 \quad \text{at} \quad x_2 = \pm h. \tag{2.29}$$

Substituting from (2.22) into (2.7), integrating and satisfying (2.29), we obtain

$$\varphi = (e_{22j}/\epsilon_{22})u_j + Cx_2, \tag{2.30}$$

where

$$e_{22j} = Q_{jr}\hat{e}_{22r}, \quad C = -(e_{22j}/\epsilon_{22})u_j(h)/h. \tag{2.31}$$

In the case of the unelectroded plate because of the reasoning leading to (2.6) we can neglect the electric field outside the plate[1] and the electrical boundary condition becomes

$$D_2 = 0 \quad \text{at} \quad x_2 = \pm h. \tag{2.32}$$

Hence, from (2.32) with (2.6), for the unelectroded plate, we obtain

$$\varphi = (e_{22j}/\epsilon_{22})u_j, \tag{2.33}$$

which is the same as (2.30) but with $C = 0$.

The traction boundary conditions on the major surfaces of the plate employ the constitutive equations in (2.8) with $m = 2$. Transforming (2.8) for $m = 2$ with Q_{ja}, substituting from (2.22), (2.26), (2.30), (2.31), and (2.11) and employing (2.14) with (2.18) and (2.19) and $(2.28)_1$, we obtain

$$T_{2j} = \bar{c}^{(j)}u_{j,2} + e_{22j}C + c_{aij2}u_{i,a}, \tag{2.34}$$

which holds for the electroded plate and for the unelectroded plate provided $C = 0$. The traction continuity conditions along the surfaces separating the electroded from the unelectroded regions of the plate are obtained from (2.8) with $m = w$. Transforming (2.8) for $m = w$ with $R_{aw}Q_{ja}$, substituting from (2.22) and (2.26), employing (2.28) and ignoring the piezoelectric coupling terms, which are negligible in the lateral continuity conditions in quartz for the modes of vibration of interest in this work[11], we obtain

$$T_{aj} = c_{aij2}u_{i,2} + c_{ajib}u_{i,b}. \tag{2.35}$$

3. Dispersion Equations

The differential equations expressed in terms of the thickness eigen-displacements in the transformed coordinate system are given in (2.27). The boundary conditions on the major surfaces of the unelectroded plate are

$$T_{2j} = 0 \quad \text{at} \quad x_2 = \pm h, \tag{3.1}$$

where the constitutive equations for T_{2j} expressed in the same way are given in (2.34) with $C = 0$. Since we are interested in the solution for plate waves in the vicinity of the thickness frequencies for the thickness eigen-displacement u_1 with wave, or decay, numbers in the plane of the plate that are much smaller than the thickness wavenumbers and a number of the transformed elastic constants are very small, the differential equations (2.27) and constitutive equations (2.34) and (2.35) can be simplified considerably. Accordingly, we neglect[8] the relatively small unimportant elastic constants c_{15}, c_{17}, c_{86}, c_{85}, c_{87}, c_{82}, c_{51}, c_{56}, c_{14}, c_{59}, c_{58}, c_{76}, c_{72}, c_{16}, c_{36} and c_{57}, the first six of which cause coupling to the u_3 thickness eigen-displacement field in the absence of any x_3

dependence and one of the remainder of which couples the symmetric to the antisymmetric solutions for the plate, and we have introduced the compressed index notation defined in Table I. When the aforementioned transformed small elastic constants are neglected, Eqs. (2.27) and (2.34) reveal[11] that for small wave, or decay, numbers along the plate u_3 is two orders of magnitude smaller than u_1 and, hence, negligible to the lowest order of approximation obtained here. Consequently, we can ignore u_3 and the third equation in each of (2.27), (2.34) and (3.1). Under these circumstances from (2.27) we obtain the pertinent nontrivial differential equations

$$\bar{c}^{(1)}u_{1,22} + (c_{66} + c_{12})u_{2,12} + c_{11}u_{1,11} + c_{58}u_{1,33} = \rho\ddot{u}_1 ,$$

$$\bar{c}^{(2)}u_{2,22} + (c_{12} + c_{66})u_{1,21} = \rho\ddot{u}_2 , \qquad (3.2)$$

where we have omitted writing the terms $c_{69}u_{2,11}$, $c_{74}u_{2,33}$, $2c_{62}u_{2,12}$, $c_{61}u_{1,11}$ and $c_{54}u_{1,33}$ in $(3.2)_2$ because they are at least an order of magnitude smaller than the smallest term retained in this treatment and, hence, negligible to the order of approximation being obtained and we have employed the relation

$$c_{ajib} = c_{bija} , \qquad (3.3)$$

which is a consequence of (2.28) and the standard symmetries of the elastic constants $\hat{c}_{vsrw}$. Under the circumstances outlined from (3.1) and (2.34) for the unelectroded plate, we obtain the nontrivial boundary conditions

$$\bar{c}^{(1)}u_{1,2} + c_{66}u_{2,1} = 0 \quad \text{at } x_2 = \pm h ,$$

$$\bar{c}^{(2)}u_{2,2} + c_{12}u_{1,1} = 0 \quad \text{at } x_2 = \pm h , \qquad (3.4)$$

where we have omitted the term $c_{62}u_{2,1}$ from $(3.4)_2$ because it is an order of magnitude smaller than the other terms in the equation and, hence, negligible in this lowest order approximation. In the case of the electroded plate with shorted electrodes the electrical boundary condition is given by (2.29) and under the circumstances outlined the nontrivial mechanical boundary conditions on the major surfaces are given by

$$T_{21} = \mp 2\rho' h' \ddot{u}_1 , \quad T_{22} = \mp 2\rho' h' \ddot{u}_2 , \quad \text{at } x_2 = \pm h , \qquad (3.5)$$

where ρ' and $2h'$ are the mass density and thickness of the electrode plating, respectively.

Under the circumstances outlined Eqs. (3.2), (3.4), (2.32) and (3.5), (2.29) are identical in form with Eqs. (1.7), (1.9) and (1.10) of Ref. 2 with the exception of a negligible term included in $(1.7)_2$. Consequently, the solutions presented in Sec. I of Ref. 2 are applicable here without change and we simply present the important pertinent results below. However, it should be carefully noted that the material coefficients appearing in this work have different values than those of Ref. 2 because of the transformed coordinates employed here even though the symbolism is the same.

It has been shown[2] that a typical asymptotic solution to second order in ξ and ν for plate waves in an unelectroded region can be written in the form

$$u_1 = (B_1^{(1)} \sin \eta_1 x_2 + B_1^{(2)} \sin \kappa\eta_1 x_2)e^{-\xi x_1} \cos \nu x_3 e^{i\omega t} ,$$

$$u_2 = (B_2^{(1)} \cos \eta_1 x_2 + B_2^{(2)} \cos \kappa\eta_1 x_2)e^{-\xi x_1} \cos \nu x_3 e^{i\omega t} ,$$

$$\varphi = (e_{26}/\epsilon_{22})u_1 , \qquad (3.6)$$

where

$$B_2^{(1)} = r\xi B_1^{(1)}/\eta_1 , \quad B_1^{(2)} = r\xi B_2^{(2)}/\kappa\eta_1 ,$$

$$B_2^{(2)} = (-1)^{(n+1)/2}(c_{12} + r\bar{c}^{(2)})\xi B_1^{(1)}/\bar{c}^{(2)}\kappa\eta_1 \sin \kappa\eta_1 h ,$$

$$\kappa = (\bar{c}^{(1)}/\bar{c}^{(2)})^{1/2} , \quad r = (c_{12} + c_{66})/(\bar{c}^{(1)} - \bar{c}^{(2)}) ,$$

$$\eta_1 = n\pi/2h , \quad n \text{ odd} , \qquad (3.7)$$

$$M_n\xi^2 - c_{55}\nu^2 + (n\pi\varepsilon/h)(\bar{c}^{(1)}\rho)^{1/2} = 0 , \qquad (3.8)$$

and

$$M_n = c_{11} + (c_{12} + c_{66})r + 4(r\bar{c}^{(1)} - c_{66}) \times (\bar{c}^{(2)}r + c_{12})\cot \tfrac{1}{2}\kappa n\pi/\bar{c}^{(2)}n\pi\kappa , \qquad (3.9)$$

$$\omega = \omega_n + \varepsilon , \quad \omega_n = (n\pi/2h)(\bar{c}^{(1)}/\rho)^{1/2} . \qquad (3.10)$$

Similarly, it has been shown[2] that a typical asymptotic solution to second order in $\bar{\xi}$ and $\bar{\nu}$ for plate waves in an electroded region with shorted electrodes can be written in the form

$$u_1 = (\bar{B}_1^{(1)} \sin \bar{\eta}_1 x_2 + \bar{B}_1^{(2)} \sin \kappa\bar{\eta}_1 x_2) \times \cos \bar{\xi} x_1 \cos \bar{\nu} x_3 \, e^{i\omega t} ,$$

$$u_2 = (\bar{B}_2^{(1)} \cos \bar{\eta}_1 x_2 + \bar{B}_2^{(2)} \cos \kappa\bar{\eta}_1 x_2) \times \sin \bar{\xi} x_1 \cos \bar{\nu} x_3 \, e^{i\omega t} ,$$

$$\varphi = (e_{26}/\epsilon_{22})u_1 + Cx_2 , \quad C = -(e_{26}/\epsilon_{22})u_1(h)/h , \qquad (3.11)$$

where

$$\bar{B}_2^{(1)} = r\bar{\xi}\bar{B}_1^{(1)}/\bar{\eta}_1 , \quad \bar{B}_1^{(2)} = -r\bar{\xi}\bar{B}_2^{(2)}/\kappa\bar{\eta}_1 , \quad \bar{\eta}_1 \approx n\pi/2h ,$$

$$\bar{B}_2^{(2)} = (-1)^{(n+1)/2}(\bar{c}^{(2)}r + c_{12})\bar{\xi}\bar{B}_1^{(1)}/\bar{c}^{(2)}\kappa\bar{\eta}_1 \sin \kappa\bar{\eta}_1 h , \qquad (3.12)$$

$$M_n\bar{\xi}^2 + c_{55}\bar{\nu}^2 - (n\pi\bar{\varepsilon}/h)(\bar{c}^{(1)}\rho)^{1/2} = 0 , \qquad (3.13)$$

$$\omega = \bar{\omega}_n + \bar{\varepsilon} , \quad \bar{\omega}_n = (n\pi/2h)(\bar{c}^{(1)}/\rho)^{1/2}[1 - (4k_{26}^2/n^2\pi^2) - R] ,$$

$$k_{26}^2 = e_{26}^2/\bar{c}^{(1)}\epsilon_{22} , \quad R = 2\rho' h'/\rho h , \qquad (3.14)$$

and we note that $(3.11)_3$ has been obtained from (2.30) and $(2.31)_2$ because $\bar{\xi}$ is small. It has also been shown[2] that for purposes of analysis of a trapped energy resonator operating in the vicinity of the thickness frequencies of overtones associated with the thickness eigen-displacement u_1, the

solutions in the unelectroded and electroded regions, respectively, can be represented in the form

$$u_1 = B_1^{(1)} \sin \eta_1 x_2 e^{-\xi x_1} \cos \nu x_3 e^{i\omega t},$$

$$\bar{u}_1 = \bar{B}_1^{(1)} \sin \bar{\eta}_1 x_2 \cos \bar{\xi} x_1 \cos \bar{\nu} x_3 e^{i\omega t}, \qquad (3.15)$$

where it is understood that (3.8), (3.10) and (3.13), $(3.14)_{1,2}$, respectively, are satisfied.

A schematic diagram of a cross-section and a plan view of one quadrant of a trapped energy resonator are shown in Figs.3 and 4, respectively. It has been shown[2] that at an x_1 junction between an electroded and unelectroded region of the plate, the four existing continuity conditions can be satisfied, approximately, by requiring the continuity of only the two quantities

$$u_1, \quad u_{1,1}, \qquad (3.16)$$

while at an x_3 junction it has been shown[2] that the continuity of only the two quantities

$$u_1, \quad u_{1,3}, \qquad (3.17)$$

is required. It has also been shown[2] that, since ξ and ν are small compared with η, it is more appropriate to satisfy all equations associated with a corner-type region to first order in ξ and ν pointwise than to satisfy (3.8), which contains terms of zero and second order in ξ and ν, pointwise and the continuity conditions at junctions (3.16) and (3.17) each of which has one condition of zero order and one condition of first order in ξ or ν, respectively, as integral conditions only. Consequently, in the interior of a corner-type region we do not satisfy (3.8), but at the junctions between corner-type regions and adjacent regions we do satisfy the continuity conditions consisting of either (3.16) or (3.17) pointwise. This means that in the interior of a corner-type region we satisfy the conditions to first order in ξ and ν, since first order terms do not exist in (3.8), and at junctions between corner-type regions and adjacent regions we satisfy the continuity conditions pointwise to first order in ξ and ν. Thus, all equations associated with the relatively unimportant corner-type regions are satisfied to first order in ξ and ν, while all equations associated with all other regions are satisfied to second order in ξ and ν.

4. Trapped Energy Resonator

In the vicinity of a coupled thickness-shear, thickness-twist trapped energy resonance, say the $N\tau\epsilon$th, one term in the series solution[2] dominates the others and the solution can very accurately be written in the form[11]

$$\bar{u}_1 = H^{N\tau\epsilon} \sin \frac{N\pi}{2h} x_2 \cos \bar{\xi}_{N\tau} x_1 \cos \bar{\nu}_{N\epsilon} x_3 \, e^{i\omega t} - \frac{e_{26} V x_2}{c^{(1)} 2h} e^{i\omega t},$$

$$\bar{\varphi} = \frac{V x_2}{2h} e^{i\omega t} + \frac{e_{26}}{\epsilon_{22}} H^{N\tau\epsilon} \cos \bar{\xi}_{N\tau} x_1 \cos \bar{\nu}_{N\epsilon} x_3 \times \left(\sin \frac{N\pi x_2}{2h} - (-1)^{(N-1)/2} \frac{x_2}{h} \right) e^{i\omega t}, \qquad (4.1)$$

$$\overset{S}{u}_1 = H^{N\tau\epsilon} \sin \frac{N\pi x_2}{2h} e^{-\overset{S}{\xi}_{N\tau}(x_1 - \ell)} \cos \bar{\nu}_{N\epsilon} x_3 e^{i\omega t},$$

$$\overset{S}{\varphi} = \frac{e_{26}}{\epsilon_{22}} \overset{S}{u}_1,$$

$$\overset{T}{u}_1 = H^{N\tau\epsilon} \sin \frac{N\pi x_2}{2h} \cos \bar{\xi}_{N\tau} x_1 e^{-\overset{T}{\nu}_{N\epsilon}(x_3 - b)} e^{i\omega t},$$

$$\overset{T}{\varphi} = \frac{e_{26}}{\epsilon_{22}} \overset{T}{u}_1,$$

$$\overset{C}{u}_1 = H^{N\tau\epsilon} \sin \frac{N\pi x_2}{2h} e^{-\overset{S}{\xi}_{N\tau}(x_1 - \ell)} e^{-\overset{T}{\nu}_{N\epsilon}(x_3 - b)} e^{i\omega t},$$

$$\overset{C}{\varphi} = \frac{e_{26}}{\epsilon_{22}} \overset{C}{u}_1, \qquad (4.2)$$

where

$$H^{N\tau\epsilon} = \frac{(-1)^{(N-1)/2} 4 e_{26} V 4 \sin \bar{\xi}_{N\tau} \ell \sin \bar{\nu}_{N\epsilon} b}{c^{(1)} (1 - \omega_{N\tau\epsilon}^2/\omega^2) N^2 \pi^2 \bar{\xi}_{N\tau} \bar{\nu}_{N\epsilon} L_{N\tau\epsilon}}, \qquad (4.3)$$

$$L_{N\tau\epsilon} = 4\left(\frac{\ell}{2} + \frac{\sin 2\bar{\xi}_{N\tau}\ell}{4\bar{\xi}_{N\tau}}\right)\left(\frac{b}{2} + \frac{\sin 2\bar{\nu}_{N\epsilon} b}{4\bar{\nu}_{N\epsilon}}\right) + \frac{2\cos^2 \bar{\nu}_{N\epsilon} b}{\overset{T}{\nu}_{N\epsilon}} \left(\frac{\ell}{2} + \frac{\sin 2\bar{\xi}_{N\tau}\ell}{4\bar{\xi}_{N\tau}}\right) + \frac{2\cos^2 \bar{\xi}_{N\tau}\ell}{\overset{S}{\xi}_{N\tau}} \left(\frac{b}{2} + \frac{\sin 2\bar{\nu}_{N\epsilon} b}{4\bar{\nu}_{N\epsilon}}\right) + \frac{\cos^2 \bar{\xi}_{N\tau}\ell \cos^2 \bar{\nu}_{N\epsilon} b}{\overset{S}{\xi}_{N\tau} \overset{T}{\nu}_{N\epsilon}}, \qquad (4.4)$$

and

$$c^{(1)} = \bar{c}^{(2)}(1 - k_{26}^2), \quad \omega_{N\tau\epsilon} = \omega_{N\tau\epsilon} + i\omega_{N\tau\epsilon}/2Q_{N\tau\epsilon}, \qquad (4.5)$$

in which $Q_{N\tau\epsilon}$ is the unloaded quality factor of the trapped energy resonator in the $N\tau\epsilon$th mode. As in previous work[2] the trapped energy resonant frequency $\omega_{N\tau\epsilon}$ and wave- and decay numbers $\bar{\xi}_{N\tau}$, $\bar{\nu}_{N\epsilon}$, $\overset{S}{\xi}_{N\tau}$ and $\overset{T}{\nu}_{N\epsilon}$ have been obtained by finding the fundamental roots of

$$\bar{\xi}_{N\tau} \tan \bar{\xi}_{N\tau}\ell = \left[\frac{k_N}{M_N} \Delta_N - \bar{\xi}_{N\tau}^2\right]^{1/2},$$

$$\bar{\nu}_{N\epsilon} \tan \bar{\nu}_{N\epsilon} b = \left[\frac{k_N}{c_{55}} \Delta_N - \bar{\nu}_{N\epsilon}^2\right]^{1/2}, \qquad (4.6)$$

where

$$\Delta_N = (N\pi/2h)(\bar{c}^{(1)}/\rho)^{1/2}[(4k_{26}^2/N^2\pi^2) + R], \qquad (4.7)$$

and $\bar{\xi}_{N\tau}$ and $\bar{\nu}_{N\epsilon}$ are related to the eigenfrequencies $\omega_{N\tau\epsilon}$ through (3.10) and $(3.14)_{1,2}$ and the first three of the dispersion equations

$$M_N \bar{\xi}_{N\tau}^2 + c_{55} \bar{\nu}_{N\epsilon}^2 - k_N \bar{\epsilon}_{N\tau\epsilon} = 0, \quad M_N \overset{S_2}{\xi}_{N\tau} - c_{55} \nabla_{N\epsilon}^2 + k_N \epsilon_{N\tau\epsilon} = 0,$$

$$-M_N \bar{\xi}_{N\tau}^2 + c_{55} \overset{T_2}{\nu}_{N\epsilon} + k_N \epsilon_{N\tau\epsilon} = 0, \quad M_N \overset{C_2}{\xi}_{N\tau} + c_{55} \overset{C_2}{\nu}_{N\epsilon} + k_N \epsilon_{N\tau\epsilon} = 0, \qquad (4.8)$$

where

$$k_N = (\bar{c}^{(1)}\rho)^{1/2} N\pi/h . \qquad (4.9)$$

The last dispersion equation in (4.8), which is for the corner region, is not required[7]. Some results obtained from this part of the analysis for the fundamental and an overtone mode are given in Table III. For a given ℓ:b:h ratio the condition for the existence of only one trapped mode for a given N has been obtained[2] from this part of the analysis by finding the condition for the existence of exactly two trapped symmetric modes at the upper limiting frequency $\omega = \omega_N$, as in the strip electrode case[1c,1]. From the analysis, the condition for only one trapped mode can be written in the form

$$d_B = \frac{\sqrt{2}}{\sqrt{\bar{c}^{(1)}}} (M_N\alpha^2 + c_{55}\beta^2\gamma^2)^{1/2} > N\frac{\ell}{h}\left(\frac{4k_{26}^2}{N^2\pi^2} + R\right)^{1/2} , \qquad (4.10)$$

where d_B is a particularly useful dimensionless trapping parameter since it is independent of R,

$$\alpha = \bar{\xi}_{N\tau}\ell , \quad \beta = \bar{\nu}_{N\epsilon}\ell , \quad \gamma = \ell/b , \qquad (4.11)$$

and α is either the first or second root of

$$\alpha \tan\alpha = (c_{55}/M_N)^{1/2}[(3\pi/2) - \alpha]\gamma , \qquad (4.12)$$

and

$$\beta = (3\pi/2) - \alpha , \qquad (4.13)$$

and for (4.10) we must have $\gamma < 1$. For $\gamma > 1$ the form of (4.10) changes[11]. The smaller of the two d_B thus determined is the maximum value permissible for only one symmetric trapped mode to exist. The results of some typical calculations of d_B for the fundamental and an overtone mode are given in Table II.

The admittance $Y_{N\tau\epsilon}$ of this rotated Y-cut quartz plate operating in the $N\tau\epsilon$th trapped energy mode is obtained by substituting from (4.1), with (4.3), into (2.6), which is then substituted into

$$I = -4\int_0^{\ell}\int_0^{b} \dot{D}_2\, dx_1 dx_3 , \qquad (4.14)$$

while employing (2.22) and $(2.31)_1$ and neglecting u_2 as compared with u_1, with the result

$$Y_{N\tau\epsilon} = \frac{I}{V} = \frac{4b\ell i\omega\epsilon_{22}}{2h}(1 + \hat{k}_{26}^2) + \frac{4i\omega\epsilon_{22}4\hat{k}_{26}^2 \sin^2\bar{\xi}_{N\tau}\ell \sin^2\bar{\nu}_{N\epsilon}b}{[(\omega_{N\tau\epsilon}^2/\omega^2) - 1]N^2\pi^2\bar{\xi}_{N\epsilon}^2\bar{\nu}_{N\epsilon}^2 L_{N\tau\epsilon}h} , \qquad (4.15)$$

where

$$\hat{k}_{26}^2 = e_{26}^2/c^{(1)}\epsilon_{22} . \qquad (4.16)$$

The quantities C_o and $C_{N\tau\epsilon}$ defined by

$$C_o = 4b\ell\epsilon_{22}(1 + \hat{k}_{26}^2)/2h , \qquad (4.17)$$

$$C_{N\tau\epsilon} = 16\epsilon_{22}\hat{k}_{26}^2 \sin^2\bar{\xi}_{N\tau}\ell \sin^2\bar{\nu}_{N\epsilon}b/N^2\pi^2\bar{\xi}_{N\tau}^2\bar{\nu}_{N\epsilon}^2 L_{N\tau\epsilon}h , \qquad (4.18)$$

are called the static capacitance and motional capacitance, respectively, and have been calculated for a few typical cases and are given in Table III.

Acknowledgement

This work was supported in part by the Army Research Office under Contract No. DAAG 29-79-C-0088 and The National Science Foundation under Grant No. ENG 78 27637.

References

1. H.F. Tiersten, "Analysis of Intermodulation in Thickness-Shear and Trapped Energy Resonators," J. Acoust. Soc. Am., 57, 667 (1975).
2. H.F. Tiersten, "Analysis of Trapped Energy Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist," J. Acous. Soc. Am., 59, 879 (1976).
3. E.P. Eer Nisse, "Quartz Resonator Frequency Shifts Arising from Electrode Stress," Proceedings of the 29th Annual Symposium on Frequency Control, U.S. Army Electronics Command, Fort Monmouth, New Jersey, 1 (1975).
4. A. Ballato, E.P. Eer Nisse and T.J. Lukaszek, "Experimental Verification of Stress Compensation in the SC-Cut," 1978 Ultrasonics Symposium Proceedings, IEEE Cat. No. CH 1344-1SU, Institute of Electrical and Electronics Engineers, New York, 144 (1978).
5. A. Ballato, "Doubly Rotated Thickness Mode Plate Vibrators," in Physical Acoustics, edited by W.P. Mason and R.N. Thurston (Academic, New York, 1977), Vol.XIII, Sec.V.A.2.
6. R.L. Filler and J.R. Vig, "The Acceleration and Warmup Characteristics of Four-Point-Mounted SC- and AT-Cut Resonators," Proceedings of the 35th Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey (1981).
7. Normal coordinates have been used in this manner before in the determination of the dispersion relations at the cutoff frequencies for propagation along the digonal axis in a monoclinic crystal plate by R.K. Kaul and R.D. Mindlin, "Vibrations of an Infinite, Monoclinic Crystal Plate at High Frequencies and Long Wavelengths," J. Acoust. Soc. Am., 34, 1895 (1962).
8. Not all the constants that have been neglected are actually that small. However, although including them would increase the accuracy of the results, it would complicate the solution for the plate waves in a nontrivial way.
9. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.7, Sec.1.
10. IEEE Standard on Piezoelectricity - IEEE Std. 176-1978, Institute of Electrical and Electronics Engineers, New York.

11. For more detail see D.S. Stevens and H.F. Tiersten, "An Analysis of SC-Cut Quartz Trapped Energy Resonators with Rectangular Electrodes," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.
12. W.L. Ferrar, Algebra (Oxford, London, 1941), Chap.XII, Sec.2.2.
13. Ref.9, Chap.9, Sec.2; Ref.10, Sec.4.3.
14. H.F. Tiersten, "Electromechanical Coupling Factors and Fundamental Material Constants of Thickness Vibrating Piezoelectric Plates," Ultrasonics, 8, 19 (1970).
15. Ref.5, Sec.III.
16. 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).

TABLE I

ij or kℓ	11	22	33	23	31	12	32	13	21
p or q	1	2	3	4	5	6	7	8	9

TABLE II

N	γ	α	d_B
1	.01	$> \pi$	2.743
1	1	$> \pi$	2.862
1	1.1	$> \pi$	2.620
1	1	$< 1/2\pi$	2.572
1	1.1	$< 1/2\pi$	2.477
1	100	$< 1/2\pi$	2.382*
3	1	$> \pi$	2.674
3	1.1	$> \pi$	2.455
3	1	$< 1/2\pi$	2.761
3	1.1	$< 1/2\pi$	2.705

Unelectroded thickness frequency $\omega_1/2\pi = 10$ MHz

Electrode dimensions $2\ell = 5$ mm (*$2b = 5$ mm)
NR = 0.01.

TABLE III

N	NR	$\omega_{N\tau\epsilon}/2\pi$ [a]	$C_{N\tau\epsilon}$ (ff)
1	0.01	9.93421	2.51
1	0.02	9.84291	3.27
1	0.03	9.74772	3.71
3	0.01	29.91202	0.45
3	0.02	29.81373	0.52
3	0.03	29.71459	0.57

[a] Unelectroded thickness frequency $\omega_1/2\pi = 10$ MHz

Electrode dimensions = $2\ell = 3$ mm, $2b = 3$ mm, $\tau = 1$, $\epsilon = 1$

$C_o = 1.94$ pf

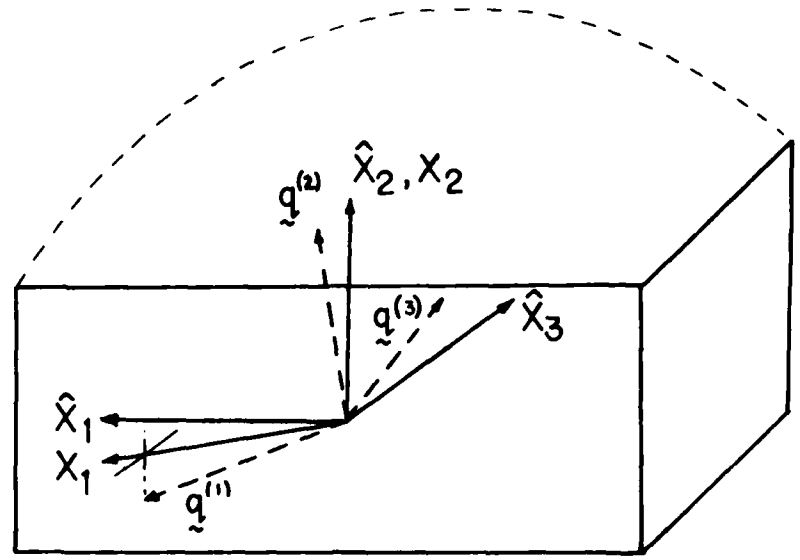


Figure 1 Schematic Diagram Showing the Original Coordinate System, the Thickness Solution Eigen-Displacement Coordinate System and the Projection of the Eigen-Displacement u_1 on the Plane of the Plate

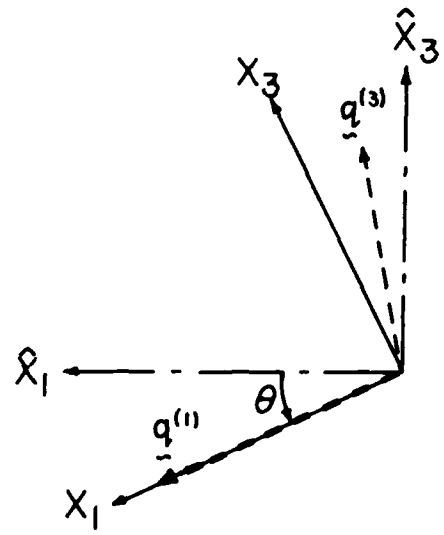


Figure 2 Plan View Showing the Original Coordinate System, the Transformed Coordinate System in the Plane of the Plate and the projection of the Thickness Solution Eigen-Displacements u_1 and u_3 on the Plane of the Plate.

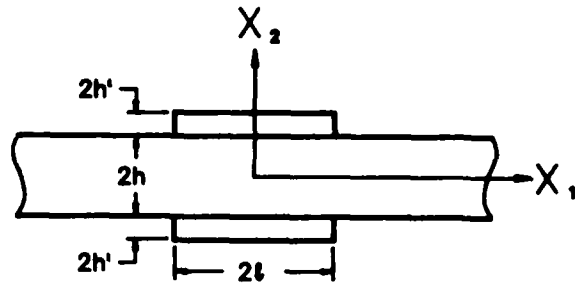


Figure 3 Schematic Diagram Showing a Cross-Section of the Trapped Energy Resonator

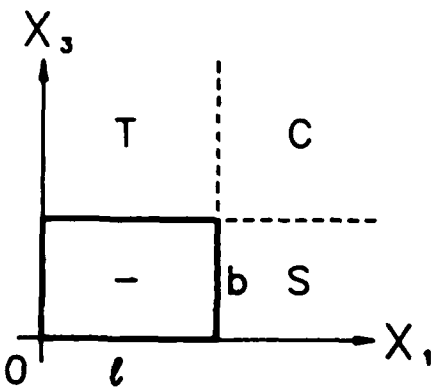


Figure 4 Plan View of the Trapped Energy Resonator Showing One Quadrant

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

STRESS COMPENSATED ORIENTATIONS FOR THICKNESS-SHEAR QUARTZ RESONATORS

Bikash K. Sinha

Schlumberger-Doll Research
P. O. Box 307, Old Quarry Road
Ridgefield, Connecticut 06877

Summary

It is well-known that a thickness-mode, quartz resonator changes its resonant frequency as a function of temperature, planar stresses in the plate and the biasing electric field (d.c. bias). While the loci of orientations which exhibit zero temperature coefficients of frequencies and zero polarizing effect have been studied in detail, the locus of orientations with zero mean force sensitivity coefficient has been, heretofore, obtained only over a small range of orientations. In this paper we report on an exhaustive study of the variation of the mean force sensitivity coefficients in the entire primitive region of the crystalline quartz. The computational results for the stress compensated orientations for thickness-shear modes show an interesting cubic behavior in the neighborhood of the SC-cut. The loci of orientations with zero temperature coefficients of frequency at 25°C and stress compensated frequency characteristics are shown to be orthogonal at the SC-cut which is in agreement with a previously reported investigation. Other orientations with large mean force sensitivity coefficients have also been identified. The computational results have been obtained from a procedure which has been successfully employed in the study of the planar stress induced frequency shifts in thickness mode quartz resonators. Both the fast and slow thickness-shear modes have been studied. Among other things, new stress compensated orientations which lie also on the first order temperature compensated locus for the fast thickness-shear mode of vibration have been identified.

Introduction

A thickness mode quartz resonator when subject to an externally applied load, a change in temperature or a biasing electric field, generally changes its resonant frequency as a result of changes in the effective elastic stiffnesses and plate geometry.[1-23] While the temperature-frequency characteristics have been studied in greater detail by various workers, a relatively lesser amount of work has been reported on the force-frequency characteristics in the existing literature. A knowledge of loci of orientations which exhibit zero mean force frequency coefficient, along with the zero temperature coefficient locus plays an important role in the design of high precision thickness-shear quartz resonators for various applications. Although the force-frequency effect in thickness-mode quartz resonators has been demonstrated both experimentally and theoretically by several workers, the loci of stress-compensated orientations has, heretofore, been reported[13] only over a small range of orientations around the well-known SC-cut.

The application of high precision quartz resonators in the development of pressure sensors with fast response time, large dynamic range and over an extended temperature range may necessitate dual mode operation of the resonator. One of the procedures for temperature compensation in pressure sensors entails two independent frequency-shift measurements on B- and C-modes of the thickness-shear resonator subject to changes in the ambient pressure and temperature. The choice of an optimum orientation for the dual mode operation of thickness-shear resonators for such pressure and temperature measurements is greatly facilitated by the availability of stress and first order temperature compensated loci of orientations. An optimum orientation may be the one which lies on the temperature locus for C-mode and stress locus for B-mode of vibrations. Thus, the former mode of vibration essentially measures the pressure acting on the crystal resonator whereas the latter mode essentially measures the temperature. It is, of course, understood that the circular plate is under radial compressive stresses resulting from radial forces acting along its periphery.[24]

The present study provides computational results for the force-frequency effect in both the fast and slow thickness-shear quartz resonators. In particular, loci of stress compensated orientations for B- and C-modes of vibration have been obtained . The mean force sensitivity curves for fast thickness-shear mode contain several zero crossings along with the maximum deviation of the mean force sensitivity coefficient $\langle K_f \rangle$ limited to about $\pm 6 \times 10^{-15}$ m.sec/N, whereas the corresponding curves for the slow thickness-shear mode of vibration exhibit relatively fewer zero crossings along with larger deviation of such force sensitivity coefficients. The maximum

mean force sensitivity coefficient has been found to occur for a rotated Y-cut quartz with the rotation angle approximately equal to 55° and for the slow thickness-shear mode of vibration.

A superposition of zero mean force sensitivity and first order temperature coefficients loci indicates that there are, indeed, orientations which are both stress and temperature compensated for the fast as well as slow thickness-shear modes of vibration. Such a stress compensated orientation (SC-cut) for the slow thickness-shear mode of vibration has been previously reported[11] and is in good agreement with the present calculations. There has been a growing interest in the stress-compensated orientation for the slow thickness-shear mode (SC-cut) because it also lies on the temperature locus, possesses faster warm up time[14] and exhibits improved aging characteristics since electrode induced isotropic stresses have no influence on the change in resonant frequency. Of the three stress and temperature compensated orientations for the B-mode of vibration in the region, $-90^{\circ}<\theta<90^{\circ}$ and $0^{\circ}<\phi<30^{\circ}$, one of the orientations ($\phi = 1^{\circ}$, $\theta = 23^{\circ}$) is in the vicinity of coalescence of B- and C-modes of vibration whereas the other two orientations ($\phi = 13^{\circ}$, $\theta = -27.5^{\circ}$) and ($\phi = 14.2^{\circ}$, $\theta = -30^{\circ}$) may have useful practical applications. The material Q for the fast thickness-shear mode is usually higher than the slow thickness-shear mode as is the case with BT-cut over that of AT-cut.[23,25] Such a consideration when applicable, would enhance the short-term frequency stability of resonators which employ the fast thickness-shear (B-) mode of vibration. The aforementioned stress and temperature compensated orientations for the B-mode of vibration would also exhibit faster warm-up time and improved aging characteristics, inasmuch as anti-symmetric biasing stresses do not affect the resonant frequency of thickness-shear mode of vibration in the first-order approximation.

An orientation ($\phi = 16.3^{\circ}$, $\theta = -34.5^{\circ}$) which lies on the temperature compensated locus for C-mode of vibration and stress compensated locus for B-mode of vibration has also been identified. Such an orientation can have useful application in the fabrication of resonators for pressure sensors over an extended temperature range and where dual mode operation is envisioned for temperature compensation. In view of the uncertainty of the basic material constants along with approximations in the analytical model, it is conceivable that the actual orientations may differ by about $\pm 2^{\circ}$. It is also important to note that a small degree of discrepancy between the temperature compensated loci for B- and C-modes of vibration and experimental data are known to exist.[26,27]

Stress Distribution in Circular Plates

A schematic diagram of a circular plate subject to a pair of diametrical forces is shown in Fig. 1. The stress distributions in such plates have been analyzed by many workers (Refs. 7,12,15,23). In the case of isotropic plates, the stresses at the center of the plate can be obtained from the solution of biharmonic equation in the Airy's stress function. The magnitude of such stresses at the center due to a pair of diametrical forces acting along the X_1 axis are given by

$$T_{11} = -\frac{6}{\pi}\frac{F}{2hD}, \quad T_{33} = \frac{2}{\pi}\frac{F}{2hD}, \text{ and } T_{13} = 0, \tag{1}$$

where T_{11}, T_{33} and T_{13} are the compressive, tensile and shear stress components, respectively; F is the magnitude of force, 2h and D are the thickness and diameter of the disk, respectively. However, in the case of anisotropic plates, the biasing stresses may be determined from the solution of generalized biharmonic equation. The values of stresses at the center are dependent on the elastic stiffnesses of material. The procedure for the determination of biasing stresses in anisotropic circular plates subject to a pair of diametrical forces has been outlined.[22] The stresses at the center of the plate have been employed to compute the baising strains from the appropriate constitutive relations for thin anisotropic plates. A homogeneous state of biasing stresses has been assumed in view of the commonly used trapped energy modes of vibration which are obtained by partial electroding and contouring of resonators. The stress distribution due to a pair of concentrated forces is nonuniform in the plane of the plate, and is, perhaps responsible for the small discrepancy between the computational and experimental results. Since an infinitesimal, homogeneous rigid body rotation does not contribute to the fractional change in the resonant frequency, the biasing displacement gradient can be simply obtained from the biasing strains. Finally, the stress induced frequency shift is expressed in terms of the force-frequency sensitivity coefficient K_f, as defined by Ratazski[3] and employed by later workers. The average edge force-frequency coefficient for uniform loading along its periphery is obtained by integrating the force-frequency coefficient due to a pair of diametrical forces over half the circumference of the circular plate as discussed in the next section.

Stress Induced Frequency Shifts

The force sensitivity coefficient K_f is defined by

$$K_f = \frac{\Delta f}{f_o}\frac{2hD}{F}\frac{1}{N}, \tag{2}$$

where $\Delta f/f_o$ is the fractional change in the resonant frequency due to the application of a pair of in-plane forces F(in Newtons) at the

circular boundary, N is the frequency constant (in m/sec), 2h and D (in meters) are the thickness and diameter of the plate resonator, respectively. When the circular resonator is subject to a uniform distribution of normal forces along its circumference, such as in quartz pressure sensors for well logging,[18,19] the total fractional change in the resonant frequency may be described in terms of the average force sensitivity coefficient $\langle K_f \rangle$. The principle of linear superposition is shown to be valid in such cases and the average force sensitivity coefficient $\langle K_f \rangle$ is given by

$$\langle K_f \rangle = \frac{1}{\pi} \int_0^{\pi} K_f(\alpha) d\alpha, \tag{3}$$

where K_f is the force sensitivity coefficient for a pair of diametrical, in-plane forces F whose direction makes an angle α with the X_1 axis in the plane of the plate.

The fractional change in the resonant frequency of a circular plate, thickness-shear quartz resonator subject to a pair of diametrical forces has been obtained from a perturbation procedure[28] for the small dynamic fields superposed on a static bias. The first perturbation of the eigenvalue due to a static bias may be written in the form

$$\omega_m = H_m^{(2)} \omega_m^2, \quad \omega_m = \omega_m - \omega_m, \tag{4}$$

where ω_m and ω_m are the unperturbed and perturbed eigenfrequencies for the m-th mode of vibration, respectively. The perturbation integral H_m is given by

$$H_m = \int_{-h}^{h} K_{L,K}^{n} K_{r,L}^{m} dX_2 \tag{5}$$

where the nonlinear portion of the Piola-Kirchhoff stress tensor has the form

$$K_{L,r}^{n} = C_{L,rMi} K_{i,M}^{m} \tag{6}$$

where

$$K_{r}^{m} = u_{r}/N, \text{ and } N^2 = \int_{-h}^{h} \rho u_r u_r dx_2 . \tag{7}$$

In Eqs. (5-6) g_r^m denotes the normalized eigensolution vector for the m-th mode of vibration as shown in Eq. (7). The nonlinear portion of the effective elastic stiffnesses is given by

$$C_{L,rMi} = T_{LM}^{1} \delta_{ri} + C_{L,rMiAB} E_{AB}^{1} + C_{L,rKM} w_{i,K} \tag{8}$$

$$+ C_{LKMi} w_{r,K} ,$$

where

$$T_{LM}^{1} = C_{LMKN} E_{KN}^{1} , \tag{9}$$

and T_{LM}^{1}, E_{AB}^{1} and $w_{i,K}$ are the static biasing stresses, strains and displacement gradients, respectively. The remaining quantities $C_{L,rKM}$ and C_{CYMiAB} in Eq. (8) are the second-order and third-order elastic stiffnesses, respectively.

The solution for thickness-shear vibrations of quartz plates with shorted electrodes on the major surfaces, which are normal to the X_2 axis (Fig. 1) has the form

$$u_Y = \sum_{n=1}^{3} C^{(n)} A_Y^{(n)} \sin \eta_n X_2 \exp(i\omega t), \tag{10}$$

$$\phi = \left(\frac{e_{22Y}}{\epsilon_{22}} \sum_{n=1}^{3} C^{(n)} A_Y^{(n)} \sin \eta_n X_2 + L X_2 \right) \exp(i\omega t),$$

where u_r and ϕ are the mechanical displacement and electric potential, respectively. The coefficient L is given by

$$L = -\frac{1}{h} \sum_{n=1}^{3} C^{(n)} \frac{e_{22Y}}{\epsilon_{22}} A_Y^{(n)} \sin \eta_n h , \tag{11}$$

and other relevant details are given in Refs. 22 and 29. Substitution of Eq. (10) into (7) yields the normalization factor

$$N^2 = \rho \sum_{m=1}^{3} \sum_{n=1}^{3} C^{(m)} A_Y^{(m)} C^{(n)} A_Y^{(n)} \left[\frac{\sin(\eta_m - \eta_n)h}{\eta_m - \eta_n} \right.$$

$$\left. - \frac{\sin(\eta_m + \eta_n)h}{\eta_m + \eta_n} \right] . \tag{12}$$

It is important to note that the biasing stresses, strains and displacement gradients must be referred to the same coordinate system as that of the dynamic eigensolution with the aid of the usual tensor transformation relations and the appropriate azimuthal angle.

Calculations have been performed for the mean force sensitivity coefficient $\langle K_f \rangle$ for thickness-shear modes of vibration and various orientations.[30] The second-order elastic stiffnesses, piezoelectric and dielectric constants employed in the calculations are those from Bechmann's work,[31] whereas the third-order elastic constants have been taken from Thurston, McSkimin and Andreatch paper.[32] Figs. 2-5 show the frequency constant and average force sensitivity coefficient $\langle K_f \rangle$ for the fast thickness-shear mode of vibration as a function of the rotation angle θ with angle ϕ fixed at intervals of 10°. The variations of corresponding quantities for the slow thickness-shear mode of vibration are shown in Figs. 6-9. The loci of stress and temperature compensated orientations for B-mode of vibration

over different ranges in orientations are depicted in Figs. 10 and 11. The region around the "horn" is of practical importance because the temperature and stress loci intersect in its vicinity. The loci of stress and temperature compensated loci for C-mode of vibration are shown in Figs. 12 and 13. It is interesting to note the cubic behavior of the stress locus in the neighborhood of the well-known SC-cut. The orthogonality of the stress and temperature loci at the SC-cut is in agreement with previous investigations. The first order temperature compensated loci for both the thickness-shear modes of vibration are from the work of Bechmann, Ballato and Lukaszek.[26] Table I contains characteristics of new orientations for thickness-shear quartz resonators with potential applications in frequency control and development of pressure and temperature sensors. As stated in the Introduction, the actual orientations may differ by about ± 2 degrees because of uncertainties in material constants and approximations in the analytical procedure.

Acknowledgments

The author wishes to thank Dr. Robert P. Porter for support and encouragement, and Mr. Thomas Muricchio for help with the calculations.

References

1. A. D. Ballato and R. Bechmann, Proc. IRE, 48, p. 261-262 (1960).

2. A. D. Ballato, Proc. 34th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, p. 89-114 (1960).

3. E. A. Gerber and M. H. Miles, Proc. IRE, 49, p. 1650-1654 (1961).

4. J. M. Ratajski, IBM J. Res. Dev., 12, p. 92-99 (1968).

5. R. M. Keys and F. W. Blair, Proc. IEEE, p. 565-566 (1967).

6. P. C. Y. Lee, Y. S. Wang and X. Markenscoff, Proc. 27th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, p.1-6 (1973).

7. P. C. Y. Lee, Y. S. Wang and X. Markenscoff, J. Acoust. Soc. Am., 57, p. 95-105 (1975).

8. A. D. Ballato, E. P. EerNisse and T. J. Lukaszek, Proc. 31st Ann. Freq. Control Symposium, USAERADCOM, Fort Monmouth, NJ, p. 8-16 (1977).

9. A. D. Ballato, in Physical Acoustics: Principles and Methods, (W. P. Mason and R. N. Thurston, Eds.), vol. 13, Chap. 5, Academic Press, New York, p. 115-181 (1977).

10. E. P. EerNisse, Proc. 29th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, p. 1-4 (1975).

11. A. Ballato, E. P. EerNisse and T. J. Lukaszek, 1978 Ultrasonics Symposium Proceedings, IEEE Cat. No. 78CH1344-ISU (IEEE, New York), p. 144-147 (1978).

12. E. P. EerNisse, T. J. Lukaszek and A. Ballato IEEE Trans. SU-25, p. 132-138 (1978).

13. E. P. EerNisse, Proc. 30th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, p. 8-11 (1976).

14. J. A. Kusters and J. G. Leach, Proc. IEEE, 65, p. 282-284 (1977).

15. D. Janiaud, L. Nissim and J. J. Gagnepain, Proc. 32nd Ann. Freq. Control Symposium, USAERADCOM, Fort Monmouth, NJ, p. 169-179.

16. C. K. Hruska, IEEE Trans., SU-25, p. 390-392 (1978).

17. C. K. Hruska, IEEE Trans., SU-27, p. 87-89 (1980).

18. D. L. Hammond and A. Benjaminson, IEEE Spectrum, 6, p. 53-58 (1969).

19. H. E. Karrer and J. Leach, IEEE Trans. IECI-16, p. 44-50 (1969).

20. P. C. Y. Lee and K-M Wu, Proc. 34th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, p. 403-411 (1980).

21. H. F. Tiersten and B. K. Sinha, J. Appl. Phys., 50, p. 8038-8051 (1979).

22. B. K. Sinha, 1980 Ultrasonics Symposium Proceedings, Cat. No. 80CH1602-2 (IEEE, N.Y.), p. 813-818 (1980).

23. A. Seed, Brit. J. Appl. Phys., 16, p. 341-1346 (1965).

24. Such a configuration has indeed been employed in a commercially available pressure sensor.[18,19]

25. J. Leach, Proc. 24th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, p. 117-125 (1970).

26. R. Bechmann, A. D. Ballato, and T. J. Lukaszek, Proc. IEEE, 50, p. 12812-1822 (1962).

27. C. A. Adams, G. M. Enslow, J. A. Kusters, and R. W. Ward, Proc. 24th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, p. 55-63 (1970).

28. H. F. Tiersten, J. Acoust. Soc. Am., 64, p. 833 (1978).

29. B. K. Sinha and H. F. Tiersten, Proc. 34th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, p. 393-402 (1980).

30. "Standards on Piezoelectric Crystals, 1949," Proc. IRE, 37, p. 1378-1395 (1949).

31. R. Bechmann, Phys. Rev., 110, p. 1060-1061 (1958).

32. R. N. Thurston, H. J. McSkimin, and P. Andreatch, Jr., J. Appl. Phys. 37, p. 267 (1966).

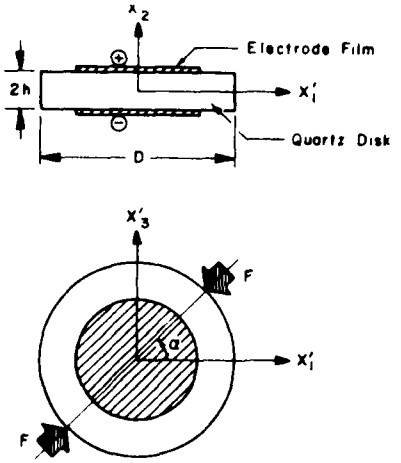


Figure 1. A Schematic Diagram of a Thickness-Shear Quartz Resonator Plate with a Pair of Diametrical Forces F Acting on it at an Angle α from the Digonal Axis.

Table I

Characteristics of New Orientations for Thickness-Shear Quartz Resonators with Potential Applications in Frequency Control and/or Pressure Sensors

Orientation		Characteristics
Φ	θ	
1°	-23°	Stress and temperature compensated for B-mode of vibration.
13°	-27.5°	
14.2°	-30°	
7°	20.2°	Stress compensated for both B- and C-modes of vibration.
24.7°	27.8°	
22.8°	-49.5°	
16.3°	-34.5°	Stress compensated for B-mode and temperature compensated for C-mode of vibration.
0°	55°	Largest mean force sensitivity coefficient ($K_f = 20 \times 10^{-15}$ m.sec/N) for C-mode of vibration.

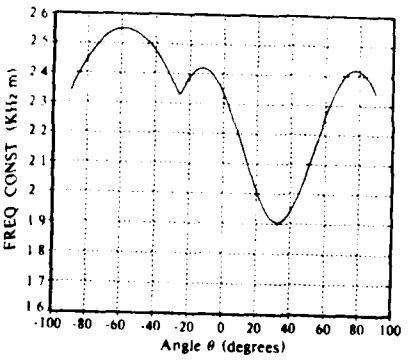


Figure 2a. Frequency Constants for B-Mode of Vibration as a Function of θ with φ = 0.

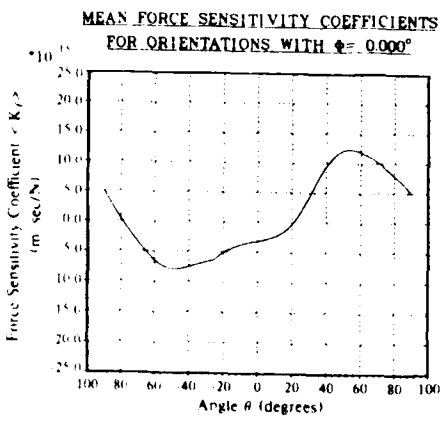


Figure 2b. Mean Force Frequency Coefficients for B-Mode of Vibration as a Function of θ with φ = 0.

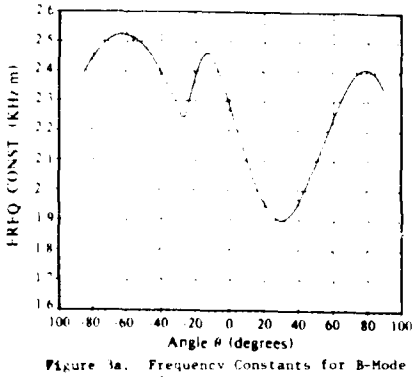


Figure 3a. Frequency Constants for B-Mode of Vibration as a Function of θ with φ = 10°.

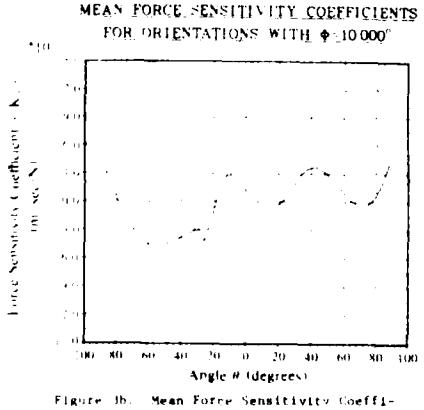


Figure 3b. Mean Force Sensitivity Coefficients for B-Mode of Vibration as a Function of θ with φ = 10°.

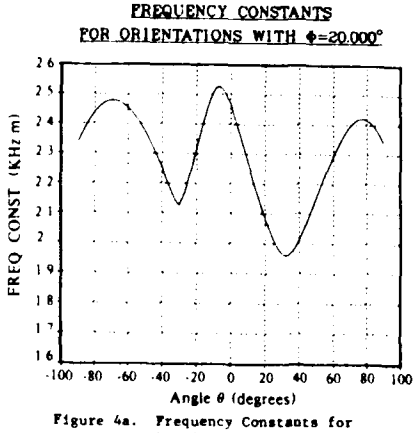


Figure 4a. Frequency Constants for B-Mode of Vibration as a Function of θ with φ = 20°.

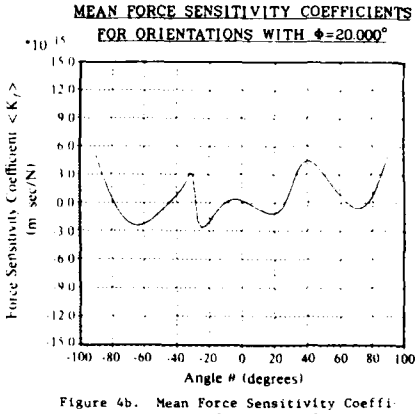


Figure 4b. Mean Force Sensitivity Coefficients for B-Mode of Vibration as a Function of θ with φ = 20°.

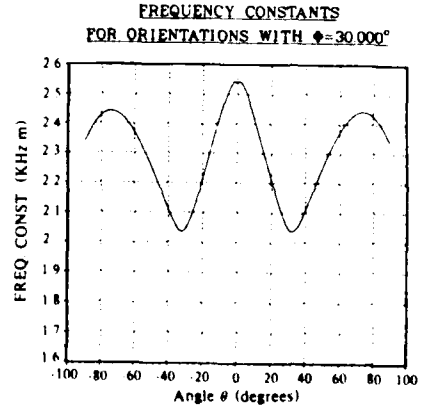


Figure 5a. Frequency Constants for B-Mode of Vibration as a Function of θ with φ = 30°.

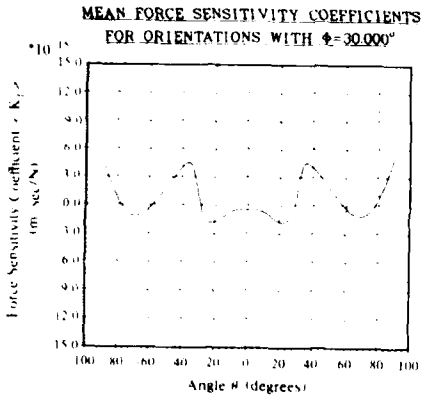


Figure 5b. Mean Force Sensitivity Coefficients for B-Mode of Vibration as a Function of θ with $\phi = 30^{\circ}$.

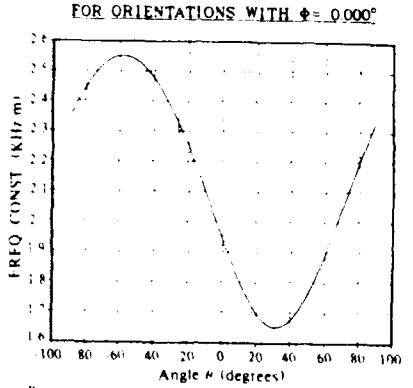


Figure 6a. Frequency Constants for C-Mode of Vibration as a Function of θ with $\phi = 0$. The curve on the left of dotted line ($\theta < -25^{\circ}$) is for the piezoelectrically active B-Mode of Vibration.

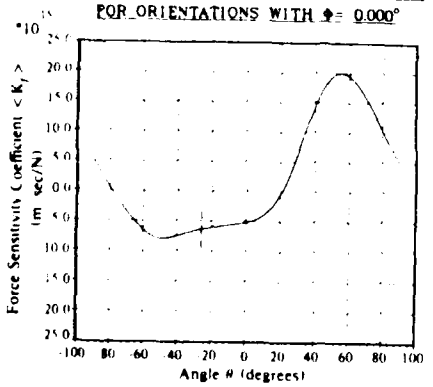


Figure 6b. Mean Force Frequency Coefficients for C-Mode of Vibration as a Function of Orientations Described in Figure 6a.

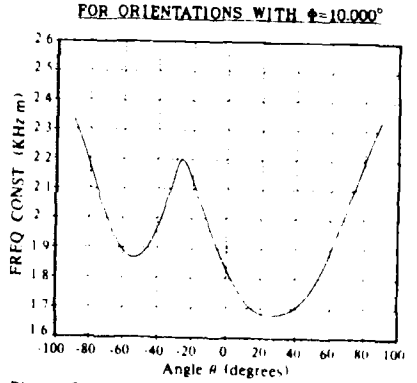


Figure 7a. Frequency Constants for C-Mode of Vibration as a Function of θ with $\phi = 10^{\circ}$.

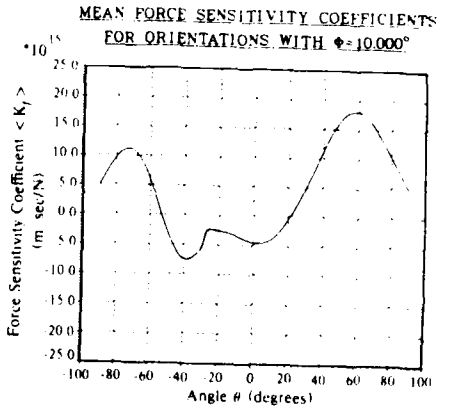


Figure 7b. Mean Force Frequency Coefficients for C-Mode of Vibration as a Function of θ with $\phi = 10^{\circ}$.

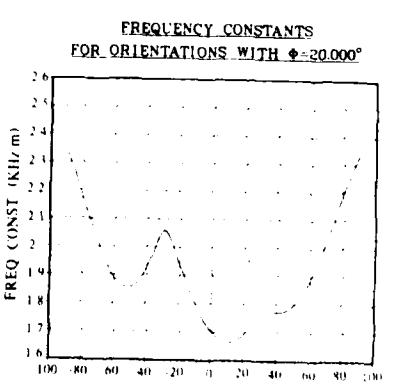


Figure 8a. Frequency Constants for C-Mode of Vibration as a Function of θ with $\phi = 20^{\circ}$.

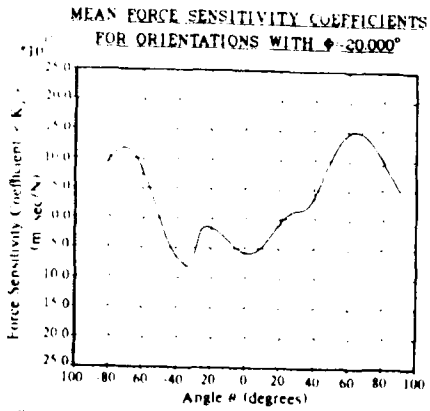


Figure 8b. Mean Force Frequency Coefficients for C-Mode of Vibration as a Function of θ with $\phi = 20^{\circ}$.

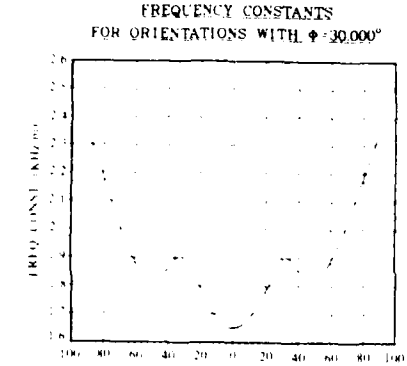


Figure 9a. Frequency Constants for C-Mode of Vibration as a Function of θ with $\phi = 30^{\circ}$.

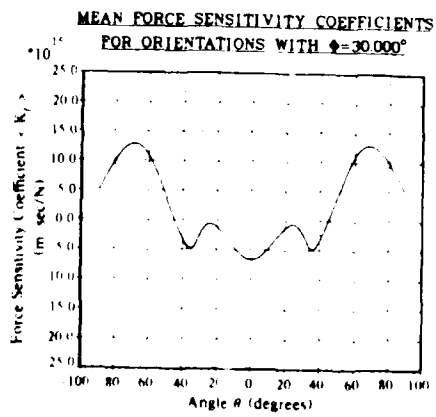


Figure 9b. Mean Force Frequency Coefficients for C-Mode of Vibration as a Function of θ with $\phi = 30^{\circ}$.

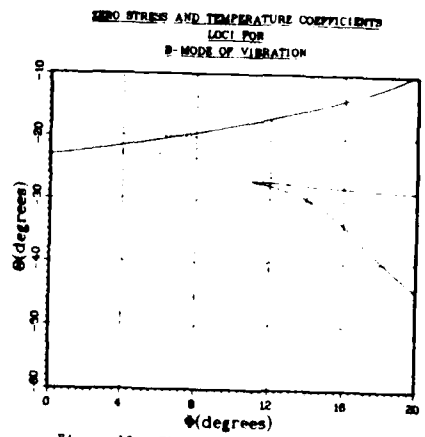


Figure 10. The dotted and solid curves represent temperature and stress compensated orientations in the region $0^{\circ}<\phi<20^{\circ}$ and $-60^{\circ}<\theta<-10^{\circ}$.

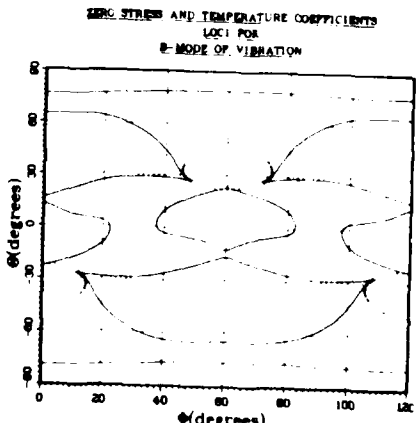


Figure 11. The dotted and solid curves represent temperature and stress compensated orientations in the region $0^{\circ}<\phi<120^{\circ}$ and $-90^{\circ}<\theta<90^{\circ}$.

ZERO STRESS AND TEMPERATURE COEFFICIENTS
LOCI FOR
C MODE OF VIBRATION

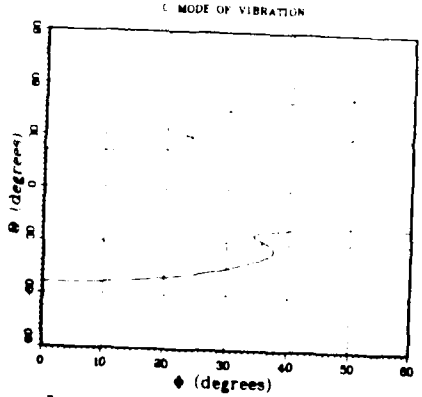


Figure 1... The dotted and solid curves represent temperature and stress compensated orientations in the region $0^\circ \le \phi \le 60^\circ$ and $-90^\circ \le \theta \le 90^\circ$.

ZERO STRESS AND TEMPERATURE COEFFICIENTS
LOCI FOR
C MODE OF VIBRATION

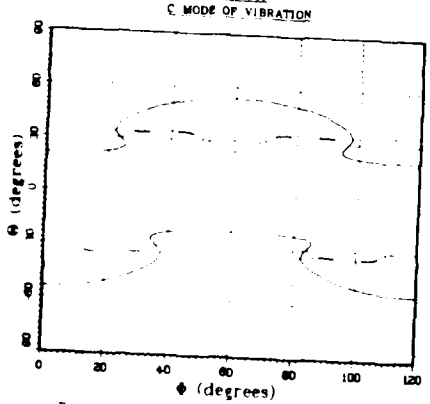


Figure 11. The dotted and solid curves represent temperature and stress compensated orientations in the region $0^\circ \le \phi \le$ and $-90^\circ \le \theta \le 90^\circ$.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

EXTENSIONAL VIBRATIONS OF RECTANGULAR CRYSTAL PLATES

P. C. Y. Lee, M. Nakazawa* and J. P. Hou

Department of Civil Engineering
Princeton University
Princeton, New Jersey 08544

Summary

A system of one-dimensional equations governing the extensional (contour) modes of vibrations in rectangular crystal plates is derived from the Cauchy-Voight two-dimensional equations of extensional motion of anisotropic, elastic plates. The system of equations is further divided into two groups of equations, governing, separately, the symmetric (contour extensional) and antisymmetric (contour shear) modes of vibrations, by setting the elastic stiffness coefficients $c_{15} = c_{35} = 0$ in order to neglect the coupling effect between the two groups.

Close form solutions of contour extensional vibrations are obtained for rectangular crystal plates satisfying the traction-free conditions at four edges. Dispersion relations and frequency equations are obtained in explicit form for the first three contour extensional modes.

Resonance frequencies, for various contour extensional modes are computed as a function of the length to-width ratio (a/c) of plates. Predicted results are compared with the detailed measurements for rectangular GT-cuts with close agreement. Two-dimensional vibration patterns have also been calculated at resonances for different values of a/c ratio.

Introduction

Extensional vibrations of thin, rectangular crystal plates have been studied by many, both analytically and experimentally.[1-5] In terms of the two-dimensional governing equations of Cauchy-Voight,[6] or Mindlin's[7] zero-order equations, of extensional vibrations of elastic, anisotropic plates, most of these solutions can be described as approximate except in very few cases in which special values of length-to-width ratio and special values of elastic constants of the plate are considered. Ekstein[4] derived his analytical solution by a perturbation method and calculated frequencies and mode shapes for square plates. Onoe[5] took an approach based on the energy principle (Lagrange's principle) so that the strain energy due to the non-zero stresses on the traction-free edges (say at $x_3 = c$) is minimized. In both of the last two papers, predicted results were compared to observed ones with good agreement.

In the present paper, we first derived a system of one-dimensional plate equations from the Cauchy-Voight two-dimensional equations by a procedure employed previously in deriving two-dimensional plate equations from the three-dimensional equations of elasticity.[8,9] Then the second-order equations for symmetrical extensional vibrations are extracted from the infinite system, and the close form solutions are obtained which satisfy the traction-free conditions on all four edges.

Dispersion relation with complex wave numbers is calculated. Two correction factors k_1 and k_2 have been introduced in the stress-displacement relations in order to improve the accuracy of the first frequency branch (the lowest branch in the ω-Reζ plane in Fig. 2). The values of k_1 and k_2 are determined by matching the slope, $d\omega/d\zeta$, to the values of that obtained from the three-dimensional theory as ζ_1 and ζ_2 either approaching zero or becoming very large as compared to 1.

Frequency spectra have been computed and compared with experimental results[10] for GT-cuts and also for square plates (a/c = 1) of Y- and Z-cuts.

Two-Dimensional Plate Equations

Consider a rectangular plate of length 2a, thickness 2b, and width 2c in a rectangular coordinate system (x_i) as shown in Fig. 1(a). The Cauchy-Voight[6] equations of extensional motion, or Mindlin's[7] zero-order stress equations of motion of crystal plates, may be written as

$$T_{11,1} + T_{13,3} = \rho\ddot{u}_1$$
$$T_{13,1} + T_{33,3} = \rho\ddot{u}_3 \qquad (1)$$

and the stress-displacement relations

$$T_{11} = \gamma_{11}u_{1,1} + \gamma_{13}u_{3,3} + \gamma_{15}(u_{3,1} + u_{1,3})$$
$$T_{33} = \gamma_{31}u_{1,1} + \gamma_{33}u_{3,3} + \gamma_{35}(u_{3,1} + u_{1,3}) \qquad (2)$$
$$T_{13} = \gamma_{51}u_{1,1} + \gamma_{53}u_{3,3} + \gamma_{55}(u_{3,1} + u_{1,3})$$

where u_i $(i = 1,3)$ is the displacement components in the x_i direction of a plate element which is originally parallel to the x_2 axis, T_{ij} the in-plane stress resultant, ρ the mass density of the plate, and γ_{pq} related to the elastic stiffness coefficients c_{pq} by the relations

*Visiting Research Fellow, on leave from Department of Electronic Engineering, Shinshu University, Nagano 380, Japan.

$$\bar{c}_{pq} = c_{pq} - c_{6p}c_{q6}/c_{66} \qquad p,q = 1,3,5$$

$$\bar{\bar{c}}_{pq} = \bar{c}_{pq} - \bar{c}_{4p}\bar{c}_{q4}/\bar{c}_{44} \qquad p,q = 1,3,5,6 \qquad (3)$$

$$\bar{c}_{pq} = c_{pq} - c_{2p}c_{q2}/c_{22} \qquad p,q = 1,3,4,5,6$$

These equations, (1) and (2), are two-dimensional and the traction-free conditions on faces $x_2 = \pm b$ have been taken into account in (1).

One-Dimensional Plate Equations

In order to approximate equations (1) and (2) by a system of one-dimensional equations, we expand the displacement field by

$$u_i(x_1,x_3,t) = \sum_{n=0}^{\infty} v_i^{(n)}(x_1,t)\cos\frac{n\pi}{2}(1-\eta) \qquad (4)$$

where $\eta = x_3/c$.[8]

The variational stress equations of motion corresponding to (1) are

$$\int_A (T_{ij,j} - \rho\ddot{u}_i)\,u_i\,dA = 0 \qquad i = 1,3 \qquad (5)$$

By substituting (4) into (5), replacing dA by $c\,dx_1\,d\eta$, integrating with respect to η from $\eta = -1$ to $\eta = +1$, and requiring that the final result must be independent of every arbitrary variation $\delta u_j^{(n)}$, we obtain the one-dimensional nth-order stress equations of motion

$$T_{1j,1}^{(n)} - \frac{n\pi}{2c}T_{3j}^{(n)} + \frac{1}{c}E_j^{(n)} = (1+\delta_{n0})\rho\ddot{v}_j^{(n)} \qquad (6)$$

where $j = 1,3$; $n = 0,1,2,\ldots$, $\delta_{n0} = 0$ when $n \neq 0$ and $\delta_{n0} = 1$ when $n = 0$; and

$$T_{ij}^{(n)} = \int_{-1}^{1} T_{ij}\cos\frac{n\pi}{2}(1-\eta)\,d\eta$$

$$T_{ij}^{(n)} = \int_{-1}^{1} T_{ij}\sin\frac{n\pi}{2}(1-\eta)\,d\eta \qquad (7)$$

$$E_j^{(n)} = \left[T_{3j}\cos\frac{n\pi}{2}(1-\eta)\right]_{-1}^{1}$$

$T_{ij}^{(n)}$, $\bar{T}_{ij}^{(n)}$ are the nth-order stress resultants and $E_j^{(n)}$ the nth-order edge-tractions on the edges $x_3 = \pm c$ (or $\eta = \pm 1$).

For convenience, we adopt the contracted notions for stresses by replacing

$$T_{11}^{(n)} \to T_1^{(n)}$$

$$T_{33}^{(n)} \to T_3^{(n)}$$

$$T_{13}^{(n)} \to T_5^{(n)}$$

and similarly for $\bar{T}_{ij}^{(n)}$. The nth-order stress-displacement relations are obtained by inserting (4) into (2) and then integrating with respect to η according to (7). Therefore we have

$$T_p^{(n)} = (1+\delta_{n0})(c_{p1}v_{1,1}^{(n)} + c_{p5}v_{3,1}^{(n)}) + \frac{1}{c}\sum_{m=0}^{\infty} B_{mn}(c_{p3}v_3^{(m)} + c_{p5}v_1^{(m)}) \qquad p = 1,5$$

$$\bar{T}_p^{(n)} = \frac{n\pi}{2c}(c_{p3}v_3^{(n)} + c_{p5}v_1^{(n)}) + \sum_{m=0}^{\infty}\frac{2}{n\pi}B_{nm}(c_{p1}v_{1,1}^{(m)} + c_{p5}v_{3,1}^{(m)}) \qquad p = 3,5 \qquad (8)$$

where

$$B_{mn} = \begin{cases} 0 & , \ m+n = \text{even} \\ \dfrac{2m^2}{m^2-n^2} & , \ m+n = \text{odd} \end{cases} \qquad (9)$$

Substitution of (8) into (6) gives the nth-order displacement equations of motion,

$$(1+\delta_{n0})c_{11}v_{1,11}^{(n)} + (1+\delta_{n0})c_{15}v_{3,11}^{(n)} - \frac{(n\pi)^2}{2c}(c_{53}v_3^{(n)} + c_{55}v_1^{(n)}) + \frac{1}{c}\sum_{m=0}^{\infty}B_{mn}(c_{13}v_{3,1}^{(m)} + c_{15}v_{1,1}^{(m)}) - \frac{1}{c}\sum_{m=0}^{\infty}B_{mn}(c_{51}v_{1,1}^{(m)} + c_{55}v_{3,1}^{(m)}) + \frac{1}{c}E_1^{(n)} = (1+\delta_{n0})\rho\ddot{v}_1^{(n)}$$

$$(1+\delta_{n0})c_{51}v_{1,11}^{(n)} + (1+\delta_{n0})c_{55}v_{3,11}^{(n)} - \frac{n\pi}{2c}(c_{33}v_3^{(n)} + c_{35}v_1^{(n)}) + \frac{1}{c}\sum_{m=0}^{\infty}B_{mn}(c_{53}v_{3,1}^{(m)} + c_{35}v_{1,1}^{(m)}) - \frac{1}{c}\sum_{m=0}^{\infty}B_{mn}(c_{31}v_{1,1}^{(m)} + c_{35}v_{3,1}^{(m)}) + \frac{1}{c}E_3^{(n)} = (1+\delta_{n0})\rho\ddot{v}_3^{(n)} \qquad (10)$$

Equations (8) and (10) form a system of one-dimensional equations approximating the two-dimensional plate equations of (1) and (2).

Symmetric Extensional Vibrations

In (10), the symmetrical modes ($v_1^{(n)}, v_3^{(n+1)}$; n = even) and the anti-symmetrical modes ($v_1^{(n)}$, $v_3^{(n-1)}$; n = odd) are coupled through the material constants c_{15} and c_{35} which relate the extensional stresses T_1 and T_3 to the shear strain $s_5 = u_{3,1} + u_{1,3}$ as it may be seen in (2). If the values of c_{15} and c_{35} are small as compared to those of c_{11}, c_{33}, c_{55}, and c_{13}, we may uncouple the symmetrical and anti-symmetrical modes of vibration in (10) by setting $c_{15} = c_{35} = 0$. In the cases of rotated Y-cuts and GT-cut ($\theta = 51°$), these two values are practically zero.

A second-order symmetric extensional theory can be extracted from the infinite system of equations (8) and (10) by

1. Setting $c_{15} = c_{35} = 0$
2. Retaining the first three symmetrical displacements $v_1^{(0)}$, $v_3^{(1)}$, and $v_1^{(2)}$.
3. Setting $v_1^{(n)} = v_3^{(n+1)} = 0$, $T_1^{(n)} = T_5^{(n)} = T_3^{(n+1)} = S_5^{(n+1)} = 0$ for $n \geq 2$ and n = even.
4. For traction-free edge conditions at $x_3 = \pm c$, setting $T_3^{(n)} = 0$.

Hence we have, for the second-order theory,

Displacement equations of motion:

$$\begin{aligned}
&\bar{c}_{11} v_{1,11}^{(0)} + \frac{2}{c}\bar{c}_{13} v_{3,1}^{(1)} = \rho \ddot{v}_1^{(0)} \\
&\bar{c}_{55} v_{3,11}^{(1)} + \frac{2}{3c}(4\bar{c}_{55} + \bar{c}_{13}) v_{1,1}^{(2)} \\
&\qquad - \frac{2}{c^2}\bar{c}_{13} v_{1,1}^{(0)} - \frac{2}{c^2}\bar{c}_{33} v_3^{(1)} = \rho \ddot{v}_3^{(1)} \\
&\bar{c}_{11} v_{1,11}^{(2)} - \frac{2}{3c}(4\bar{c}_{55} + \bar{c}_{13}) v_{3,1}^{(1)} - \frac{2}{c^2}\bar{c}_{55} v_1^{(2)} = \rho \ddot{v}_1^{(2)}
\end{aligned} \tag{11}$$

Stress-Displacement Relations:

$$\begin{aligned}
T_1^{(0)} &= 2\bar{c}_{11} v_{1,1}^{(0)} + \frac{2}{c}\bar{c}_{13} v_3^{(1)} \\
T_5^{(1)} &= \bar{c}_{55}\left(2v_{3,1}^{(1)} + \frac{8}{3c} v_1^{(2)}\right) \\
T_1^{(2)} &= \bar{c}_{11} v_{1,1}^{(2)} + \frac{2}{3c}\bar{c}_{13} v_3^{(1)} \\
T_3^{(1)} &= \frac{2}{c}\bar{c}_{33} v_3^{(1)} + \frac{4}{3c} k_1 \bar{c}_{13} v_{1,1}^{(0)} \\
T_5^{(2)} &= \frac{2}{c}\bar{c}_{55} v_1^{(2)}
\end{aligned} \tag{12}$$

In (11) and (12), two correction factors k_1 and k_2 have been introduced. It may be seen in (12) that k_1 is introduced in $T_1^{(1)}$ and $T_3^{(1)}$ and k_2 in $T_5^{(1)}$ in a similar manner done in Ref. 8 and 9. We note that (11) are essentially identical to the one-dimensional second-order extensional equations of Ref. 9 except in which c_{pq} are replaced by $\bar{c}_{pq}$. Eq. (11) is also similar to the Mindlin-Medick's extensional equations for elastic, isotropic plate.[11]

Dispersion Relation and Correction Factors

We select solution which corresponds to the extensional motion symmetrical with respect to the x_1 and x_3 axes

$$\begin{aligned}
v_1^{(0)} &= A_0 \sin \xi x_1 \, e^{i\omega t} \\
v_3^{(1)} &= A_1 \cos \xi x_1 \, e^{i\omega t} \\
v_1^{(2)} &= A_2 \sin \xi x_1 \, e^{i\omega t}
\end{aligned} \tag{13}$$

Equations (11) are satisfied by (13) provided

$$\begin{bmatrix}
\bar{c}_{11} z^2 - \Omega^2 & \frac{2}{\pi} k_1 \bar{c}_{13} z & 0 \\
\frac{4}{3\pi} k_1 \bar{c}_{13} z & k_2 \bar{c}_{55} z^2 + 1 - \Omega^2 & -\frac{4}{3\pi}(4\bar{c}_{55} + \bar{c}_{13}) z \\
0 & -\frac{4}{3\pi}(4\bar{c}_{55} + \bar{c}_{13}) z & \bar{c}_{11} z^2 + 4\bar{c}_{55} - \Omega^2
\end{bmatrix}
\begin{bmatrix} A_0 \\ A_1 \\ A_2 \end{bmatrix} = 0 \tag{14}$$

where

$$\bar{c}_{pq} = \frac{c_{pq}}{c_{33}} \quad , \quad \Omega = \frac{\omega}{\omega_0} \sqrt{\frac{\bar{c}_{33}}{\rho}}$$

$$\omega_0 = \frac{\pi}{2c} \quad , \quad z = \frac{\xi}{\pi/2c} \tag{15}$$

The vanishing of the determinant of the coefficient matrix of (14) for nontrivial solutions leads to the dispersion relation which may be written as

$$P_6 z^6 + P_4 z^4 + P_2 z^2 + P_0 = 0 \tag{16}$$

where

$$\begin{aligned}
P_6 &= k_2^2 \bar{c}_{11} \bar{c}_{55} \\
P_4 &= \bar{c}_{11}[\bar{c}_{11} + 4k_2^2 \bar{c}_{55} - 16(4\bar{c}_{55} + \bar{c}_{13})^2/9\pi^2 - 8k_1^2 \bar{c}_{13}^2/\pi^2 \\
&\quad - \Omega^2(\bar{c}_{11} + 2k_2^2 \bar{c}_{55})] \\
P_2 &= [\bar{c}_{11} - 8k_1^2 \bar{c}_{13}^2/\pi^2 - \Omega^2(\bar{c}_{11} + k_2^2 \bar{c}_{55})][(4\bar{c}_{55} - \Omega^2) \\
&\quad + 16(4\bar{c}_{55} + \bar{c}_{13})^2 \Omega^2/9\pi^2 - \bar{c}_{11}(1 - \Omega^2)\Omega^2 \\
P_0 &= (1 - \Omega^2)(\Omega^2 - 4\bar{c}_{55})\Omega^2
\end{aligned} \tag{17}$$

For a given value of Ω which is real for elastic system, (16) may be regarded as a bicubic equation

of z with real coefficients P_j. Hence it has three pairs of roots, designated by

$$z_1, \quad z_2, \quad z_3$$

where z_j can be real, imaginary, or complex. If complex roots exist, they must appear in pairs which are complex conjugates to each other.

Dispersion curves are computed from (16) and are shown in Fig. 2 for a GT-cut plate of quartz. A GT-cut is a doubly rotated cut which, according to the 1949 IRE Standard, may be obtained by a rotation of 51° about the X axis from a Y-cut position, and then by a second rotation of 45° about the plate normal as shown in Fig. 1(a).

The values of γ_{ij} of GT-cut ($\varphi = 51°$) referred to the plate axes with the x_1 axis corresponding to the diagonal axis of quartz are computed by (3) from Bechmann's[12] values and are listed below, in 10^9 N/m^2,

$$\gamma_{11} = \gamma_{33} = 114.54 \quad , \quad \gamma_{55} = 30.00$$

$$\gamma_{13} = -5.07 \quad , \quad \gamma_{15} = \gamma_{35} = 0.43$$

Figure 2 shows that there are three cut-off frequencies. The values of the cut-off frequencies wit' zero wave-numbers can be obtained from the determinant in (14) by setting $z = 0$, and their values are

$$\Omega = 1, \quad \text{and} \quad 2\sqrt{\bar{\gamma}_{55}} \tag{18}$$

or in dimensional form

$$\omega = \frac{\pi}{2c}\sqrt{\gamma_{33}/\rho} \quad , \text{ and } \quad \frac{\pi}{c}\sqrt{\gamma_{55}/\rho} \tag{18'}$$

These values agree well with those from the three-dimensional theory of elasticity.

The asymptotic expression for the first branch for $|\Omega| \ll 1$ and $|z_j| \ll 1$ is obtained from (16)

$$\Omega^2 = \left(\bar{\gamma}_{11} - \frac{8}{\pi^2} k_1^{2-2}\gamma_{13}\right) z^2 \tag{19}$$

For (19) to agree with the corresponding expression $\Omega^2 = (\bar{\gamma}_{11} - \bar{\gamma}_{13}^2/\bar{\gamma}_{55})z^2$ from the three-dimensional theory, we set

$$k_1 = \pi/\sqrt{8} \tag{20}$$

The slope of the first branch for both $|\Omega| \gg 1$ and $|z| \gg 1$ can be approximated by ($\Omega^2 \doteq k_2 \bar{\gamma}_{55} z^2$)

$$\frac{d\Omega}{dz} = \frac{1}{\sqrt{k_2\bar{\gamma}_{55}}}\left[k_2\bar{\gamma}_{55} + \frac{\bar{\gamma}_{11} - k_2\bar{\gamma}_{55} - \bar{\gamma}_{13}^2}{(\bar{\gamma}_{11} - k_2\bar{\gamma}_{55})z^2 + 1}\right]$$

The second term of the above expression can be further simplified by dropping $\bar{\gamma}_{13}^2$ as compared to $(\bar{\gamma}_{11} - k_2\bar{\gamma}_{55})$ and by choosing z^2 so large that

$$(\bar{\gamma}_{11} - k_2\bar{\gamma}_{55})z^2 \gg 1 \tag{21}$$

Therefore,

$$\frac{d\Omega}{dz} = \frac{1}{\sqrt{k_2\bar{\gamma}_{55}}}\left(k_2\bar{\gamma}_{55} + \frac{1}{z^2}\right) \tag{22}$$

It is known that, for large z, the slope of the first extensional branch approaches to surface wave velocity, i.e., in dimensionless form

$$\frac{d\Omega}{dz} \doteq v \quad \frac{c}{\sqrt{\gamma_{33}/\rho}} \tag{23}$$

where c is the surface wave velocity.

By substituting (22) into (23) and solving for k_2, we have

$$k_2 = \frac{1}{2\bar{\gamma}_{55}}\left[\left(v^2 - \frac{2}{z^2}\right) + v^2\sqrt{1 - \frac{4}{z^2v^2}}\right] \tag{24}$$

Values of k_2 are computed from (24) and listed as follows

Plate Orienta.	v^2	$\bar{\gamma}_{11}$	$\bar{\gamma}_{55}$	z	k_2
GT-cut (4=51°)	0.2235	1.0	0.2619	7	0.7274
GT-cut (4=61°)	0.2414	1.0	0.2878	7	0.6896
Z-cut	0.3502	1.0	0.4298	7	0.7428
Y-cut	0.3664	0.7517	0.4731	10	0.7315

The values of v^2 for GT-cut are computed according to the solution of Deresiewicz and Mindlin,[13] and those for Y- and Z-cuts are obtained from Ref. 14. The values of z are chosen so that (21) is satisfied by letting $(\bar{\gamma}_{11} - k_2\bar{\gamma}_{55})z^2 \doteq 40$.

Frequency Spectra

The appropriate form of solution of (11) in case of an alternating and uniform electric field is impressed over the faces $x_2 = \pm b$ of the rectangular plate can be written as

$$v_1^{(0)} = \sum_{s=1}^{3} \alpha_{0s} A_{1s} \sin \xi_s x_1 \, e^{i\omega t}$$

$$v_3^{(1)} = \sum_{s=1}^{3} A_{1s} \cos \xi_s x_1 \, e^{i\omega t} \tag{25}$$

$$v_1^{(2)} = \sum_{s=1}^{3} \alpha_{2s} A_{1s} \sin \xi_s x_1 \, e^{i\omega t}$$

where the amplitude ratios are obtained from (14)

$$\alpha_{0s} \quad \frac{A_{0s}}{A_{1s}} = \frac{\frac{2}{\pi}k_1\bar{\gamma}_{13}z_s}{\Omega^2 - \bar{\gamma}_{11}z_s^2}$$

$$\frac{A_{2s}}{A_{1s}} = -\frac{\frac{4}{3}(4\bar{c}_{55} + \bar{c}_{13})z_s}{2 - \bar{c}_{11}z_s^2 - 4\bar{c}_{55}} \qquad (26)$$

and $z_s = \gamma_s/(\pi/2c)$, $s = 1,2,3$ are the roots of (11).

The traction-free edge conditions at $x_3 = \pm c$ have been taken into account in (11) by setting $\xi_i^{(n)} = 0$. The traction-free conditions at edges $x_1 = \pm a$ require that

$$T_1^{(0)}\Big|_{x_1=\pm a} = 0, \quad T_5^{(1)}\Big|_{x_1=\pm a} = 0, \quad T_1^{(2)}\Big|_{x_1=\pm a} = 0 \qquad (27)$$

Substitution of (25) into the first three equations of (12) and imposition of conditions (27) lead to the simultaneous homogeneous equations of A_{1s} ($s = 1,2,3$)

$$\begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} A_{11} \\ A_{12} \\ A_{13} \end{bmatrix} = 0 \qquad (28)$$

where

$$\begin{aligned} M_{1s} &= (\bar{c}_{11}\gamma_{0s}z_s + \frac{2}{\pi}k_1\bar{c}_{13})\cos(z_s\frac{\pi}{2}\frac{a}{c}) \\ M_{2s} &= (\frac{16}{3}\bar{c}_{55}\gamma_{2s} - k_2z_s)\sin(z_s\frac{\pi}{2}\frac{a}{c}) \\ M_{3s} &= (\bar{c}_{11}\gamma_{2s}z_s - \frac{4}{3\pi}\bar{c}_{13})\cos(z_s\frac{\pi}{2}\frac{a}{c}) \end{aligned} \qquad (29)$$

The consistency of (28) requires that the determinant of coefficients vanishes, yielding the frequency equation

$$F(\Omega, a/c, \bar{c}_{ij}) = |M_{rs}| = 0, \quad r,s = 1,2,3 \qquad (30)$$

In (30), $\bar{c}_{ij}$ can be calculated for given orientations of the cut, and a/c is determined by the dimensions of the rectangular plate. The roots satisfy (30) and (16) simultaneously are the resonance frequencies of the extensional vibrations of rectangular plate with all four edges free. A series of values computed as a function of a/c ratio gives the frequency spectrum.

Resonance frequencies are calculated for square plates of Y- and Z-cuts with $2a = 1$ mm. The presently calculated values are compared with those [illegible][4] and experimental values by Bechmann[2] as shown below.

	f(Ekstein)	f(Bechmann)	f(Present)
f_1 (KHz)	2.66 x 10^3	2.55 x 10^3	2.53 x 10^3
f_2	2.73	2.72	2.75
f_3	2.86	2.90	2.91

Y-Cut	f(Ekstein)	f(Bechmann)	f(Present)
f_1 (KHz)	2.64 x 10^3	2.68 x 10^3	2.67 x 10^3
f_2	2.90		2.94
f_3	3.17	3.19	3.19

Figure 3 presents the measured resonances versus the length-to-width ratio a/c of GT-cut (φ=51°) plates with thickness 2b = 1.065 mm, and width 2c = 11.91 mm.[10] The strong resonances are indicated by hollowed circles and the weak resonances by solid dots. Among the three branches of strong resonance, branches a'-a'' and b'-b'' are stronger than the branch c'-c''.

Frequency spectra are computed by the present theory for GT-cut plates, and are compared with experimental results[10] as shown in Figs. 4 and 5, respectively, for φ = 51° and 61°.

Mode Shapes

Once the resonance frequencies are determined, it is of interest to identify the mode shape corresponding to each resonance and to examine the effect of a/c ratio on the changes of mode shapes as well as on the frequencies.

By (28), the ratios along the amplitudes A_{11}, A_{12}, A_{13} can be calculated. Then $v_1^{(0)}$, $v_3^{(1)}$, and $V_1^{(2)}$ are determined by (25) within an arbitrary constant of multiplication. By substituting (25) into (4) and dropping the factor $e^{i\omega t}$, the two-dimensional mode shapes are obtained.

$$\begin{aligned} u_1(x_1,x_3) &= \sum_{s=1}^{3} A_{1s}\sin(z_s\frac{\pi}{2}\frac{x_1}{c})[\gamma_{0s} - \gamma_{2s}\cos(\frac{\pi x_3}{c})] \\ u_3(x_1,x_3) &= \sum_{s=1}^{3} A_{1s}\cos(z_s\frac{\pi}{2}\frac{x_1}{c})\sin(\frac{\pi x_3}{2c}) \end{aligned} \qquad (31)$$

The resonance frequencies of GT-cut (φ=51°) with a/c = 0.80, 1.00, and 1.20 are designated by points a_i, b_i, and c_i (i = 1,2,3), respectively, in Fig. 4. The corresponding mode shapes at points a_i, b_i, and c_i are calculated by (31) and shown, respectively, in Figs. 6, 7, and 8.

By examining the modes corresponding to points a_1, b_1, and c_1 on the lower branch in Fig. 4, we see that mode a_1 (a/c = 0.80) is essentially extensional in x_3 and slightly dilatational in character; mode b_1 (a/c = 1) has an equal amount of extensions in x_1 and x_3 and is strongly dilatational; mode c_1 (a/c = 1.20) becomes essentially extensional in x_1 and slightly dilatational in character. Similar observations can be made for modes at points a_3, b_3, and c_3 on the upper branch. Modes at a_3 and c_3 are essentially extensional in x_3 and x_1, respectively, and are slightly equivoluminal in character, while mode b_3 is strongly equivoluminal. In contrast, the modes at points a_2, b_2, and c_2 on the middle branch have strong coupling between the x_1- and x_3-extension and are strongly equivoluminal in charac-

ter, and they seem to be less dependent on the a/c ratio.

Mode shapes for square z-cut plates are obtained. They are very similar to those shown in Fig. 7 for GT-cut. However, the sequence of occurrence is reversed, i.e., for z-cut, the mode b_3 occurs at a lower frequency than that for mode b_2, and mode b_2 has a lower resonance than that for mode b_1. Both the mode shapes and sequence for z-cut are in agreement with Ekstein's[4] computation.

We note that the near symmetry (or anti-symmetry) with respect to both the x_1 and x_3 axes of the vibrational patterns shown in Fig. 7 for square plates is due to the elastic constants $\gamma_{11} = \gamma_{33}$ for GT- and Z-cuts. Computed mode shapes for a square Y-cut plate, for which $\gamma_{11} \neq \gamma_{33}$, confirm this observation.

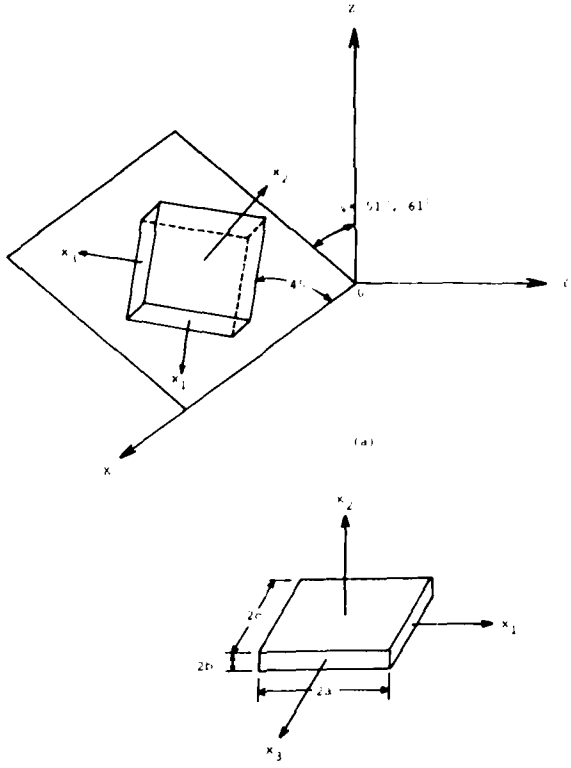
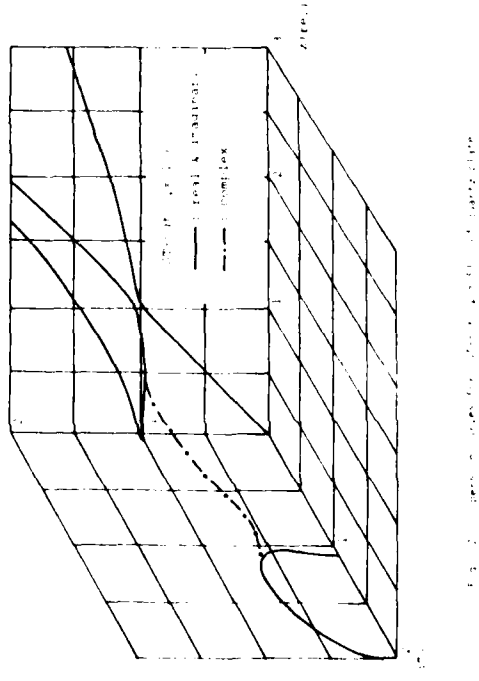


Fig. 1 A rectangular plate referred to x_i coordinate system and angles of rotation for GT-cut plate



References

1. W. P. Mason, Bell Sys.Tech.J., Vol. 13, 405 (1934).
2. R. Bechmann, Zeits.f.Physik, Vol. 117, 180, (1941); Vol. 118, 515, (1941); Proc.Phys.Soc. (London), Vol. 864, 323, (1951).
3. I. Koga and J. Fukuyo, J.Inst.Elec.Comm. Engrs. Japan, Vol. 36, 59, (1953).
4. H. Ekstein, Phys.Rev., Vol. 66, 108, (1944).
5. M. Onoe, J.Acoust.Soc.Am., Vol. 30, 1159, (1958).
6. A. L. Cauchy, Exercices de Mathematiques, Paris, Vol. 4, 13 (1829); W. Voigt, Lehrbuch de Kristallphysik, 2nd Ed., Leipzig, 1928, p. 698.
7. R. D. Mindlin, "An Introduction to the Mathematical Theory of Vibrations of Elastic Plates," U.S. Army Signal Corps Engrg. Lab., Fort Monmouth, NJ, 1957, Chap. 4.
8. P. C. Y. Lee and Z. Nikodem, Int.J.Solids Structures, Vol. 8, 581, (1972).
9. Z. Nikodem and P. C. Y. Lee, Int.J.Solids Structures, Vol. 10, 177, (1974).
10. M. Nakazawa and S. Kozima, IEE Japan, US76-6, 7, (1972); IEE Japan, Technical Report of Standard Freq. and Circuits Committee, No. 25-3, (1978). (Both in Japanese.)
11. R. D. Mindlin and M. A. Medick, J.Appl.Mech. Vol. 26, 561, (1959).
12. R. Bechmann, A. D. Ballato, and T. J. Lukaszek, Proc. IRE., Vol. 50, 1812, (1962); R. Bechmann, Phys.Rev., Vol. 110, 1060, (1958).
13. H. Deresiewicz and R. D. Mindlin, J.Appl.Phys., Vol. 28, 669, (1957).
14. T. C. Lim and G. W. Farnell, J.Appl.Phys. Vol. 39, 4319, (1968).

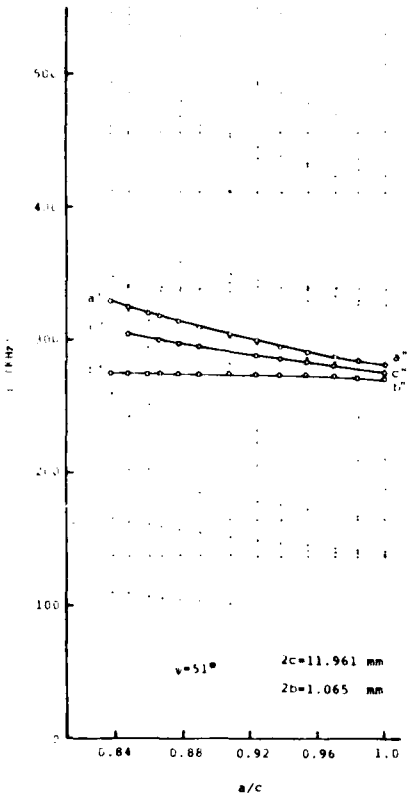


Fig 3 Measured resonance frequencies vs the length-to-width ratio of rectangular GT-cut (ϕ = 51°) plates.

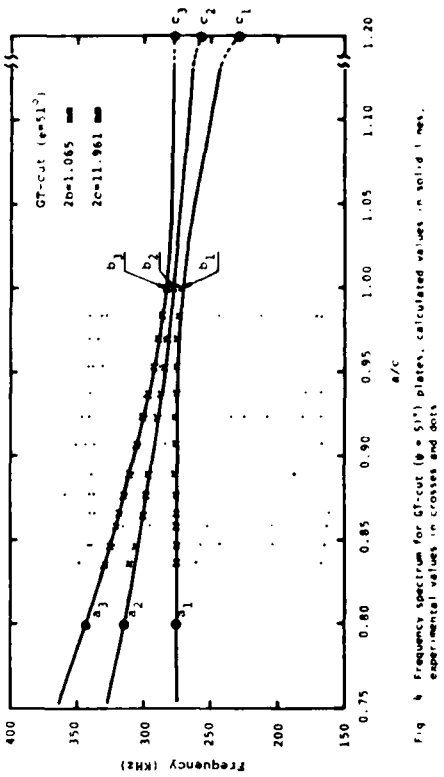


Fig 4 Frequency spectrum for GT-cut (ϕ = 51°) plates, calculated values in solid lines, experimental values in crosses and dots

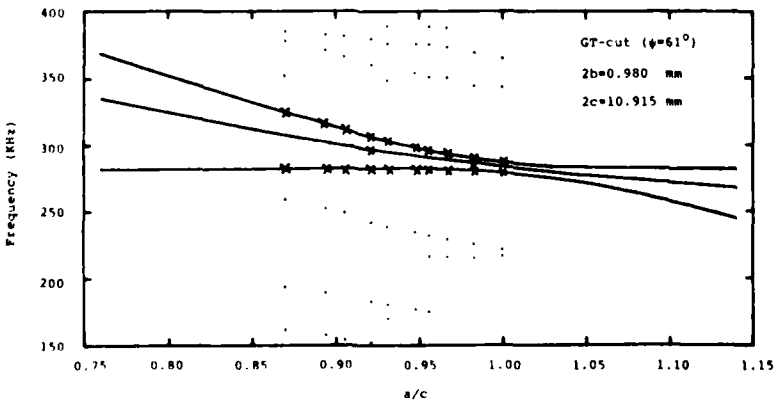


Fig 5 Frequency spectrum for GT-cut (ϕ = 61°) plates; calculated values in solid lines, experimental values in crosses and dots

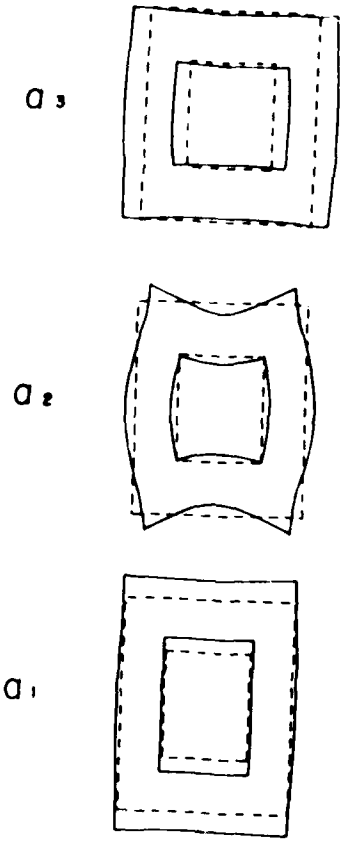


Fig. 6 Mode shapes for rectangular GT-cut (ψ = 51°) plate, a/c = 1.86

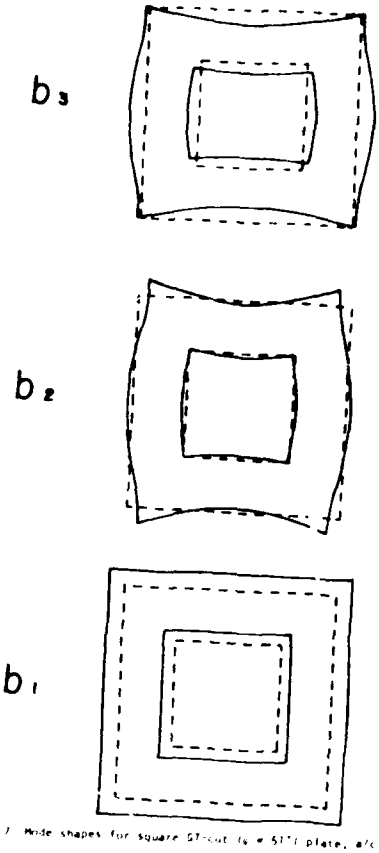


Fig. 7 Mode shapes for square GT-cut (ψ = 51°) plate, a/c = 1.00

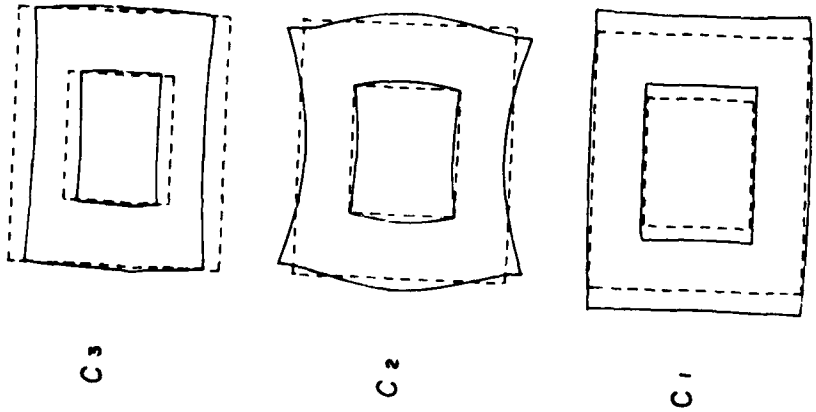


Fig. 8 Mode shapes for rectangular GT-cut (ψ = 51°) plate, a/c = 1.20

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

COUPLED THICKNESS-SHEAR AND THICKNESS-TWIST RESONANCES IN UNELECTRODED RECTANGULAR AND CIRCULAR AT-CUT QUARTZ PLATES

H.F. Tiersten
Department of Mechanical Engineering, Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, New York 12181

R.C. Smythe
Piezo Technology Incorporated
Orlando, Florida 32804

Abstract

The equation of coupled thickness-shear and thickness-twist vibrations is applied in the analysis of unelectroded rectangular and circular AT-cut quartz plates with free-edges. In the case of the rectangular plate the appropriate edge conditions are found and the solutions are obtained. In the case of the circular plate the two coefficients appearing in the planar differential operator are written as a sum of an isotropic part plus a deviation. The eigensolutions for the nearby isotropic case are found and the resonant frequencies for the actual anisotropic case are obtained from the equation for the first perturbation of the eigenfrequency from the isotropic solution. For the circular plate the results have been compared with measurements of the fundamental, and the third harmonic overtones and the agreement is quite good.

1. Introduction

Both trapped energy resonators and monolithic crystal filters are fabricated from rectangular and circular AT-cut quartz blanks. Before the deposition of the electrodes the resonant frequency of the fundamental or some other convenient harmonic overtone is measured. The thickness frequency, which is employed in the design of the device, if analytical design procedures are used, must be obtained from the measurement. Consequently, the relation between the thickness frequency and the measured frequency of the unelectroded rectangular or circular plate with free edges is of practical interest. In this paper the equation of coupled thickness-shear and thickness-twist vibrations[1,2] is applied in the analysis of unelectroded rectangular and circular AT-cut quartz plates with free edges. In the case of the rectangular plate a simple equation relating the measured frequency to the thickness-frequency is obtained. In the case of the circular plate the solution for the nearby isotropic plate is found and the corrected frequency is obtained from the equation for the first perturbation of the eigenfrequency due to the anisotropy and modified free edge condition. However, it turns out that the perturbation of the eigenfrequency resulting from the anisotropy vanishes for all modes except those with one angular nodal plane. Furthermore, the perturbation term due to the modified free-edge condition is essentially negligible for all modes for which the equations of coupled thickness-shear and thickness-twist are valid.

The results of calculations are presented for the circular plate and measurements for a practical case are compared with calculated results.

2. Basic Equations

Plan views of the unelectroded rectangular and circular AT-cut quartz plates are shown in Figs.1 and 2, respectively, along with the associated coordinate systems. The x_2-axis points out of each figure and the x_1-axis is in the digonal direction. In Fig.2 the cylindrical coordinates r and θ are shown in addition to the Cartesian coordinates x_1 and x_3. Note that θ is measured from the x_3-axis.

The differential equation of coupled thickness-shear and thickness-twist vibrations for the nth harmonic of an unelectroded AT-cut quartz plate may be written in the form[1,2]

$$M_n u_{,11} + c_{55} u_{,33} - \frac{n^2\pi^2}{4h^2}\bar{c}_{66} u + \rho\omega^2 u = 0, \quad n = 1,3,5,\ldots, \tag{2.1}$$

when the forcing term is omitted and where

$$M_n = c_{11} + (c_{12} + c_{66})r + \frac{4(r\bar{c}_{66} - c_{66})(c_{22}r + c_{12})\cot \kappa n\pi/2}{c_{22} n\pi\kappa}, \tag{2.2}$$

with

$$\bar{c}_{66} = c_{66} + \frac{e_{26}^2}{\epsilon_{22}}, \quad r = \frac{c_{12} + c_{66}}{\bar{c}_{66} - c_{22}}, \quad \kappa = \left(\frac{\bar{c}_{66}}{c_{22}}\right)^{1/2}. \tag{2.3}$$

In (2.1) and throughout it is understood that the time dependence of the standing wave has been factored out and the subscript 1 on the displacement field $u \equiv u_1$ has been omitted. We employ the convention that a comma followed by an index denotes partial differentiation with respect to a space coordinate and the usual compressed notation for tensor indices. In (2.1) ω is the circular frequency of the mode and 2h is the thickness of the plate. Equation (2.1) satisfies the differential equations and boundary conditions on the major surfaces of the plate to second order in the small wavenumbers along the plate[2] in the absence of certain relatively small unimportant elastic constants and because of small piezoelectric coupling.

In the case of the rectangular plate the boundary conditions on the free edges normal to x_1 and x_3 must still be considered. It has been shown that to lowest order, which is consistent with (2.1), the non-negligible pertinent constitutive equations take the form[1,3]

$$T_{11} = c_{11}u_{1,1} + c_{12}u_{2,2},$$

$$T_{12} = \bar{c}_{66}u_{1,2}, \quad T_{31} = c_{55}u_{1,3}, \tag{2.4}$$

basically since u_2 is one order of magnitude smaller than u_1 and u_3 is two orders of magnitude smaller than u_1 for small wavenumbers along the plate. Further consideration of this ordering and the form of the uniqueness theorem[4] reveals that on a free-edge normal to x_1 we need not satisfy any condition on T_{13} and should satisfy

$$T_{11} = 0, \quad T_{12} = 0, \quad \text{at } x_1 = \pm \ell, \tag{2.5}$$

only. For the same reasons[5] and, in fact, as already utilized[1], on a free-edge normal to x_3 we need not satisfy any conditions on T_{32} and T_{33} and should satisfy

$$T_{31} = 0, \quad \text{at } x_3 = \pm w, \tag{2.6}$$

only. Since in the approximation of coupled thickness shear and thickness-twist we have eliminated the flexural branch and, in fact, in all instances all other branches except the one retained, on an edge normal to x_1 we can satisfy only one condition and we have the two conditions in (2.5). However, an examination of (2.4) along with the solution[1,3] for small wave numbers along the plate reveals[3] that T_{11} is of the order of $u_{1,1}$ which is small, while T_{12} is of the order of u_1, which is large. Clearly then, we take $(2.5)_1$ to be satisfied approximately because of the small wavenumber along the plate and require the solution to satisfy $(2.5)_2$ only, which with $(2.4)_2$ yields

$$u_{,2} = 0 \text{ at } x_1 = \pm \ell. \tag{2.7}$$

Equation (2.6) with $(2.4)_3$ yields

$$u_{,3} = 0, \quad \text{at } x_3 = \pm w. \tag{2.8}$$

Although Eq. (2.1) is a useful form for the rectangular plate it is not a particularly useful form for the circular plate. For the circular plate we rewrite Eq. (2.1) in the form

$$c_n u_{,aa} + A_n + \rho(\omega^2 - \omega_n^2)u = 0, \tag{2.9}$$

where

$$\omega_n^2 = (\bar{c}_{66}/\rho)(n^2\pi^2/4h^2), \quad A_n = a_n(u_{,11} - u_{,33}), \tag{2.10}$$

and

$$c_n = \frac{M_n + c_{55}}{2}, \quad a_n = \frac{M_n - c_{55}}{2}, \tag{2.11}$$

in which c_n is the mean or isotropic constant, a_n is the anisotropic constant, A_n is the anisotropy term and we have introduced the convention that the subscript a takes the values 1 and 3 and skips 2 and employed the summation convention for repeated indices. We will obtain the natural frequencies of the modes of the anisotropic circular plate by finding the perturbations in eigenfrequencies from the solutions for the nearby isotropic plate due to the anisotropy. To this end we omit A_n and write

$$c_n v_{,aa} + \rho(\omega_\nu^2 - \omega_n^2)v = 0, \tag{2.12}$$

as the equation for the nearly unperturbed isotropic problem, in which v is the unperturbed isotropic variable and ω_ν is the unperturbed eigenfrequency of the nearby isotropic problem. Since we are concerned with not too large anharmonics of thin plates, for which $h/R \ll 1$, the large condition in (2.7) can be shown[1] to dominate except in the vicinity of $\theta = 0$ and $\theta = \pi$. Consequently, for the unperturbed nearby isotropic problem the traction-free edge condition is taken in the form

$$v = 0, \quad \text{at } r = R, \tag{2.13}$$

and we recognize that we must consider some form of edge perturbation term due to the use of the incorrect edge condition in (2.13) over a small portion of the free cylindrical surface in this unperturbed problem.

It can be shown[6] that the equation for the perturbation in the eigenfrequency can be written in the form

$$\Delta_\nu = \frac{H_\nu}{2\omega_\nu N_\nu^2}, \quad \omega = \omega_\nu - \Delta_\nu, \tag{2.14}$$

where ω_ν and ω are the unperturbed and perturbed eigenfrequencies, respectively, and

$$H_\nu = -\oint_C n_a c_n v_{\nu,a} u\, dS + \int_S A_n v_\nu\, dS, \tag{2.15}$$

where n_a is the outwardly directed unit normal to the circular edge C enclosing the area S and we have taken the liberty of omitting the irrelevant x_2-dependence. Hence

$$N_\nu^2 = \int_S \rho v_\nu^2\, dS, \tag{2.16}$$

and, of course, from $(2.10)_2$

$$A_n = a_n(v_{\nu,11} - v_{\nu,33}), \tag{2.17}$$

and the variable u is to be selected in order to compensate for (2.13) being incorrect over a small portion of the edge.

3. Rectangular Plate

The four independent solutions for the different modes of coupled thickness-shear and thickness-twist vibrations for the nth harmonic of the rectangular plate with free-edges may be written in the forms

$$u = B \sin \frac{n\pi x_2}{2h} \cos \xi x_1 \cos \nu x_3 ,$$

$$u = C \sin \frac{n\pi x_2}{2h} \cos \xi x_1 \sin \nu x_3 ,$$

$$u = D \sin \frac{n\pi x_2}{2h} \sin \xi x_1 \cos \nu x_3 ,$$

$$u = E \sin \frac{n\pi x_2}{2h} \sin \xi x_1 \sin \nu x_3 , \qquad (3.1)$$

each of which satisfies (2.1) provided

$$\omega^2 = \frac{1}{\rho}\left[\bar{c}_{66} \frac{n^2\pi^2}{4h^2} + M_n \xi^2 + c_{55} \nu^2\right] . \qquad (3.2)$$

Each of Eqs. (3.1) respectively satisfies the edge conditions in (2.7) and (2.8) provided

$$\xi = \frac{m\pi}{2\ell} , \ m = 1,3,5,\ldots , \ \nu = \frac{p\pi}{2w} , \ p = 0,2,4,6,\ldots ,$$

$$\xi = \frac{m\pi}{2\ell} , \ m = 1,3,5,\ldots , \ \nu = \frac{p\pi}{2w} , \ p = 1,3,5,\ldots ,$$

$$\xi = \frac{m\pi}{2\ell} , \ m = 2,4,6,\ldots , \ \nu = \frac{p\pi}{2w} , \ p = 0,2,4,6\ldots ,$$

$$\xi = \frac{m\pi}{2\ell} , \ m = 2,4,6,\ldots , \ \nu = \frac{p\pi}{2w} , \ p = 1,3,5,\ldots . \qquad (3.3)$$

Clearly, the substitution of (3.3) into (3.2) yields

$$\omega^2 = \frac{1}{\rho}\left[\bar{c}_{66} \frac{n^2\pi^2}{4h^2} + M_n \frac{m^2\pi^2}{4\ell^2} + c_{55} \frac{p^2\pi^2}{4w^2}\right] ,$$

$$n = 1,3,5,\ldots , \ m = 1,2,3,\ldots . \ p = 0,1,2,3,\ldots . \qquad (3.4)$$

Since the relation between the measured frequency ω and the thickness-frequency ω_n in (2.10) is wanted, from (3.4) we write

$$\omega^2 = \omega_n^2 + \frac{M_n}{\rho} \frac{m^2\pi^2}{4\ell^2} + \frac{c_{55}}{\rho} \frac{p^2\pi^2}{4w^2} . \qquad (3.5)$$

For the fundamental mode of any odd harmonic, $p = 0$ and $m = 1$, and from (3.5) we obtain

$$\omega_F^2 = \omega_n^2 + (M_n/\rho)(\pi^2/4\ell^2) . \qquad (3.6)$$

Furthermore, from $(3.1)_1$ and $(3.3)_1$ we find for any n that for the fundamental mode the solution function takes the form

$$u = B \sin \frac{n\pi x_2}{2h} \cos \frac{\pi x_1}{2\ell} , \qquad (3.7)$$

which shape for the fundamental mode was identified by Sykes[5] and observed by Van Dyke[7] as discussed recently by Sherman[11].

4. Circular Plate

In the case of the circular plate we begin by obtaining the unperturbed solution to the isotropic equation (2.12), which may be written in cylindrical coordinates thus

$$\frac{\partial^2 v}{\partial r^2} + \frac{1}{r}\frac{\partial v}{\partial r} + \frac{1}{r^2}\frac{\partial^2 v}{\partial \theta^2} + \beta^2 v = 0 , \qquad (4.1)$$

where

$$\beta_n^2 = \frac{\rho}{c_n}(\omega_\nu^2 - \omega_n^2) . \qquad (4.2)$$

The solutions satisfying (4.1) may be written in the form

$$v = e^{im\theta} f(r) , \ m = 0,1,2,3,\ldots . , \qquad (4.3)$$

the substitution of which in (4.1) yields

$$f'' + \frac{1}{r} f' + \left(\beta_n^2 - \frac{m^2}{r^2}\right) f = 0 , \qquad (4.4)$$

which is Bessel's differential equation. From (4.3) and (2.13) we obtain the boundary condition

$$f = 0 \text{ at } r = R . \qquad (4.5)$$

The solutions satisfying (4.4) and not diverging at $r = 0$ may be written in the form

$$f = J_m(\beta_n r) , \qquad (4.6)$$

where J_m is the Bessel function of order m. The substitution of (4.6) into (4.5) yields

$$J_m(\beta_n R) = 0 , \qquad (4.7)$$

which enables us to write

$$\beta_{nmp} = Z_{nmp}/R , \qquad (4.8)$$

where the Z_{nmp} denote the zeros of J_m for nonzero Z_{nmp}, and we note that for the solution to be valid we must have

$$\beta_{nmp} \ll n\pi/2h . \qquad (4.9)$$

Since the edge condition (2.13) is incorrect only over a small portion of the edge in the vicinity of $\theta = 0$ and $\theta = \pi$ and from (2.7) and (2.8) the size of the region increases with β_p and decreases with $n\pi/2h$, the edge perturbation term is selected in the form

$$u = \pm \kappa e^{im\theta} w(r) \quad \begin{matrix} -\frac{\pi}{2} < \theta < \pi/2 \\ \frac{\pi}{2} < \theta < \frac{3\pi}{2} \end{matrix} , \qquad (4.10)$$

where

$$\kappa = \frac{\beta_{nmp} 2h}{n\pi} \cos\theta , \ w = J_m(\alpha_n r) , \qquad (4.11)$$

with

$$\frac{dJ_m}{dr} = 0 \text{ at } r = R . \qquad (4.12)$$

Equation (4.12) yields

$$\alpha_{nmp} = \zeta_{nmp}/R , \qquad (4.13)$$

where the ζ_{nsp} denote the zeros of J_n' with $\zeta_{n00} = 0$ and all other $\zeta_{nsc} \neq 0$ $(m \neq 0)$, but rather determined by the first nonzero point for zero J_n'. Substituting from (4.3) into (2.17) and utilizing the transformation relations between Cartesian and cylindrical coordinates, we find that the anisotropy term in the perturbation integral (2.15) takes the form[6]

$$A_n = a_n\Big[\Big(\frac{2\beta_{nmp}}{r}J_m' + \Big(\beta_{nmp}^2 - \frac{2m^2}{r^2}\Big)J_m\Big)\cos 2\theta + 2im\Big(\frac{\beta_{nmp}}{r}J_m' - \frac{J_m}{r^2}\Big)\sin 2\theta\Big]e^{im\theta}, \quad (4.14)$$

where $\mathrm{Re}A_n$ comes from the unperturbed solution function

$$v_m = \cos m\theta J_m(\beta r), \quad (4.15)$$

which corresponds to a nodal plane along x_1 and $\mathrm{Im}A_n$ comes from the unperturbed solution function

$$v_m = \sin m\theta J_m(\beta r), \quad (4.16)$$

which corresponds to a nodal plane along x_3. Substituting from (4.8) into (4.2), we obtain the eigenfrequency ω_{nsp} of the unperturbed isotropic solution in the form

$$\omega_{nmp}^2 = \omega_n^2 + \frac{c_n}{\rho}\beta_{nmp}^2. \quad (4.17)$$

From (2.14) we see that the perturbation in eigenfrequency can be written in the form

$$\omega = \omega_{nmp} - \Delta_{nmp}, \quad \Delta_{nmp} = \frac{H_{nmp}}{2\omega_{nmp}N_{nmp}^2}, \quad (4.18)$$

where from (2.15) we can write

$$H_{nmp} = H_{nmp}^S + H_{nmp}^A, \quad (4.19)$$

in which H_{nsp}^s is the first, or edge perturbation, term in (2.15) and H_{nsp}^A is the second, or anisotropy perturbation, term in (2.15). Substituting from (4.10) and (4.11) into the first term in (2.15), we obtain[6]

$$H_{nmp}^S = -c_n\beta_{nmp}J_m'(\beta_{nmp}R)\frac{\beta_{nmp}2h}{n\pi}J_m(\alpha_{nmp}R)R\mathcal{J}_m^S, \quad (4.20)$$

where

$$\mathcal{J}_m^S = \int_{-\pi/2}^{\pi/2}\cos\theta\begin{bmatrix}\cos^2 m\theta\\ \sin^2 m\theta\end{bmatrix}d\theta - \int_{\pi/2}^{3\pi/2}\cos\theta\begin{bmatrix}\cos^2 m\theta\\ \sin^2 m\theta\end{bmatrix}d\theta, \quad (4.21)$$

and the upper terms in the brackets are from (4.15) while the lower terms are from (4.16). Substituting from $\mathrm{Re}A_n$ in (4.14) and (4.15) and $\mathcal{J}_m A_n$ in (4.14) and (4.16), respectively, in the second term in (2.15), we obtain the two equations

$$H_{nmp}^A = a_n\int_0^R rdr\int_0^{2\pi}d\theta\Big[\Big(2\frac{\beta_{nmp}}{r}J_m' + \Big(\beta_{nmp}^2 - \frac{2m^2}{r^2}\Big)J_m\Big)\cos 2\theta\begin{pmatrix}\cos m\theta\\ \sin m\theta\end{pmatrix} \pm \Big(\frac{J_m}{r^2} - \frac{\beta_{nmp}}{r}J_m'\Big)\sin 2\theta\begin{pmatrix}\sin m\theta\\ \cos m\theta\end{pmatrix}\Big]J_m\begin{pmatrix}\cos m\theta\\ \sin m\theta\end{pmatrix}, \quad (4.22)$$

in which the upper equation is from the former substitution and the lower equation is from the latter. Substituting from (4.15) into (2.16), we obtain

$$N_{nmp}^2 = \rho\int_0^{2\pi}d\theta\int_0^R rdr\,J_m^2(\beta_{nmp}r)\cos^2 m\theta\,d\theta, \quad (4.23)$$

and when (4.16) is substituted the cos is replaced by a sin. Evaluation of the angular integrals in (4.22) reveals that[5]

$$H_{nmp}^A = 0 \quad \text{unless } m = 1, \quad (4.24)$$

while for $m = 1$ the radial integral in each of (4.22) reduces[6] to the well-known normalization integral[10] for Bessel functions and we obtain

$$H_{n1p}^A = \pm a_n\frac{\pi}{2}\beta_{n1p}^2\frac{R^2}{2}J_2(\beta_{n1p}R)J_0(\beta_{n1p}R). \quad (4.25)$$

This very fortuitous circumstance that arises in (4.24) and (4.25) means that except for $m = 1$ the perturbation in the eigenfrequency ω_{nsp} due to the planar anisotropy vanishes and for $m = 1$ we obtain a very simple form to evaluate for H_{n1p}. Furthermore, forgetting the very small perturbation due to the edge condition, for $m \neq 1$ the eigenfrequency ω is influenced only by the mean (or isotropic) constant c_n and is independent of the deviation (or anisotropy) a_n. However, the real anisotropic mode shape differs from the isotropic mode shape we have found even though the eigenfrequencies are the same to lowest order in a_n/c_n. Substituting from (4.15) into (4.23), we obtain

$$N_{\substack{nmp\\ m\neq 0}}^2 = -\rho\pi\frac{R}{2}J_{m+1}(\beta_{nmp}R)J_{m-1}(\beta_{nmp}R), \qquad N_{nop}^2 = \rho\pi R^2J_1^2(\beta_{nop}R). \quad (4.26)$$

5. Results

Calculations for unelectroded circular AT-cut quartz plates with free-edges can now readily be performed using Eqs. (4.17) - (4.21) and (4.24) - (4.26). However, we first wish to indicate the degree of anisotropy present in the families of anharmonics for each of the first few harmonics. Accordingly, we first note that[11]

$$c_{55} = 6.88\times 10^{10}\ \mathrm{N/m^2}, \quad (5.1)$$

and in Table I we show the values of M_n, c_n, a_n and a_n/c_n for $n = 1, 3, 5$ for AT-cut quartz. The table shows that for the third harmonic the AT-cut is nearly isotropic in the $x_1 - x_3$ plane for the anharmonics, while for the fundamental it is quite anisotropic. Since the perturbation analysis in Sec. 4 shows that for $m \neq 1$ there is no perturbation in eigenfrequency linear in the anisotropy a_n/c_n, the results should be quite accurate even for the quite anisotropic fundamental for $m \neq 1$ because the first correction to the eigenfrequency for the

nearby isotropic plate would be proportional to $(a_n - c_n)^2$, which from the last line in Table I is always quite small. Nevertheless, the theory is obviously more accurate for the fifth harmonic than for the fundamental.

Typical calculations for an unelectroded circular AT-cut quartz plate with free-edges are shown in Tables II and III. The plate has a nominal infinite plate thickness-frequency of 15000 kHz, which corresponds to a thickness 2h of 4.34×10^{-3} in., and a diameter 2R of .310 in. Table II is for n = 1 and Table III is for n = 3. The columns labeled frequency contain $\omega/2\pi$ from $(4.18)_1$ and the columns labeled delta contain Δ_{mn}. It is clear from the tables that the values of Δ_{mn} for m ≠ 1 are much smaller than those for m = 1. This means that the perturbation in frequency due to the anisotropy (when nonzero) is much larger than the perturbation due to the edge condition. In addition, the perturbation due to the edge condition very rapidly gets smaller with increasing harmonic. In fact the perturbation due to the edge condition is negligible for measurement purposes (because of much greater scatter in the data) except for high anharmonics of low harmonics, which are not of practical interest. It is also clear from Tables II and III that for m = 1 the perturbation in eigenfrequency for the anharmonics of the third harmonic are much less than for those of the fundamental. This is a result of the fact that the anisotropy of the third harmonic is so much smaller than that of the fundamental coupled with the larger order of the overtone as shown by (4.18). Table IV shows the frequencies of the fundamental f_{100} and third harmonic f_{300} measured on ten nominally identical units .310 in. in diameter. The fixture electrode diameter is .09 in. and the air gap is .012 in. Note the typical scatter in the measurements. The mean of the measured values of $3f_{100} - f_{300}$ is 25.4 kHz. From Tables II and III for a nominal infinite plate fundamental thickness frequency of 15000 kHz, which corresponds to a thickness 2h of 4.34×10^{-3} in., and a plate diameter 2R of .310 in., we obtain

$$3f_{100} - f_{300} = 28.33 \text{ kHz}, \qquad (5.2)$$

in quite good agreement with the measured mean of 25.4 kHz in view of the scatter in the data.

Acknowledgements

This work was supported in part by the Army Research Office under Contract No. DAAG 29-79-C-0088.

References

1. H.F. Tiersten, "Analysis of Trapped Energy Resonators Operating in Overtones of Coupled Thickness-Shear and Thickness-Twist," J. Acoust. Soc. Am., 59, 879 (1976).
2. H.F. Tiersten and R.C. Smythe, "An Analysis of Contoured Crystal Resonators Operating in Overtones of Coupled Thickness-Shear and Thickness-Twist," J. Acoust. Soc. Am., 65, 1455 (1979).
3. H.F. Tiersten, "Analysis of Intermodulation in Thickness-Shear and Trapped Energy Resonators," J. Acoust. Soc. Am., 57, 667 (1975).
4. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.5, Sec.4, Chap.15, Sec.5.
5. For more detail see H.F. Tiersten and R.C. Smythe, "Coupled Thickness-Shear and Thickness-Twist Vibrations of Unelectroded AT-Cut Quartz Plates," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, NY 12181.
6. R.A. Sykes, "Modes of Motion in Quartz Crystals," Bell Syst. Tech. J., 23, 52 (1944). See Eq.(6.9).
7. K.S. Van Dyke, "Piezoelectric Studies of Measurements," Proceedings of the 10th Annual Symposium on Frequency Control, Signal Corps Engineering Laboratories, Fort Monmouth, New Jersey, 2 (1956).
8. J.H. Sherman, "Measurement of the Characteristic Frequency of an AT-Cut Plate," Proceedings of the 31st Annual Symposium on Frequency Control U.S. Army Electronics Command, Fort Monmouth, New Jersey, 108 (1977).
9. We cannot evaluate the edge perturbation term in a neat way because we do not have a direct means for the determination of u. Rather than introducing another nonperturbation procedure for the determination of the small correction in eigenfrequency resulting from the edge condition being incorrect over a very small region, we prefer to use the perturbation procedure with the selection of u in (4.10), which we believe overestimates the essentially negligible perturbation in eigenfrequency somewhat because of the behavior of cos θ.
10. P.M. Morse and H. Feshbach, Methods of Theoretical Physics (McGraw-Hill, New York, 1953), p.765.
11. Ref.4, Chap.7, Eqs.(7.36). The constants were obtained from R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 110, 1060 (1958).

TABLE I

n		1	3	5
M_n	10^{10} N/M^2	11.00	7.58	9.01
c_n	10^{10} N/M^2	8.94	7.23	7.95
a_n	10^{10} N/M^2	2.06	.35	1.06
a_n/c_n		.23	.05	.13

TABLE II

n = 1

Nodal Axis	Indices m	Indices p	Frequency (kHz)	Delta (kHz)
	0	0	15010.392	0.233
	0	1	15055.146	0.753
	0	2	15135.290	1.719
	0	3	15250.324	3.077
x	1	0	15023.269	3.690
	1	1	15078.266	11.921
	1	2	15164.293	24.738
	1	3	15280.860	41.933
z	1	0	15029.467	-2.507
	1	1	15098.956	-8.769
	1	2	15207.518	-18.487
	1	3	15354.351	-31.559
x	2	0	15047.660	0.736
	2	1	15128.068	1.587
	2	2	15243.288	2.847
	2	3	15392.639	4.460
z	2	0	15047.554	0.842
	2	1	15127.841	1.814
	2	2	15242.882	3.254
	2	3	15392.001	5.098

TABLE III

n = 3

Nodal Axis	Indices m	Indices p	Frequency (kHz)	Delta (kHz)
	0	0	45002.845	0.021
	0	1	45015.032	0.068
	0	2	45036.945	0.156
	0	3	45068.578	0.281
x	1	0	45007.047	0.229
	1	1	45023.656	0.732
	1	2	45049.745	1.523
	1	3	45085.300	2.598
z	1	0	45007.399	-0.123
	1	1	45024.834	-0.447
	1	2	45052.222	-0.954
	1	3	45089.545	-1.647
x	2	0	45013.004	0.066
	2	1	45034.958	0.144
	2	2	45066.610	0.260
	2	3	45107.962	0.411
z	2	0	45012.994	0.076
	2	1	45034.937	0.164
	2	2	45066.573	0.297
	2	3	45107.903	0.469

TABLE IV

Unit No.	Measured Frequencies (kHz) f_{100}	f_{300}	$3f_{100} - f_{300}$
1	15063.7	45168.1	23.0
2	15068.2	45180.7	23.9
3	15061.9	45164.0	21.7
4	15070.7	45183.0	29.1
5	15065.9	45173.0	24.7
6	15070.7	45186.9	25.2
7	15070.6	45181.0	30.8
8	15069.8	45182.9	26.5
9	15063.8	45169.0	22.4
10	15071.3	45187.1	26.8

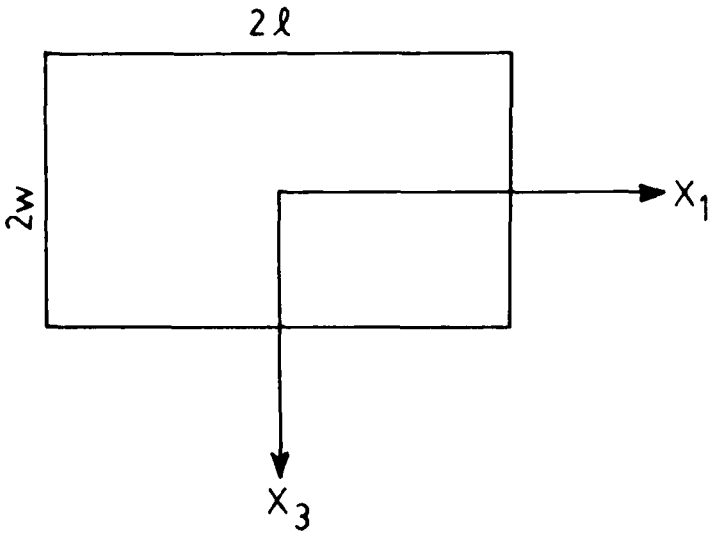


Figure 1 Plan View of the Rectangular Plate

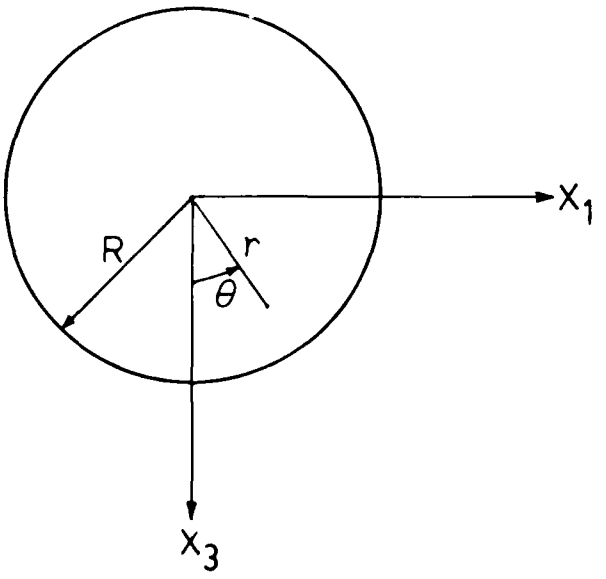


Figure 2 Plan View of the Circular Plate

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

AIR-GAP PROBE EVALUATION OF THIN QUARTZ PLATES

Lawrence Dworsky
Motorola Inc.
Schaumberg, Illinois

and

Garth Kennedy
Motorola Inc.
Franklin Park, Illinois

Summary

The electrical response of a small air-gap probe placed over a convex AT-cut quartz plate is analyzed as a function of frequency and of probe location. It is shown that the response is directly derivable from Tiersten and Smythe's analysis of a fully electroded convex quartz plate.

A model is derived which allows for identification of the (rotated) x and z crystallographic axes, and measurement of the radius of curvature and location of the thickest point of the crystal plate. Experimental results are shown which demonstrate the validity of the model.

An automated system is described which characterizes crystal plates in terms of radius of curvature and location of the thickest part of the plate. This system basically consists of an automatic network analyzer and a motor-driven-stage scanning air-gap probe which operates under network analyzer control. Sample data demonstrating the repeatability and accuracy of this system is presented.

Introduction

In the manufacture of thin AT-cut quartz plates for piezoelectric resonators or coupled resonator (monolithic) filters, it is necessary to measure the thickness of each plate by suspending the plate between the electrodes of an air dielectric capacitor and measuring the input impedance seen by this capacitor (Figure 1). Since the resonant frequencies of the plate vary essentially as the inverse of the plate thickness, resonances observed at the capacitor terminals (using appropriate impedance matching circuits as required) can be correlated to the plate thickness. The data obtained can then be used to predict the mass and lateral dimensions of metal electrodes which, when deposited onto the plate, would produce a crystal resonator (or filter) at a desired frequency.

The manufacture of crystal devices always includes the provision for a final electrode mass correction while monitoring the actual resonator frequency(ies). This allows for an overall tolerance in the electrode dimensions, upper and lower electrode alignment, deposited metal mass, variations in crystal angle, and variations in the thickness of the crystal plate itself. So long as the effects of the latter variation are small compared to the (sum of the) others, it is not necessary to investigate the mechanism of the air-gap frequency measurement too closely. However, a general trend toward more precise control in the manufacture of quartz devices, particularly for high frequency, high overtone use is changing this conclusion. Also, when building multi-resonator devices, it is important to inspect the thickness profile of the quartz plate in detail so that the relative frequencies of the resonators on the plate may be predicted.

The intuitive approach to measuring local thickness precisely is to construct a probe whose lateral dimensions are much smaller than those of the quartz plate and to mount it in a fixture which is provided with a means of transversing the surface of the plate while keeping the electrode separation constant. When this is done the results are at first both surprising and disappointing. Figure 2 shows oscilligraphs of the reflection coefficient magnitude obtained at several locations about the surface of a 3mm radius, ~38μ thick plate, measured at the 5th overtone. As may be seen, there are multiple resonances being excited, and there is no simple correlation between the data and the thickness profile of the plate.

The purpose of this paper is to develop a model for the air-gap probe measurement system which will explain results such as those shown in Figure 2, and then to propose a measurement procedure which will facilitate the extraction of the desired information from the data. Comparison of measured data such as that shown in Figure 2 with predictions of the model will be

shown.

The Convex Plate

In 1979, Tiersten and Smythe published an analysis of fully electroded convex surface AT-quartz resonators[1]. Their results showed that an infinite number of modes (at each overtone) are "trapped" beneath the convex surface(s). We can use these results to describe the resonance modes of an air-gap structure by setting the electrode mass equal to zero:

$$\omega^2 = \left[\frac{n\pi}{2h}\right]^2 \frac{\bar{C}_{66}}{\rho}\left\{1 + \frac{1}{N\pi}\left(\frac{2h}{R}\right)^{\frac{1}{2}}\left[\left(\frac{M_m}{\bar{C}_{66}}\right)^{\frac{1}{2}}(2m+1) + \left(\frac{C_{55}}{\bar{C}_{66}}\right)^{\frac{1}{2}}(2p+1)\right]\right\} \quad (1)$$

where all terms are defined as in reference 1.

The eigensolutions which Tiersten and Smythe found consisted of the separated-variable product of a sinusoid in y and Hermite functions in x and z. These Hermite functions are:

$$X_{mn} = \exp(-\alpha x^2/2) H_m(\alpha^{\frac{1}{2}} x) \quad (2)$$

$$Z_{pn} = \exp(-\beta z^2/2) H_p(\beta^{\frac{1}{2}} z) \quad (3)$$

H_m and H_p are Hermite Polynomials of order m and p respectively[2]. Hermite polynomials are even or odd functions of x (or z) depending upon whether m (or p) is even or odd. Also,

$$\alpha^2 = n^2\pi^2 C_{66}/8Rh^3 M_m \quad (2a)$$

$$\beta^2 = n^2\pi^2 C_{66}/8Rh^3 C_{55} \quad (3a)$$

Now, consider the situation depicted in Figure 1. A probe whose surface area is small compared to that of the quartz plate is exciting the resonant modes of the plate. Due to the fact that the ratio of the electrode surface area to the plate surface area is small, especially at higher overtones we may assume that the resonant frequencies (and modes) are determined by the plate itself and are not noticeably perturbed by the location of or even the existence of the probe. The location and size of the probe will of course determine the "strength" of the excitation, which may be measured in terms of the motional capacitance seen by the probe. Using Lewis' relation for motional capacitance[3],

$$C_{mot} = \frac{\left[\iint D_y(x,z)\,ds\right]^2}{\rho\omega^2 \iiint u_x^2(x,y,z)\,dv} \quad (4)$$

The limits of integration in the denominator above are the entire volume of the plate. The limits of integration in the numerator are the surface area of the plate under the probe. From reference 1, the electric displacement current, D_y, is given by

$$D_y = e_{26}\frac{\partial u}{\partial y} - \epsilon_{22}\frac{\partial \phi}{\partial y} \quad (5)$$

Since there is no free charge in the quartz,

$$\nabla \cdot \underline{D} = \frac{\partial D_y}{\partial y} = e_{26}\frac{\partial^2 u}{\partial y^2} - \epsilon_{22}\frac{\partial^2 \phi}{\partial y^2} = 0 \quad (6)$$

Equation 6 suggests that the potential function, ϕ, has the form

$$\phi = \frac{e_{26}}{\epsilon_{22}} u(x,y,z) + yA(x,z) + B(x,z) \quad (7)$$

where A and B are unknown at present.

As was stated above, u is given by

$$u(x,y,z) = \sin\left(\frac{n\pi y}{2h}\right) X(x) Z(z) \quad (8)$$

Referencing our potential function to the plane y = 0, therefore,

$$B(x,z) = 0 \quad (9)$$

The homogeneous (shorted electrode) boundary condition is that

$$\phi(x,\pm h,z) = 0 \quad (10)$$

Therefore, from (7),

$$0 = \frac{e_{26}}{\epsilon_{22}} u(x,\pm h,z) \pm hA(x,z) \quad (11)$$

or,

$$A = \pm\frac{e_{26}}{\epsilon_{22} h} u(x,\pm h,z) \quad (12)$$

Differentiating (7) with respect to y,

$$A = \frac{\partial \phi}{\partial y} - \frac{e_{26}}{\epsilon_{22}}\frac{\partial u}{\partial y} \quad (13)$$

and from (5), therefore,

$$-\epsilon_{22} A = e_{26}\frac{\partial u}{\partial y} - \epsilon_{22}\frac{\partial \phi}{\partial y} = D_y \quad (14)$$

Comparing equations (14) and (12)

$$-D_y = \pm\frac{e_{26}}{h} u(x,\pm h,z) \quad (15)$$

Returning to the expression for motional capacitance (equation 4),

$$C_{mot} = \frac{e_{26}^2}{\rho\omega^2 h}\,\frac{\left[\iint u(x,h,z)\,ds\right]^2}{\iiint u^2(x,y,z)\,dv} \quad (16)$$

The above equation may be simplified by using (8)

$$C_{mot} = \frac{e^{2}\varepsilon_{6}}{\rho\omega^{2}h^{2}} \cdot \frac{\left[\iint X(x)Z(z)\right]^{2}}{\iint X^{2}(x)Z^{2}(z)\,dxdz} \qquad (17)$$

In the case of the convex plate, h in (17) is replaced by h_o.

In general, due to the nature of X(x) and Z(z), the integrals in (17) must be found numerically. In order to be able to compare analytic to experimental results, it is necessary to assume that, for the crystal plate under consideration, the radius of curvature is small enough and/or the diameter of the plate is large enough so that edge effects may be ignored.

Experimental Results

Consider the lowest few modes of the integrand of the numerator of (17):

m,p	
0,0	$[\exp(-\alpha x^2/2)\exp(-\beta z^2/2)]$
0,1	["] $2\beta^{\frac{1}{2}}z$
1,0	["] $2\alpha^{\frac{1}{2}}x$
0,2	["] $(4\beta z^2-2)$
1,1	["] $(4\alpha^{\frac{1}{2}}\beta^{\frac{1}{2}}xz)$
2,0	["] $(4\alpha x^2-2)$

The first three x terms of the above are plotted in Figure 3. From the form of (18), we may expect several things: Noting that the denominator of (17) is independent of the probe location, we see that

a) C_{mot} for the (n,0,0) mode is largest when the (circular) probe is centered at (x,z) = (0,0). The (n,0,0) mode is of course the lowest frequency n^{th} overtone mode.

b) C_{mot} for the (n,0,1), the (n,1,0), and all other modes for which m and/or p is odd are zero when the probe is located at (0,0).

c) Locating the probe along the z (x) axis will allow the excitation of odd order modes in z (x) but not in x (z).

d) Locating the probe arbitrarily in x and z will, in general, allow excitation of all possible modes.

e) Higher order terms are more dependent upon edge conditions than lower order terms and therefore will not fit the model as well.

Now consider again the crystal plate used in generating Figure 2. The (reflection coefficient) response at (x,z) = (0,0) is shown in Figure 4a. Figure 4b shows the response at (x,z) = (.18,.20) mm, a location which was chosen to maximize C_{mot} for the (5,0,0) mode - that is, to maximize the depth of the lowest frequency resonance (dip) shown. This means that the geometric center of this plate does not correspond to the (x,z) = (0,0) location referred to in the analysis above. If we denote the point (.18,.20) as (x_o,z_o), and since we can always describe an off-centered parabola by

$$f(x) = f_o + a(x-x_o)^2$$
$$= (f_o+ax_o^2) + ax^2 - 2(ax_o)x \qquad (19)$$

we may interpret the above result by saying that the plate has a linear or "wedge" contribution to its thickness variation in addition to its convexity. So long as $(x^2 + y^2)^{\frac{1}{2}}$ is sufficiently less than the radius of the plate that edge effects can still be neglected, we may expect the analytic predictions to hold - provided that we reference our (x,z) system to (x_o,z_o).

Returning to Figure 4b, note that there are two resonant modes, spaced close to each other, at 31 and 35 kHz above the (5,0,0) mode. Figure 5 is a plot of the frequency separation between f_{5mp} and f_{500} (at f_{500} = 100 mHz) versus radius of curvature of the plate (R). Remembering that at $(x,z) = (x_o,z_o)$, only even order modes (in both x and z) are excited, we may identify these modes as the (5,0,2) and the (5,2,0) modes for a plate with a radius of curvature of ≈65 meters.

Offsetting the probe by .2 mm in the z direction, we see a new mode appear beside the (5,0,0) mode, ≈15 kHz above it (Figure 4c). Referring again to Figure 5, this is the (5,0,1) mode. Similarly, offsetting the probe by .2 mm in the x direction we see the (5,1,0) mode (Figure 4d). If the probe were offset in both x and z we would see the (5,0,1) and (5,1,0) modes, as well as the (5,1,1) mode appearing between the (5,0,2) and (5,2,0) modes. Note that the (5,1,1) mode has a very small motional capacitance as compared to the other modes observed, and will therefore be more difficult to find.

Using (17) and (18) we may predict how C_{mot} will vary with probe location over the crystal surface. Figure 6 shows C_{mot} variation as the probe is moved along the z axis, at $x = x_o$. The probe location at which the (5,0,1) mode peaks can be used to calculate, through (17), the radius of curvature of the plate. This procedure has proven to be more rel-

table in practice than that of examining mode frequency separations.

Automated Measurement System

A system has been built which automatically examines a crystal plate and determines (x_0, z_0), and R. With reference to Figure 7, the system operates using the following procedure:

1. Orient the crystal plate so that the x axis of the plate in on an axis of the test fixture.

2. Scan the plate in the x and z directions to find the lowest frequency mode, (n,0,0), at the operating overtone. Find the location on the plate at which the response of this mode is strongest. This is (x_0, z_0).

3. Scan the z axis of the plate, keeping $x = x_0$, and locate the point at which the response of the (n,0,1) mode is the strongest. The (n,0,1) mode is the mode which is closest in frequency to that of the (n,0,0) mode.

4. Calculate R from the separation along the z axis of the response peaks of the (n,0,1) and (n,0,0) modes.

5. If reasonable data cannot be obtained after some predetermined maximum time, discontinue operation.

So long as the electrical resistance presented by the crystal plate at series resonance of all of the modes of interest (after impedance matching, if appropriate), is greater than the characteristic impedance of the reflection coefficient measurement equipment (typically 50 ohms), the magnitude of the reflection coefficient at resonance is monotonically proportional to C_{mot}. The relationship between the plate radius of curvature and the mode separation as described above, found by numerical integration of (17) is approximately

$$R = \frac{(S - a)^4}{b} \qquad (20)$$

where S = separation along the z axis between the (n,0,0) and the (n,0,1) mode peaks.

R = radius of curvature

a,b = constants, valid for ±1% frequency range about a nominal frequency.

A block diagram of the automated test system is shown in Figure 8. The test is run under microcomputer control. The test fixture translates the crystal plate under the probe, in response to measurements made by the (scalar) network analyzer. If a test runs too long or if the results are not reasonable the test is discontinued. Since the test requires locating response maximums and frequencies, substantial translational movement and several detailed frequency searches are required. Test times typically range from 15 to 35 seconds.

The accuracy of the automatic measurement is limited by four principal sources of error. These are:

1. Equipment errors - The resolution limit of the (digitized) network analyzer measurement.

2. Plate placement error - Caused by nonuniformities in the plate and the fixturing. This error only affects location of the point (x_0, z_0).

3. Frequency measurement error.

4. Translation error - Caused by the finite stepping motor step size and the control algorithms.

Estimates of the above errors were made for crystal plates operating at a frequency $f_{500} \approx 100$ mHz. One hundred plates were each measured one hundred times. The predicted and measured repeatabilities (one standard deviation) were

	(x_0, z_0) (location)	Mode Separation (peaks)
Anticipated:	.047 mm	.037 mm
Measured:	.067	.031
Minimum measured:	.034	.028
Maximum measured:	.081	.037

The mode separation errors can of course be related to radius of curvature errors through (20).

The usefullness of the system turned out to be limited by several factors. For plates with R > 1000 meters edge effects perturbed the data and the measurements did not fit the model well. For plates with R < 40 meters measurement errors become severe. Some plates simply do not have plano-convex or convex-convex surfaces.

Summary

The air-gap response of convex surface AT quartz plates has been analyzed and the results used to describe a set of measurements from which the surface

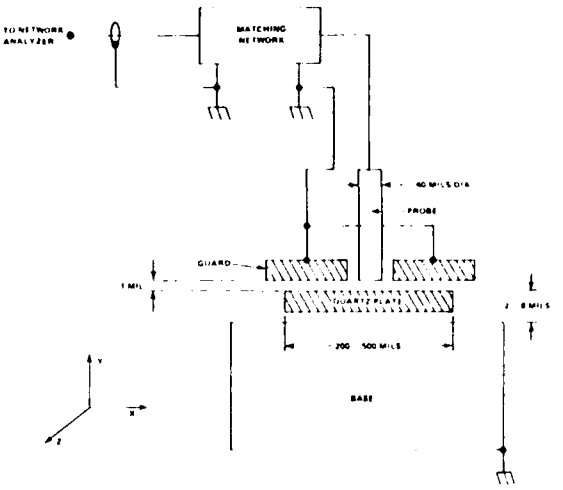
profile of the plate may be ascertained. An automated test system for making the measurements and performing the necessary calculations has been described. This system has proven to be very useful in characterizing quartz plates to be used for many tight tolerance applications and also for understanding the effects of quartz plate manufacturing parameters.

References

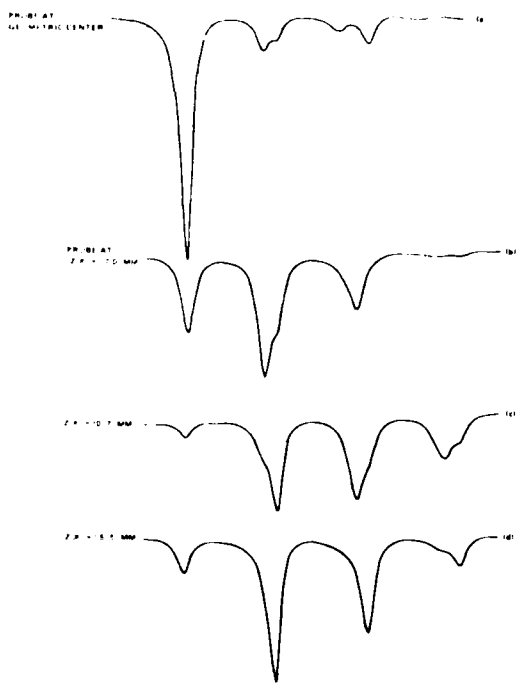
1. H. F. Tiersten & R. C. Smyther, "An Analysis of Contoured Crystal Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist." J. Acoust. Soc. Am., Vol. 65, #6, 6/79.

2. J. Matthews & R. Walker, "Mathematical Methods of Physics." W. A. Benjamin, Inc., New York, 1965.

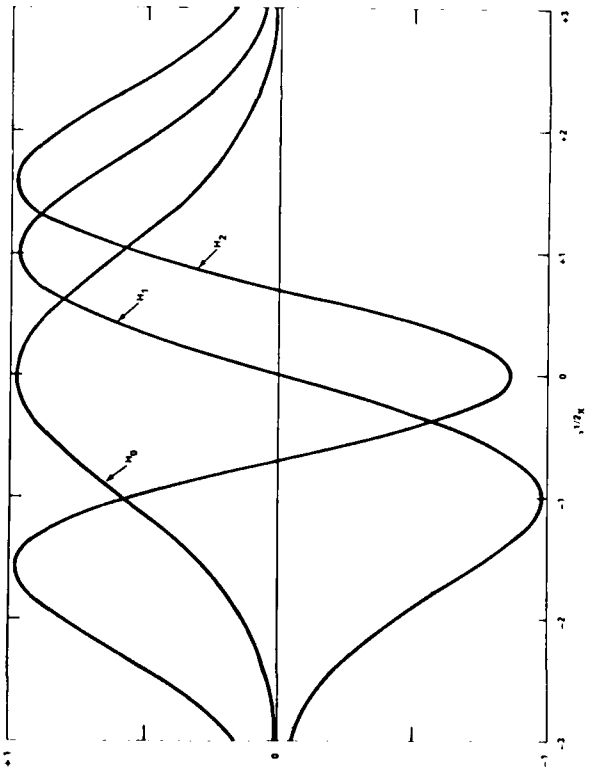
3. J. Lewis, "The Effect of Driving Electrode Shape on the Electrical Properties of Piezoelectric Crystals." Bell System Tech. Journal, Vol. 40, 9/61.



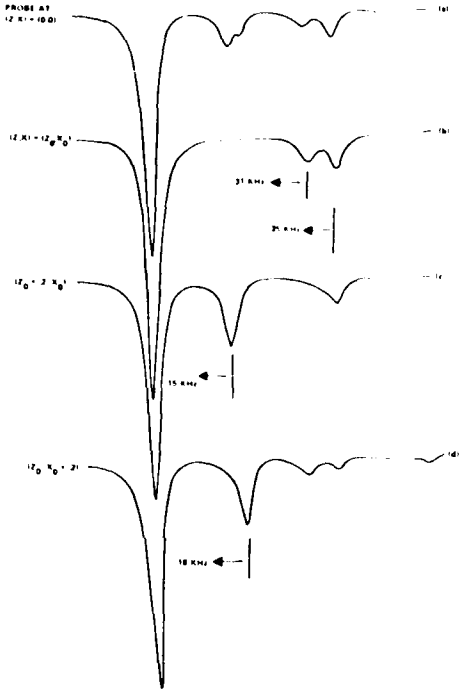
1. Schematic Drawing of Experimental Air-Gap Probe.



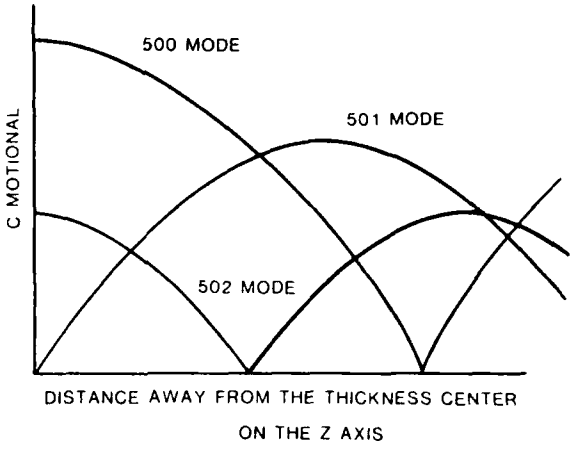
2. Magnitude of Reflection Coefficient vs. Frequency for Several Arbitrary Probe Locations, 1mm dia. probe.



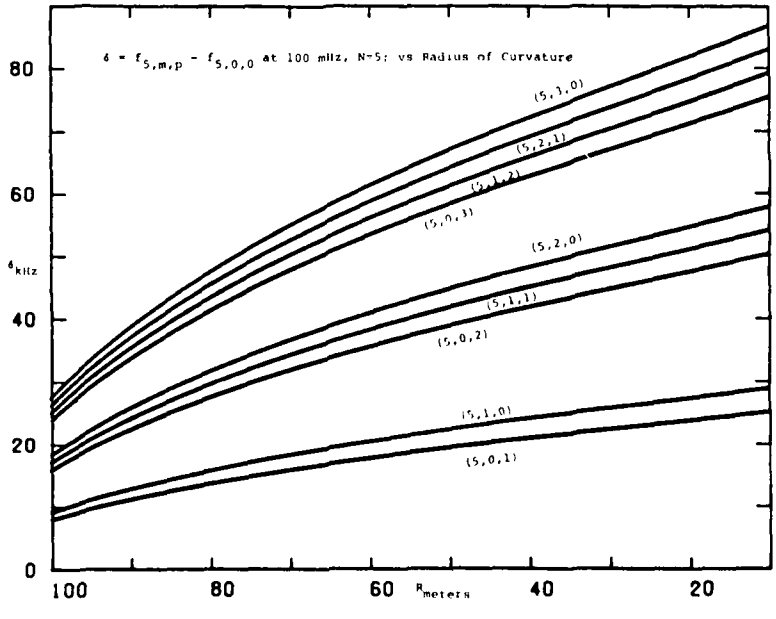
3. 3 Lowest Order Hermite Polynomials.



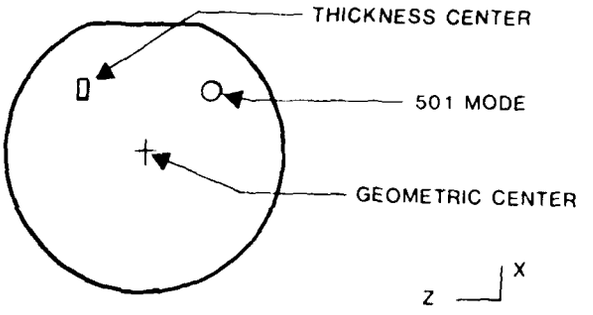
4. Magnitude of Reflection Coefficient vs Frequency for Several Probe Locations about (z_o, x_o).



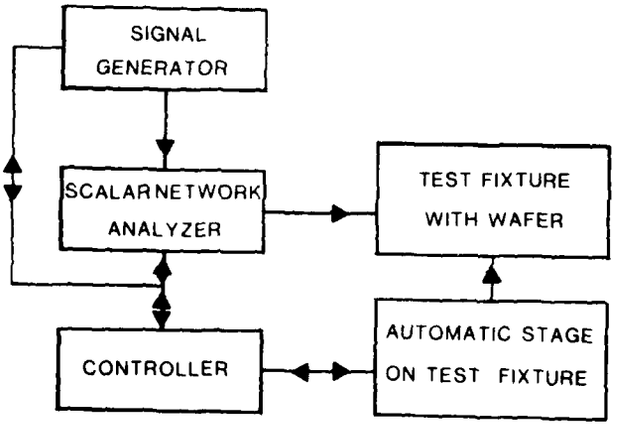
6. Motional Capacitance vs Probe Position (z, x_o).



5. Frequency Separation between (5,m,p) and (5,0,0) Modes vs. Radius of Curvature at f_{500} = 100 mHz.



7. Locations of Interest during Automatic System Operation.



8. Block Diagram of Automatic Evaluation System.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

LASER PROCESSED VHF MONOLITHIC CRYSTAL FILTERS WITH ON PLATE INTEGRATED MATCHING IMPEDANCES

René LEFEVRE

C.N.E.T. 196 Rue de Paris
92220 BAGNEUX

Summary

In this paper we describe VHF monolithic crystal filters with their thin film matching impedances integrated on the piezoelectric substrate and an original method in the design and tuning of such devices.

With regard to classical methods, we start from a single resonator, which electrode sizes take into account the inter-resonator spacing of the future filter, and separate them with a YAG Laser. Adjustment of inter-resonator spacing leads simultaneously to calculated bandwidth and center frequency of the filter with a very good accuracy.

Thin film resistances are then trimmed with the same laser to achieve correct in band ripple and desired 3dB bandwidth.

By this method, very good symmetry of the passband is achievable.

This process, which can be fully automatized, is well adapted to VHF monolithic crystal filter because of the very small spot size of the laser (about 10 µm) while integrated resistances reduce influence of parasitic capacitances and inductances of the can and leads on the end-resonators.

Finally, influence of the laser output power on typical parameters are discussed.

Introduction :

Monolithic crystal filters are now well-known and widely used in telecommunication systems[1]. Their cost are strongly coupled to the manufacture method which complexity and difficulty increase with frequency and order of the filter. The method, that we present, uses widely a laser as tool for adjustment reducing to a minimum, process trimming with masks.

After a recall of the classical method of tuning and its drawback, we describe in detail the new laser-processed method while specifying its limits. Crystal damages due to laser beam are discussed and experimental curves of a 62 MHz filter tuned by this method are shown.

Classical method

In the classical method, most part of adjustment is made in a high vacum apparatus and laser is only used for fine trimming[2]. Movable slot in front of a mask must be used to select different areas on the plate and a precise and elaborate mechanism is necessary when center-frequency and order of filter increase. On the other hand frequency measurements are not accurate due to the presence of metallic masks or slot.

For VHF monolithic filters masks must be more accurate because electrode size and inter-resonator spacing are small to prevent unwanted inharmonic modes while keeping a proper thickness to coatings for a good electrical conductivity. In this case photoetching processes are better and interesting to use but one must tune each resonator and coupling through a mask which leads to an increasing number of handling. On the other hand, electrodes misalignment induces asymmetry in electrical behaviour[3] which is sometimes difficult to suppress. Finally external matching resistances introduce parasitic impedances reducing maximum relative bandwidth achievable ; center frequency tuning is also more critical.

Laser processed method

This method suppresses design and manufacture of complex masks and simplify tuning process by use of laser. On the other hand, since start of process, we can take a particular resonant frequency as reference because its value is weakly modified while tuning is running. More, if, during adjustment, this frequency becomes sensitive tuning of filter is not possible.

Various main steps can be separated :

a) From electrical synthesis, typical values of parameters in equivalent circuit are calculated. Electrodes sizes and plate back are carried out from inductances values and trapping factor [3,4] which formula is recalled :

$$\Omega = \frac{\pi}{2} \frac{Li}{h} \sqrt{\frac{C_{66}}{Cii}} \sqrt{\frac{F_o^2}{F_R^2} - 1} \qquad (1)$$

where Li represents electrode size in coupling direction
h is the thickness of the plate
C_{66} and Cii are the piezoelectrically stiffened values of elastic constants

F_o and F_R are cut off frequencies of unelectroded and electroded areas

Value of this trapping factor is taken equal to 3 to reduce unwanted inharmonic activity[3,4] at a minimum. Measured inductances are within 5% of theoretical values.

Electrical synthesis gives also eigenfrequencies of the structure when resonators are all short circuited.

b) With a bi-dimensionnal elastic thickness-twist model[5,6] we calculate inter-resonators spacing.

c) With these data we generate a single resonator - so called "equivalent resonator" -which electrode size, along the coupling axis of future filter, takes into account calculated values of electrode size and inter-resonator spacing of this filter as is shown in Fig.1.

On can see that, for N poles filter, resonant frequency of the N^{ieme} trapped inharmonic mode in this "equivalent resonator", and resonant frequency of the N^{ieme} main mode of the filter have the same value when inter-resonator spacings are decreasing to zero (Fig.2). On the other hand it is well-known that, in a filter, this resonant frequency is weakly modified when inter-resonator spacing is varying, due to the presence of node line. This theoretical predicted result has been well verified on three typical two poles monolithic filters enhanced by this method :

1) a 41 MHz fundamental filter
2) a 62 MHz third overtone filter
3) a 91 MHz third overtone filter

In each case, when gap between resonators was equal to 20 µm, difference between the two frequencies was not more than 0.2 KHz. Consequently this frequency serves as reference during all adjustment process.

So, for an even N poles monolithic crystal filter the N^{ieme} trapped inharmonic mode of the initial "equivalent resonator" is an odd mode which has not, theoretically, electrical activity[3]. But, in pratice, especially in VHF filters, there is always cause of asymmetry and this mode will be piezoelectrically active. Now for an odd N poles filter this mode is an even mode which electrically active. Accordingly, this Laser processed method will give better results for an odd N poles monolithic crystal filter.

On the Figure 1 one can see also two strips (noted R and R') which are the future matching resistances.They are of same material and same thickness (about 200 Å of Ni-Cr) that the undercoating of electrodes. Their strip shape ensures a quasi resistive behaviour up to about 600 MHz. Their width are calculated to obtain, in a first step, a very low value. Their weak thickness does not disturb the mechanical behaviour of filter. Figure 1 shows a design, but one can imagine various others.

d) In the case of a two poles monolithic filter, resonant frequency of the 2^{nd} trapped inharmonic mode in the initial "equivalent resonator" is adjusted at the following F value :

$$F = F_c - B/2 - \Delta \qquad (2)$$

- where F_c is the desired center frequency of filter
- B is the frequency difference between the two main modes of the final filter without matching impedances.
- Δ is a term that takes into account slight rise of the reference frequency during adjustment process.

Then, plate, ready for laser trimming is mounted on its can and connected to a network analyzer. With a pulsed YAG Laser, electrodes of equivalent resonator are divided (in the case of an N poles filter, N-1 gap. would be made).

e) Inter-resonator spacing is then stretched to achieve the calculated Bandwidth value ; measurements of main resonant frequencies are made after each removal of metal. So we note that reference frequency (resonant frequency of the 2^{nd} main mode) does not vary much. Fig.3 and 4 show evolution of these frequencies in the case of the 41 MHz fundamental filter and the 62 MHz third overtone filter.

f) Finally thin film resistances are trimmed by reducing their width to keep them in a strip form. This adjustment is very accurate and leads to a very fine ripple in-band adjustment and also a very good symmetry of the insertion loss curve versus frequency. Figure 5 and 6 shows the stop band and passband response of a 62 MHz third overtone two poles filter. The two maximums are balanced at about 0,01 dB which was the limits of the network analyzer.

Final characteristics of this filter are :
Center frequency : 62097.5 KHz (for 62097.6 KHz calculated).

3 dB bandwidth : 17.5 KHz.

This leads to an error of 0.57 % (related to the 3 dB bandwidth).

This case was the better. For the two other cases (41 MHz and 91 MHz) center frequencies and 3 dB bandwidth was tuned at about 1.5 % of the correct value, due to a more inaccurate measurement of the reference frequency in the initial equivalent resonator. (for reason exposed in c).

One must note that if the reference frequency becomes increasing too much, process will not converge. Correct value of inter-resonator spacing is exceeded and bandwidth cannot more be reduced. This due to an extreme decrease of dimension size of resonators which leads to a less effective trapping of main modes.

Laser beam interaction

At this time it remains one drawback : laser YAG does damages to the quartz crystal. One thinks that it will be possible to cancel them or reduce strongly their effects, because various filters had no visible defects. These damages are superficial melting of quartz crystal which affect Q of the first main mode. More often than not, damages are on the side exposed to laser beam and not on the opposite

side which is machined simultaneously.

We think that there are various possible ways to solve this problem.

1) by defocussing laser beam to reduce its energy density to the required vaporization metal value.
2) by reducing pulse duration.
3) by vaporizing with a string of pulses which first of them decreases the reflexion coefficient of material that will be vaporized, by the following pulses[7].
4) by modifying laser characteristics such as wavelength.

Conclusion and further developments

A new method for monolithic crystal filters adjustment using widely laser machining has been described. This process suppresses manufacture of delicate masks because we start from a single resonator and leads to the correct final characteristics of the filter with a very good accuracy because we take as reference a trapped mode of initial resonator which frequency does not vary much during the process.

On the other hand we integrate matching resistances on the plate. In this way we get truly monolithic crystal filter by opposition to the Hybrid integrated monolithic crystal filter presented by Kazuo Okuno and Takaya Watanabe[8].

We think that this powerfull method, well adapted to VHF filters, is very interesting for case of higher order filters because leading to a more simple tuning process.

Acknowledgement

The author gratefully acknowledges MM. L. Jenselme and D. Servajean for their experimental work.

References

1. R.C. SMYTHE "Modern crystal filters". Proc. of the 33rd Ann. Symp. on Freq. Control ; pp. 209-213 (1979).

2. P. LLYOD "Monolithic crystal filters for frequency division multiplex". Proc. of the 25th Ann. Symp. on Freq. Control ; pp. 280-286 (1971).

3. A. GLOWINSKI, R. LANCON, R. LEFEVRE, "Effects of asymmetry in trapped-energy piezoelectric resonators". Proc. of the 27th Ann. Symp. on Freq. Control ; pp. 233-242 (1973).

4. R. LEFEVRE, "Monolithic crystal filters with high Q factor and low spurious level". Proc. of the 33rd Ann. Symp. on Freq. Control ; pp. 148-158 (1979).

5. N.H.C. REILLY, M. REDWOOD, "Wave propagation analysis of the monolithic crystal filter". Proc. IEEE, Vol.116, N°15, pp. 653-660 (May 1969).

6. A. GLOWINSKI and R. LANCON, "Resonance frequencies of Monolithic Quartz structures". Proc. of 23rd Ann. Symp. on Freq. Control ; pp. 39-55 (1969).

7. T.R. ANTHONY, P.A. LINDNER, "The reverse laser drilling of transparent materials". J.A.P. , Vol.51, N°11, Nov. 1980, pp. 5970-5975.

8. KAZUO OKUNO, TAKAYA WATANABE, "A Hybrid integrated monolithic crystal filter". Proc. of 30th Ann. Symp. on Freq. Control ; pp. 109-118 (1976).

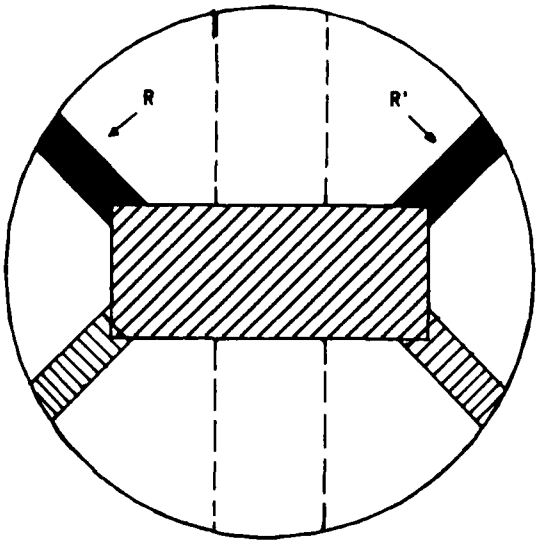


Figure 1

Initial "equivalent resonator" with on plate NiCr thin film integrated matching resistances. (R and R').

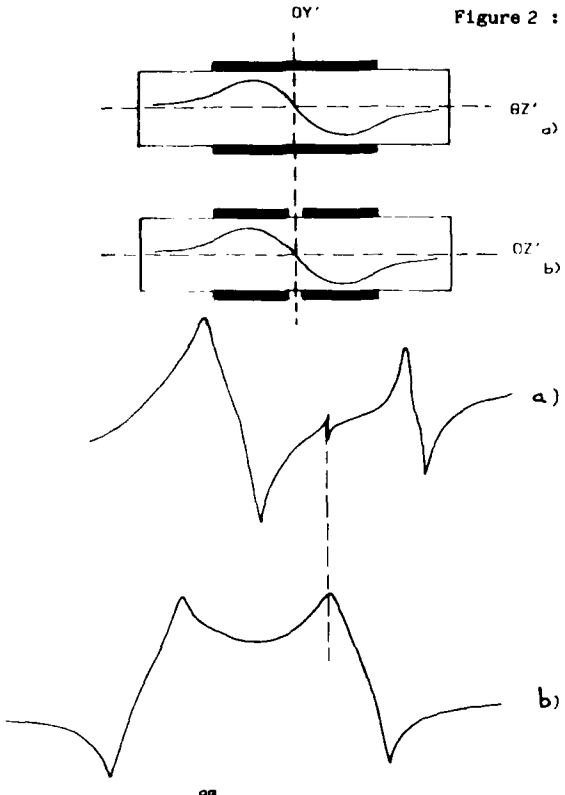


Figure 2 : Thickness-twist displacement amplitude distribution of the 2nd trapped inharmonic mode in the "equivalent resonator" in a), and of the 2nd main mode in the corresponding filter in b), when gap between the two resonators is small. In this case frequency is not modified by the gap due to the presence of the node line.

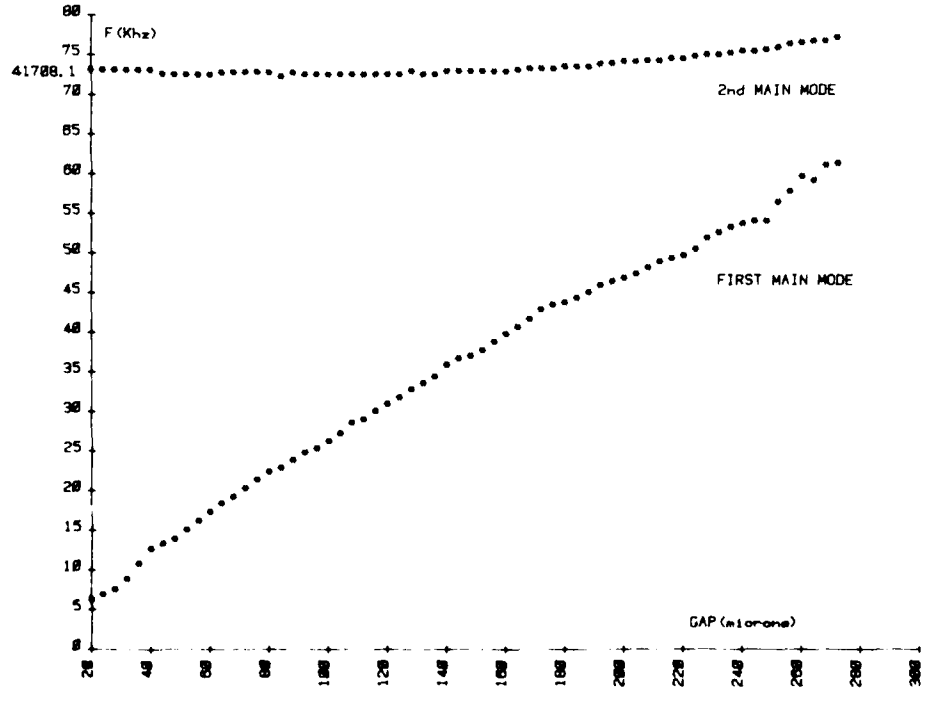


Figure 3 : Evolution of the two main mode resonant frequencies versus inter-resonator spacing for a 41 MHz fundamental filter.

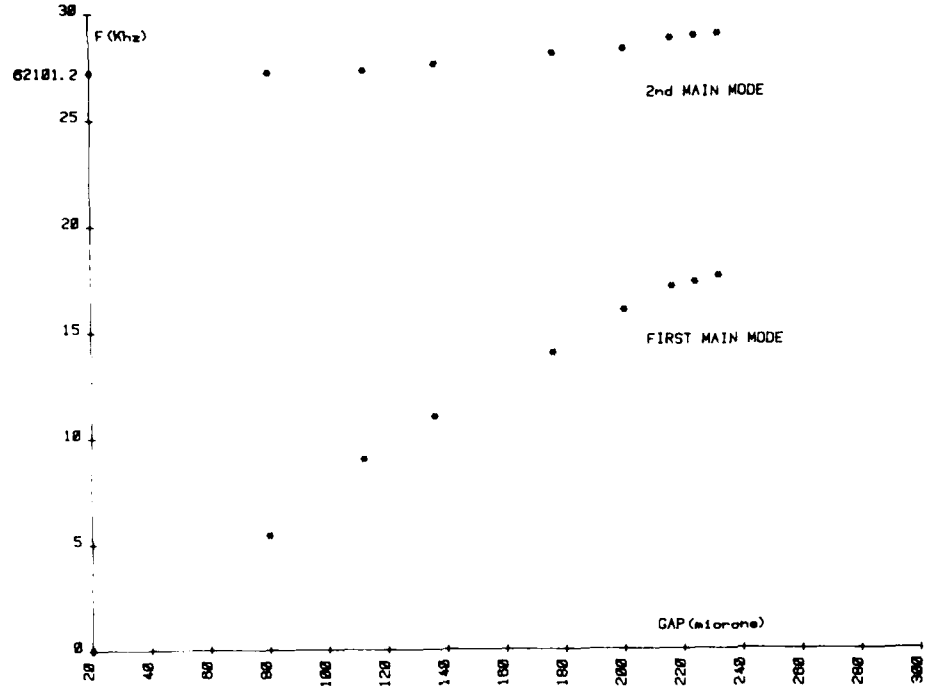


Figure 4 : Same curve as in Figure 3 but for a 62 MHz third overtone filter.

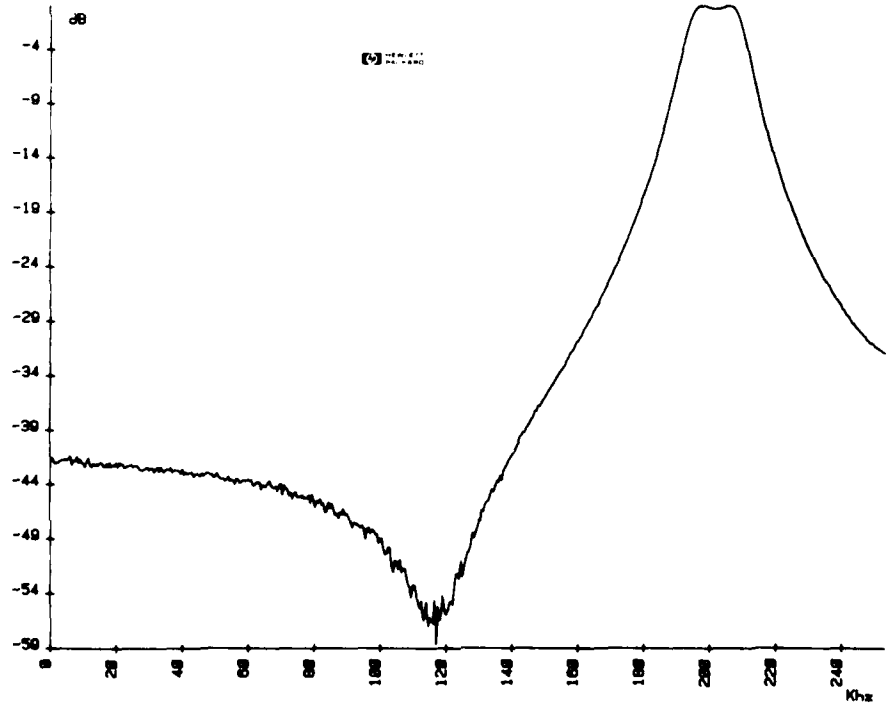


Figure 5 : Aspect of the stopband of a 62 MHz third overtone laser processed monolithic filter.

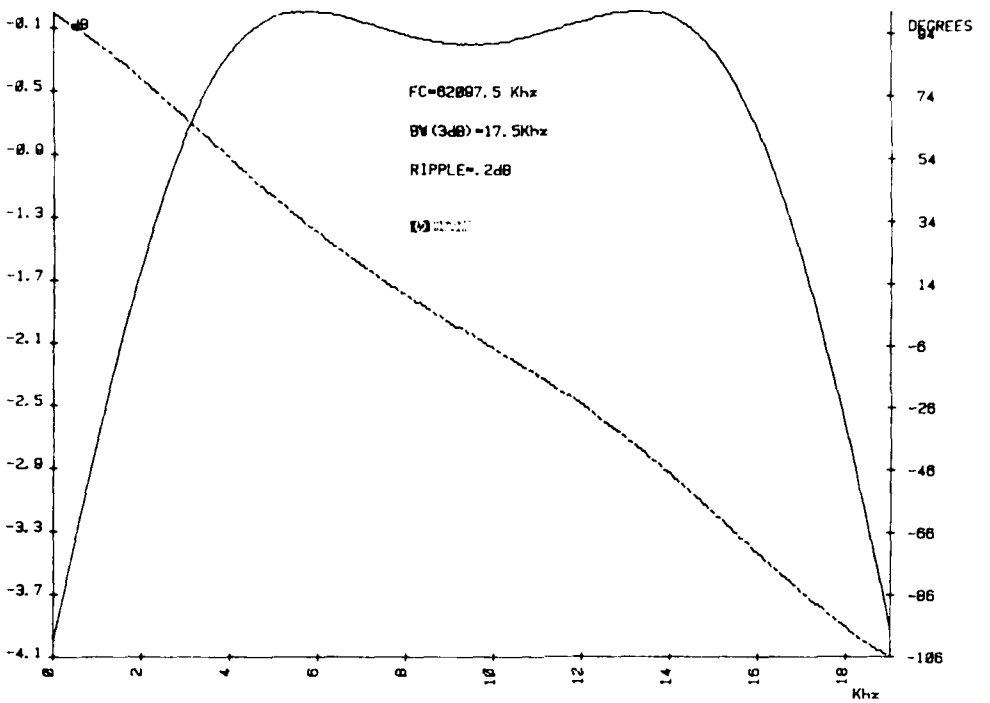


Figure 6 : Amplitude and phase in the passband of the same filter.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

UNWANTED MODES IN 5° X-CUT CRYSTAL UNITS

J. J. Boyer
Bell Telephone Laboratories, Incorporated
555 Union Boulevard
Allentown, Pennsylvania 18103

Summary

Groups of 5° X-cut resonators with common width or thickness dimensions and with w/l ratios between 0.1 and 0.2 have been frequency scanned in a full lattice bridge for the purpose of locating and identifying their unwanted modes. The fourth length-thickness flexure and second length torsion modes have been identified. Also, several modes have been found to be independent of plate thickness and plate width. They intercept the main length extensional mode at .112 and .164 w/l.

The mode data obtained have been plotted in a convenient form and should prove useful in designing extensional resonators to be free of unwanted modes that can occur near the main resonance frequency.

Introduction

Occasionally difficulties are experienced with unwanted resonances in crystal filters and oscillators. These unwanteds are traced to the offending crystals whose plate dimensions are then changed slightly to move them to a less sensitive part of the frequency spectrum. Air loading is an alternative method of dealing with unwanted modes in crystal units. Here the crystal enclosure is sealed at partial or full atmospheric pressure to suppress the modes if the requirements on Q permit.

The frequencies of most unwanteds are controlled by plate width or by plate thickness. Both these parameters are normally used by the crystal designer to achieve desired equivalent circuit values in the resonator and to standardize quartz plate dimensions over a range of frequencies. Thus the probability of an encounter with these unwanted modes is good.

Measurements

The conventional measurement scheme for frequency scanning crystal units is shown on Figure 1. It consists of a signal generator, attenuator, half lattice bridge, and detector. The crystal is placed in one arm of the bridge. The adjustable capacitor in the other arm is used to balance the static capacitance of the crystal. The major disadvantage of this scheme is that within about 10% of the main resonance the crystal no longer behaves like the capacitor, and thus the bridge becomes unbalanced. The resultant level of signal transmission tends to hide all but the strongest unwanted modes in the vicinity of the main resonance.

This shortcoming was overcome by substituting a crystal that was identical to the one being measured for the adjustable capacitor in the bridge, converting it to a full lattice configuration. High loss levels were now attainable at frequencies very close to resonance, and unwanteds were readily located within these regions. One important bonus from this measurement scheme is that both the bridge crystal and the crystal under test are scanned simultaneously. Whenever possible three crystals of each design were measured and the responses averaged and plotted. If a response was found in only one crystal it was not weighted as heavily unless it was located on a locus of points. The detected resonance frequencies were expressed as follows:

$$\frac{\text{Freq. Unwanted Mode}}{\text{Freq. Main Extension Mode}} \times 2810 \text{ kHz mm}$$

and are plotted versus the plate width-to-length ratio.

Discussion

The first group of crystals scanned were samples of the development models used for signaling filters. The mode spectrum is shown in Figure 2. Notice the coupling between what has been tentatively identified as the 6th l-t flexure and the extensional mode (main mode) at 0.12 w/l and 2810 kHz mm. This is obviously a w/l to be avoided when using .381 mm thick plates. Also of interest is the mode which parallels the extensional mode. This mode appears in all of the crystals scanned. At first it was thought to be some kind of intermodulation product of the test equipment; however, the mode does couple to the 6th l-t flexure just below w/l = .12. This is therefore a

characteristic of a plate mode and not equipment related. This mode exhibits a temperature coefficient of frequency similar to that of the main mode; thus, it never crosses the path of the main mode as the temperature is varied. It is undoubtedly the principal reason why channel bank filter crystal units are sealed at atmosphere rather than evacuated. Evacuation tends to enhance the strengths of unwanted modes because of the absence of air loading on the quartz plate, and these modes would then appear in the passband of these wideband filters.

No unwanted resonances were found immediately above the main mode at the w/l ratio of .159. This is noteworthy because this design exhibited an unwanted mode during initial manufacture by both the Western Electric Company and outside suppliers, whereas the development models had no unwanted modes at this location. A plate width change eliminated the problem. The difference between the development and production groups was in the method of tuning, and this may have provided a significant insight into how some unwanted modes can be driven in these resonators. The development models were adjusted to frequency by hand abrading the length end of the plate which results in a reasonably orthogonal end surface when finished. The production models, however, were tuned to frequency with an airborne abrasive which resulted in a predominantly wedge shaped end of the plate. Miller[7] has found that wedge shaped ends do indeed change both the amplitude and location of unwanted modes. It is therefore possible that one of the adjacent unwanted modes was moved closer to the main mode by this tuning method.

Another possible explanation of the difference in the unwanted mode spectra between the development and production units must be accommodated. Meeker[5] has calculated that there are many combinations of overtones and anharmonics of known modes that at least are mathematically possible. Some occur near the main mode. Most of these modes are not driven because they are charge canceled by the symmetry of the device. However, this symmetry can be upset by tuning the plate with abrasive, especially the airborne variety. This can result in a small net charge on the plate which may weakly drive these modes.

Figure [illegible] is a plot of a group of resonators with a slightly different plate width dimension. Note the unwanted mode that intersects the extensional mode at .165 w/l. This ratio should be avoided in designs.

Figure 4 was generated from a third group of plates .[illegible] mm thick and [illegible] mm wide. The low width-to-thickness ratio of this design permitted the application of a finite element analysis program that was developed for bars to the range of w/l's covered.

The computed results are plotted in Figure 5. An origin was added because it is a data point for all length dependent flexures when length is the variable. The plot includes the experimental data points (x) of Figure 4.

There appears to be very good agreement on the locations and on the couplings between the calculated and experimental data leading one to conclude that the loci are those of the fourth 1-t flexure and the second torsion modes. (The coupling can be related to the distance of closest proximity between modes; therefore, the second torsion and fourth 1-t flexure couple more strongly than do the fourth 1-t flexure and the fundamental extensional modes.)

A second analysis was performed on the same plate except the width of the plate was varied instead of the length. This was done to demonstrate how the torsional modes in the plate move relative to the change in the plate width. Figure 6 is a plot of the results. As a point of reference, w/l = .1422 is common to both Figures 5 and 6. It shows that a w/l from .14 to .18 would be free of torsion modes. (There is a coupling at w/l = .164 which the analysis does not predict. It will be discussed later.)

Figures 7, 8 and 9 are plots of 5° X-cut resonators covering the same w/l ratios and having the same thickness. Each group, however, has a slightly different width. From this one should get an indication of how unwanted modes vary as a function of plate width. For instance, on Figure 7, the mode whose locus of points crosses w/l = .18 and 2710 kHz mm does not appear on Figures 8 and 9 at that location; therefore, it can be assumed to be a width dependent mode. A mode of similar slope does appear at .18 w/l and 2970 kHz mm on Figure 8 and at .18 w/l and 3185 kHz mm on Figure 9. The shifts in loci are not, however, inversely proportional to the plate widths as would be expected for a width flexure. The identity of these modes will therefore have to wait for further analysis.

There are a considerable number of randomly located modes in these last three figures that may be related to asymmetry in the plate geometry as mentioned earlier. These modes appear to be more prevalent in the thinner plates possibly because of the relatively larger damage sustained by the plate during frequency adjustment.

As suggested in Figure 7, w/l = .185 should be avoided when designing a quartz plate .274 mm thick and 5.26 mm wide. Figure 9 indicates that w/l = .177 should not be used when designing a plate .274 mm thick and 2.72 mm wide.

Figure 10 is a plot of the modes in a group of resonators whose dimensions were t=.284 mm and w = 5.67 mm. There appear to be many modes in this plot, especially below 2800 kHz mm, and some data points can be connected several ways. There appears, however, to be a mode with a negative slope intersecting the extensional mode at w/l = .175, as well as one with a positive slope crossing over at w/l = .164.

If the preceding experimental data figures are overlaid, some interesting information emerges. For instance, regardless of the plate widths and thicknesses, four unwanted modes have the same loci. These modes are shown on Figure 11. The third 1-w flexure and the mode below and parallel to the extensional mode, as expected, are at the same place in all figures, but two other modes that intersect the main mode at w/l = .112 and .164 are also present. They have been compared with calculated mode spectra, but they do not match. Fortunately these modes have fixed locations of w/l regardless of the thickness. They can therefore by easily avoided when designing resonators.

In summary, this investigation has uncovered many weak unwanted modes in 5° X-cut resonators; some of which have been identified. It has also suggested that some methods of frequency tuning these resonators can actually generate these modes. It follows that proper choice of plate dimensional ratios and care in the frequency adjustment operation will result in improved resonators.

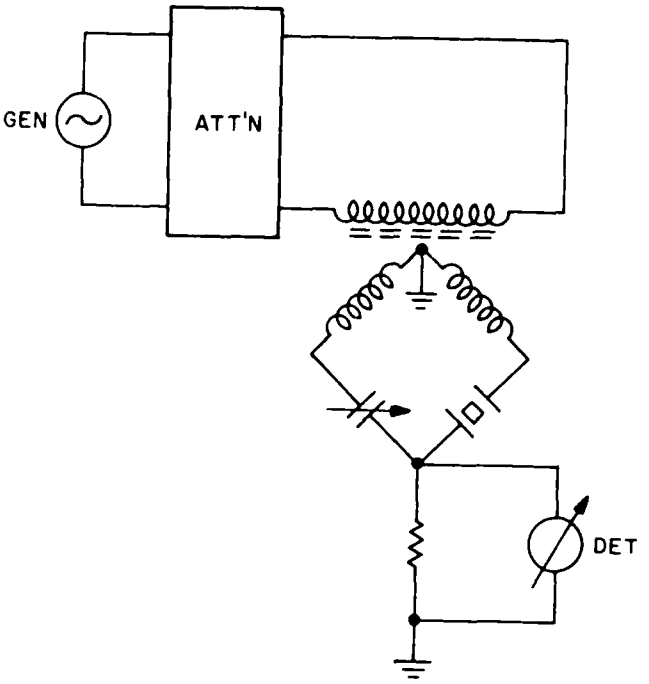


Figure 1 Bridge schematic for measurement of unwanted modes.

References

[1] Miller, A. J., Private Communication.

[2] T. R. Meeker, "Extension, Flexure and Shear Modes in Rotated X-Cut Quartz Rectangular Bars," 33rd Proceedings Annual Symposium on Frequency Control, U.S. Army Electronics Command, Ft. Monmouth, N.J., pp. 286-292. Copies available from Electronic Industries Association, 2001 Eye Street, N.W., Washington, D.C. 20006.

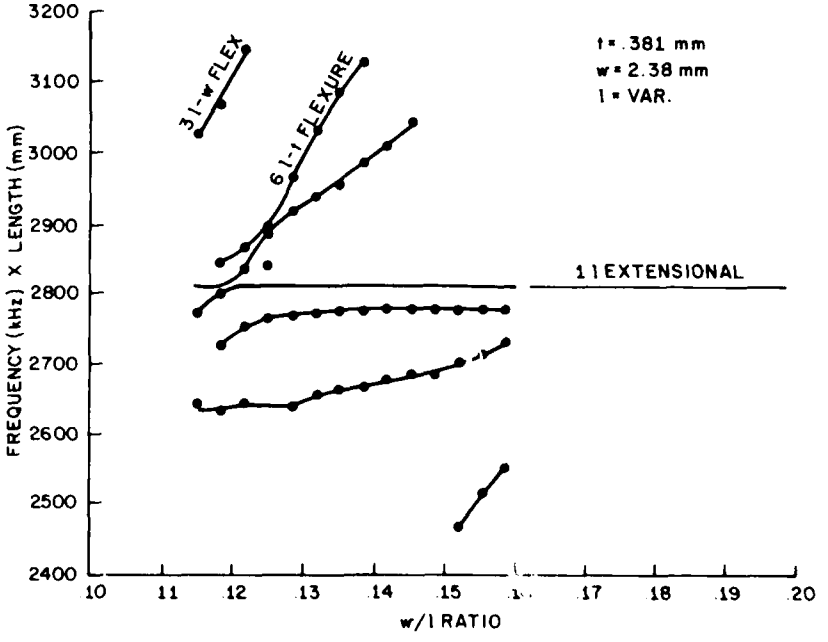


Figure 2 Mode spectrum of 5° X-cut. w = 2.38 mm, t = .381 mm.

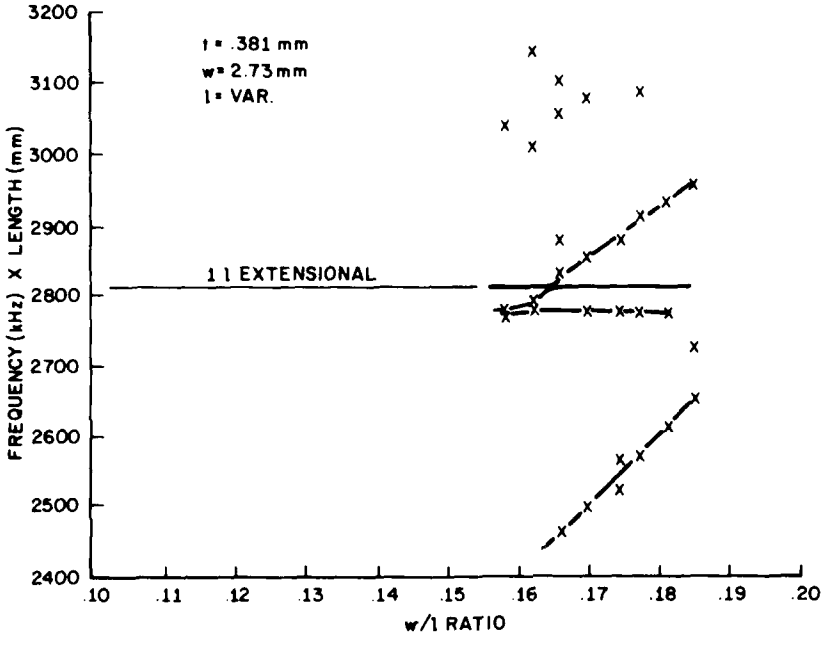


Figure 3 Mode spectrum of 5° X-cut. w = 2.73 mm, t = .381 mm.

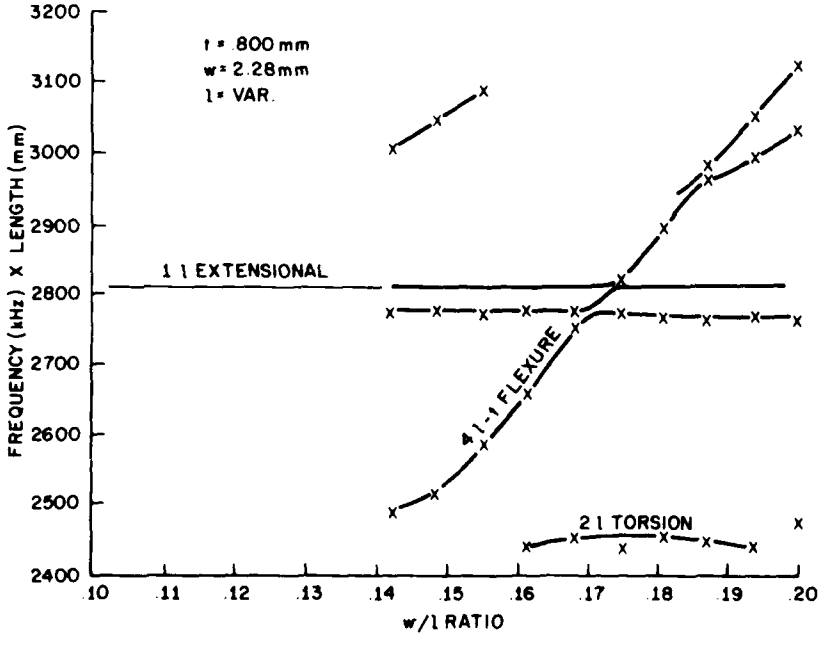


Figure 4 Mode spectrum of 5° X-cut. w = 2.28 mm, t = .800 mm.

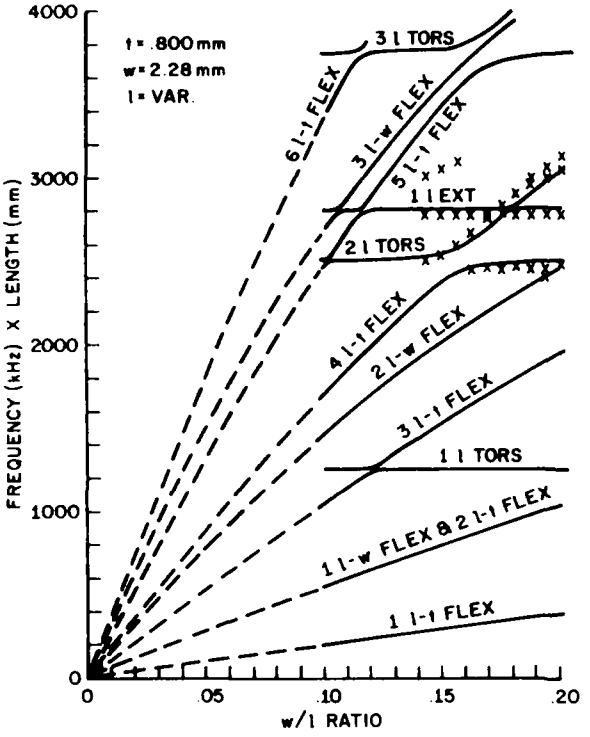


Figure 5 Finite element analysis of 5° X-cut w = 2.28 mm, t = .800 mm, = variable.

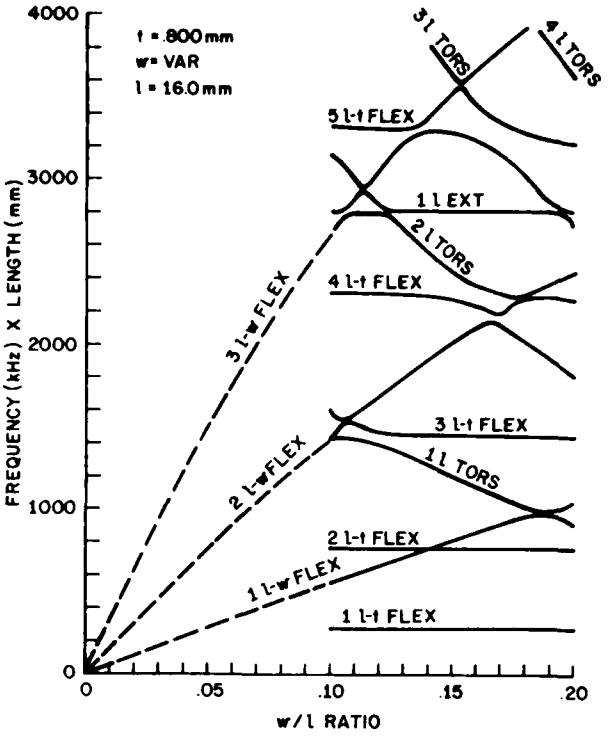


Figure 6 Finite element analysis of 5° X-cut w = variable, t = .800 mm, = 16.0 mm.

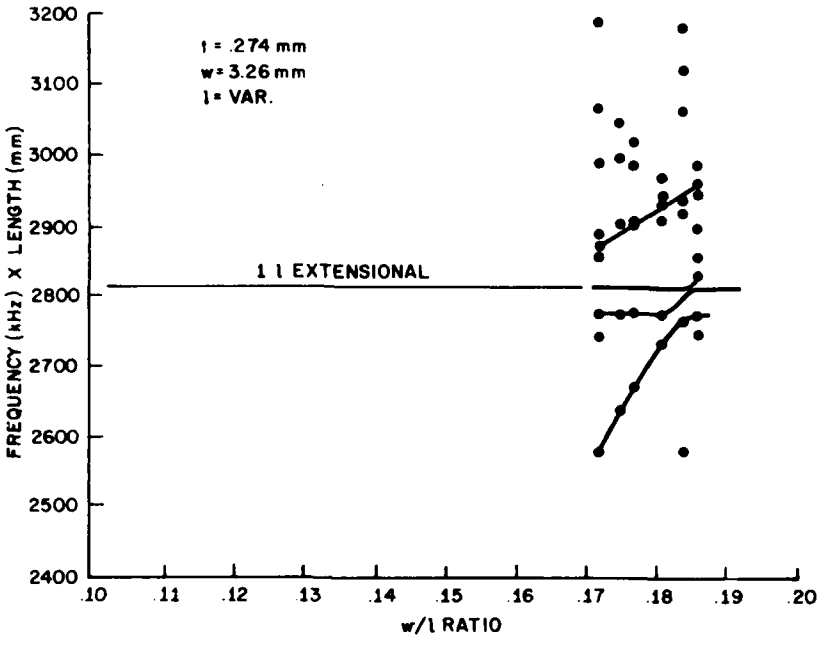


Figure 7 Mode spectrum of 5° X-cut. w = 3.26 mm, t = .274 mm.

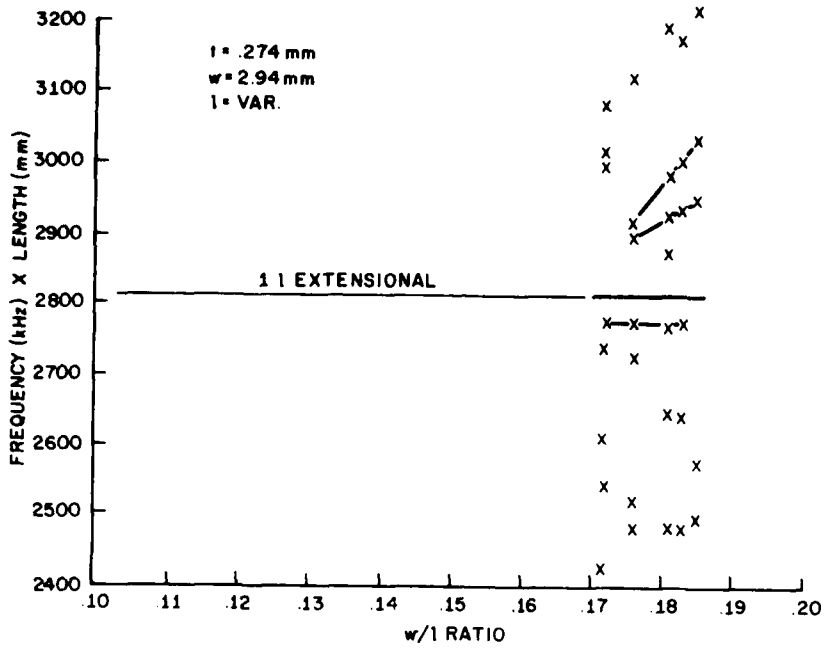


Figure 8 Mode spectrum of 5° X-cut. w = 2.94 mm, t = .274 mm.

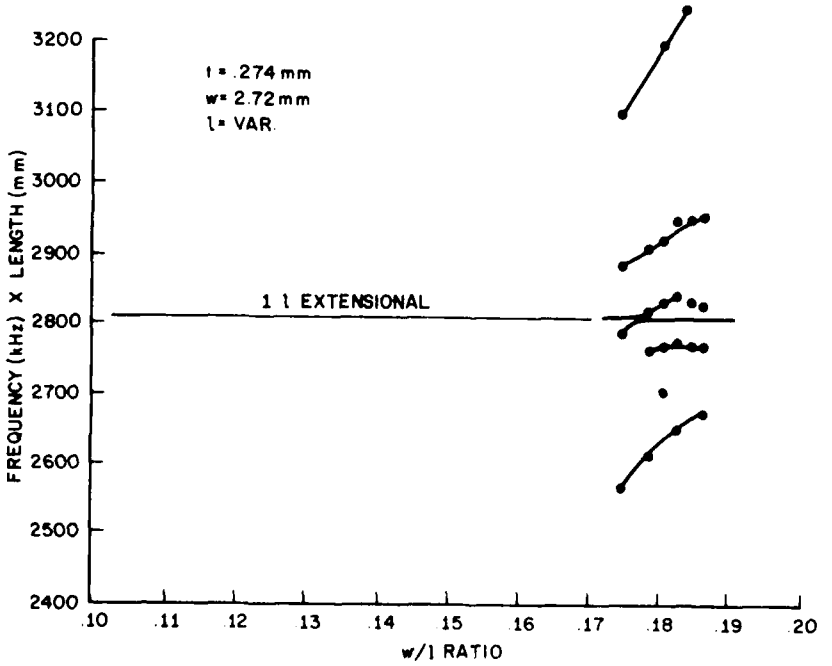


Figure 9 Mode spectrum of 5° X-cut. w = 2.72 mm, t = .274 mm.

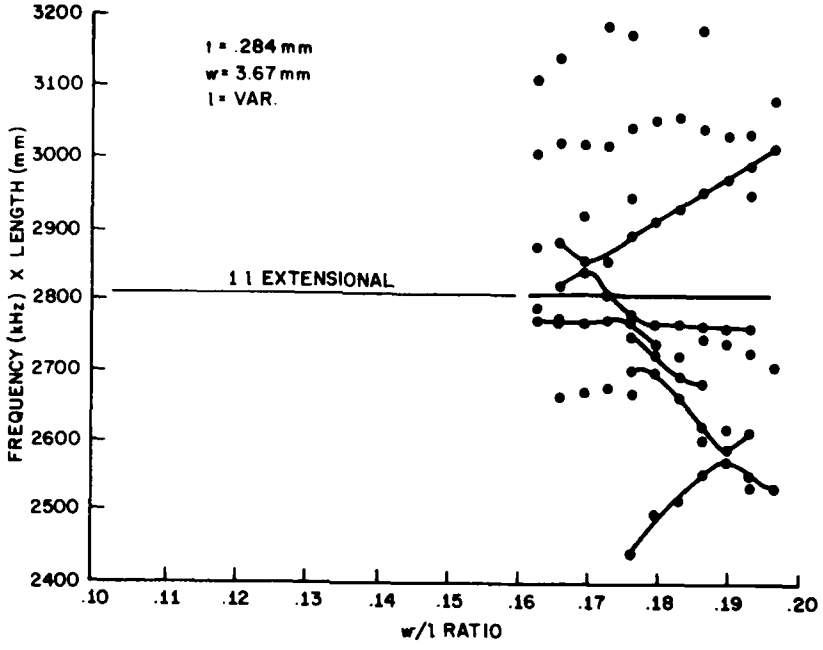


Figure 10 Mode spectrum of 5° X-cut. w = 3.67 mm, t = .284 mm.

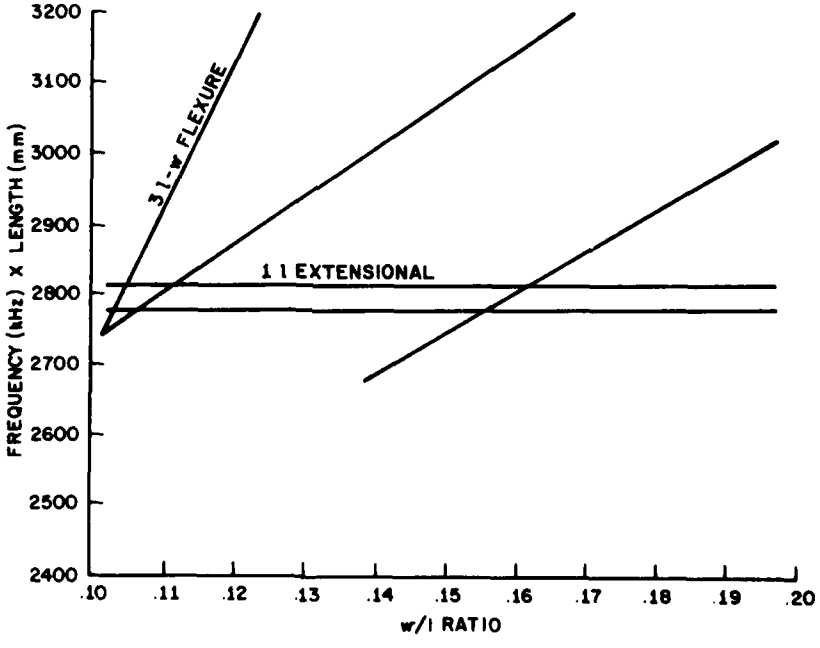


Figure 11 Composite of common data points.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

EQUIVALENT CIRCUIT MODELING OF STACKED CRYSTAL FILTERS

K. M. Lakin
215 Reactor Building
Ames Laboratory - USDOE
Iowa State University
Ames, Iowa 50011

Abstract

The traditional stacked crystal filter (SCF) is composed of two bonded piezoelectric plates with a common ground plane at the bond line. Thus, the filter input and output terminals are largely isolated by the ground plane electrode to provide a high degree of off resonance rejection. In operation, the system is driven to resonance by the input resonator (transducer) and the resulting energy is radiated to the output plate which acts as a receive transducer to couple the energy to the external electrical circuit.

In the SCF the input to output coupling is much stronger than in the case of monolithic filter because radiated fields are involved rather than trapped energy modes. This implies that the external circuit may have a more significant affect on the filter pass band shape.

Considerable analysis of this filter geometry has been carried out by Ballato using a wave propagation and boundary value problem approach. In order to gain some insight into the device characteristics, we have adopted an equivalent circuit model approach based upon the Mason model for a single pure mode in a piezoelectric disk. For the case of a simple symmetrically bonded pair of disks the resultant lumped element equivalent circuit may be readily evaluated to determine the filter response in terms of Butterworth-VanDyke equivalent circuit parameters.

In order to implement the SCF configuration at high frequencies, using microelectronic techniques, a different physical configuration has been investigated. This configuration is similar to the composite resonator described previously and offers significant advantages in terms of implementation and device operating characteristics.

Introduction

The stacked crystal filter, SCF, is a relatively simple device composed of two piezoelectric plates configured such that the driven fields from one plate are directly radiated to the second plate.[1] The basic device configuration is shown in Fig. 1 and can be compared to the monolithic crystal filter depicted in Fig. 2. Of critical importance to device implementation and operation is the bond region between the two piezoelectric plates. This region must be a conductor in order to form the ground plane electrode yet must have the appropriate acoustic properties to insure low propagation loss and desirable bandpass filter characteristics. In the lower HF range the technology exists for implementing this configuration and successful devices have been demonstrated.[2-6] However, there is much less demand for low loss filters in this frequency range as compared to the VHF, UHF and microwave frequency range. At these higher frequencies the receiver front end filter has a significant affect on the overall noise figure of the system. In addition, the need for low loss miniature channel filters has been left unfulfilled.

Unfortunately, the problem of crystal plate bonding in the SCF is extremely difficult at these high frequencies. However, the thin film technologies of chemical vapor deposition and sputtering of piezoelectric films offers a clear solution to the bonding problem. The only drawbacks to the thin film approach is the fact that, in the VHF region, the available technology may produce films of less than the required thickness and the films may lack the mechanical strength required for device fabrication and mounting. For example, the basic SCF configuration of Fig. 3a would require a total plate thickness of approximately 50 μm at 100 MHz. Such a thin structure composed of non-single crystal materials would be fragile and difficult to mechanically support.

For a practical filter implementation at high frequencies we have proposed two other SCF configuration that promise a greater likelihood of successful implementation, Fig. 3b,c. In Fig. 3c the extended SCF configuration resembles an ordinary acoustic delay line except that the total delay would correspond to approximately one half acoustic wavelength. The problem with this configuration is that the delay line region can produce an impedance transformation which significantly alters the filter response, and requires a two-sided fabrication process. The composite stacked crystal filter, CSCF, of Fig. 3c offers some significant fabrication and operational advantages for the higher frequency SCF implementations. Both of these configurations could avoid the problem of mechanical strength by using single crystal material for the delay path. The configuration of Fig. 3c is suggested by the composite resonator devices recently reported.[7,8]

Based upon the previous discussion it is clear that there are a large number of SCF device configurations and material combinations that could be investigated. With this fact in mind, the research reported here was initiated with the goal of developing a simple circuit model, the SCF, which could be analyzed for the purpose of finding a relation between material parameters which could be used to assess material requirement in terms of filter performance. The primary concern was the coupling coefficient and mechanical Q required to achieve a particular minimum insertion loss in the filter.

Equivalent Circuit Modeling

In order to obtain a first order knowledge of SCF operation the Mason lumped element circuit model was applied to the configurations of interest. This model only applies exactly to the case of single mode plane wave propagation yet gives a good picture of device operation when the resonator transverse dimensions are finite so long as only a single mode resonance is observed. The Mason model may be applied in a form suggested by Fig. 4. Here two Mason models are shown in a back-to-back configuration as a LEFT model and RIGHT model. The right acoustic port of the LEFT model is connected directly or through a transmission line to the left acoustic port of the RIGHT model. The other acoustic ports are assumed to be terminated in passive loads denoted by ZL and ZR. The circuit parameters, without subscripts, are given by,

C = constant strain capacitance

N^2 = $CZoK^2$

Z2 = jZo csc

Z1 = jZo tan

Zo = characteristic impedance of the piezoelectric

K^2 = electromechanical coupling coefficient

= phase across piezoelectric plate

= f/fo

fo = parallel resonant frequency of plate

= /2.

For analysis purposes it is useful to transform the network into a totally electrical circuit by "drifting" the left-most transformer to the right to form the circuit of Fig. 4b. In this manner, all acoustical impedances are changed to electrical impedances and the transformer in Fig. 4b has a dimensionless turns ratio. The new impedances are now given by:

Z2' = -jZo csc ,
Z1' = jZtan ,
ZL' = Z(ZL/Zo) ,
Z2' = -jZ(Zoz/Zol)csc 2,
Z1' = jZ(Zoz/Zol)tan ,
ZR' = Z(ZR/Zol),
' = n ,

where Z = 1/(K² C),
and n = N1/N2.

Once in the form of an electrical lumped element network there are numerous circuit analysis techniques that can be used to determine the system transfer function. Since the Mason model is now in the form of a two-port network, a simple ABCD matrix cascading scheme may be employed.

In the case of the most basic SCF configuration, both piezoelectric plates are identical, the passive acoustic loads are of zero impedance, and the bond region between the two plates is perfect and of zero thickness. In this case the network reduces to the simple configuration of Fig. 5. This network could be easily analyzed using conventional techniques. However, it is useful to simplify the network by a series of transformation between PI and T networks as suggested in Fig. 5. After considerable algebra the network of Fig. 6 results with the simple PI network values given by,

$$Zr_2 = \frac{1}{j\omega C}\left(1-K^2\,\frac{\tan\phi}{\phi}\right),$$

$$Zc = Zs\,(1-\cot^2\phi),$$

where $Zs = 2Zr_1Xc/(2Zr_1+Xc)$,

$$Zr_1 = \frac{1}{j\omega C}\left(1-K^2\,\frac{\tan\phi}{\phi}\right),$$

and $$Xc = \frac{1}{j\omega(-C/2)}.$$

Note that Zr_2 represents a free plate resonator of the combined thickness of the two plates and that Zr_1 is the impedance of a single plate resonator.

The network in Fig. 6a is exact within the Mason model approximation but may be simplified for those frequencies around the filter transmission peak near fo. Near fo Zc = Zs and the network in Fig. 6b results where $2Zr_1$ is in parallel with a capacitor of value -C/2. In order to gain a further physical insight into the filter operation it is useful to expand $2Zr_1$ into the Butterworth-VanDyke equivalent circuit and note that only the series arm of the model is left when the -C/2 capacitor is substracted. The exact form for Zs,

$$Zs = \frac{2}{j\omega C}\left(\frac{\phi}{K^2}\cot\phi-1\right)$$

may be expanded about $\phi=\pi/2$ directly as an alternative approach. In either case the network of Fig. 7 results. The loss term is of major interest and is given by

$$Ra = 1/(\omega CK^2Q)$$

where $Q = \frac{fr}{2}\left.\frac{d\phi}{df}\right|_{fr}$.

Here ϕ is the electrical phase of the impedance of a single free plate resonator and fr is the parallel or series resonant frequency.

The simple PI network in Fig. 7 may be further analyzed to find the power transfer function or insertion loss at series resonance,

$$IL_{dB} = 20\text{Log}[1+(q+1/q)/p]$$

where $q = \omega CRg$

and $p = K^2Q$.

The insertion loss is a minimum when $q = \omega CRg=1$. This also corresponds to the case of maximum bandwidth where the source and load resistances significantly load the series resonant circuit. For a resonant IL of less than 3dB the required material parameters are

$$K^2Q(\frac{8}{1+q^2}) >4;\ IL < 3dB.$$

These material parameters imply rather modest material requirements at high frequencies where q can approach unity without serious design problems. At low frequencies the capacitance is likely to be small, (AT quartz, 3pf at 3 MHz), and the reactance quite large (18K ohms) so that material Q would have to be much higher ($Q>10^5$). At VHF and UHF the material requirements are considerably eased due to the decreased reactance, assuming that the resonator area need not scale in inverse proportion to the frequency.

Calculations of IL for some representative cases have been carried out using the two port ABCD matric technique. Plots of IL versus normalized frequency are shown in Figs. 8 and 9. Calculations were for a material K^2 of 10^{-2}, Q of 1000, source and load resistances of 50 ohms, and with capacitor reactance values labeled next to the curves. Figure 8 shows the relative narrow bandwidth result when the reactance is large. The poor skirt selectivity is due to the tight acoustic coupling achieved by the SCF configuration. In Fig. 9 is shown the affect of capacitive reactance, and hence circuit coupling, on the filter bandwidth, insertion loss, and center frequency. The frequency shift is due to the series transformed equivalent of C that depends upon q.

The composite stacked crystal filter, CSCF response is shown in Fig. 10. Here the delay line region was assumed to have the same material parameters as the piezoelectric plates. The piezoelectric plates were of equal thickness and the delay line eight times thicker. Thus, the fundamental response occurs at near one tenth the plate resonant frequency, fo. In accordance with composite resonator theory, both even and odd order resonances occur with the effective coupling coefficient increasing with harmonic order. Since the effect K^2 is increasing, the IL decreases and the bandwidth increases with harmonic order as shown in Fig. 10. For this case the capacitance reactance was chosen to be 50 ohms and the Q=1000 at the fundamental (f/fo=0.1). The Q was chosen to decrease as the square of the frequency.

Summary

The results of this study of the equivalent circuit model of the SCF has produced simple analytical results that reflect the physical operation of the device and which can be used to evaluate candidate material configurations for experimental implementation. For the VHF and higher frequencies the material Q requirement is relaxed by the allowed and physically realizable lower device capacitor reactance and K^2 obtainable with thin film materials.

In the upper frequencies of interest the composite stacked crystal filter, CSCF, was proposed as a practical configuration for implementation in view of the work already done on the composite resonator. Work is continuing on the modeling and physical implementation of the CSCF using thin films and microelectronic fabrication techniques.

Acknowledgement

The author wishes to acknowledge the continued support of the Air Force Office of Scientific Research.

References

1. "A novel frequency selective device: The stacked crystal filter", A. Ballato and T. Lukaszek, Proc. 27th Annual Frequency Control Symposium, June 1973, pp. 262-269.

2. "Transmission-line analogs for stacked piezoelectric crystal devices", A. Ballato, Proc. 26th Annual Frequency Control Symposium, June 1972, pp. 86-91.

3. "Systematic design of stacked-crystal filters by microwave network methods", A. Ballato, H. L. Bertoni, and T. Tamir, IEEE Trans., Vol. MTT-22, January 1974, pp. 14-25.

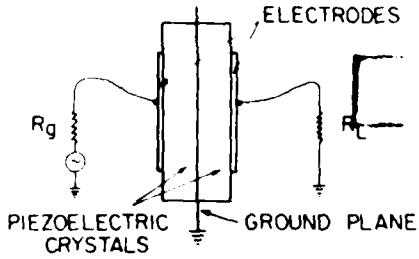
4. "Stacked-crystal filters", A. Bellato and T. Lukaszek, Proc. IEEE, Vol. 51, October 1973, pp. 1495-1496.

5. "The stacked-crystal filter", A. Ballato, Proc. IEEE Intl. Symposium on Circuits and Systems, April 1975, pp. 301-304.

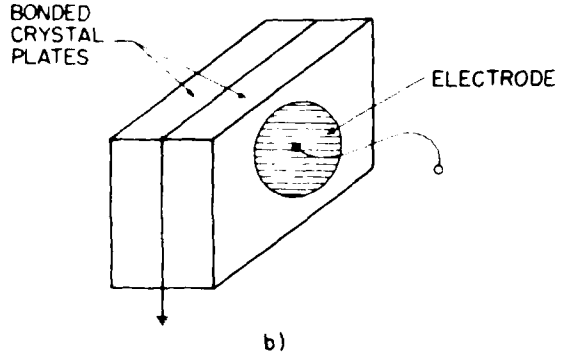
6. "Distributed network modeling of bulk waves in crystal plates and stacks", A. Ballato and T. Lukaszek, Technical Report ECOM-4311, May 1975, U.S. Army Electronics Command, Fort Monmouth, N.J. 07703.

7. "Acoustic bulk wave composite resonators", K. M. Lakin and J. S. Wang, Appl. Phys. Lett. 38 (3), 1 Feb., 1981.

8. "Fundamental mode VHF/UHF bulk acoustic wave resonators and filters on silicon", T. W. Grudkowski et al., 1980 Ultrasonics Symposium Proceedings, Vol. 2, 80CH1602-2, p. 829.

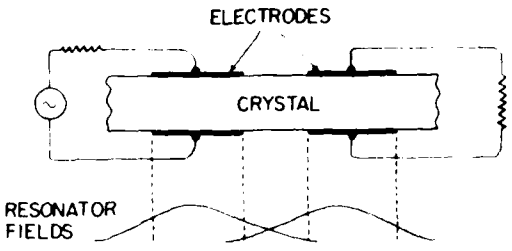


a)

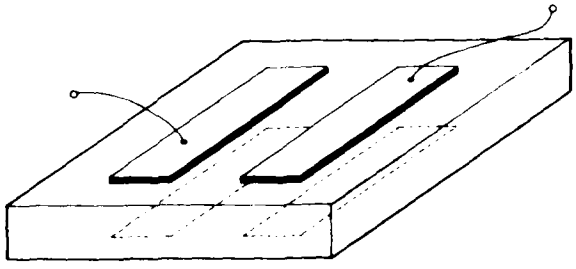


b)

Fig. 1. Basic stacked crystal filter, SCF, configuration, a) end view, b) perspective view. The transducer radiates sound directly to the right transducer.



a)



b)

Fig. 2. Monolithic crystal filter. a)side view, b) perspective view. Filter transmission is through the coupled trapped energy fields.

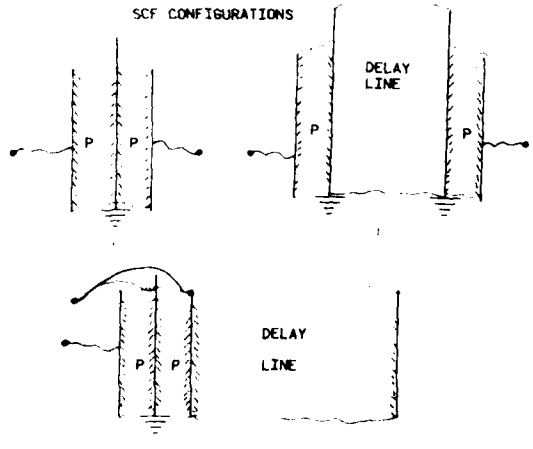


Fig. 3 Extended SCF configurations, a)basic SCF with possible unequal thickness and different material plates, b) short delay line configuration with total length of the order one half wavelength at resonance, c) composite stacked crystal filter, CSCF, having both piezoelectric plates on one side of a crystal plate delay line.

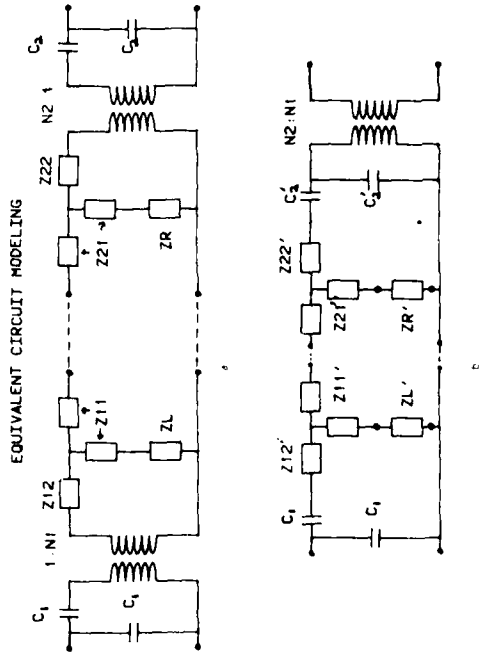


Fig. 4. Schematic for equivalent circuit modeling, a) two Mason models arbitrarily coupled (dashed lines) having opposite acoustic ports terminated in passive loads ZL and ZR, b) all electrical network formed by drifting the transformers to the right hand side.

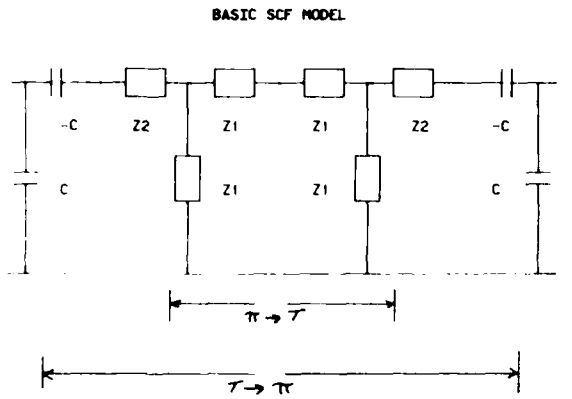


Fig. 5. Schematic of basic ideal SCF composed of identical plates perfectly bonded and having zero electrode thickness.

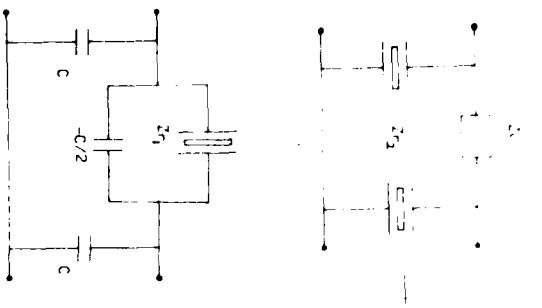


Fig. 6. Equivalent circuit resulting from a simplification of the circuit in Fig. 5 through the use of network transforms; a) the exact equivalent PI network, b) the network near the filter passband frequency.

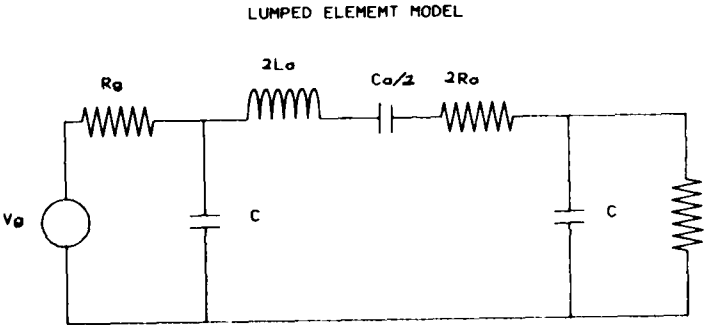


Fig. 7 Constant lumped element equivalent circuit resulting from the replacement of the resonator in the series arm of Fig. 6b. with a Butterworth-VanDyke model.

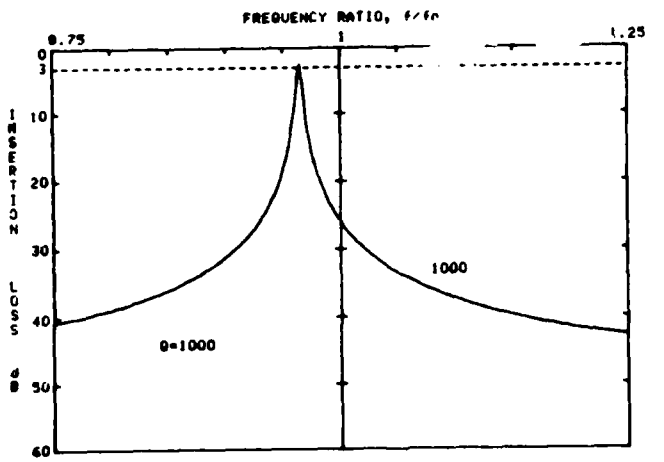


Fig. 8 Response of an ideal basic SCF configuration demonstrating the apparent tradeoff between material Q, K^2 and capacitive reactance. Insertion loss calculated by ABCD matrix cascading and analysis for K^2 = 10%. Here f/fo is frequency normalized by the parallel resonant frequency. Source and load impedances were 50 ohms. The capactive reactance is labeled next to the curve.

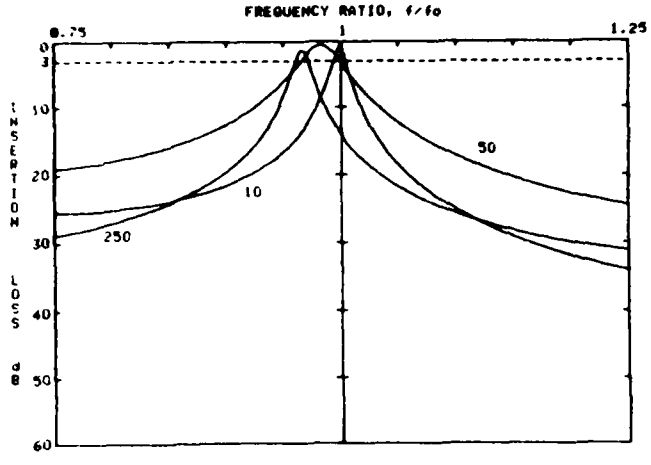


Fig. 9 Insertion loss under same conditions as Fig. 8 except for different capactive reactance values.

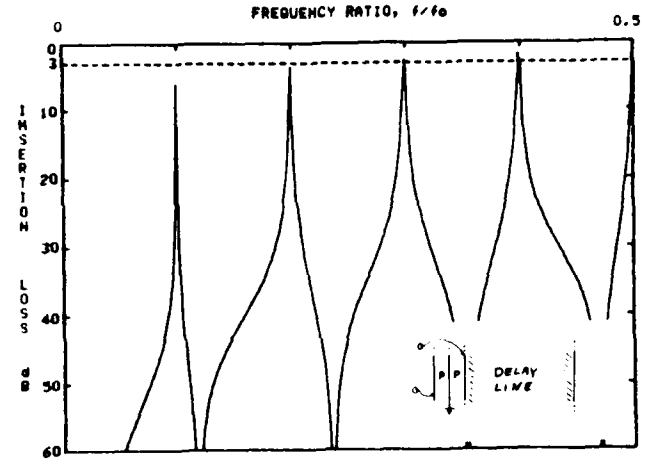


Fig. 10 Insertion loss of composite stacked crystal filter, CSCF. The delay line has the same impedance and velocity of the piezoelectric plates and is eight times as thick. The equal thickness plates were assumed to have K^2 of 10%, Q = 1000 and capacitive reactance of 50 ohms at f/fo = .1 with Q decreasing as frequency squared.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

THE QUARTZ RESONATOR AUTOMATIC AGING MEASUREMENT FACILITY

Dale E. Beetley, Baynard R. Blitch, and Thomas M. Snowden
General Electric Company, Neutron Devices Department*
St. Petersburg, Florida

Summary

A multi-channel, high precision aging** measurement facility has been developed for high volume test capability with ovenized oscillators. The frequency reference is traceable to the National Bureau of Standards (NBS). Computer-controlled data acquisition is used for unattended operation.

The following features are considered unique by the authors for production aging systems:

1. Loran-C/DTF Oscillator Frequency Standard
2. Direct Current Power Bus Design.
3. Measurement and Switching Techniques.
4. High Volume Automatic Resonator Aging.

Key words (for information retrieval) Production Oriented, Loran-C, Fractional Frequency, Allen Variance, Aging.

Introduction

This paper will review the design considerations applied to achieve a frequency drift measurement of parts in 10^{-11} per day in a production oriented test facility. The major sub-systems of the test facility include the frequency standard, measurement scanning console, aging consoles, power system, computer system and oscillator "set-up" console.

Discussion

Loran-C transmission is utilized to provide NBS frequency traceability.[1] The receiver produces a "raw" frequency reference phase-locked to a Loran-C signal. The signal is further conditioned with a Disciplined Time-Frequency standard oscillator, or DTF[2]. The resulting phase-locked output of the DTF has excellent long-term and short-term frequency stabilities which are typically 5×10^{-12} per day[1] and (1 second) Allen Variance of 2×10^{-12} respectively.[2]

The aging console is comprised of thirteen test drawers (Figure 1), each containing positions for 20 ovenized oscillators (Figure 2). Direct current power is obtained from high stability, high current supplies that power a maximum of 80 oscillators each. A special dc bus was fabricated for each supply. The bus consists of short, heavy, tin-plated copper bars with voltage and sense leads for the associated power supply, connected at the center (refer to Figure 3). Since individual power leads are run for each oscillator, this minimizes the effect of bus voltage variation due to bus loading with oscillators. The bus is also shielded and radio frequency decoupled for each chassis grouping of oscillator power source leads. Power is distributed from the bus to individual oscillators through unbroken, foil-shielded, twisted pair cables (see Figures 4 and 5). The power is also radio-frequency decoupled at each oscillator socket. In addition, the signal output for each oscillator is switched through radio-frequency matrix switches that provide a constant 50-ohm load to the oscillators and exhibit radio-frequency isolation of 80 dB between channels. The signal is then routed to a precision frequency counter that has a resolution of 0.0001 Hz (see Figure 6). The frequency reference for the counter is supplied by the DTF. The oven in each oscillator contains a thermistor for temperature monitoring. This output is measured at the test socket using a four-terminal measurement, essentially eliminating errors due to lead resistance (refer to Figure 6). Care must be taken when choosing the ohmmeter range used for reading the thermistor. If the range chosen generates too high a current, then self-heating in the thermistor can occur. This self-heating can cause large temperature offsets (several tenths of a degree) where oven stability may be ±50 milli-degrees or less. Also, since the wait states in the computer are not exact, a variation in timing will occur between readings. This will cause a variation in the temperature reading due to the variable length of time current is applied to the thermistor. A test current of 5 microamperes was found to be adequate when measuring the thermistor (5000 ohms) to overcome errors due to the ohmmeter output current. Channel selection for 260 channels is provided by seven commercial scanner units capable of the required four-terminal

*Operated for the U. S. Department of Energy by General Electric Company under Contract No. DE-AC04-76DP00656

**Aging, as referred to in this report, is the frequency drift per day from an initial value.

switching (refer to Figures 1 and 7). Good radio-frequency grounding and shielding techniques are used throughout each test chassis (see Figures 6 and 8).[3] All non-radio-frequency wiring is made of foil-shielded, twisted-pair conductors to minimize noise problems. Extreme care in this area is essential since radio-frequency coupling through a poorly designed direct current bus, such as one that does not incorporate "single point feed, single point ground", may cause fractional frequency pulling[4]. This problem, manifested by similar aging characteristics for all units, could precipitate complete dismantling of a test station to correct.[4] Spectrum Analyzer results reveal the content of the (shielded vs the non-shielded) noise or electromagnetic interference (EMI) that was observed on a supply bus. For an unshielded* bus with 22 oscillators operating (see Figure 9), noise and signals from other oscillators is down 40 dB_m from the reference. However, with the shielded bus (see Figure 10), noise is down 50 dB_m giving an additional 10 dB_m noise margin. Additionally, extremely good source voltage stability must be maintained over the test period to prevent power change influence on the natural crystal aging; Figure 12 shows this effect. A dc bus, powering 17 oscillators which had undergone long term aging (greater than 60 days), was loaded with 40 additional oscillators. Following the increased bus loading, a decrease was noted in the aging rate of the 17 oscillators (approximately 1 ppb/day) which later recovered when the bus was unloaded. The effect was not, however, sensed by the highly aged oscillators, as 200 additional oscillators were added onto other power buses in the system.

To ensure uninterrupted operation, even in the event of power failure, an uninterruptible power supply (UPS) is used. The UPS is a charger/battery/inverter unit capable of supplying 10 kW for 15 minutes, which is sufficient time for emergency generators to come on-line and take up the load. The system has saved several critical tests from interruption when thunderstorms caused many power interruptions in the local area and in the plant (which has emergency generators).

The computer system (Figure 11) consists of a desktop calculator, two floppy discs, a real time clock, and digital plotter. The system interface is the IEEE-488 instrument control bus. Once started, all measurements are made daily, automatically, for a specified length of time, typically 30 to 45 days. Daily measurements of 260 oscillators usually can be completed in 45 minutes; subsequently the computer system may be applied to other testing tasks or graphic plotting.

*The cover of the bus was removed; all power distribution wiring being foil-shielded twisted pair.

An oscillator set-up station was developed to permit precise manual adjustment of an oscillator/oven to the upper temperature "turning point" of the resonator, prior to installation of the oscillator into the aging console. The station has the capacity to set-up 20 oscillators, simultaneously. It has the same type of power source and bus as the aging console. It also has a precision ohmmeter (for temperature measurement), a computing counter and a keyboard. The computing counter has been used to verify (using "two-sample" Allen Variance) basic oscillator stability for a one-second "gating" time (which is used for the aging measurements).

The test facility has been in use for over two years. During this time, several hundred 20-MHz test bed oscillators have been operated successfully for 30-day crystal aging determinations. One prolonged test was conducted with a group of seventeen 20-MHz oscillators that were operated without interruption for over 500 days. Aging rates at the end of this period for most of the units, were predicted[5] and measured to be a few parts in 10^9/day[6] (see Figure 12). Several very stable 5-MHz oscillators have been directly measured on the system with aging rates of parts in 10^{11}/day (a typical plot may be seen in Figure 13). One data point on the plot represents an average of ten (one second) frequency measurements.

The aging model[5] currently being utilized for "ceramic flatpack"[7-10] quartz resonators is:

$$\frac{\Delta f}{f} = a + b \ln T, \qquad (1)$$

where T represents total days of oscillator operation (when T is greater than or equal to one). Aging rate per day is defined as b/T. Note the aging plot in Figure 13 for the relationship of the best fit of the "normalized" aging vs the actual data points. Also note the trace of the oven temperature regulation over the test period. This trace helps to resolve anomalies in frequency behavior of a resonator or in the malfunction of the test bed oven/oscillator.

References

1. USCG HQ, Loran-C System Performance Inquiry, April 1981.

2. D. Babitch, J. Ho and M. Bloch, Frequency Electronics Inc., "The Disciplined Time/Frequency Standard: A New Multi-Function Crystal Oscillator," Proceedings, AFCS 26, 1972.

3. R. Morrison, "Grounding and Shielding Techniques in Instrumentation," John Wiley and Sons, Inc., 1967.

[4] R. Greene, private communication, Greenray Industries, August 1978.

[5] J. A. Koehler, Test Requirements, Sandia National Laboratories, Albuquerque, N. M., December 1978.

[6] R. J. Filler, E. J. Kerres, T. M. Snowden and J. R. Vig, "Ceramic Flatpack Enclosed AT and SC-cut Resonators," Proceedings, 1980 IEEE Ultrasonics Symposium, 1980.

[7] E. Hafner and J. R. Vig, "Crystal Resonator Housing Configuration," U. S. Patent No. 3,991,388, Jan. 6, 1976.

[8] J. R. Vig and E. Hafner, "Packaging Precision Crystal Resonators," Technical Report ECOM 4154, U. S. Army Electronics Command, Fort Monmouth, N. J., July 1973.

[9] P. Wilcox, G. Snow, E. Hafner and J. R. Vig, "A New Ceramic Flatpack for Quartz Resonators," Proceedings, AFCS 28, June 1974, p. 207-218.

[10] R. L. Peters, "Ceramic Flatpack Enclosures for Precision Quartz Crystal Units," Proceedings, AFCS 30, June 1976, p 22-31.

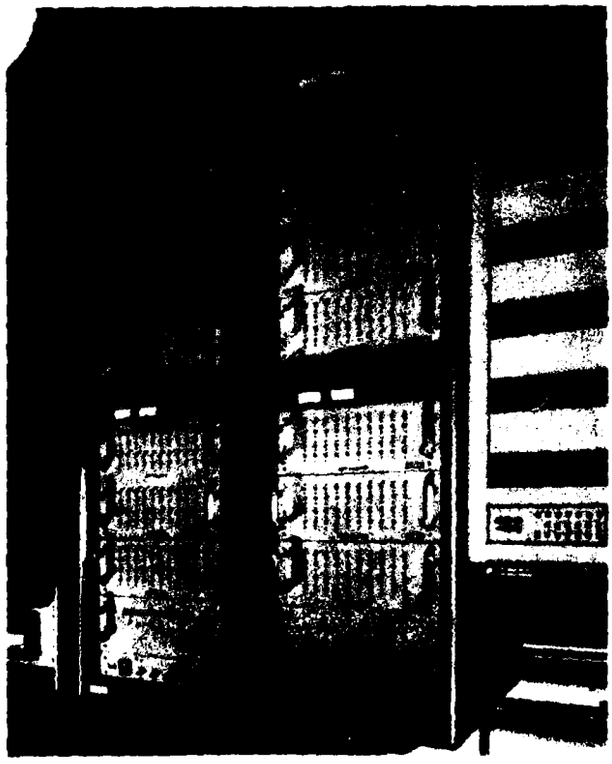


Figure 1. Aging Consoles

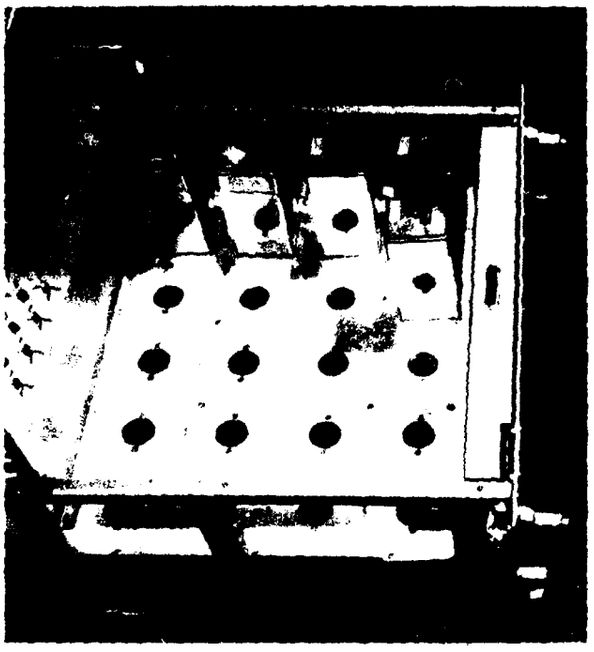


Figure 2. Aging Chassis (Top View)

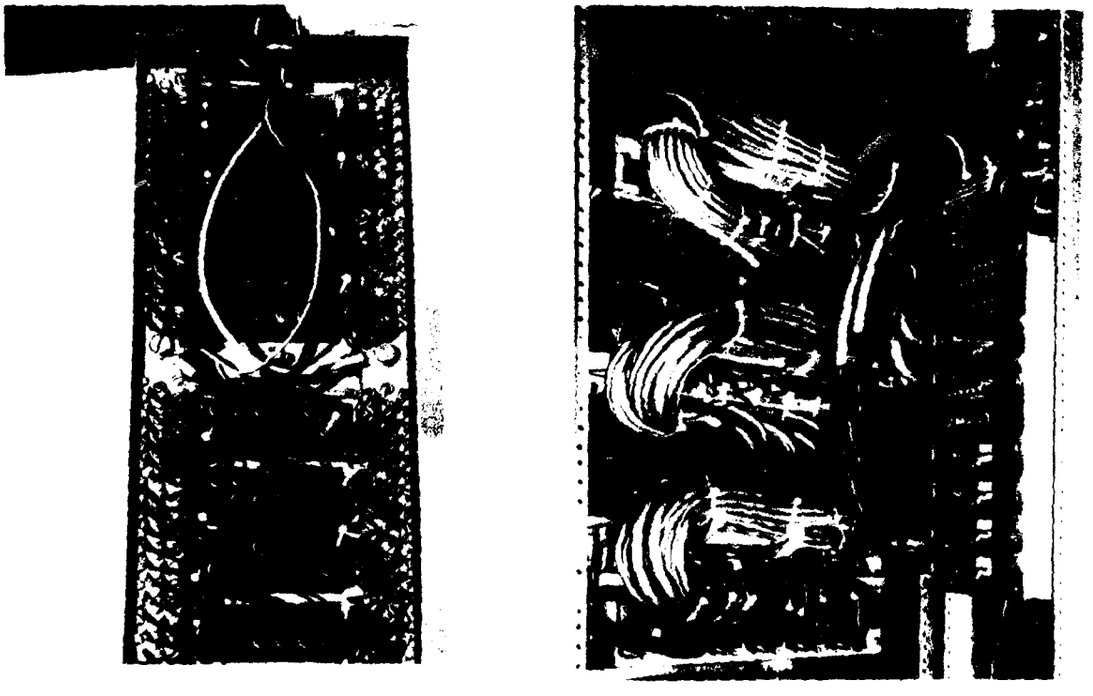


Figure 3. Burned Current Bus Module

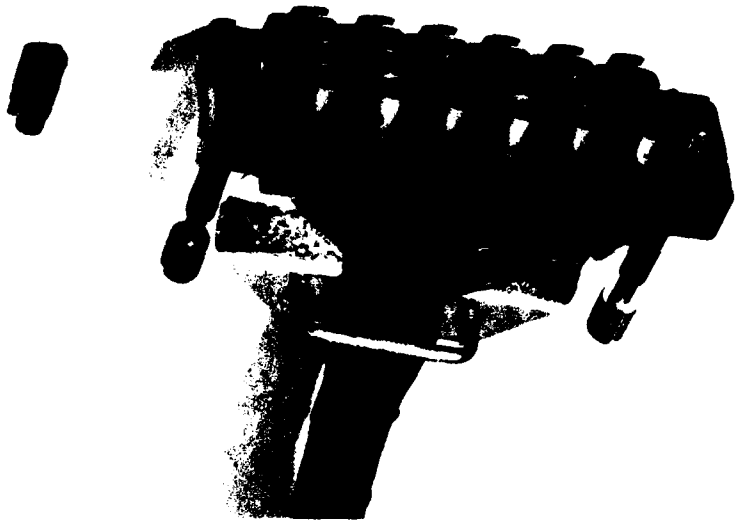


Figure 5. Typical Connector (Power Distribution and Temperature Measurement)

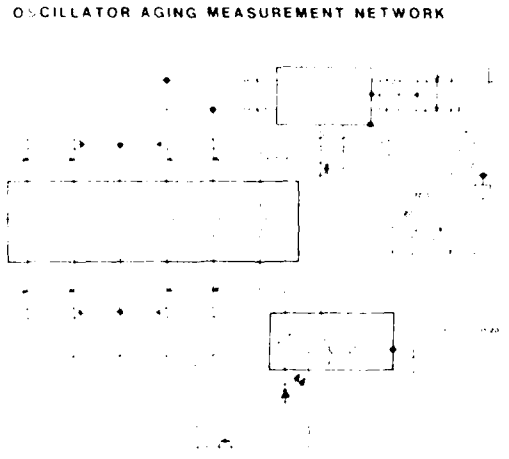


Figure 6. Oscillator Aging Measurement Network

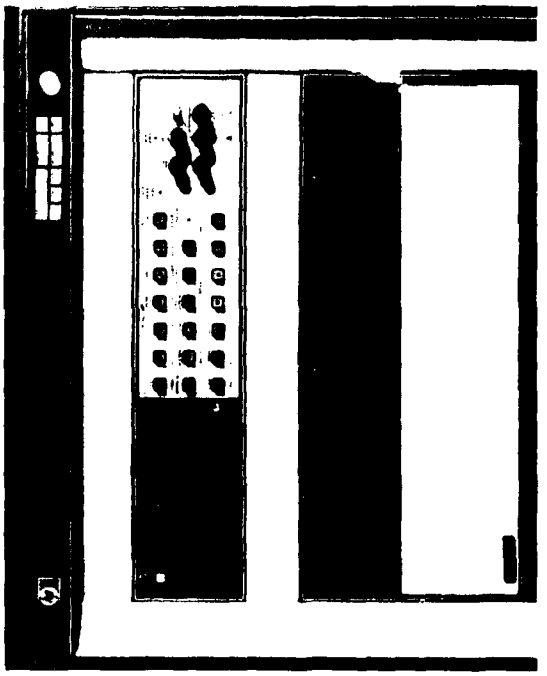


Figure 7. Measurement/Scanner Console



Figure 8. Aging Chassis (Bottom View)

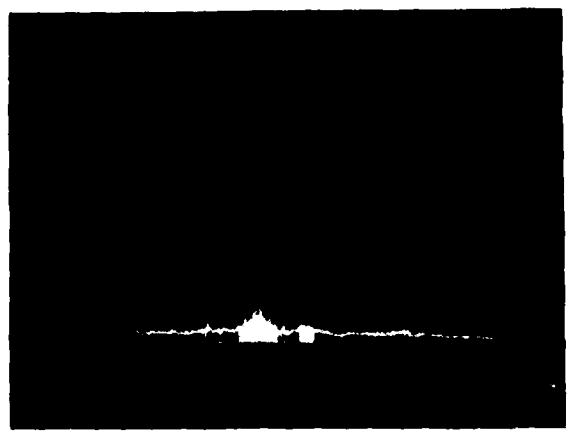


Figure 9. Power Bus Noise (Unshielded)

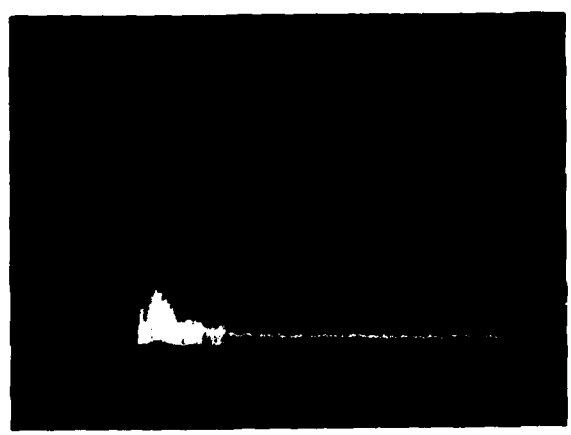


Figure 10. Power Bus Noise (Shielded)



Figure 11. System Computer

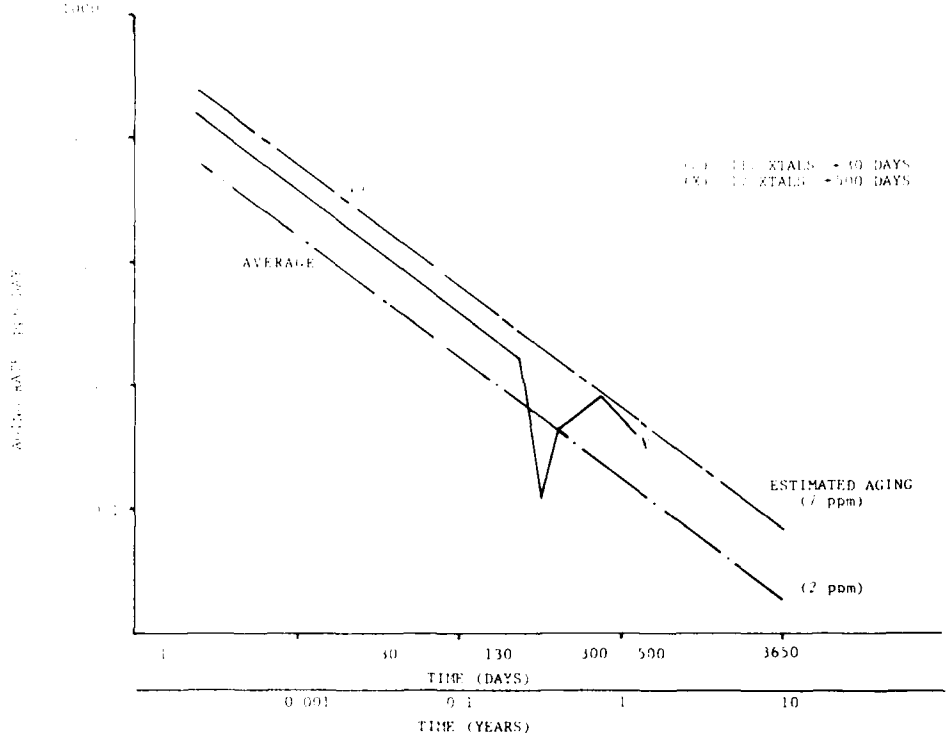


Figure 12. 20-MHz Ceramic Flatpack Quartz Resonator Aging Study

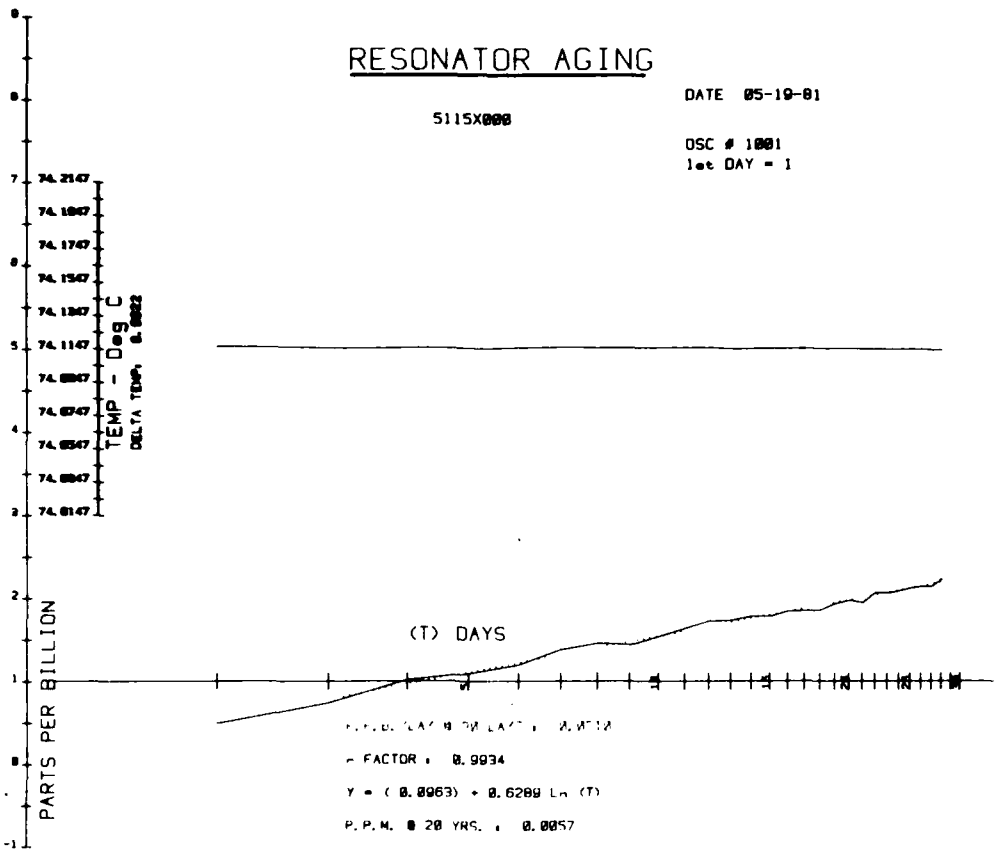


Figure 13. Commercial 5-MHz High-Precision Oscillator Aging

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

COMPARISON OF METHODS FOR MEASUREMENT OF QUARTZ CRYSTAL RESONATORS WITH LOAD CAPACITANCE

W.H. Horton, T.S. Payne, R.C. Smythe & D.A. Symonds

PIEZO TECHNOLOGY INC.
ORLANDO, FLORIDA 32804

Summary

Measurements on crystal resonators are usually performed with the crystal unit connected in series with a specified load capacitor for two reasons: first, to simulate the conditions under which the crystal will be operating when used in an oscillator; second, as one method of determining crystal motional parameters. This paper concerns itself primarily with the first of these.

Accurate measurement of the load resonance frequency and resistance is considerably more difficult than measurement of series resonance frequency and resistance. In load capacitor measurement systems errors due to the effect of known stray capacitances on frequency are straightforward to accommodate. The effect of stray capacitance on resistance measurement is more subtle. In currently used CI meters this type of resistance error can be significant.

In a passive measurement system the technique of phase or reactance offset can be used to overcome these difficulties. However, this indirect method is not applicable to CI meter measurements.

This paper reviews and compares measurement problems and techniques associated with the following methods:

1. CI meter.
2. Pi-network transmission method.
3. T-network & inverted-L network transmission methods.
4. V - I Method.
5. Reflection coefficient brige.

Design and construction of a special load-capacitance fixture is discussed. Such a fixture may be used for both CI meter and transmission measurements and facilitates correlation between methods. It is especially valuable in obtaining correlation between manufacturer and customer when they use different measurement systems.

1.0 Introduction

Crystal units are measured with a series load capacitor for two reasons:

1.1 To determine the crystal equivalent circuit motional inductance and capacitance or related properties such as pullability.

1.2 To determine effective frequency (f_L) and resistance (R_L) under load conditions which are intended to simulate those existing when the crystal unit is incorporated into an oscillator circuit. It is this measurement which is the main concern of this paper.

It is possible to calculate f_L & R_L using the crystal equivalent circuit values measured by other means[1]. However significant errors may arise due to nonlinearity, coupled modes or errors in the crystal equivalent circuit model. Hence it is good practice to determine load frequency and resistance by direct measurement.

This paper reviews methods for the direct measurement of f_L and R_L. Section 2 reviews some equivalent circuit fundamentals. Sections 3 and 4 outline the advantages and limitations of the load capacitor method and the reactance offset method. Section 5 analyses the effect of parasitic elements on measurements with a load capacitor and describes a load capacitor fixture having low parasitic capacitance.

Finally, Section 6 discusses types of measurement systems in current use, while Section 7 gives a comparison of measurements of f_L and R_L by the two methods using several measurement systems.

2.0 Impedance at Load Resonance Frequency (f_L)

Consider the simplified equivalent circuit of a crystal resonator, figure 1a. Under certain usually valid assumptions, the impedance of the resonator can be represented by a circle in the R-X plane, figure 1b. A definitive treatment of the resonator circle diagrams has been given by Hafner[1]. In figure 1l, f_r is the series resonance frequency, at which, by definition, the resonator impedance is resistive, of value R_r. If a load capacitor, C_L, is now connected in series with the resonator, figure 2a, the impedance of the series combination is represented by a new circle whose center is shifted by an amount equal to the reactance of C_L, figure 2b. The series resonance frequency of the crystal unit - load capacitor combination becomes f_L, the load resonance frequency, the frequency at which

the total impedance is resistive of value R_L. It is important to note that if the same crystal without a load capacitor was measured at the same frequency f_L its impedance would be R_L in series with $-C_L$ as is shown in Fig. 3. Such a measurement will be referred to as a reactance offset measurement. The advantages and disadvantages of the load capacitor and the reactance offset methods of measurement are discussed in the next two sections.

3.0 Load Capacitor Measurements:

Load capacitor measurements are performed with a physical load capacitor connected in series with the crystal unit. This technique has the following features:

3.1 It is only feasible for negative reactance loads because of the unavoidable and excessive loss effects associated with an inductor used as a positive reactance load.

3.2 Fixtures most conveniently have only one value of load capacitance.

3.3 Careful design is needed to obtain a fixture that is free of stray impedances.

3.4 When load capacitor measurements are used to determine motional parameters the crystal C_o must be measured, or else measurements must be made with more than one load capacitor value so that the effect of C_o may be cancelled. The normal procedure is to measure f_r, C_o, f_L (for known C_L) and calculate

$$L_1 = \frac{1}{8\pi^2 f_r (f - f_r)(C_o + C_L)}$$

3.5 The measurement can be made in an oscillator, as f_L is a frequency where there is both zero phase shift across the crystal - load capacitor combination and a high phase slope exists.

3.6 Because of the high phase slope only small errors in resistance and frequency result from quite large phase changes. Hence, passive measurement systems do not require high resolution of phase.

3.7 High drive levels can be obtained easily because the impedance at f_L is R_L. In the reactance offset method the impedance at f_L is R_L in series with $-X_L$. Usually, X_L is much larger than R_L and limits the maximum drive level that can be obtained in the latter method.

4.0 Reactance Offset Measurements

Reactance offset measurements[2] are made possible by the availability of instrumentation which can determine the real and imaginary components of impedance. The method has the following features:

4.1 It can be used for positive or negative reactance loads.

4.2 Any number of arbitrary values of load reactance can be used.

4.3 Fixture design is relatively simple as only the basic measurement network is used; consequently, there are no strays which cannot easily be measured and cancelled.

4.4 In motional measurements crystal Co does not have to be determined with great accuracy as errors in its measurement tend to be self-cancelling. Normally Co is determined by measuring the impedance at frequencies above and below f_r. In subsequent measurements, the susceptance of Co is subtracted from the total resonator admittance, giving the motional arm admittance. An iterative frequency search is performed until the motional arm is purely resistive, giving f_s and R_1. Then motional reactance values X_u and X_l are measured at frequencies $f_u = f_s + \Delta f$ and $f_l = f_s - \Delta f$.
The motional inductance is determined as:

$$L = \frac{1}{4\pi} \frac{Xu - X}{Fu - F}$$

Note that for resonators having adequate figures of merit it is possible to measure motional parameters without measuring Co, but the above procedure is preferred, and the value of Co is generally required for other purposes.

4.5 The technique cannot readily be adapted for use with oscillators because measurements must be made with non-zero phase shift across the crystal.

4.6 The measurement technique requires high resolution and accuracy of phase, and to a lesser extent, amplitude. Consequently the measurement system is expensive.

4.7 High signal source amplitudes are required to obtain typically specified drive levels because of the high equivalent reactance in series with R_L.

5.0 Load Capacitor Fixture Design

5.1 Stray Analysis

Numerical analysis of the generalized load capacitor network (fig. 4) was performed to determine the effects of stray capacitance, and certain non-ideal, components. A source and load resistance of 20 ohms was assumed, but the measurement errors depend only weakly on this value. In figure 4,

- $C_{1,3}$ represents the load capacitor.
- $C_{1,4}$ and $C_{2,4}$ represent input and output strays to ground.
- $C_{1,2}$ is a bridging capacitor
- $C_{3,4}$ is stray C from the load-crystal junction to ground.
- $C_{2,3}$ represents stray capacitance in parallel with the crystal Co.
- $R_{1,3}$ is the loss of a switch typically included to short-circuit the load capacitor.

The results, tables 1 and 2, show that the only strays which cause significant errors are as follows:

5.1.1 C_{34} (Resistance error). It is assumed that the load capacitor $C_{1,3}$ was adjusted so that the capacitive load is correct, and hence no f_L error is introduced.

5.1.2 R_{13} (Resistance error). Note that the error in ohms is independent of the crystal resistance.

5.1.3 C_{23} (Frequency and resistance errors). The more significant frequency error may be corrected but a small resistance error will remain.

5.2 Physical Arrangement

The fixture shown in Fig. 5 was designed taking into account the above analysis. Important features of the fixture are:

5.2.1 It can be used in both oscillator and Pi- network type measurement systems.

5.2.2 Calibration can be performed using a 3-terminal capacitance bridge.

5.2.3 The capacitance to ground at the load capacitor - crystal junction is very small.

5.2.4 The load capacitor shorting switch has low loss.

5.2.5 There is a defined reference plane connected to terminal 2.

5.2.6 Crystal units with wire or pin leads can be accommodated.

5.2.7 Its reproducibility makes it useful for measurement correlation between different organizations.

5.2.8 The maximum useable frequency is limited, depending on dimensions. (Table 3).

5.2.9 Sockets are susceptible to wear so the fixture may not be suitable for high-volume production use.

6.0 Measurement Systems

Systems for resonator measurement may be classified into two categories -- oscillators or CI meters, and passive or transmission systems.

6.1 C.I. Meter

Oscillators suitable for crystal measurement are commonly referred to as crystal impedance meters (C.I. Meters)[3]. A greatly simplified C.I. meter is shown in figure 6. Important characteristics are:

6.1.1 The maximum value of R_L which can be measured is determined by the requirement of unity loop gain.

6.1.2 Accurate measurement of frequency and resistance requires the use of calibrated substitution resistors. Since the phase shift of the active circuitry may be level-dependent, this substitution must be performed at the desired drive level.

6.1.3 For f_L and R_L measurements a load capacitor is used; hence there will be errors due to C_s and other stray capacitances as discussed previously.

6.1.4 At f_L the impedance between nodes 1 and 2 is R_L (neglecting C_s), which facilitates the use of higher drive levels then are available in the reactance offset method (see 3.7.).

6.2 Passive Systems

The essential features of passive systems for resonator measurement are described by the block diagram of Fig. 7. A tunable source produces two phase and amplitude related signals,one of which is applied to the crystal through a fixture or test network. This network may be as simple as a pi-network, or as complex as an impedance or admittance bridge. The signal output from the fixture is compared with the other, reference, signal in a vector ratio detector. The impedance can be determined from the reading of the ratio detector. A wide variety of test networks can be used, each having its own advantages and disadvantages. With this type of system it is possible either to insert a load capacitor and search for points of pure resistance or to search for reactance of a specific value. As the measurement frequency is not determined by the device being tested it is possible to measure the parameters of unwanted modes.

Commonly used networks are:

6.2.1 Pi Network (Fig. 8)

The most common form of the pi-network conforms to IEC Standard 444,[4] with R_1 = 15.9 ohm R_2 = 14.2 ohm R_3 = 66.2 ohm. When used in a 50 ohm system R_M is 12.5 ohm. Some properties of the Pi are:

Its maximum useful frequency is limited. IEC Standard 444 covers only the frequency range from 1 MHz to 125 MHz, although a method for extending this to 250 MHz has been proposed[5].

The detected signal amplitude is nearly maximum at zero phase.

The measurement relies on the accuracy of the Pi-network impedance, although by measuring standards impedances corrections can be made. Suitable standards are not easily obtainable however, especially for high-frequency use.

The crystal unit has neither pin grounded.

The IEC pi-network requires 50 ohm source and load impedances.

6.2.2 Inverted L - Network (Fig. 9)

Properties of the L-network are:
Its maximum useful frequency is somewhat greater than for the pi-network. By adding an adjustable tank circuit or stub line across the measurement port, the frequency range may be extended still further.

The detected signal amplitude is nearly minimum at zero phase.

The measurement relies on the accuracy of R_f, although by measuring standards corrections can be made.

The crystal unit has one pin grounded.

The network requires a low impedance (50 ohm) source and a high impedance detector to measure V_2.

6.2.3 T - Network (Fig. 10)

The properties of the T-network are similar to those of the L-network, except that the T permits the use of a low impedance detector.

6.2.4 V-I Method (Fig. 11).

In this technique a current is passed through both the device to be measured and a reference resistor and from vector voltage measurements of V_1 and V_2 the impedance of the device is calculated.

The technique relies on the accuracy of R_{ref}.

The device being measured is floating, although by suitable arrangement one terminal may be grounded .

It is not a 50 ohm system.

The Hewlett Packard 4192A Impedance analyzer implements this approach and includes Kelvin connections to the unknown impedance. The block diagram of this instrument is shown in Fig. 12. By means of internal processing the instrument can display the impedance in series or parallel form. The upper frequency limit of the instrument is 13 MHz.

6.2.5 Reflection Coefficient Bridge

A system incorporating a reflection coefficient bridge is available as the Hewlett Packard 4191A Impedance Analyzer. Its overall block diagram is also represented by Fig. 12.
The use of this instrument for resonator measurement is treated in a companion paper.[2]

A simplified schematic of the bridge and associated circuitry is given in Fig. 13, which also shows the external frequency synthesizer required to obtain frequency resolution adequate for resonator measurements.
Some properties of the 4191A are:

It is useful to 1000 MHz.

The detected signal is proportional to the reflection coefficient of the impedance under test and is therefore a minimum at 50 ohms resistive.

It is a coaxial instrument. Hence, the device has one pin grounded.

It relies on the accuracy of coaxial impedance standards, short circuit, open circuit and 50 ohms.

6.2.6 Impedance and Admittance Bridges

R.F. impedance and admittance bridges have been used for many years for the determination of crystal resonator equivalent circuit parameters. Because of their limited availability, the use of these bridges will not be considered here. Moreover, those commercially available bridges of which the authors are aware, besides being tedious to use, possess range limitations which limit their use for reactance offset measurements. Also beyond the limited scope of this paper is the automated microcircuit bridge system being developed for precision resonator measurements[6]. This system uses a Schering bridge, and incorporates standard load capacitors for the measurement of f_L and R_L.

6.2.7 Resolution

To determine the resolution obtainable in measurements of f_L & R_L for the different networks by the reactance offset method a simulation was carried out for a typical 4 MHz crystal for a unit error in the least significant digit of display resolution of suitable instruments. No allowances were made for fixture errors, accuracy errors, or noise.

The results (Table IV) show that all instruments have sufficient display resolution to obtain the load frequency for this unit to better than 1 Hz. The measurement of R_L is more sensitive. For the example given, the best resolution is obtained on the HP 4192 A system.

6.2.8 Instrument Accuracy

Reliable information on the accuracy of measured parameters is difficult to estimate from manufacturers' data. In all instruments usable accuracy tends to be better than specified accuracy, partly because worst case is used for obtaining specifications and also because in the measurement offsets are made so that relative accuracy is more important than absolute accuracy.

For this measurement the HP 3042 system has an amplitude accuracy of about 0.02

dB/10 dB if calibrated by the synthesizer, and a phase accuracy of about 0.30°. This gives, for the pi-network, load frequency to within 0.4 Hz and load resistance to within 7 ohms. (All figures in this section refer to reactance offset measurements of the 4 MHz crystal described in Table IV).

The HP 4192A has a C accuracy of about 0.22 pf and a G accuracy of about 0.4 mS giving load frequency to within 0.4 Hz and resistance to within 0.9 ohms for this measurement according to its specification.

At low frequencies the HP 4191A has a guaranteed accuracy of Γx & Γy of about 0.007 and typically 0.0018. This gives load frequency to within 56 Hz and resistance to within 119.7 ohms for guaranteed values or frequency within 9.6 Hz and resistance within 20 ohms for typical values. Our results[2] suggest that the f_L and R_L errors are no more than one-tenth of the latter figures.

It should again be emphasized that the errors in measuring f_L and R_L are in general much greater than in measuring f_r and R_r or the equivalent circuit parameters.

7.0 Measurement Results (Table V)

A crystal unit was measured in several different measuring systems. All measurements with a physical load capacitor were made using the load capacitance fixture described earlier in this paper. Since the measurement temperature was not held constant from system to system, the frequency difference $\Delta f = (f_L - f_r)$ rather than f_L, must be examined. The first reading of Table V was arbitrarily chosen as the reference to obtain the variation in Δf in ppm. Excluding the reactance offset measurement in the HP 3042 system, the Δf variation was less than one ppm.

The value of R_L using the PTI load capacitance fixture in the Saunders 110HF and the RFL 459A CI meter could be determined only to the nearest 5 ohms. The reading for HP 3042 reactance offset measurement varied between 19 and 24 ohms. The readings obtained in the HP4191A, and HP4192A by reactance offset and in the pi-network with PTI load capacitor fixture were within one ohm of each other and repeatable.

8.0 Errors due to the 3 terminal nature of the crystal. (Fig. 14).

The value of f_L and R_L are affected by the effective value of Co of the crystal. The value of Co in the circuit is dependent on the type of connection to the can (Fig. 15). This connection must be specified before the load values can be correctly measured. If the connection is not specified, correlation problems can exist between measurement procedures which are otherwise identical.

To solve this problem a standard measurement procedure is required with allowances for offsets. Connection to the can should be made in a specified manner. To facilitate this a third pin connected to the can may be added to the crystal unit. Alternatively, one pin may be connected to the can.

9.0 Conclusion

We have described various methods of obtaining load resonance frequency & resistance measurements of crystal resonators and introduced a fixture design to make possible correlation measurements between oscillator and transmission measurements. We have shown that the introduction of two new R.F. Impedance Analyzers makes the measurement by the reactance offset method a viable alternative to the load capacitor method.

REFERENCES

1. Hafner, E., "The Piezoelectric Crystal Unit - Definitions and Methods of Measurements", Proc. IEEE, v. 57 No. 2, pp. 179-201; Feb. 1969.

2. Smythe, R. C., "An Automated Resonator Measurement System using a Reflection Coefficient Bridge", these Proceedings.

3. Gerber, E. A., "A Review of Methods for Measuring The Constants of Piezoelectric Vibrations". Proc. I.R.E., v. 41, No. 9, pp. 1103-1112; Sept., 1953.

4. International Electrotechnical Commission, "Basic method for the measurement of resonance frequency and equivalent series resistance of quartz crystal units by zero phase technique in a - network". Publication 444 1st Ed.; 1973.

5. Fischer, R., & L. Schulzke, "Extending the Frequency Range of the Transmission Line Method for the Measurement of Quartz Crystals up to 250 MHz". Proc. 31st. A.F.C.S., pp. 96-101; 1977.

6. Malinowski, G., & E. Hafner, "Automatic Microcircuit Bridge for Measurements on Quartz Crystal Units". Proc. 32nd A.F.C.S., pp. 354-364; 1978.

FIGURE 1

CRYSTAL WITHOUT LOAD CAPACITOR

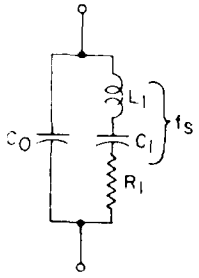


FIG 1a

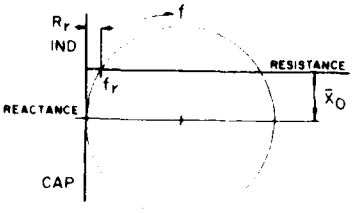


FIG 1b

FIGURE 2

CRYSTAL WITH LOAD CAPACITOR

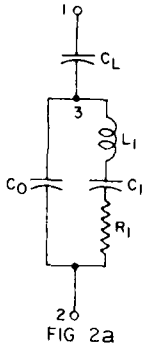


FIG 2a

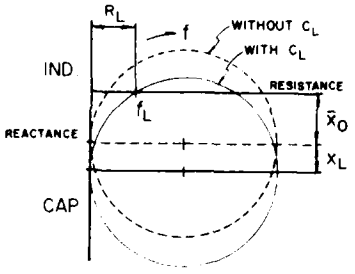
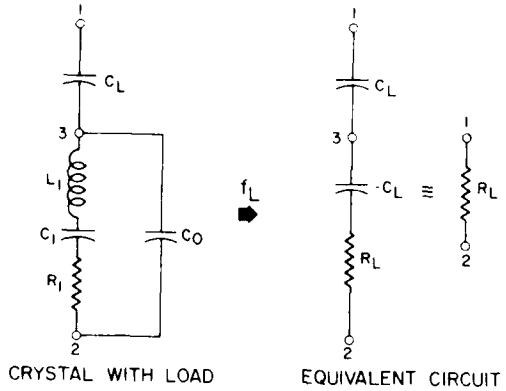


FIG 2b

FIGURE 3

EQUIVALENCE AT f_L

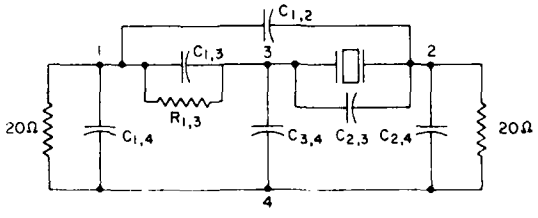


CRYSTAL WITH LOAD

EQUIVALENT CIRCUIT AT f_L

FIGURE 4

SCHEMATIC USED FOR ANALYSIS



f_S	=	6397930	HZ
L_1	=	261	MH
C	=	62	PF
R_1	=	10	Ω

FIGURE 5

LOAD CAPACITANCE FIXTURE

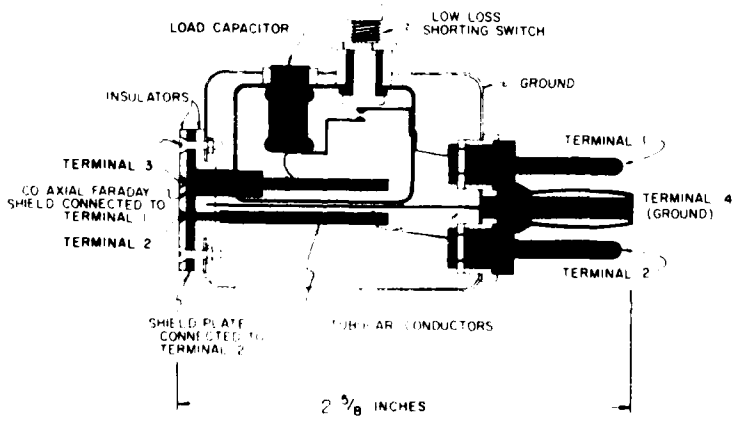


FIGURE 6

SIMPLIFIED C.I. METER

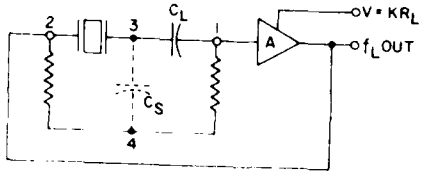


FIGURE 7

PASSIVE MEASUREMENTS

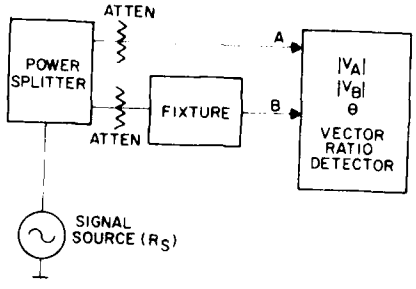
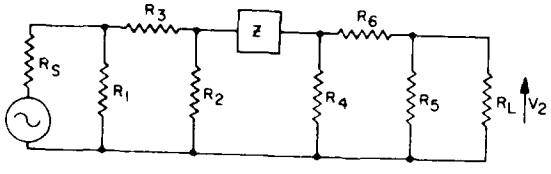
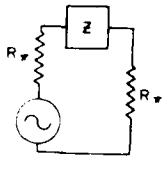


FIGURE 8

π NETWORK





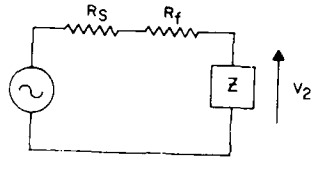
$Z = 2R_\pi(V-1)$

WHERE $V = \frac{V_{20}}{V_2}$

$V_{20} = V_{(Z=0)}$

FIGURE 9

L NETWORK

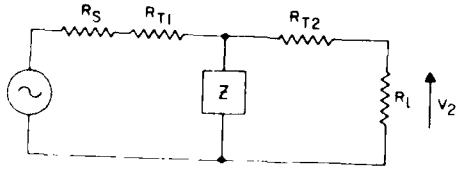


$Z = \frac{R_f + R_S}{(V-1)}$ WHERE $V = \frac{V_{20}}{V_2}$

$V_{20} = V_{(Z=\infty)}$

FIGURE 10

T NETWORK



$Z = \frac{R_T + R_S}{2(V-1)}$ WHERE $R_T = R_{T1} = R_{T2}$

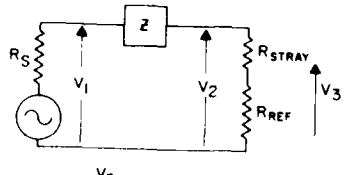
$R_S = R_L$

$V = \frac{V_{20}}{V_2}$

$V_{20} = V_{(Z=\infty)}$

FIGURE 11

V-I MEASUREMENT METHOD



$I = \frac{V_3}{R_{REF}}$

$Z = \frac{V_1 - V_2}{I} = \frac{(V_1 - V_2) R_{REF}}{V_3}$

FIGURE 12

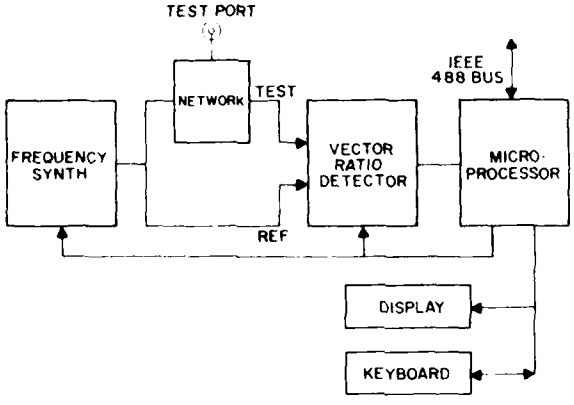


FIGURE 13

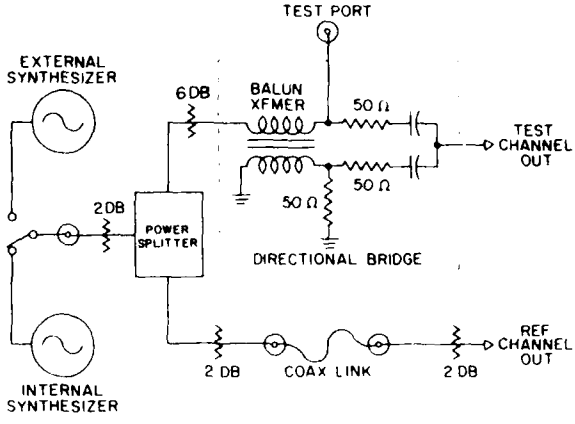
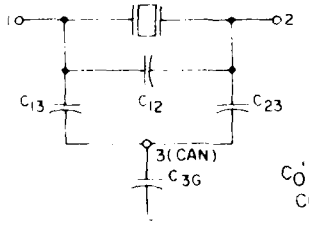


FIGURE 14

ERRORS DUE TO 3 TERMINAL NATURE OF CRYSTAL



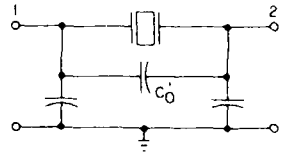
$$\Delta f = \frac{1}{8\pi^2 L f_S (C_O' + C_L)}$$

C_O' VARIES DEPENDING ON CONNECTION TO CAN

FIGURE 15

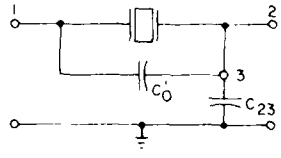
CAN FLOATING

$$C_O' = C_{12} + \frac{C_{13}\ C_{23}}{C_{13}+C_{3G}+C_{22}}$$



CAN CONNECTED TO PIN 2

$C_O' = C_{12} + C_{13}$



CAN GROUNDED

$C_O' = C_{12}$

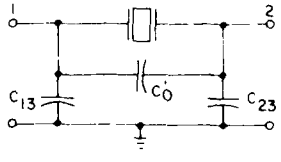


TABLE I

EFFECT OF STRAY CAPACITANCE
(RESULTS FROM COMPUTER ANALYSIS)

STRAY CAPACITANCE	$C_{1,4}$ (PF)	$C_{1,2}$ (PF)	$C_{1,3}$ (PF)	$R_{1,3}$ Ω	$C_{2,3}$ (PF)	$C_{3,4}$ (PF)	$C_{2,4}$ (PF)	R_1 XTAL Ω	f_L (HZ)	R_{EQUIV} Ω
ZERO	0	0	32	10^{12}	0	0	0	10	6399915	14.3
IN/OUT, SHUNT	15	0	32	10^{12}	0	0	10	10	6399914	14.3
IN/OUT, BRIDGING	0	5	32	10^{12}	0	0	0	10	6399915	14.3
LOAD/XTAL JUNCTION, TO GROUND	0	0	27	10^{12}	0	5	10	10	6399914	17.5
ACROSS XTAL SOCKET	0	0	32	10^{12}	0.5	0	0	10	6399890	14.6

TABLE II

EFFECT OF LOSS IN SERIES CAPACITOR

(RESULTS FROM COMPUTER ANALYSIS)

	$C_{1,4}$ (FF)	$C_{2,4}$ (PF)	$C_{1,3}$ (PF)	$R_{1,3}$ Ω	$C_{3,4}$ (PF)	$C_{2,4}$ (PF)	R_I XTAL Ω	f_L (HZ)	R_{EQUIV} Ω
REFERENCE	0	0	32	10^{12}	0	0	10	6399915	14 3
WITH LOSS	0	0	32	10^{5}	0	0	10	6399915	20 3
REFERENCE	0	0	32	10^{12}	0	0	80	6399921	114 0
WITH LOSS	0	0	32	10^{5}	0	0	80	6399921	170 5

TABLE III

LOAD CAPACITANCE FIXTURE

C_L = 32 PF C_{14} = 13 57 PF C_{24} = 6 18 PF

LOAD CAPACITANCE VS FREQUENCY

F MHz	C_L pF
001	32 00
010	31 93
100	31 91
1 000	31 88
5 000	31 83
10 000	32 08
13 000	32 30

TABLE IV

REACTANCE OFFSET RESOLUTION LIMITS FOR f_L & R_L

	INSTRUMENT RESOLUTION	Δf_L (HZ)	ΔR_L (Ω)
12 5Ω π NETWORK (HP 3042 SYSTEM)	01°	- 01	- 22
	01 DB	71	05
5000Ω L-NETWORK or 10000Ω T-NETWORK (HP 3042 SYSTEM)	01°	05	22
	01 DB	71	56
C//G (HP 4192A)	1 μS	0	15
	01 PF	- 018	014
REFLECTION COEFFICIENT BRIDGE 50Ω (HP4191A)	ΓX + 0001	1	- 1 54
	ΓY + 0001	7	- 17

FOR A UNIT WITH f_S = 4 000 MHz AT C_L = 32 PF (X = -J1243Ω)
R_S = 20Ω R_L = 23 93Ω
L = 140 MH Δf = 646 HZ

TABLE V

MEASUREMENT RESULTS

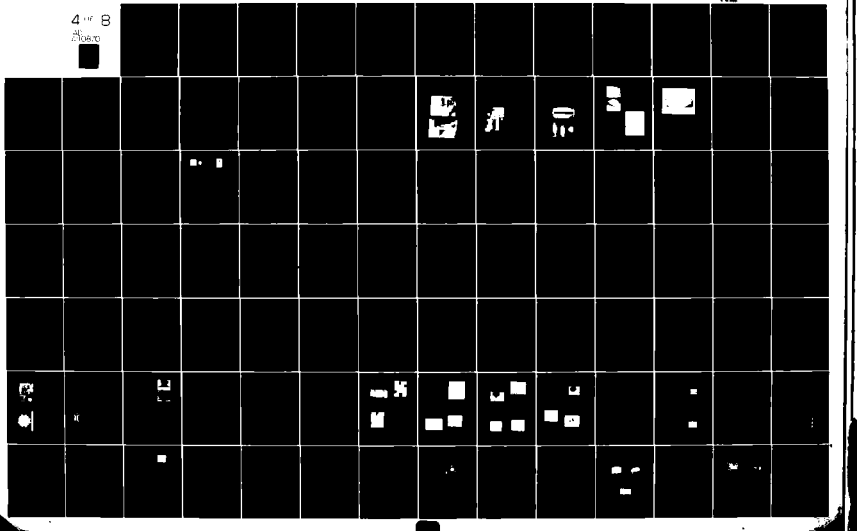
TYPE OF MEASUREMENT	INSTRUMENT SYSTEM	f_r (HZ)	R_r Ω	f_L (HZ)	R_L Ω	Δf (HZ)	Δf_{VAR} (PPM)
PHYSICAL C_L *	SAUNDERS 110 HF	4000191		4000829	25	638	0
PHYSICAL C_L *	CI METER 459A	4000194		4000830	30	636	- 5
REACTANCE OFFSET	HP 4191A	4000196	21 28	4000837	25 1	641	+07
REACTANCE OFFSET	HP 4192A	4000198	21 27	4000841	25 3	643	+12
REACTANCE OFFSET	π NETWORK HP 3042 SYS	4000193	21 59	4000820	23 8	627	- 27
PHYSICAL C_L *	π NET W/C_L HP 3042 SYS	4000189	21 3	4000828	26 0	639	+ 25

* USING LOAD CAPACITANCE FIXTURE

AD-A110 870 ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND AD--ETC F/G 9/5
PROCEEDINGS OF THE 35TH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL. --ETC(U)
1981

UNCLASSIFIED

NL



Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

AN AUTOMATED RESONATOR MEASUREMENT SYSTEM USING A REFLECTION COEFFICIENT BRIDGE

R. C. Smythe

Piezo Technology Inc.
Orlando, Florida 32804

SUMMARY

An automated measurement system has been developed for the precise determination of the equivalent circuit parameters of quartz resonators. The system is based on the H-P 4191A RF Impedance Analyzer[1], a programmable instrument using a reflection coefficient bridge and having a frequency range of 1 to 1000 MHz. Using certified impedance standards the measurements can be made traceable to national standards.

The complete measurement system consists of the impedance analyzer, a programmable frequency synthesizer, and a suitable controller. Special fixtures have been made to accommodate standard crystal packages. The system hardware and principal software features are described, and measurement examples are given.

Besides obtaining the equivalent circuit parameters, the system permits measurement of resonance frequency and resistance with a simulated load capacitance. Parameters of unwanted modes may also be measured. The system is suitable for measurement of SAW resonators as well as bulk-wave resonators.

Hardware Description

The system block diagram of Figure 1 shows that the system consists of the H-P 4191A RF Impedance Analyzer together with a suitably chosen programmable frequency synthesizer (we use a PTS-200 synthesizer[2] for most measurements below 200 MHz) and an appropriate controller (we use a H-P 9825 desktop computer). In addition a suitable test fixture and a set of reference impedances (a set is supplied with the 4191A) are required.

The instruments are interfaced via the IEEE -488 bus. Using the bus, or other interfaces, any desired peripheral devices, such as printers, plotters, or mass memory devices, may readily be added to the system.

The H-P 4191A measures the reflection coefficient, with respect to 50 ohms, of an unknown impedance. An internal microprocessor converts the measured value into any of several forms -- for example, into R + jX or into G in parallel with C. The frequency range of the instrument is 1 MHz to 1000 MHz. The microprocessor allows the instrument to be calibrated against reference impedances -- an open circuit, a short circuit and a 50 ohm termination. The microprocessor stores the calibration information and uses it to correct future measurements[1].

The measurement port of the 4191A is an APC-7 precision coaxial connector; it is therefore necessary to provide an appropriate test fixture for resonator measurement. Moreover, the contribution of the fixture to the measured reflection coefficient must be taken into account. This may be done in either of two ways.

First, if the reference impedances can be applied at the resonator measurement port instead of at the instrument port, then a calibration may be performed which includes the fixture. Figure 2 shows a fixture which allows this to be done, and which accepts HC-18 style resonator packages.

If this approach cannot be followed, then it is necessary to determine the equivalent circuit of the fixture and correct the measurements appropriately. Figure 3 shows a typical equivalent circuit. As an example, for the H-P 16092A spring clip fixture, the equivalent transmission line length is .34 cm. Z_1 represents a 1.3 nH. inductor and a small series resistance, while Y_2 represents a capacitance, typically 1.3 pF., in parallel with a small conductance.

Using this system, all the usual resonator measurements can be performed at frequencies from 1 to 1000 MHz. Of particular interest are the measurements of the equivalent circuit parameters and the measurements of the load frequency and resistance, f_L and R_L. These measurements are described in the next two sections.

Measurement of Equivalent Circuit Parameters

The equivalent circuit of Figure 4 is assumed[3]. C_o is determined as the average of shunt capacitance measurements at 1.1 f_{nom} and .8888 f_{nom} (Appendix A) where f_{nom} is the nominal value of f_s.

At any frequency, f, the admittance of the motional branch is just the total resonator admittance less the admittance of the shunt capacitance, $2\pi f C_o$. The series resonance frequency, f_s, is found by changing the measurement frequency until the impedance of the motional branch

is a pure resistance, R_1. The search for f_s proceeds in two stages. In the first stage, starting at f_{nom}, the frequency is increased or decreased in constant increments, R, until the sign of the motional reactance changes, thus bracketing f_s within an interval R. In the second stage, an interpolation routine takes over, causing rapid convergence to f_s.

Next, the motional inductance is found from motional reactance measurements at $f_s \pm \Delta f$, using the relation (Appendix A).

$$L_1 = \frac{X_1(f_s+\Delta f)-X_1(f_s-\Delta f)}{8\pi\,\Delta f} \qquad (1)$$

For the equivalent circuit assumed, the value of L_1 from (1) is independent of Δf within the accuracy of the narrow-band approximation. In practice, limitations exist due to measurement noise*. Table 1 shows the mean and standard deviation for measured values of L_1 for a 4.192 MHz fundamental mode resonator, using Δf's from 2 Hz to 2 kHz. The table shows that over this range the measured value of L_1 changes by less than .1%.

Load Measurements

The problems associated with measuring f_L and R_L, the resonance frequency and equivalent resistance of a piezoelectric resonator in series with a load capacitor, C_L, by various methods have been discussed elsewhere[4]. For the reflection bridge system, we have implemented the offset method[4], in which no physical load capacitor is used. Instead, the test frequency is adjusted until the equivalent series capacitance of the crystal unit is $-C_L$, the negative of the desired load capacitance. The equivalent series resistance is then R_L and the frequency, f_L.

Using this method, the problems of realizing a known load capacitance and minimizing or accounting for stray capacitances[4] are avoided. The disadvantages arise from the fact that, typically, the reactance of the load capacitance is large with respect to both R_L and the 50 ohm bridge impedance. Consequently, the measurement resolution of the equivalent resistance is degraded. In addition, the crystal current at f_L is much lower than at f_s or f_R.

For example, consider a 10 MHz crystal and a 32 pF load capacitance, test data for which appear at Table II. At f_s, the crystal current is approximately**.

$$I = 45/(50 + 18) = .66 \text{ mA}$$

while at f_L the dominant impedance is $-X_{CL} = 497$ ohms, so that

$$I = 45/497 = .08 \text{ mA}$$

The load frequency and resistance can also be calculated from the measured equivalent circuit parameters, using relations given by Hafner[3]. While Table II shows good agreement between the two methods, direct measurement at the load frequency is the preferable one.

VHF/UHF Measurements

An important feature of the 4191A is its ability to make high resolution reflection coefficient measurements at frequencies up to 1000 MHz. Table III shows a series of measurements made on a 145.6 MHz fifth overtone filter crystal. Table IV shows a similar series performed on a 375 MHz SAW resonator. These examples illustrate the measurement resolution obtainable.

Table V shows measurements of successive harmonic overtones of a 121.5 MHz fifth overtone oscillator crystal, demonstrating the ability of the system to measure modes with very low figures of merit, Q/r. Similarly, Figure 5 and Table VI illustrate the measurement of unwanted modes.

Resolution and Accuracy

The 4191A measures Γ_x and Γ_y, the real and imaginary components of the reflection coefficient, Γ, with a resolution of $\pm$ 0.0001. Tables VII & VIII show the calculated effect of Γ_x and Γ_y errors of -0.0001 on measurements of f_s, R_1, L_1, C_o, and (Table VII) f_L and R_L. In these tables the errors are in the same units as the measured quantities; e.g., Hz, Ohms. In Table VII, Meas. f_L and R_L refer to the values obtained by offset measurements, while Calc. f_L and R_L refer to the values calculated from the (measured) equivalent circuit parameters.

The estimation of accuracy is more difficult. If error calculations are performed using the manufactures's values of typical Γ_x and Γ_y errors[5], one obtains R_L errors which are inconsistent with experimental results. An example is given in Table VII, where the calculated R_L error for "typical" Γ_x error is 23 ohms. Based on actual measurements, a more realistic value is less than 2 ohms. Clearly, the manufacturer's "typical" error values do not apply in all regions of the reflection coefficient plane. The "typical" accuracy of series resonance frequency and equivalent circuit parameter measurements is, however, excellent, as these and other calculations illustrate.

*Moreover, the range of validity of the equivalent circuit is limited by the presence of other modes as well as by possible non-linear effects[3]. Also, for SAW resonators an additional shunt loss element must be included.

**The H-P 4191A measurement port has an open-circuit voltage of 45 mV and an impedance of 50 ohms.

Conclusions

The system described permits precise measurements of bulk-wave and surface acoustic wave resonators over the frequency range from 1 to 1000 MHz. The measurements can be made traceable to national standards. Hence, the system should be given consideration as a standard for piezoelectric resonator measurement.

Acknowledgement

The author wishes to thank Dr. E. J. Staples, Rockwell Electronic Research Center for providing the SAW resonator used to obtain the data in Table IV.

APPENDIX A

Motional Inductance Measurement

For the resonator equivalent circuit of figure 4, the reactance of the motional branch is given by

$$X_1(\omega) = L_1(\omega^2 - \omega_s^2)/\omega \qquad \text{(A-1)}$$

Making the usual narrowband approximation

$$X_1(\omega) \simeq 2L_1(\omega - \omega_s) \qquad \text{(A-2)}$$

At $\omega_s \pm \Delta\omega$

$$X_1(\omega_s \pm \Delta\omega) = 2L_1 \Delta\omega$$

from which

$$L_1 = \frac{X_1(\omega_s + \Delta\omega) - X_1(\omega_s - \Delta\omega)}{4\Delta\omega}$$

or

$$L_1 = \frac{X_1(f_s+\Delta f) - X_1(f_s-\Delta f)}{8\pi\Delta f} \qquad \text{(A-3)}$$

Shunt Capacitance Measurement

The (frequency-dependent) equivalent series capacitance of the motional branch is

$$C(\omega) = \frac{-1}{\omega X_1(\omega)}$$

Substituting (A-1),

$$C(\omega) = \frac{-1}{L_1(\omega^2 - \omega_s^2)} \qquad \text{(A-4)}$$

Ignoring R_1, at ω the total equivalent shunt capacitance of the resonator is

$$C'(\omega) = C_o + C(\omega) \qquad \text{(A-5)}$$

To measure C_o, choose two frequencies, ω_a and ω_b, such that

$$C'(\omega_a) + C'(\omega_b) = 2\,C_o$$

Substituting (A-4) and (A-5) and solving,

$$\omega_b^2 = \omega_s^2 - \omega_a^2 \qquad \text{(A-6)}$$

If we select

$$\omega_a = 1.1\ \omega_s$$

Then

$$\omega_b^2 = .79\ \omega_s^2$$

$$b = .8888\ \omega_s$$

References

1. Ichino, T., et al., "Vector Impedance Analysis to 1000 MHz", Hewlett-Packard J., v. 31, no. 1, pp. 22-32; Jan., 1980.

2. Programmed Test Sources, Littleton, Mass.

3. Hafner, E., "The Piezoelectric Crystal Unit - Definitions & Methods of Measurement", Proc. IEEE, v. 57, no. 2, pp. 179-201: Feb., 1969.

4. Horton, W. H., et al, "Comparison of Methods for Measurement of Quartz Crystal Resonators with Load Capacitance", these Proceedings.

5. No Author, "Operating Manual, 4191A RF Impedance Analyzer", Yokagawa H-P, Ltd; Apr. 1980.

TABLE I

EFFECT OF ΔF ON MOTIONAL INDUCTANCE MEASUREMENT

(NO MEASURE = 5 F = 4.192 MHz)

ΔF (Hz)	$\bar{L}_1$ (mH)	σ_{L_1}
2	117.14	.06
20	117.11	.00
200	117.07	.02
2000	117.11	.09

Measured Motional Inductance vs. Frequency Increment.

TABLE II

LOAD CAPACITOR MEASUREMENT

	F_L (KHz)	R_L (OHMS)
MEASURED	10000.988	18.03
CALCULATED	10000.997	17.61

CL = 32 PF

F_S = 9999.038 KHz C_0 = 3.72 PF

R_1 = 14.14 OHMS L_1 = 18.03 mH

Q = 80.1 K

Calculated and Measured Load Frequency and Resistance.

TABLE III

FILTER CRYSTAL MEASUREMENT

MEAS NO	FREQ (KHz)	R_1 (OHMS)	L_1 (MH)	C_0 (PF)
1	145640.107	251.74	17.91	1.19
2	145640.106	251.67	17.91	1.19
3	145640.109	251.67	17.91	1.19

N = 5 M = 3.6

Equivalent Circuit Parameter Measurements, 145.6 MHz 5th OT Filter Crystal.

TABLE IV

SAW RESONATOR MEASUREMENT

MEAS NO.	FREQ. (KHz)	R_1 (OHMS)	C_1 (FF)	C_0 (PF)
1	375163.1	67.15	.8907	2.07
2	375163.2	67.14	.8907	2.07
3	375163.2	67.15	.8888	2.07
4	375163.2	67.13	.8946	2.07
5	375163.2	67.14	.8945	2.07

MEAN C_1 = .892 FF Q = 7.1 K

E [σ] = .003

Equivalent Circuit Parameter Measurements, 375 MHz SAW Resonator.

TABLE V

MEASUREMENT OF NTH OT USING HP-4191A 121.5 MHz, N=5 CRYSTAL

N	F_S (KHz)	R_1 (OHMS)	L_1 (mH)	Q	Q/r
5	121499.01	30.5	2.75	68.8×10^3	6.9
7	170100.64	57.3	3.05	56.9	2.6
9	218701.05	96.1	2.92	41.7	1.1
11	267299.05	137.8	2.33	28.4	0.6
13	315896.75	143.6	1.51	20.8	0.5
15	364491.46	130.7	0.75	13.2	0.4
17	413083.66	90.9	0.24	6.8	0.4

Nth Overtone Measurements (N = 5 to 17), 121.5 MHz 5th OT Oscillator Crystal.

TABLE VI

UNWANTED MODE MEASUREMENTS

121.5 MHz 5TH OT

MODE REF. NO	F_S (KHz)	R_S (OHMS)	L_S (mH)	Q/r
0	121500.302	32.3	2.57	6.8
1	530.770	298	25.4	.74
2	536.63	869	35.7	.25
3	543.32	102	9.12	2.2
4	588.23	347	28.8	.63
5	606.97	317	13.3	.69
6	649.46	1305	34.3	.17
7	664.55	642	52.0	.34
8	685.51	454	38.0	.48
9	738.55	1170	55.4	.19
10	757.53	1131	64.8	.19

Unwanted Mode Measurements, 121.5 MHz 5th OT Oscillator Crystal.

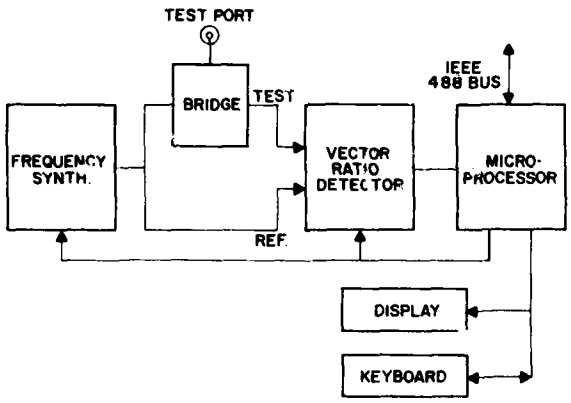
Table VII Calculated Measurement Errors (C_L = 32 pF.)

		Resolution Error		Mfgr's. "Typical" Error	
	Nominal Values	Γx-.0001	Γy-.0001	(1-.0018)Γx	(1-.0018)Γy
f_s (Hz)	4,192,000	< .001	< .001	< .001	< .001
R_1 (Ohms)	20.00	< .01	< .01	.04	< .01
C_o (pF.)	3.50	< .01	+ .03	< .01	- .01
L_1 (mH.)	120.00	- .02	< .01	.06	- .15
Meas. f_L (Hz)	4,192,709	.10	- .75	2.00	-1.14
R_L (Ohms)	24.61	1.41	.18	23.16	+ .26
Calc. f_L (Hz)	4,192,709	.09	- .76	- .49	-1.01
R_L (Ohms)	24.62	< .01	.05	+ .06	-0.01

Table VIII Calculated Measurement Errors

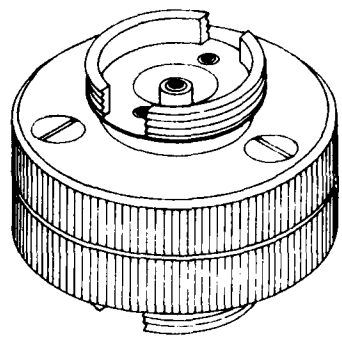
		Resolution Error		Mfgr's. "Typical" Error	
	Nominal Values	Γx-.0001	Γy-.0001	(1-.0021)Γx	(1-.0021)Γy
f_s (Hz)	145,600,000	< 1	< 1	< 1	< 1
R_1	250	-.09	+.01	-1.25	- .01
C_o (pF.)	1.20	<.01	<.01	< 1	< 1
L_1 (mH.)	18.00	-.01	<.01	- .16	- .03

FIGURE 1



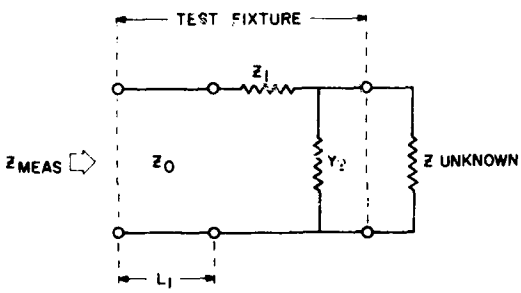
System Block Diagram

FIGURE 2



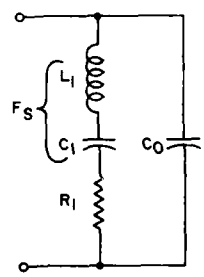
HC-18 Test Fixture

FIGURE 3



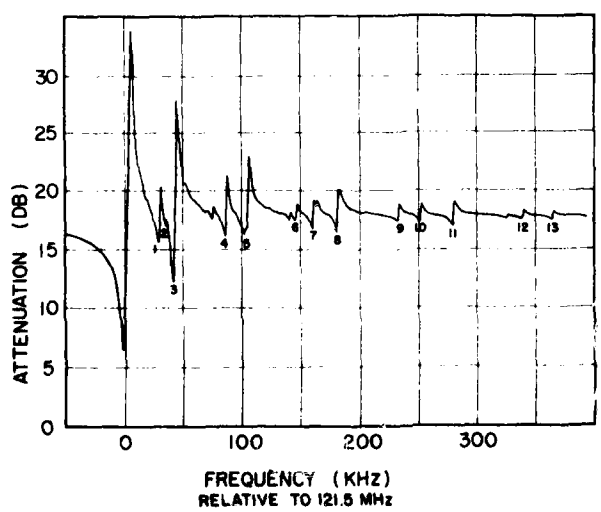
Typical Fixture Equivalent Circuit

FIGURE 4



Resonator Equivalent Circuit

FIGURE 5



Mode Plot, 121.5 MHz, 5th OT Oscillator Crystal
(12.5/12.5 Ohm Pi-Network).

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ, 07703. May 1981

IMPLEMENTATION OF AN AUTOMATIC MICROCIRCUIT MEASURING SYSTEM FOR QUARTZ CRYSTALS

GREGORY J. MALINOWSKI
US Army Electronics Technology & Devices Laboratory (ERADCOM)
Fort Monmouth, New Jersey 07703
and

GEORGE L. SNIDER & CHARLES NYHOLM
Hughes Aircraft Co.
500 Superior Avenue
Newport Beach, California 92663

ABSTRACT

The Automatic Bridge Measuring System (AMBS) is designed to serve as the reference test set for the measurement of electrical characteristics of quartz crystal units covering the frequency range of 0.8 MHz to 220 MHz. Quartz crystal parameters to be measured include parallel resonant frequency (f_p), static capacitance (C_o), motional capacitance (C_1), quality factor (Q), series resonant resistance (R_r) and parallel resonant resistance (R_p).

The system consists of an electronically tunable microcircuit admittance bridge (EMCB) which is designed to measure the resistive and reactive component of the quartz crystal impedance by adjusting, in a calibrated manner, two (2) electrically orthogonal internal elements until the output of the bridge is zero; a single crystal oven which controls the temperature surrounding the crystal to within .02°C in the range of 40°C to 95°C; a tracking servo bridge detector (TSBD) which incorporates the features necessary to permit automating the operation of the electronically tunable microcircuit bridge; an offset local oscillator which provides the RF signal for the tracking servo bridge detector and a bridge balance control unit which provides the interface between the electronically tunable bridge and the tracking servo bridge detector to effect semi-automatic or automatic balancing of the bridge.

The heart of the system is the electronically tunable microcircuit bridge head of the Schering type with the "unknown" in the parallel arm. The variable elements for the capacitive and resistive values are varactors whose capacitance versus voltage characteristic is described by the equation

$$c = \frac{k}{(V+a)^n}$$

where k, a and n are constants. The DC bridge control signals are supplied by the bridge balance control unit (BBCU). The bridge itself contains no active elements.

A set of calibration elements is provided with each EMCB for continuous traceability to NBS.

NBS is presently evaluating the AMBS with the intention of establishing a secondary standard for measuring passive components at any frequency in the range of 0.8 to 220 MHz. Test data demonstrating the accuracy, resolution and repeatability of the measurement system is presented.

Introduction

A comprehensive program for the development of a new generation quartz crystal measurement system is being pursued by the US Army Electronics Research and Development Command (ERADCOM). The goal of the program is a modern system capable of evaluating all crystal parameters automatically, including changes with temperature and time. The primary purpose of the system is to replace the family of CI meters currently used as the government reference test sets for crystal units.

Toward this goal, an electronically tunable microcircuit admittance bridge was designed and demonstrated at the 30th AFCS (1976)[1]. Included in the 1976 report was a description of a sensitive detector (tracking servo bridge detector - TSBD) for which a pilot production line was established at the General Radio Corp. under an Army Manufacturing Methods and Technology (MM&T) contract. The system at that stage of development was suitable for manual operation. However, further development work to optimize the bridge was considered desirable before mounting a production engineering effort.

An R&D contract was awarded to GenRad in June, 1976 for work on the electronically tunable bridge as well as on an interface unit between the bridge and the TSBD. The purpose of the interface unit was to effect automatic control of the voltages applied to the bridge for balancing it in the proper sequence, as required for the various measurements in the course of a crystal evaluation. The feasibility of the concept was further confirmed under the GenRad contract,[2] and the task has been continued at ERADCOM for the detailed implementation of the automatic measurement system.

The results of the work at ERADCOM were reported at the 32nd AFCS (1978)[3]. This report concluded that an automated test set for measuring quartz crystal units could be established as a reference standard. The system would be based on the use of an electronically tunable microcircuit admittance bridge.

To establish a production source for the system and to establish traceability of the system calibration to NBS, an MM&T contract on a single crystal measuring system was awarded to Hughes Aircraft Co., Newport Beach (HAC/NB), in October 1979.

A second (Phase II) contract has been awarded to HAC/NB for the production of a multicrystal system for automatic frequency temperature measurement of up to 200 crystal units per load. Work on this phase is currently underway.

The Automatic Microcircuit Bridge System (AMBS)

The Automatic Microcircuit Bridge System was developed to meet the following specifications:

Frequency Range: 0.8 to 220 MHz

Resistance Range: 2 to 20,000 ohms

Capacitance Range: 40 pf

Resistance Accuracy: $\pm$ 4%

Capacitance Accuracy: $\pm$ 2%

Drive Level Range (across unknown): 1 mv to .2Vrms

Control Voltage Range: 0 to 30 VDC X&Y

Input, Output Impedances: 50 ohms

Load Capacitors: 20, 30 and 100 pF $\pm$ 1%

Temperature Range: -60°C to +105°C

Frequency Resolution: better than 1 X 10^{-8}

The complete system is capable of automatically measuring the following crystal parameters, at any temperature, in the above stated range:

F_L Load Capacitance

F_r Series Resonant Frequency

Fp Parallel Resonant Frequency

R_r Resistance at Series Resonance

C_o Static Capacitance

C_1 Motional Capacitance

Q Quality Factor

In addition, the system permits viewing and measurement of spurious modes and harmonics.

With the inclusion of the temperature chambers and transport mechanisms now being developed, the system is intended to perform frequency vs. temperature measurements on 200 crystals at a time over the military temperature range.

A block diagram of the AMBS is shown in Figure 1.

The basic elements of the system consist of the electronically tunable microcircuit bridge (EMCB), tracking servo bridge detector (TSBD) with its offset local oscillator (OSLO), a single crystal oven (SCO), an oscilloscope, a frequency counter including an RF switch, the bridge balance control unit, (BBCU) and a calculator/controller. The calculator/controller, frequency counter, frequency synthesizer and the oscilloscope are off-the-shelf items. An RF Switch has been added to the frequency counter to permit the direct or beat frequency of the synthesizer to be applied to the counter. The remaining elements; the microcircuit bridge, the tracking servo bridge detector, the offset local oscillator, the single crystal oven, and the bridge balance control unit were designed and fabricated especially for the subject system.

The Bridge Balance Control Unit (BBCU)

The design concepts used in the BBCU are described in reference 3. The BBCU acts as the interface and control element between the calculate/controller and the crystal measurement system. It has the following functions in the measurement system:

(a) Interface to the system calculator via the IEEE-488 Bus

(b) Control of all of the remote TSBD Functions

(c) Generation and control of the bridge X&Y Voltages

(d) Control and measurement of the SCO temperature

(e) Selection of the RF input to the frequency counter

The BBCU is composed of functional elements on printed circuit board which are the:

(a) BCU Controller (MC6800 CPU)

(b) TSBD Interface

(c) Bridge Control (X,Y Voltage)

(d) SCO and RF Switch

In addition, the BBCU also contains its own power supply and front display panel.

The Electronically Tunable Microcircuit Admittance Bridge

The entire bridge is contained in a rectangular brass block, the physical size of which is approximately 1.25 x 1.75 x 7.25 cm. The solid coax lines connecting the bridge to the TSBD are feedlines only and not part of the measuring circuit. The bridge is self contained and can be used remotely from the TSBD, such as in a single crystal oven or a multiple crystal oven.

The configuration is that of a traditional Schering Bridge with the additional elements required for biasing the varactors which serve as the variable bridge elements. All resistors are part of a thick film substrate. The fixed capacitors are all high quality porcelain chips.

The bridge is designed for measurement of crystal units with resistive values greater than two ohms. The capacitance measurement range is approximately 40 pF.

All major bridge elements, with the exception of the varactor bias resistors, bypass capacitors, and

transformers are mounted on a 21 x 30 mm alumina substrate. The substrate with the components mounted on it is attached to a gold plated insert which is then fitted into an intricately worked cavity in the bridge block. After the insert is soldered into place, each component is essentially housed in its own individual cavity for maximum shielding. Connections to the transformers are made by means of 0.5 mm O.D. semirigid coaxial cables. The terminal block for the Unit Under Test (UUT) is mounted on the exposed side of the bridge substrate and connected to the bridge network by a short post extending through a hole in the substrate. Load capacitor chips of 20pF, 30pF and 100pF are mounted alongside the terminal block.

Mathematical modeling of the circuit has been performed at ERADCOM and the results are reported in reference 3.

The Tracking Servo Bridge Detector (TSBD)

The TSBD was designed to be used in conjunction with RF bridges and frequency synthesizers to measure quartz crystals with high reproducibility and accuracy. The unit serves as a very sensitive detector for the bridge unbalance signal, permitting measurements at power levels ranging from several milliwatts to well below a nanowatt. High signal detectivity (-155 dBm) is achieved by super-heterodyne conversion, low noise IF amplification and synchronous detection. The latter attribute not only provides independent reactive and resistive unbalance detection, but also permits the use of a control loop to lock the synthesizer frequency to the crystal under test, automatically maintaining reactive bridge balance. This key feature makes practical the widespread use of bridge measurements of crystals.

Other features include X10 frequency multiplier (45-220 MHz) to extend synthesizers of limited frequency range, a broadband power amplifier with attenuator (+30 to -70dBm) for control of measurement power level, linear and log detector response modes, accurate adjustment of synchronous detector phase, a wide band of swept frequency displays for initial set up and inspection of spurious modes, and a versatile servo amplifier for locked operation and automatic lock acquisition. The frequency lock servo system not only permits fast, automatic determination of a certain characteristic frequency of a crystal resonator (series resonance, for example) but also can follow the impedance circle to determine crystal inductance, or track frequency variations versus temperature.

The frequency range of the instrument is 0.8 to 220 MHz. The instrument, together with its companion offset local oscillator unit, RF bridge, frequency control synthesizer, and oscilloscope, provides all needed functions for crystal measurement.

The Offset Local Oscillator (OSLO)

The OSLO produces a signal offset of +80KHz from the synthesizer input for use as the local oscillator in the receiver section of the TSBD. The receiver input mixer then produces a constant 80KHz I.F. as the synthesizer varies over the specified frequency interval.

It is important that the OLSO signal be free of spurious frequency components which could result in any 80KHz I.F. signal in the absence of receiver R.F. input. Any such spurious response will limit the ability of the receiver to detect and lock to the correct bridge balance condition. The ideal situation exists when the receiver detectivity is limited by noise. Towards this end, a new design has been developed which has advantages over the original OSLO design. The major improvement is the operation of the internal frequency at the single sideband generator in a higher range than input or output frequency. This arrangement reduces the interference of the harmonic frequencies to the output signal.

The input and output levels are dictated by those specified for the tracking servo bridge detector; zero dBm generator input and +10 dBm OSLO output. The OSLO unit obtains power, control, and alarm signals from the TSBD.

The approach taken in this OSLO design, based on the results of a study described in reference 4, first generates a crude offset signal by phasing type single sideband techniques and then uses a phase locked loop as a filter to achieve the required spectral purity.

Generation of the SSB offset reference signal is done at double-frequency to reduce the harmonic generator component on the OSLO output.

The Single Crystal Oven

The single crystal oven is designed to maintain the crystal temperature to an accuracy of $\pm$ 0.02°C over the temperature range of +35°C to +95°C. A maximum of 15 minutes is required for stabilization to within $\pm$ .02°C of the temperature set point.

The mechanical package consists of a round, 11 cm diameter, clam shell fixture fabricated of aluminum. The EMCB and the temperature control electronics, surrounded by foam insulation, are contained within the shell. The temperature control electronics are proportional, consisting of a temperature monitor and an adjustment bridge which drives an operational amplifier controlling a power transistor. The collector resistor of the power transistor is the heating element. Feedback from the transistor collector to the operational amplifier provides the proportional control and enables the circuit to closely control the oven temperature.

Calibration of the AMBS System

Calibration of the AMBS system is performed using a group of resistive and capacitive elements calibrated at the National Bureau of Standards, Boulder, Colorado. The element set consists of seven discrete values of resistance and four of capacitance, covering the specified AMBS resistance range of 2 ohms to 20,000 ohms and the capacitance range of 40 picofarads. The resistive and reactive components of each element are measured and recorded by NBS over the system frequency range of 0.8 MHz to 220 MHz. This data will accompany each set of calibration elements. These elements will be returned to NBS periodically for recalibration. (The period has not yet been established.)

The resistive elements are fabricated using thin film techniques in seven discrete values including: 3, 5, 10, 20, 50, and 400 ohms (which use the same geometry), and 10,000 ohms which requires a different geometry. To promote ease of fixturing and handling, gold plated end caps were added to the chips so that the element when completed is a rectangular block approximately 5 x 1.5 x 1.6 mm.

The capacitive elements are fabricated using high quality porcelain chips in four discrete values 5, 10, 20, and 40 picofarads. Again, to promote ease of fixturing and handling, gold plated end caps are added so that the element, when completed, is a rectangular

block approximately 4 x 1.5 x 1.6 mm.

The elements are calibrated at NBS at room temperature at the following discrete frequencies: 0.8, 1.0, 2.0, 5.0, 10.0, 20.0, 50.0, 100.0, 150.0, 200.0, and 220.0 MHz.

The AMBS system calibration will consist of measuring and recording the system bridge values for each resistive and capacitive calibration element at the eleven frequency test points listed in the preceding paragraph. The system is required to maintain a measurement accuracy within ± 4% of the NBS calibrated value for the resistive elements and ± 2% for the capacitive elements.

System Performance

The AMBS performance was checked using a set of calibration elements that had been measured by the National Bureau of Standard.

The plots of frequency vs resistance and frequency vs capacitance are shown in Figures 2-9. Four resistor values, 5, 10, 20 and 50 ohms, and capacitors of 5, 10, 20 and 40 pF, were employed. The NBS data is plotted with an asterisk symbol, and measured results with dots. The frequency range over which the elements were tested was 0.8 MHz to 220 MHz, and the tests were performed at room temperature. The vertical scale represents 1% variations in resistance per tic and 2% variation per tic for capacitance.

The results of these tests indicate that the AMBS performance is within the specified accuracy ranges of ± 4% for resistance, and ± 2% variation per tic for capacitance.

Conclusion

A system to perform precision measurements on quartz crystals has been described. The results of tests, reported here and in previous symposia, have verified the accuracy of the system.

It has been demonstrated that the results obtained with this system can be traced to measurements made by the National Bureau of Standards.

Acknowledgements

The dedicated efforts of Dr. E. Hafner (ERADCOM) are truly appreciated. Without his foresight and continued perserverance, the development of this system would not have been possible.

References

1. E. Hafner and W.J. Riley, "Implementation of Bridge Measurement Techniques for Quartz Crystal Parameters", Proc. 30th AFCS May 1976, pp. 92-102

2. W.F. Byers, R.J. Hanson and W.J. Riley; "Automatic Microcircuit Bridge and Auto Balancing Unit", Final Report ERADCOM Contract NR. DAAB07-76-C-1380, June 1978 ADA075921

3. G. Malinowski and E. Hafner; "Automatic Microcircuit Bridge for Measurements on Quartz Crystal Units". Proc. 32nd AFCS, May 1978, pp 354-364

4. W.J. Riley, W.F. Byers and R.A. Soderman; "Automatic Microcircuit Bridge", Final Report ERADCOM Contract NR. DAAB07-75-C-1341, Aug 1976, ECOM-75-1341-F, ADA047929

5. W.J. Riley and W.T. Wadell; "Tracking Servo Bridge Detector, Final Report, Contract NR. DAAB-07-73-C-0609 , ADA055083, ADA055084, ADA055085

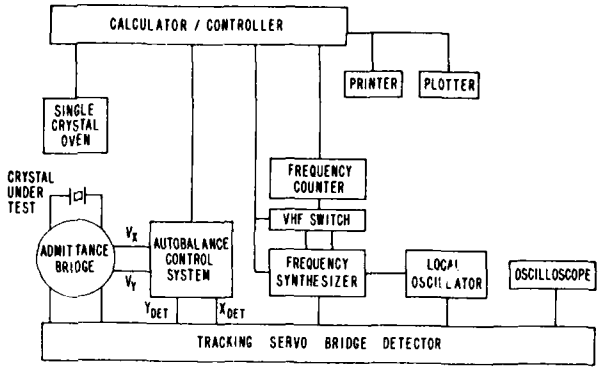


Fig. 1 AUTOMATIC MICROCIRCUIT BRIDGE TEST SYSTEM (SINGLE CRYSTAL)

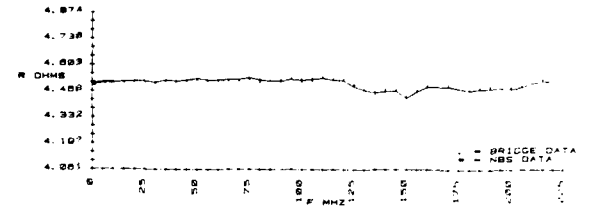


Fig. 2: Frequency vs. 5 ohm Resistor

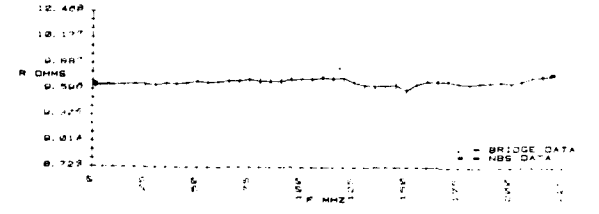


Fig. 3: Frequency vs. 10 ohm Resistor

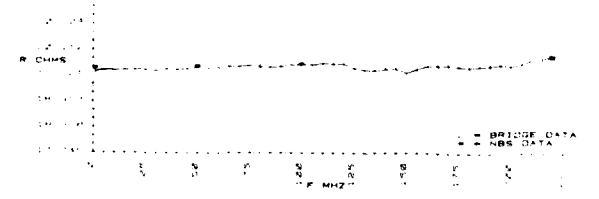


Fig. 4: Frequency vs. 20 ohm Resistor

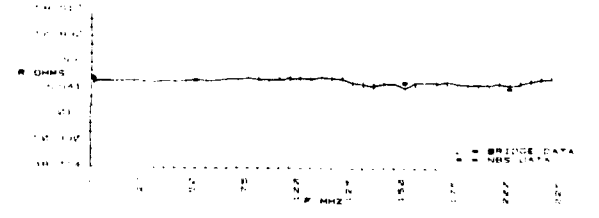


Fig. 5: Frequency vs. 50 ohm Resistor

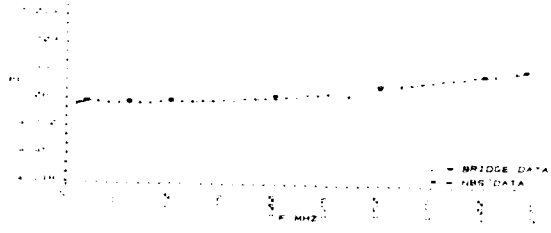


Fig. 6: Frequency vs. 5 pF Capacitor

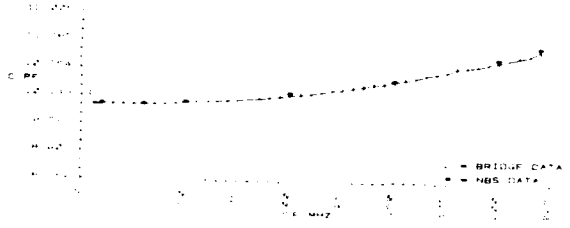


Fig. 7: Frequency vs. 10 pF Capacitor

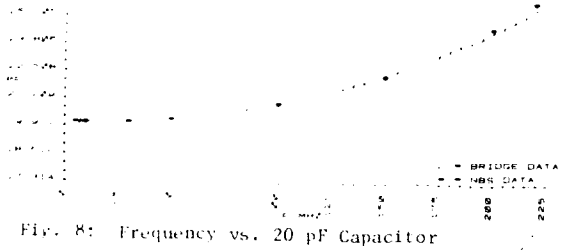


Fig. 8: Frequency vs. 20 pF Capacitor



Fig. 9: Frequency vs. 50 pF Capacitor

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

RECENT PROGRESS ON ALUMINUM PHOSPHATE CRYSTAL GROWTH

E. D. Kolb and R. A. Laudise

Bell Laboratories
Murray Hill, New Jersey

Summary

Progress on $AlPO_4$ is summarized in four areas: (a) the discovery of a new hydrothermal solvent for $AlPO_4$, HCl, and a report of solubility and growth relations in this solvent. (b) the measurement for the first time of p-V-T relationships in H_3PO_4 saturated with $AlPO_4$ and their use to rationalize solubility data in H_3PO_4. (c) the use of p-V-T data for H_3PO_4 and HCl to choose high fill - high growth rate conditions for the growth of $AlPO_4$ single crystals. (d) coupling constant and dielectric constant measurements for $AlPO_4$ which confirm its device potential and demonstrate a novel loading technique which only requires small samples for piezoelectric measurements.

Introduction

This paper summarizes in a form useful to device and device materials people recent work on four separate studies on $AlPO_4$. Those interested in experimental and other details on the physical, chemical, crystal growth, or solid state aspects of these studies may wish to consult the appropriate more detailed papers. The subjects discussed are HCl as a solvent (1), p-V-T relations in H_3PO_4 (2), high growth rates in H_3PO_4 and HCl (3) and coupling and dielectric (4) constants.

HCl - A New Solvent for $AlPO_4$ (1)

In a search for a new hydrothermal crystal growth solvent for $AlPO_4$ a variety of solvents were investigated (Table 1). Of these possible solvents only HCl appeared promising. Systematic solubility studies in HCl, Figs. 1,2,3, showed the dependence on temperature to be retrograde but with a considerably higher solubility for a comparable mineralizer concentration when compared to H_3PO_4 (5), previously the only known solvent for $AlPO_4$ growth.

Preliminary growth experiments were conducted at the following conditions

Conc. HCl - 1.525 m and 3.05 m.
Seed region temperature at beginning of run - 150-152°C
Growth region temperature at beginning of run - 151-154°C
Rate of temperature increase - 2-20°/da
% fill - 82

Seeds were suspended in the lower (hotter) part of the autoclave and excess solute nutrient in a basket in the upper (cooler) part of the autoclave. The method of growth and nutrient preparation was that which we first used in growing $AlPO_4$ in H_3PO_4 (6).

Growth rates increased when the heating rate was increased up to 10°/da but no substantial improvement occurred at 15°/da in 1.525 m HCl while in 3.05 m HCl a rate of temperature increase of 20°/da proved practicable. There was no evidence of homogeneous or wall nucleation of $AlPO_4$ even at the highest heating rates in 1.525 m HCl while in 3.05 m, the spurious nucleation was not excessive. Under comparable conditions of heating rate in H_3PO_4, $AlPO_4$ spontaneous nucleation regularly occurs on seeds resulting in poor quality growth. Growth rates are given in Table 2. Rates in 3.05 m HCl are higher than for comparable conditions in H_3PO_4. The quality as judged by visual and microscopic inspection in an immersion oil seemed approximately comparable to H_3PO_4 grown crystals and the growth rates in 3.05 m begin to approach those for quartz. *Therefore, the most promising solvents for $AlPO_4$ are H_3PO_4 and the new solvent, HCl.*

p-V-T Relations in H_3PO_4 (2)

One of the principal problems in systematizing $AlPO_4$ crystal growth rate and quality relationships in H_3PO_4 is a lack of good solubility data at the *constant fill* conditions which are used for growth (6). Good solubility data is available at *constant pressure* (5) but p-V-T data are needed to convert it to constant fill. We have now measured the equation of state of H_3PO_4 saturated with $AlPO_4$ over the pressure-volume(fill)-temperature - H_3PO_4 concentration regions of interest as shown in Figs. 4,5,6. As can be seen, the slope of p-T curves for 7.58 m H_3PO_4 at constant percent fill saturated with $AlPO_4$, $\left(\frac{\delta p}{\delta T}\right)_{\%f}$ (equivalent to constant density and constant specific volume), is independent of temperature at temperatures above the coexistence curve. Thus, data may be extrapolated to higher temperatures and pressures without significant error. As shown elsewhere (2), pressures in H_3PO_4 are substantially depressed from pure water and pressures in the system $H_2O-H_3PO_4-AlPO_4$ are slightly depressed from H_3PO_4. The pressure data are used to convert $AlPO_4$ solubility data at constant pressure to data at constant percent fill. Fig. 7 shows that in general the temperature coefficient of solubility at constant percent fill $\left(\frac{\delta s}{\delta T}\right)_{\%f}$ is negative.

The temperature at which the autoclave fills with one phase (Fig. 5) can be used to avoid two fluid phase regions where growth is slow because of poor convection and where quality is poor because of bubble inclusions.

High Growth Rates in H_3PO_4 and HCl (3)

Our pVT data for H_3PO_4 and HCl pVT data from the literature (7), can be used to choose conditions where only one fluid phase is present. For continuous growth where supersaturation is caused by a temperature differential, seeds are in the lower region of the autoclave (hotter) and nutrient in an upper region in a basket using the procedures which we have first reported previously (6). Table 3 gives typical conditions and results. As can be seen, growth rates in excess of 50 mil/da can be obtained, under conditions of *no* spontaneous nucleation. These results are quite promising for further study.

Coupling Constants and Dielectric Constants (4)

A novel loading technique (4,8) and dielectric constant methods (4) were used to measure the properties of $AlPO_4$. Tables 4 and 5 give typical results. As a further check of our procedures the same dielectric constant method was used to measure the dielectric constants of vitreous SiO_2 and crystalline quartz and it gave good agreement with literature values. Similarly, we used our loading procedures to measure the coupling contact of crystalline quartz and obtained reasonable agreement with literature values.

As can be seen, our values of coupling constant and dielectric constant for $AlPO_4$ agree with literature values and confirm the attractiveness of $AlPO_4$ for device applications. The loading technique we have used for piezoelectric constant measurements is attractive because it permits measurements on samples as small as ~2x2x2 1/2 mm. The alternative resonance - antiresonance method typically requires plates ~1cm in diameter. Thus our method permits evaluation of small specimens before cyrstal growth is scaled up and can be used to map homogeneity. Finally, we should point out that our dielectric constant measurements are the only experimental confirmation of an early literature value (9) and permit the use with confidence of a dielectric constant of 6 in calculations of the properties of $AlPO_4$ devices.

Conclusions

On the basis of a comparatively short experimental evaluation, a new solvent, HCl, gives quality at best equivalent to and possibly superior to H_3PO_4, the previous solvent which has been evaluated for 20 years. Thus, further tuning of the conditions used with this solvent is warranted.

Systematic p-V-T measurements have provided background for interpreting solubility results, can be used for the choice of both low pressure and high pressure growth conditions and have pointed the way to new growth conditions.

New high pressure growth conditions in the one-fluid phase region give the highest rates without spontaneous nucleation ever seen for both H_3PO_4 and HCl.

A novel loading technique suitable for small sample specimen, piezoelectric constant evaluation, and new dielectric constant measurements confirm the promising physical properties of $AlPO_4$ for devices and show that present growth conditions can produce material suitable for further study.

Acknowledgments

We would like to acknowledge A. J. Caporaso's experimental assistance and A. M. Glass, E. Johnson and R. L. Rosenberg for discussions. B. H. T. Chai, Corporate Research Center, Allied Chemical Corp., Morristown, NJ kindly provided several specimens for dielectric constant measurements as did W. R. McBride, Naval Weapons Center, China Lake, California for loading experiments.

References

(1) E. D. Kolb, J. C. Grenier and R. A. Laudise, J. Cryst. Gr. *51* 178 (1981).

(2) E. D. Kolb and R. A. Laudise, J. Cryst. Gr. (in press).

(3) E. D. Kolb and R. A. Laudise, (to be published).

(4) E. D. Kolb, A. M. Glass and R. A. Laudise, (to be published).

(5) E. D. Kolb, R. L. Barns, J. C. Grenier and R. A. Laudise, J. Cryst. Gr. *50* 404 (1980).

(6) E. D. Kolb and R. A. Laudise, J Cryst. Gr. *43* 313 (1978).

(7) R. W. Bach, H. A. Friedrichs and H. Rau, High Temperature - High Pressure *9* 305 (1977).

(8) S. C. Abrahams and J. L. Bernstein, J. Appl. Cryst. *12* 425 (1979).

(9) W. P. Mason, Piezoelectric Crystals and Their Applications to Ultrasonics, Van Nostrand, New York, 1950, p. 84.

Table 1 - $AlPO_4$ Results in Various Solvents

Run No.	Solvent	Temperature (°C)	Pressure (kpsi)	Time (days)	Observations
1	3.05m NaOH	170	10	3	Residue not $AlPO_4$
2	4.62m Na_3PO_4	300	10	3	Residue not $AlPO_4$
3	3.0m Na_2HPO_4+ 4.5m H_3PO_4	300	10	5	~1% solubility
4	3.05m CH_3COOH	170	10	3	Negligible solubility
5	3.05m NaCl	170	10	3	~4% solubility
6	3.05m NaCl	300	10	4	Negligible solubility
7	3.05m NH_4HF_2	300	10	5	Residue not $AlPO_4$
8	3.05m NH_4HF_2	170	10	5	Residue not $AlPO_4$
9	3.05m HNO_3	170	10	5	Negligible solubility

Table 2 - $AlPO_4$ Growth in HCl

	1.525 m HCl *(10°/day - 14 days)*	3.05 m HCl *(20°/day - 8 days)*
basal seed (0001)	4.3 mil/da	15.4 mil/da
Y cut seed $(10\bar{1}0)$	0.3	1.7
X cut seed $(11\bar{2}0)$	2.6	10.8

Table 3 - High Fill Conditions Based on p-V-T Measurements

Solvent	*% Fill*	*Crystallization Temp. (Bottom of Autoclave)*	*ΔT*	*Rate (0001)*	*Time of Expt.*
7.58m H_3PO_4	95	185°C	15°C	43 mil/da	5 days
6.0m HCl	92	190°C	20°C	50 mil/da	5 days

Table 4 - $AlPO_4$ Dielectric Constants

	~5mHz	1mHz	100kHz	10kHz	1kHz
ξ_{11}^T (field parallel to x)	____	5.919 ± .02	6.025 ± .02	6.022 ± .02	6.024 ± .02
ξ_{11}^T Mason (9) (Resonance-antireson.)	6.05	____	____	____	____

Table 5 - Piezoelectric Coupling Constants

$AlPO_4$

d_{11}	Measured Directly	Meas. from d_{31}^1 and d_{32}^1 *	Meas. from d_{12}*
Sample			
2401-3a-1a	$3.84\times10^{-12}CN^{-1}$		$3.62\times10^{-12}CN^{-1}$
2401-3a-1b	3.50		3.46
2401-3b-1		$3.22\times10^{-12}CN^{-1}$	
Avg. d_{11}	3.67	3.22	3.54
Grand avg. d_{11}	$3.52\pm.25$		
Mason (9) (resonance antiresonance)	3.33×10^{-12}		

*$d_{11}=-d_{31}^1-d_{32}^1$

$d_{11}=-d_{12}$

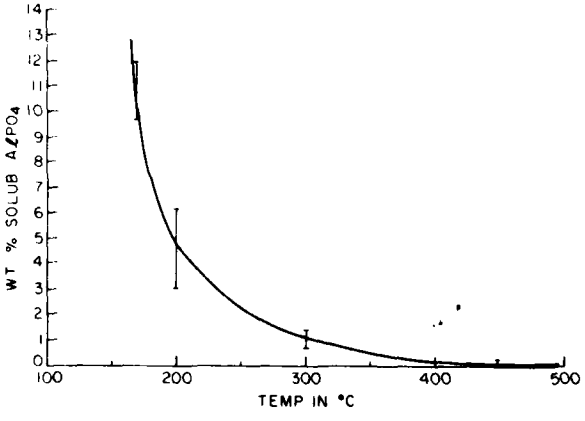


Fig. 1. Solubility of $AlPO_4$ in 3.03m HCl at 10 kpsi.

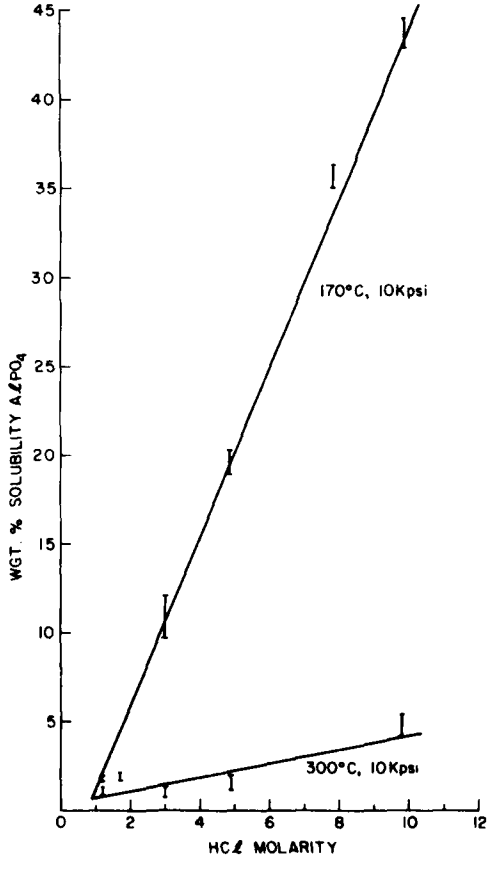


Fig. 2. Solubility of $AlPO_4$ in HCl at 10 kpsi.

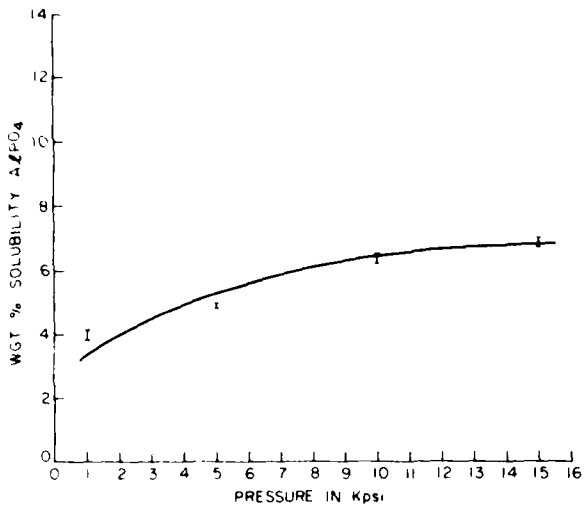


Fig. 3. Solubility of Al_{x} $_{4}$ in 3.03m HCl at 200°C.

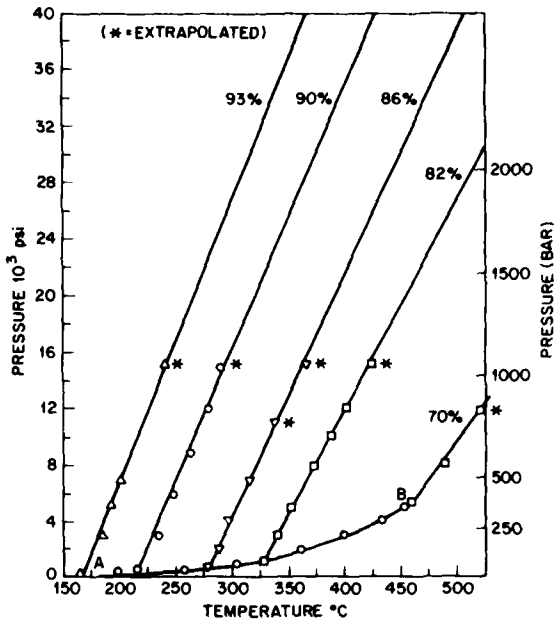


Fig. 4. Pressure vs. Temperature - 7.58m H_3PO_4–$AlPO_4$.

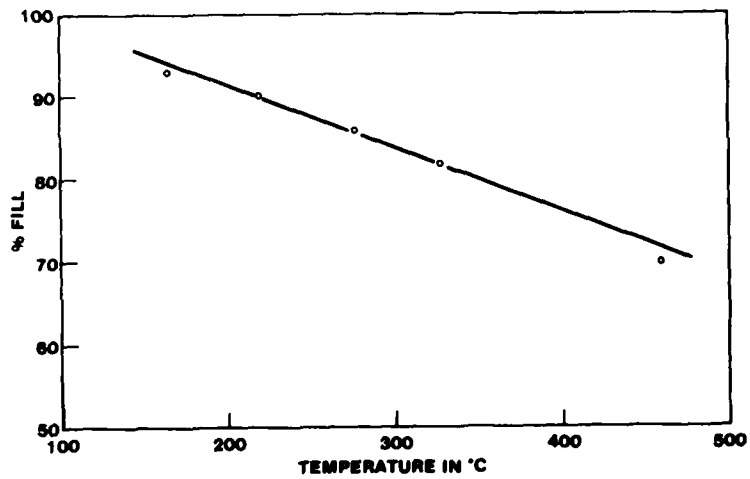


Fig. 5. Temperature where autoclave fills vs. initial % fill -7.58m H_3PO_4–$AlPO_4$.

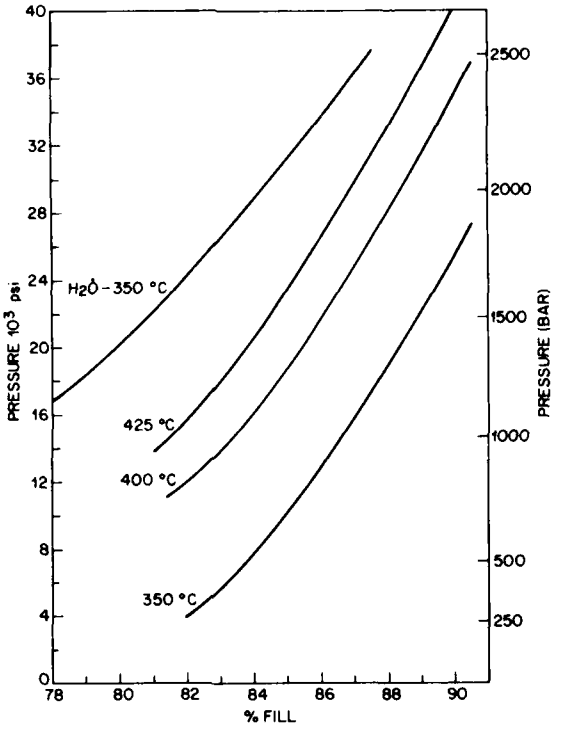


Fig. 6. Pressure vs. % fill - 7.58m H_3PO_4–$AlPO_4$.

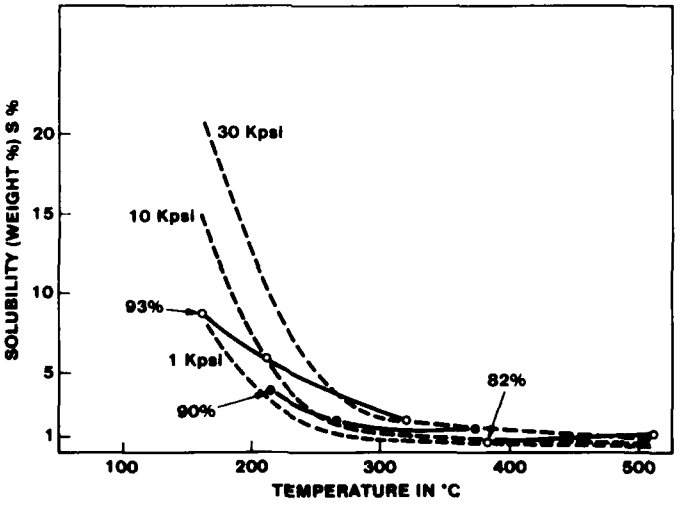


Fig. 7. Solubility of $AlPO_4$ in 7.58m H_3PO_4.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703 May 1981

INITIAL RESULTS WITH THE AIR FORCE HYDROTHERMAL FACILITY

A. F. Armington, J. J. Larkin, J. J. O'Connor, J. A. Horrigan

RADC/ESM, Hanscom AFB, Bedford, Mass.

Summary

This presentation describes the automated hydrothermal system constructed and operated by RADC/ESM. Initial results using this facility for quartz growth will be presented. Crystals containing less than one ppma aluminum have been produced. There is some indication that the growth mechanism at low rates is different than at higher growth rates. Some material has been swept and is being fabricated into resonators.

Introduction

A completely computerized hydrothermal facility, consisting of ten crystal growth autoclaves, has been constructed at RADC/ESM. It is essentially a research operation and has primarily been used to investigate growth conditions for the production of high quality, low drift, radiation tolerant quartz. Initial emphasis has been placed on improving the purity of quartz by modification of growth conditions and nutrients, by examining the effects of seed quality, and by using noble metal liners.

Facility

Some of the crystal growth autoclaves are shown in Figure 1. There are four autoclaves with interior dimensions 1¼" X 1 foot, four at 3" by three feet and two with dimensions 4" X 4 feet. Those autoclaves are equipped with modified Bridgman seals and can be operated at 400 degrees C and 30,000 psi.

Silver and platinum inserts are often used to isolate the growth from the steel autoclave chamber. These consist of tubular cans, prefilled and then welded. Some of these liners have been totally sealed, while in others a capillary opening has been left to prevent crushing due to differences in vapor pressure during heatup and cooldown.

Primary long term control, data keeping, and alarm processing is managed by a Hewlett Packard Series 1000 computer (Figure 2). This computer has a real time operating system and schedules a control, data, and alarm program based on its internal clock for each autoclave system (Figure 3). Basic run parameters for a run are input through an initialization program, and the entire course of the run, including alarm management and data file creation is then under computer control (Figure 4).

Each autoclave has four thermocouples, an embedded thermocouple and a skin thermocouple for each zone, as well as a pressure transducer. The computer accesses these through a Leeds and Northrup Trendscan 1000 scanner, and controls in a supervisory mode through Leeds and Northrup Centry setpoint controllers (Figure 5). All temperatures and the pressure value are stored periodically in disc files, and checked more frequently to insure that no alarm value has been exceeded. When the control program executes, the imbedded thermocouples are compared to values projected by the program at that time and the centry setpoints are adjusted to correspond. Archival storage of data and computer system backup is done on magnetic tape, and an X-Y plotter is incorporated into the system for rapid data examination.

In case of computer failure, a number of backup systems exist. The Trendscan is interfaced into an alarm processor which itself sets digital alarms on all four temperatures and pressure. The pressure transducer is fed into a readout which has an internal alarm entirely independant of all other components of the system. The effect of all electronic alarms is to turn off the power to the system. In addition there are two 30,000 psi rupture discs on each autoclave. Data can be printed out periodically with the Trendscan internal printer if the computer is not maintaining files. The Centry controllers maintain the last set values, and can be operated manually during an extended computer outage.

A 65 Kw generator backs up the system in case of power failure. The computer and the Trendscan have battery backup systems to protect against short duration loss of power, and will restart upon resumption of power. A telephone dialer is wired to the system and can be used to alert personnel of alarms and other abnormal conditions.

The overall control and reliability of the system is quite good. Figure 6 shows a projected run and Figure 7 actual graphical data from the run. Since the imbedded thermocouples are not actually in contact with the solution, overall control may be slightly different than shown.

Results

The initial work with this facility involves the preparation of high purity quartz in sizes sufficiently thick for resonator fabrication. When this phase of the work has been completed, attempts will be made to improve the seed quality.

In the purification portion of the work the primary emphasis has been the use of cultured starting material and inert liners. Ultrapure mineralizer and deionized, distilled water have been used, but it is not believed that this has a serious effect on the quartz purity. Generally the mineralizer has been 0.5 N sodium hydroxide with a small amount of lithium hydroxide added. In some cases 0.5 N sodium carbonate has been used as will be discussed later in this presentation. The per cent fill is in the region of 75 - 82%. The standard procedure has been the growth of crystals in the small, one-inch autoclaves and to take the optimum results to use in runs performed in three inch autoclaves. We are just starting the runs in four inch autoclaves where we can produce a sufficient number of crystals for extensive evaluation and resonator fabrication. Evaluation techniques include X-ray topography, atomic absorption, EPR, low temperature IR and some SEM.

The growth rate has been varied by altering the temperature difference between the upper and lower portions of the autoclave which are separated by a baffle. The design of this baffle, which has about a twenty per cent open area, does not significantly affect the growth rate. The top temperature has been held at 348 - 350 degrees centigrade with the bottom temperature being varied to change the gradient. It must be emphasized that the gradients shown in Figure 8 are specific for our autoclaves and are only relative values related to our thermocouple placement and heat losses inherent in our systems. It is estimated that the lower temperature (the nutrient area) is reading about 30 degrees too low because of its position in the autoclave. There is considerable scatter in the points but it is evident that the rate vs. gradient is quite close for the unlined and platinum lined runs but that the gradient must be higher for the silver liners because of the high thermal conductivity of that material.

The purity has been evaluated as a function of the type of liner used. Unlined runs are also compared (Figure 9). The iron and carbon results shown on the slide are not correct since these elements are the result of contamination during preparation for the analysis. Later analyses using a different preparation indicate the iron content to be in the one PPMA range in lined runs. The best results here, in terms of aluminum which is the chief consideration, indicate that the purest material is produced in silver liners with hydroxide mineralizer. The results for platinum liners are based on a single analysis and could be too high. The use of a silver liner with carbonate results in higher values for both silver and aluminum, with some indication of platinum contamination with the carbonate in a platinum liner. These results indicate that the silver liner is most effective. We are still performing runs in platinum as well as silver liners however, to get more data on platinum. Subsequent results also show the alkali metal impurities to be too high in the early results shown here.

Several runs were performed using high purity glass as well as the sand used for the preparation of the glass. Both of these materials are about as pure as the cultured quartz used as our standard starting material. All of these runs resulted in very slow growth and poor crystals which we believe was due to the much higher solubility of these silicas in the solution. This could result in a more viscous solution which could inhibit the transfer of the silica from the nutrient to the crystal. An attempt was made to reduce this effect by using 0.1 N sodium hydroxide in place of the 0.5 sodium hydroxide usually employed. This resulted in increased growth and a fairly well formed crystal. X-ray topographs of the crystal indicated however, that the strain pattern was different than normally grown crystals (Figure 10). Unlike the usual strain lines, which tend to propagate through the Z growth area at angles to the Z axis (Figure 11), the line in this sample all appear parallel to the Z axis. The crystal is more strained than usual also. It seems possible that the growth mechanism had been altered from a screw dislocation type to a stepwise type growth. SEM pictures of the grown Z face also were different. The bottom picture on Figure 12 is an SEM of a normal crystal while the top picture is that of the sand nutrient crystal. Again one could assume step growth as the mechanism. The possibility of producing a purer crystal by this mechanism was shattered by the purity analysis of the material (Figure 13). Both the EPR and atomic absorption showed a massive aluminum purity. The source of the contamination must be from the autoclave walls or the holder materials used for the seed. Perhaps the increased impurity contamination with slow growth rates reported in literature (Ref. 1) may be attributed to some step growth at low growth rates. SEM photos of the X surfaces (Figure 14) show a type of sheet formation also. Since this growth area is also higher in aluminum than the Z area it indicates that the screw dislocation mechanism is desirable for the growth of pure quartz.

The true test of impurity effects on a resonator or oscillator is the evaluation of the device itself. Until we have a thorough evaluation, no attempt will be made to further lower the impurity content. Ten megaherz resonators of this material are being fabricated.

The scaling up of the one inch autoclave runs to three inch was easily accomplished. Figure 15 shows a seven inch crystal grown at a fast rate with no liner. We have produced crystals in lined runs also which are being fabricated in resonators.

Since we have accomplished an improvement in purity and have demonstrated that we can produce large crystals, we are attempting to reduce the

dislocation density of the crystals, primarily by the use of good seeds or altering available seeds as shown in Figure 16. At present we are testing X growth seeds and seeds swept in air and vacuum. We hope to test the other type seeds in the next few months.

Reference

1. N.C. Lias, et al, J. Crystal Growth, 18, 1(1973).



Figure 1 Autoclave Facility

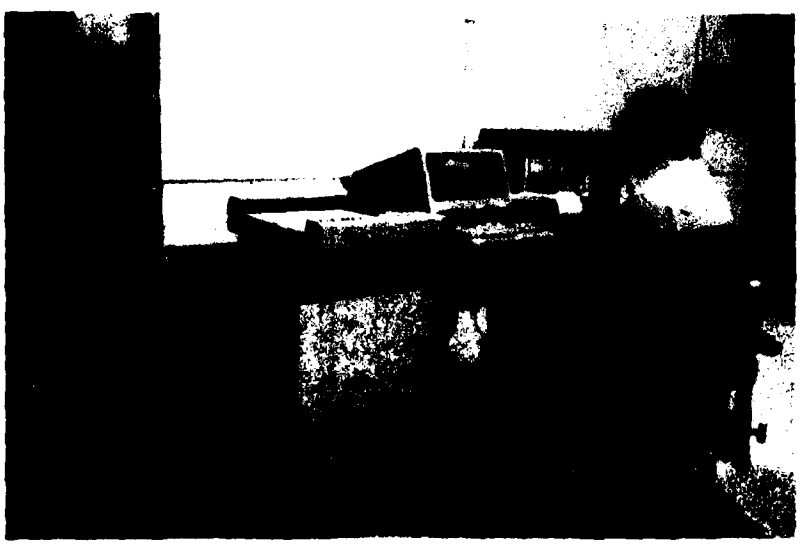


Figure 2 HP 1000 Computer used as Process Control

```
11:15:  4:200
*****************************************************************
PT ST PRGRM, T, PRIOR*DRMT*SCHD*I/O *WAIT*MEMY*DISC*OPER * NEXT TIME   *
*****************************************************************
 0 ** MEMX  *1*09999 ***** 1 ***********************************11:15:  4:220
 0 ** WHZAT*2*00010 ***** 1
 0 ** BELL  *3*00030   0 ***********************************11:15:56:040
 0 ** S1PRG*2*00050   0 ***********************************11:27:  0:  00
 0 ** S1FIL*2*00070   0 ***********************************11:20:13:990
 0 ** ALRM1*2*00060   0 ***********************************11:15:59:140
 0 ** S4PRG*2*00053   0 ***********************************11:19:  0:  00
 0 ** S4FIL*2*00073   0 ***********************************11:20:49:870
 0 ** ALRM4*2*00063   0 ***********************************11:16:40:630
 0 ** S6PRG*2*00055   0 ***********************************11:27:  0:  00
 0 ** S6FIL*2*00075   0 ***********************************11:29:20:890
 0 ** ALRM6*2*00065   0 ***********************************11:16:20:870
 0 ** S7PRG*2*00056   0 ***********************************11:24:  0:  00
 0 ** S7FIL*2*00076   0 ***********************************11:41:14:450
 0 ** ALRM7*2*00066   0 ***********************************11:16:41:  50
 0 ** TOALA*2*00011   0 ***********************************11:16:13:830
*****************************************************************
DOWN LU'S
*****************************************************************
DOWN EQT'S
*****************************************************************
11:15:  5:140
```

Figure 3 Operating Programs active during Run



Figure 5 L&N Transcan Supervisory Control

```
RU, S4UP
SYSTEM NUMBER?  1
AVERAGE ROOM TEMPERATURE?  25
SCHEDULING FREQUENCY OF CONTROL PROGRAM IN MINUTES?  15
SYSTEM TURNUP TIME IN HOURS?  72
INITIAL TURNUP TEMPERATURE?  375
ETCH DURATION IN MINUTES?  0
NEGATIVE GRADIENT FOR ETCH [DEGREES]?  0
RUN TIME IN DAYS?  45
FINAL BOTTOM TEMPERATURE?  380
INITIAL GRADIENT?  25
FINAL GRADIENT?  30
RUN DOWN TIME IN HOURS?  72
MAINTAIN GRADIENT DOWN [YES OR NO]?  NO
DEVIATION REQUIRING ADDITIONAL CONTROL?  20
SCHEDULING FREQUENCY OF DATA PROGRAM IN MINUTES?  15
ALARM VALUE FOR TOP CONTROL?  399
ALARM VALUE FOR TOP MEASUREMENT?  399
ALARM VALUE FOR BOTTOM CONTROL?  399
ALARM VALUE FOR BOTTOM MEASUREMENT?  399
PRESSURE ALARM IN PSI?  20000
SCHEDULING FREQUENCY OF ALARM PROGRAM IN MINUTES?  3
PROGRAM VERIFIED GRAPHICALLY [YES OR NO]?  YES
```

Figure 4 Input Program for Run Initiation

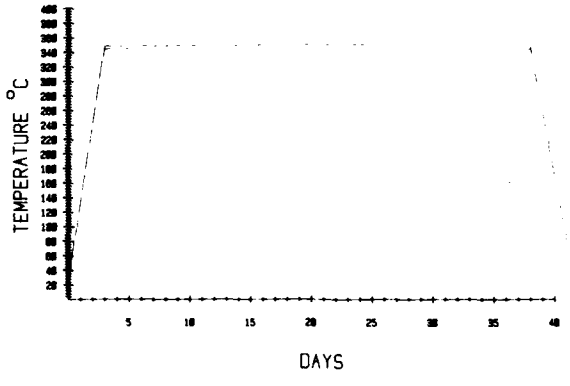


Figure 6 Run Projection Produced During Initiation

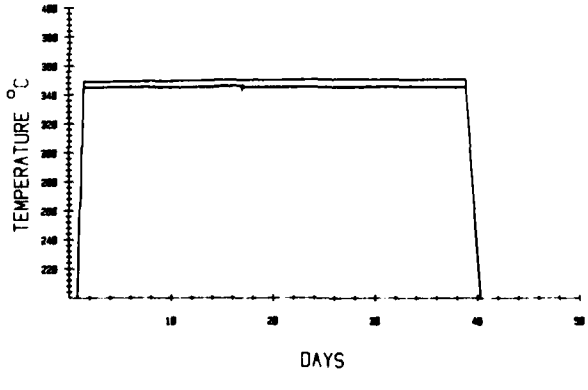


Figure 7 Actual Run Profile

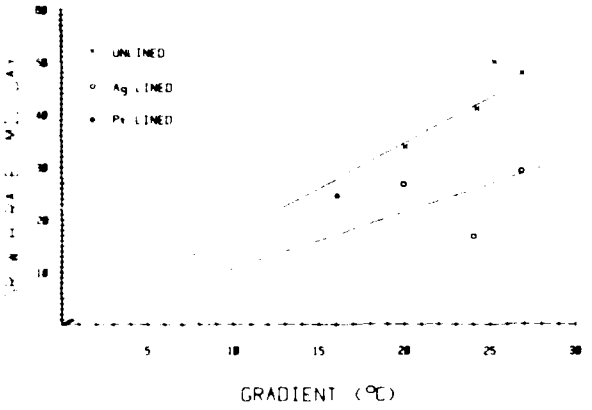


Figure 8
Graph of Temperature Gradient Against Growth Rate

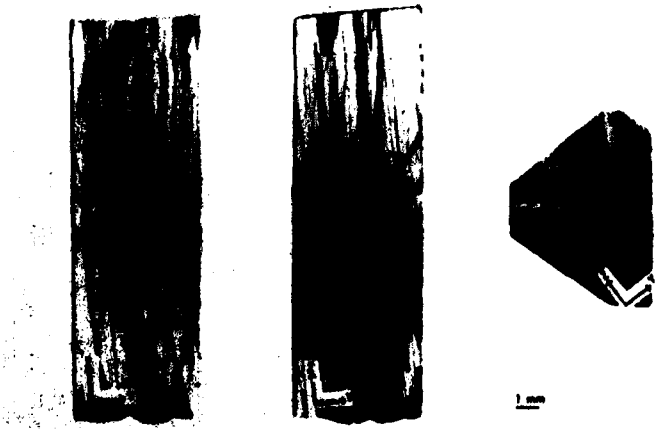
IMPURITIES IN QUARTZ GROWTH RUN

	SILVER CARBONATE	PLATINUM CARBONATE	PLATINUM HYDROXIDE	UNLINED : CARBONATE	UNLINED HYDROXIDE	SILVER HYDROXIDE
Al	18	22	8	10	10	0.8
Na	42	28	<1	30	27	30
Fe	561	280	700	87	200	
C	180	570	290	220	220	
Ag	18	-	-	10	-	<1
Pt	-	<10	<10	-	-	
K	-	2.7	8.1	<10	10	<1
Li	?	8	93	-	-	

Figure 9 Purity as a Function of Liner Material



Figure 10 Topograph of Crystal Grown from Sand



X-RAY TRANSMISSION TOPOGRAPHS OF QUARTZ CRYSTALS GROWN ON SEEDS OF DIFFERENT CRYSTALLOGRAPHIC ORIENTATION

Figure 11 Topograph of Normal Crystal



Figure 12

Scanning Electromicroscope View of Z Surfaces (Top-Sand Grown Crystal)

Figure 13

Aluminum Content of Crystal Grown from Sand

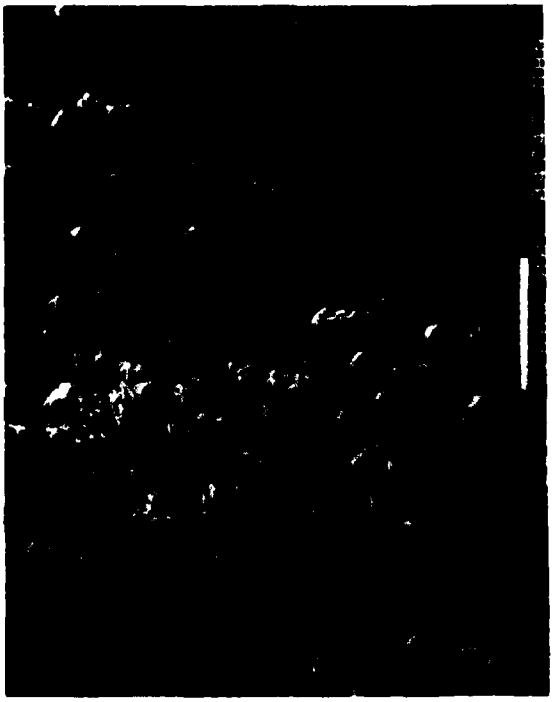


Figure 14 SEM photograph of X Surface of Crystal



Figure 15 Crystal Grown in Three Inch Autoclave

Possible Seed Variations

Use of X Growth Area

Sweeping - Air
- Vacuum

Good Synthetic Seeds

Natural Seeds

Figure 16 Possible Seed Variations to Reduce Dislocation Density

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703 May 1981

SYNTHETIC QUARTZ CRYSTALS GROWN IN NaCl, KCl SOLUTIONS AND PURE WATER, AND THEIR LOW TEMPERATURE INFRARED ABSORPTION

M. Hosaka and S. Taki
Institute of Inorganic Synthesis, Yamanashi University, Kofu, Yamanashi, Japan

and

K. Nagai and J. Asahara
Toyo Communication Equipment Co., Ltd., Kanagawa, Japan

ABSTRACT

Fairly large transparent crystals were grown hydrothermally in NaCl, KCl solutions and pure water. The effects of low temperature (~78°K) in infrared absorption were studied by illuminating the samples with polarized light with the electric vectors parallel (E//C) and perpendicular (E⊥C) to the C-axis and the effects of radiation were studied exposing the samples to the neutron flux in a 100KW nuclear reactor. We also performed nondestructive analysis of the Al and Na impurity contents by neutron activation and γ-ray spectrometry. Absorption bands due to the Al-OH^- defect centers were observed at 3364 and 3304 cm^{-1} in the samples of "as grown" crystals grown in both NaCl and KCl solution and pure water. The crystals grown in the NaCl solution did not exhibit the absorption bands at 3580, 3395, and 3345 cm^{-1} which are generally exhibited by crystals grown in an alkaline solution. In this respect, the profiles of absorption spectra of crystals grown in an NaCl solution are very similar to those of high quality natural quartz crystals.

INTRODUCTION

Alkaline solutions, such as NaOH or Na_2CO_3, are presently used as the mineralizer for growing commercially available synthetic quartz crystals for both electronic devices (e.g., high precision oscillators, filters, etc.) and optical purposes. However, the chemical composition of fluid inclusion in natural quartz shows that the mineralizer usually consists of Na^+, K^+, Cl^-, and HCO_3^- ions, etc. Therefore, natural quartz is grown hydrothermally in a solution containing such ions.

The 4000 to 3000 cm^{-1} infrared absorption and morphology of synthetic and natural quartz crystals show remarkable differences. These differences are caused by the different growing conditions such as the mineralizer (solution), pressure, temperature, and so forth.

Recently, Hosaka and Taki have grown crystals in NaCl, KCl solutions and pure water[1-3].

Meanwhile, many researchers have investigated the infrared absorption of quartz crystal in the 3μm region. Lipson et al.[7] has shown that, at low temperature infrared absorption, high grade quartz exhibits four principal band peaks and that additional bands (3305 and 3366 cm^{-1}) can be induced by subsequent electrodiffusion of the samples along the C-axis. Recently, Halliburton[9] and Sibley et al[8] systematically investigated the radiation effect on various quartz crystals by low temperature infrared absorption in connection with experiments on electron spin resonance (ESR) and proved that the absorption bands at 3307 and 3367 cm^{-1} induced by electrodiffusion and x-ray, γ-ray, and electron irradiation are caused by the forming of Al-OH^- centers. Halliburton et al.also showed that the defect configuration of substitutional aluminum can be represented by three models, e.g. Al-OH^-, Al-M^+ (M: Alkaline ions), and [Al e^+]° (hole trapped center), as charge compensated centers.

However, the bands at 3580, 3440, 3400, and 3348 cm^{-1} are due to OH vibrations that have not yet been identified.

Therefore, investigation of the point defects in quartz crystal by using samples grown in NaCl, KCl solution and pure water is very useful.

This paper reports on studies on the 3μm region spectra of transparent crystals grown in NaCl, KCl, and H_2O. To compare the differences in the infrared absorption, synthetic crystals were hydrothermally grown in NaCl, KCl, H_2O and in an alkaline solution (Na_2CO_3, K_2CO_3, Na_2CO_3-D_2O solution).

This report also describes nondestructive and reliable analytical methods for impurity aluminum and sodium. The quantitative impurity of aluminum below 1μg can be detected by exposing the crystals to thermal neutrons. The radiation effect of the above samples was studied by low temperature infrared absorption.

EXPERIMENTS

1. Method of growing crystals

Quartz crystals were hydrothermally grown in NaCl-SiO_2-H_2O, KCl-SiO_2-H_2O, K_2CO_3-SiO_2-H_2O, SiO_2-H_2O, Na_2CO_3-SiO_2-D_2O, and Na_2CO_3-SiO_2-H_2O systems in platinum or gold capsules placed in autoclaves of various sizes. The seeds used were X-cut ($11\bar{2}0$), R-cut ($10\bar{1}1$), r-cut ($01\bar{1}1$), AT-cut, and Y-bar. Three types of nutrients, transparent Brazilian natural quartz crystals, synthetic quartz crystals and high purity synthetic fused silica glass were used. In particular, the fused silica glass was used to grow purer crystals to prevent contamination from the nutrient. Table 1 shows the analytical data for synthetic fused silica. The fused silica was transformed to α-quartz by hydrothermal treatment before being used as a nutrient. The solutions employed to grow the quartz crystals were 10wt% NaCl and KCl solution or 1N Na_2CO_3 and K_2CO_3 solutions. In case of the Na_2CO_3-SiO_2-D_2O system, D_2O purity of 99.8wt% was used. The capsule was

filled with the desired solution to 70% of its free volume and placed in an autoclave. The growth temperature was varied from 350°C to 500°C. The temperature gradient between the growth and dissolving regions was chosen within 10°C to 50°C. The baffle was opened 30% in all the experiments. The temperatures were measured and controlled with chromel-alumel thermocouples attached to the outside of the autoclaves. The growth period for the alkaline solutions was from several days to two weeks. However, because of the slower growth rates, 2 to 3 months was required for the NaCl, KCl solutions and pure water.

Table 1

Impurity	Co	Cr	Fe	Na	Zn
Contents (ppb)*	<0.01	<0.05	<10.0	<0.35	<0.01

* Determined by neutron activation analysis.

2. Infrared absorption

Low temperature (near 78°K) infrared data were obtained by a metal Dewer with CaF_2 windows as the light path. All measurements were made after temperature equilibrium with liquid nitrogen. Polarized infrared light was observed with an AgCl polarizer having an efficiency of 95%. The polarizer was placed in front of the thermocouple detector. Polarized dichromism was detected by the electric vector (E) of the polarized light parallel (E//C) or perpendicular (E⊥C) to the C(Z)-axis. A Japan Spectroscopic Co., Ltd. Model DS-701G grating infrared spectrophotometer with 300 lines/mm grating from 4000 to 2300 cm^{-1} was used for all infrared absorption measurements. Purging of the spectrometer with very dry air was essential to reduce the noise pattern caused by the water vapor in the light path. Dry air having a dew point below -73°C at 1 atm was produced with a General Cable Corp. Pure-Gas Heatless Dryer. When 2.0 to 3.0 mm diameter metal windows were attached to some samples, the resolution of the spectrophotometer dropped to 4 cm^{-1} at wavenumber 3400 cm^{-1}. Generally, a 6.5 mm diameter metal window provided a resolution within 2.5 cm^{-1} at wavenumber 3400 cm^{-1}.

This infrared spectroscopy allowed calibration of the discrete absorption bands of the quartz crystal by wavenumber. Table 2 is a list of the bands proposed by some authors and the results of our calibration at 78°K spectra.

Table 2

Lipson et al.[7]	Halliburton et al.[9]	Brown et al.[5] (12°K)	Our calibration (78°K)	Remarks (Polarization)
3581 (S4)	3581	3583	3580	E⊥C
3438 (S3)	3437	3434	3435	E⊥C
3396 (S2)	3400	3399	3395	(E⊥C) = (E//C)
3348 (S1)	3348	3350	3345	E//C
3366 (e2)	3367	3367	3364	E⊥C
3305 (e1)	3306	3306	3304	E⊥C

3. Neutron irradiation and activation analysis

The impurity aluminum and sodium in quartz crystals were nondestructively analyzed by neutron activation and γ-ray spectrometry.

Neutron irradiation was performed with the 100kW nuclear reactor (TRIGA-Mark II) at the Atomic Research Center of Rikkyo University, Yokosuka. The samples were exposed to neutron flux in T.C. (Thermal Column) for impurity Al analysis and subsequently irradiated in R.S.R. (Rotary Specimen Rack) for Na analysis.

The flux of the thermal and fast neutrons in the T.C. and R.S.R. are as follows:

	Thermal Column	Rotary Specimen Rack
Flux of thermal neutron	3.0×10^{10}	$4 \sim 5 \times 10^{9}$
Flux of fast neutron	1×10^{8}	6.5×10^{10}
Cd ratio	~ 30	3.5
γ-ray dose	2.2×10^{4} r/hr	1.4×10^{6} r/hr

γ-ray spectrometry was performed with a Ge (Li) scintillator with multichannel pulse height analyzer (Canberra Model 8100).

(A) Analysis of aluminum

Radioactive ^{28}Al was produced through the reaction with thermal neutrons,

$^{27}Al\ (n, \gamma)\ ^{28}Al$

The half life and γ-ray energy of ^{28}Al are 2.31 minutes and 1.776 MeV, respectively.

In the past, quantitative analysis of Al in excessive Si was considered to be extremely difficult because of the transmutation reaction of Si to Al caused by the coexistence of fast neutrons,

$^{28}Si\ (n, p)\ ^{28}Al$

However, irradiation in a thermal column made detection of less than 1μg of Al in pure 1g SiO_2 possible by limiting the transmutation reaction from the Si. Five minutes of irradiation provided enough activity to analyze the Al content.

(B) Analysis of sodium

Quantitative analysis of the Na content was easier than that of Al because the transmutation reaction from ^{24}Mg and ^{27}Al shown below could be ignored.

$^{24}Mg\ (n, p)\ ^{24}Na$

$^{27}Al\ (n, \alpha)\ ^{24}Na$

The half life and γ-ray energy of ^{24}Na are 15 hours and 1.36 MeV and 2.75 MeV, respectively. One hour of irradiation in an R.S.R. was enough to obtain sufficient activity for analysis of Na below 3μg. However, annealing for two hours is necessary to reduce the activity from Bremsstrahlung and other radioactive nuclei having a shorter life. The samples available after impurity analysis were used to measure the infrared absorption to detect the radiation effect.

Results

1. Growth of crystals

Results of crystal growth in dependence of each solution, growth temperature and growth rate are summerlized in Table 3(a). Table 3(b) lists the impurity Al and Na contents of some crystals obtained by neutron activation analysis. Detailed results and discussion about crystal growth have been described in References1 - 3.

Transparent crystals grown in NaCl by using X-cut and Z-cut seed are shown in Fig. 1.

Striations, which corresponded to those of natural quartz were seen on the prism faces, whereas rectangular or hexagonal growth pyramids were generally observed on the surface of quartz crystals grown in alkaline solutions.

In case of KCl solution growth, growth conditions and these effect were nearly identical to the result of NaCl growth.

Solubility in pure water is extremely low; i.e., 0.1-0.3wt , and is from one tenth to one hundredth of the solubility in alkaline solutions, such as NaOH or Na_2CO_3. We succeeded in growing large crystals in pure water having such low solubility, but the growth rates obtained were very slow.

For example, the growth rate for Z and X faces were 0.01, 0.001, respectively, under the conditions of a growth temperature 425°C and a dissolving temperature 450°C. These were from one tenth to one hundredth of the growth rates of quartz crystals grown in alkaline solutions. Fig. 2 shows quartz crystal grown on Y-bar (lower part) and X-cut (upper part) seeds.

2. Infrared absorption

Crystals grown in NaCl solution

Figure 3 shows the polarized infrared absorption spectra of crystals grown in an NaCl solution. Two absorption bands of 3364 and 3304 cm^{-1} caused by Al-OH centers were clearly observed in "as grown" crystals. The concentration of OH^- was determined with the equation presented by Kats[6], and then applied by Sibley and Halliburton. For example,

$$NH = 2.16 \times 10^{16}\ H.\alpha m$$

Where OH^- concentration NH (defects/cm^3), half band width H, and maximum absorption coefficient αm (cm^{-1}) are substituted from the absorption bands. The concentration of OH^- was calculated to about 7ppm from Table 4.

Table 4

Al-OH^- band	H	αm
3364 cm^{-1}	8.5	0.97
3304 cm^{-1}	7.5	0.31

A remarkably intense absorption ($\alpha \sim 0.7$) having a 7 to 8 cm^{-1} half band width was found at 3345 cm^{-1}. The dotted line in Figure 4 shows the absorption spectra after irradiation of crystals. The Al-OH^- concentration increased about 50% and the absorption intensity decreased to about one-third at the 3450 cm^{-1} band. Table 5 lists the principal absorption bands changed by irradiation.

Table 5

Wavenumber	Before irradiation (α cm^{-1})	After irradiation (α cm^{-1})
3450	0.76	0.25
3430	0.20	0.25
3364	0.97	1.41
3304	0.31	0.56

Furthermore, the superimposed background absorption decreased remarkably after irradiation.

Crystals grown in KCl solution (Figure 5)

A discrete absorption band was observed at 3364 cm^{-1} in crystals grown in a KCl solution. However, another band at 3304 cm^{-1} could not be clearly distinguished because of the superimposed background absorption. The striking change of the absorption profiles taken at various temperatures (300°K to 78°K) is noteworthy. This tendency is also evident in the purer crystals grown in H_2O in Figure 9.

Crystals grown in H_2O

Figure 6 shows the polarized infrared absorption spectra of crystals grown in pure water. Strong discrete absorption bands were observed at 3364 and 3304 cm^{-1} from both crystals grown in H_2O and those in NaCl and KCl solutions. Additional bands at 3580 (E⊥C), 3435 (E⊥C), 3390 (E⊥C, E//C), and 3345 (E//C) commonly found in the spectra of crystals grown in an alkaline solution were also found in these crystals.

However, the most distinguishing feature of crystals grown in pure water was their almost identical spectra before and after irradiation as shown in Figure 7. Table 6 shows the changes of the extinction coefficient in each band before and after irradiation.

Table 6

Wavenumber	Before irradiation (α cm^{-1})	After irradiation (α cm^{-1})
3580	0.76	0.83
3435	1.13	1.01
3390	0.71	0.41
3364	1.84	2.38
3304	0.69	0.64

The absorption intensity changes at 3364 and 3304 cm^{-1} were much smaller than those of crystals grown on alkaline or NaCl, KCl solutions. We also grew purer crystals in water by using synthetic fused silica as the nutrient. Figure 8 shows the polarized infrared absorption of the purer crystals. These crystals exhibited an intense superimposed background absorption at the 3364 and weaker 3304 cm^{-1} bands. The calculated OH^- concentration for the 3364 and 3304 cm^{-1} bands was about one seventh that of the crystals shown in Figure 6. The superimposed absorption was sensitive to temperature changes, the same as crystals grown in KCl solution, as shown in Figure 9.

Crystals grown in Na_2CO_3-SiO_2-D_2O

Figure 10 shows the 4000 to 2200 cm^{-1} polar-

ized infrared absorption of crystals grown in Na_2CO_3 with D_2O as the solvent.

As can be seen from Figure 10, neither discrete bands due to OH^- vibrations nor superimposed absorption due to clusters of impurities are observed in the 4000 to 3000 cm^{-1} range. Therefore, the intensity of the absorption in the 3μm region can be estimated from the intrinsic absorption from the Si-O lattice vibration. Table 7 compares the absorption at typical wavenumbers of these crystals and very high grade synthetic quartz having a Q of over 3 million.

Table 7

Wave-number	D_2O quartz	Very high grade synthetic quartz grown in an NaOH solution
3585	0.0055	0.026
3390	0.082	0.084
3290	0.14	0.15
3190	0.13	0.14

Figure 11 shows the effects of radiation on crystals grown in Na_2CO_3-D_2O. Neither enhanced absorption in the 4μm region nor radiation induced absorption at 3364 and 3304 cm^{-1} were observed in these spectra after irradiation. Table 8 lists the wavenumbers of typical absorption due to OD- vibration and corresponding OH- vibration.

Table 8

OD	OH	$\nu OH/\nu OD$
2645	3580	1.353
2550	3430	1.345
2525	3395	1.345
2495	3345	1.341

Purer quartz crystals grown in Na_2CO_3-SiO_2-H_2O and K_2CO_3-SiO_2-H_2O

Figure 12 and 13 show the polarized infrared absorption of crystals grown in Na_2CO_3 and K_2CO_3, respectively. The four absorption bands, 3580, 3435, 3395, and 3345 cm^{-1}, generally exhibited by crystals grown in an alkaline solution were also observed in these crystals. However, additional smaller bands were also found at 3540 and 3465 cm^{-1}.

DISCUSSION

1. Since some of the potential configurations of the charge compensators for point defects of substitutional Al in crystals grown in an NaCl solution may be only Al-OH^- and Al-Na^+ centers, the absorption band at 3450 cm^{-1} may be closely related to Al-Na^+ defects. Radiation, which serves to remove Na^+ from the vinicity of the Al-site, reduced the absorption at 3450 cm^{-1} and enhanced the absorption caused by Al-OH^- centers.

2. The three absorption bands at 3580, 3395, and 3345 cm^{-1} commonly observed in crystals grown in an alkaline solution were not detected in the crystals grown in NaCl. However, the "as grown" crystals did exhibit absorption bands at 3364 and 3304 cm^{-1}. In this respect, the profile of absorption spectra of "as grown" crystals grown in NaCl is very similar to that of high quality natural quartz as shown in Figure 14. However, as shown in Figure 15, crystals grown in a coexistence solution with Na^+ and K^+, for example, NaCl-KCl-LiCl-H_2O systems, did not exhibit the remarkable features of the crystals grown in an NaCl solution.

3. The 3364 and 3304 cm^{-1} absorption bands of crystals grown in pure water showed a strong resistance against irradiation. This probably due to the fact that all the substitutional aluminum was formed as Al-OH^- defect centers because there were no alkaline ions near the Al-site. However, these crystals did exhibit absorption bands at 3580, 3435, 3395, and 3345 cm^{-1}. Therefore, these four bands may not be related to the Al-M^+(M: alkaline ion) centers. The four identical absorption bands found when the interstitial alkaline ions were changed from Na^+ to K^+ also support this assumption.

4. Table 9 lists the dichroic factors (dichroic factors are defined as E⊥C / E//C)empirically confirmed at each band. The dichroic factors listed below were found in K_2CO_3-SiO_2-H_2O systems.

Table 9

Wave-number	(α cm^{-1}) E⊥C	E//C	E⊥C / E//C	Normal dichroic factor
3580	1.78	0.81	2.2	>10
3435	1.91	1.30	1.5	≳ 5
3395	0.95	1.78	0.5	≲ 1

The absorption bands of a crystal grown in K_2CO_3 solution exhibit a relatively strong extinction coefficient and irregular dichroic factor as shown in Figure 15. A higher concentration of impurities may cause action in both E⊥C and E//C absorption at 3580, 3435, and 3395 cm^{-1}, but the absorption at 3364 and 3304 cm^{-1} did not show any irregular dichroic factor.

CONCLUSION

1. Flawless transparent quartz crystals were grown by the hydrothermal gradient method in the following systems:
   1) NaCl-SiO_2-H_2O, KCl-SiO_2-H_2O
   2) SiO_2-H_2O
   3) Na_2CO_3-SiO_2-D_2O, Na_2CO_3-SiO_2-H_2O, K_2CO_3-SiO_2-H_2O by using pure synthetic fused silica as a nutrient.

2. "As grown" crystals grown in NaCl, KCl solution, and pure water exhibited absorption bands at 3364 and 3304 cm^{-1} due to Al-OH^- centers. A crystal grown in NaCl shows an exceptional absorption spectra similar to that of high grade natural quartz at 3580, 3395, and 3345 cm^{-1}. However, it does not exhibit the 3580, 3395, and 3345 cm^{-1} absorption bands commonly found in crystals grown in an alkaline solution (i.e., NaOH, Na_2CO_3, K_2CO_3), KCl, and pure water.

More detailed investigations on the exceptional absorption spectral of crystals grown in NaCl solution are required. The reason why crystals grown in, for example, NaCl/KCl=85/15 solution do not exhibit this remarkable infrared absorption while natural crystals that grow in a similar coexistence

ions solution (Na^+ and K^+) do and what types of potential chemical species exist in the NaCl-SiO_2-H_2O system are also subjects for further study.

3. The four absorption bands at 3580, 3434, 3395, and 3345 cm^{-1} normally observed in crystals grown in an alkaline solution may not be related to Al-M^+ (M: alkaline ion) centers, but may be caused by interstitial H_2O or its derivatives.

REFERENCES

1. M. Hosaka and S. Taki, "Hydrothermal growth of quartz crystals in NaCl solution", J. Cryst. Growth 52, 837 (1981).
2. M. Hosaka and S. Taki, "Hydrothermal growth of quartz crystals in pure water", J. Cryst. Growth 51, 640 (1981).
3. M. Hosaka and S. Taki, "Hydrothermal growth of quartz crystals in KCl solution", J. Cryst. Growth, To be published.
4. L. E. Halliburton, N. Koumvakalis, M. E. Markes, and J. J. Martin, J. Appl. Phys., To be published.
5. R. N. Brown and A. Kahan, J. Phys, Chem. Solid, 36, 467 (1975).
6. A. Kats, Philips Res. Repts, 17, 133 (1962)
7. H. G. Lipson, F. Euler, and A. F. Armington, Proc. 32nd Annual Symposium on Freq. Control, p. 11, (1978).
8. W. A. Sibley, J. J. Martin, M. C. Wintersgill, and J. D. Brown, J. Appl. Phys., 50 (8), 5449, (1979).
9. L. E. Halliburton, M. E. Markes, and J. J. Martin, Proc. 34th Annual Symposium on Freq. Control, p. 1, (1980).

Table 3(a) Summerized data of crystal growth

Run	Solvent	Nutrient	Growth temp. (°C)	Temp. gradient (°C)	days	Growth rate(10^{-2} mm/days) X	Z	R	r	Remarks
A-7	NaCl	Natural quartz	400	25	8	1.3	37.5	0.97	0.10	Growth on R, r faces and Y-bar seed. Some crevice flawing in Z sector.
A-6	NaCl	Natural quartz	420	25	8	2.6	58.0	1.60	0.20	
A-5	NaCl	Natural quartz	440	25	8	5.0	92.0	2.90	0.60	
C-25	NaCl	Natural quartz	450	25	111					Growth on AT cut seed. (0.01 mm/day) Transparent
G-3	NaCl	Natural quartz	450	25	20	38.8				Growth on X cut seed. Transparent
D-31	KCl	Natural quartz	400	25	6	0.5	34.0	1.02	0.16	Growth on R, r faces and Y-bar seed. Some crevice flawing in Z sector.
D-32	KCl	Natural quartz	420	25	6	0.8	64.0	1.70	0.27	
D-30	KCl	Natural quartz	440	25	5	3.2	115.0	2.76	0.59	
B-6	KCl	Natural quartz	400	25	24	6.5	69.0			Growth on Y-bar seed.
I-10	H_2O	High purity fused silica	490	25	100		7.4			Growth on Y-bar seed.
I-30	H_2O	Synthetic quartz	465	25	109	0.8	6.5			
20	K_2CO_3	Synthetic quartz	330	40	14	12.7	99.0			Transparent
14	K_2CO_3	High purity fused silica	350	10	7	19.0	76.0			Growth on Y-bar seed. Transparent D_2O was used in No. 12.
12	K_2CO_3	High purity fused silica	350	10	12	32.0	54.0			
3	Na_2CO_3	High purity fused silica	350	10	10	40.0	60.0			High purity quartz crystal.

In G-3, B-6, C-25, and I-30, crystals were grown in a platinum capsule(V=100 ml)placed in a autoclave(V=157 ml). In other runs, crystals were grown in a platinum capsule(V=18.73 ml) placed in a autoclave(V=28.23 ml) or a gold capsule(V=12.7 ml) placed in a autoclave(V=31.9 ml).

Table 3(b)

Impurity aluminum and sodium contents obtained by neutron activation analysis.

Sample	(Run)	Concentration of Al(ppm)	Concentration of Na(ppm)
NaCl soln.	(G-3)	163	216
KCl soln.	(B-6)	-	-
H_2O soln.	(I-30)	50.7	0.67
D_2O soln.	(No.13)	14.5	80.3
Na_2CO_3 soln.	(No.3)	2.07	-
K_2CO_3 soln.			
(1) normal	(No.20)	-	-
(2) purer	(No.14)	6.63	1.25



(a) (b)

Fig. 1 Crystals grown in NaCl solution. Crystal (a) is grown on X-cut seed, crystal (b) is grown on Z-cut seed. The striations perpendicular to Z-axis are observed on a prism face in similar to those of natural quartz.

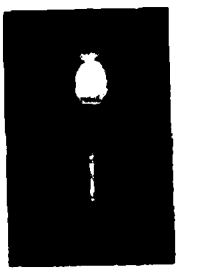


Fig. 2. Crstals grown in pure water on a platinum frame, using X-cut (upper) and Y-bar (lower) seed. The striations on a prism face are observed.

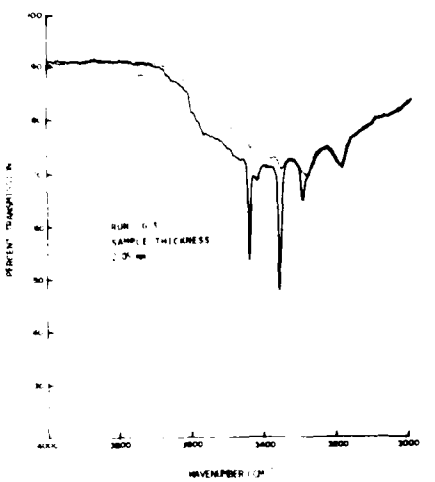


Fig. 3 E⊥C (Solid line) and E//C (Dotted line) absorption spectra of crystal grown in NaCl solution at 78°k. Intense absorption at 3364 and 3304 cm^{-1} are exhibited from "as grown" crystal.

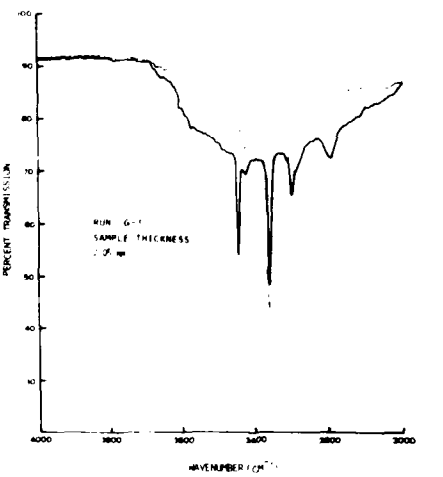


Fig. 4. Radiation effects of crystals grown in NaCl in E⊥C absorption spectra at -78°k as comparing with before (Solid line) and after (Dotted line) neutron irradiation.

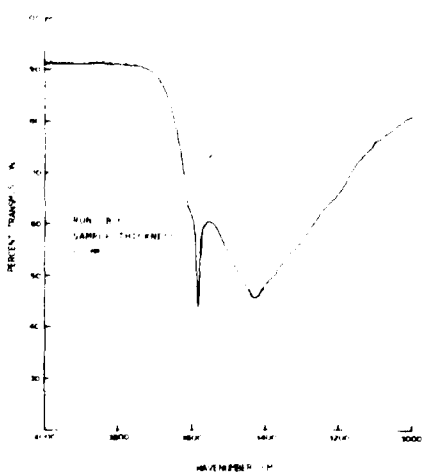


Fig. 5 Temperature dependence of absorption spectra of crystals grown in KCl solution at 300°k (Solid line) and 78°k (Dotted line). Absorption band at 3364 cm^{-1} is observed at low temperature infrared

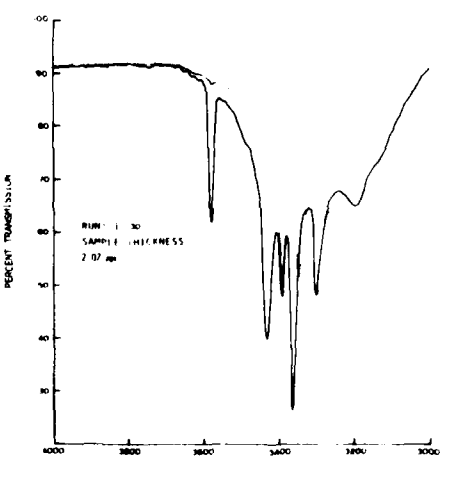


Fig. 6. E⊥C (Solid line) and E//C (Dotted line) absorption spectra of crystals grown on pure water at -78°k. Absorption bands at 3364 and 3304 cm^{-1} are observed from "as grown" crystal and additional four bands at 3580, 3395 and 3345 cm^{-1} which are exhibited commonly by crystals grown in alkaline solution are also found.

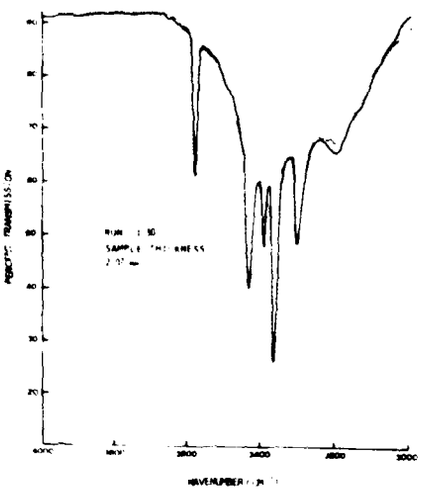


Fig. 7. Radiation effects of crystals grown in pure water in E⊥C absorption spectra at 78°k as comparing with before (Solid line) and after (Dotted line) neutron irradiation.

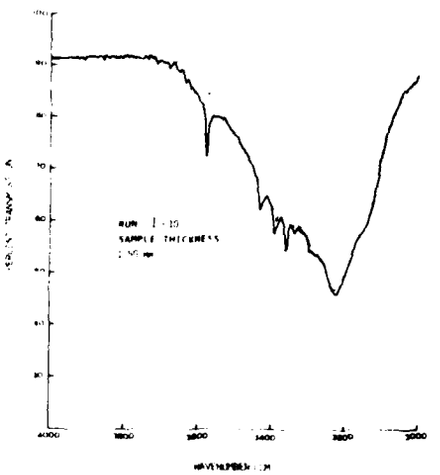


Fig. 8. E⊥C (Solid line) and E //C (Dotted line) absorption spectra of purer crystals grown in pure water. Absorption intensity of discrete bands are weakened, however strong superimposed background is exhibited.

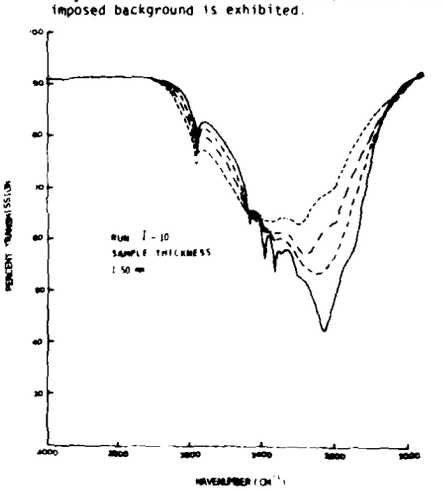


Fig. 9. Changes of absorption spectra as a function of temperature from 300 to 78°k in case of purer crystals grown in pure water.

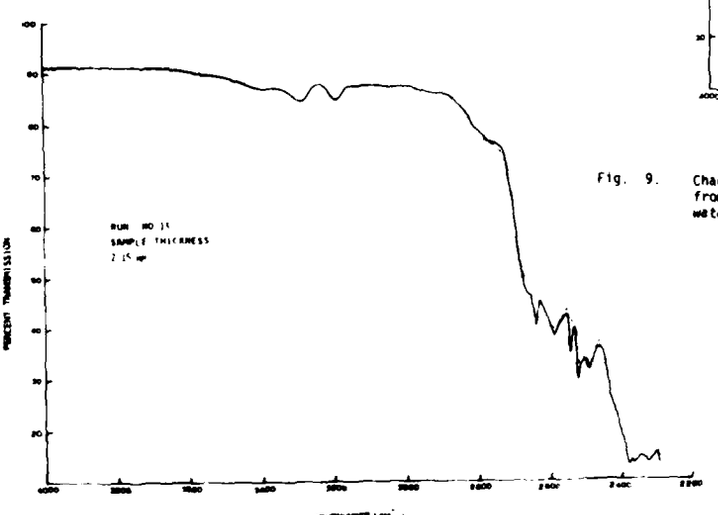


Fig. 10. E⊥C (Solid line) and E//C (Dotted line) absorption spectra of crystals grown in Al_2O_3-SiO_2-H_2O system. Four principal OH-vibration bands are exhibited in 4 μm region.

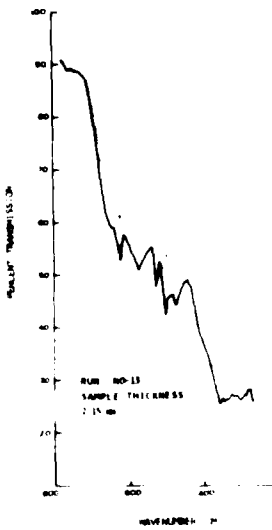


Fig. [illegible] Radiation effects of crystals grown in Na2CO3-SiO2-D2O in E⊥C absorption spectra at 78°k as comparing with before (Solid line) and after (Dotted line) neutron irradiation. All of discrete OD-vibration bands are weakened by neutron irradiation.

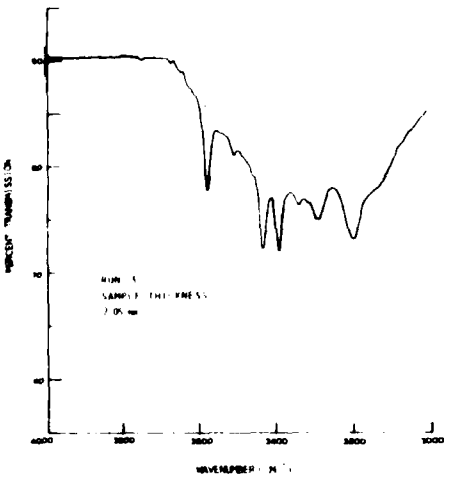


Fig. 12 E⊥C (Solid line) and E//C (Dotted line) absorption spectra at 78°k of purer crystals grown in Na2CO3 solution.

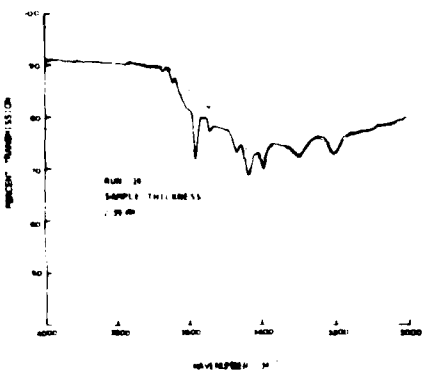


Fig. 13 E⊥C (Solid line) and E//C (Dotted line) absorption spectra at 78°k of purer K2CO3 solution

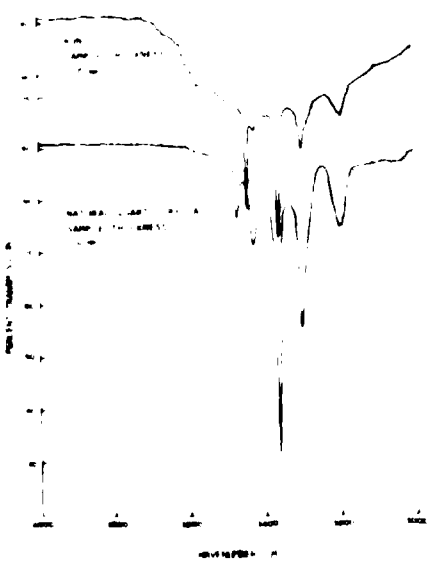


Fig. 14 E⊥C (Solid line) and E//C (Dotted line) absorption spectra at 78°k as comparing with natural quartz crystals and crystals grown in NaOH solution

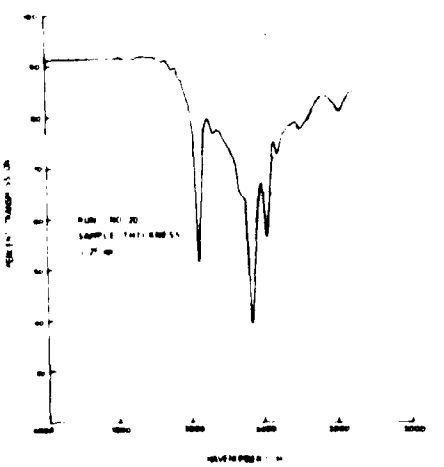


Fig. 15 Infrared absorption of crystal grown in K_2CO_3 solution after neutron irradiation. Irregular dichroic factors are exhibited between E⊥C (Solid line) and E//C (Dotted line).

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703 May 1981

THE INFLUENCE OF THE QUALITY FACTOR OF QUARTZ ON SOME DEVICE PROPERTIES

J.C. Brice*, J. Dowsett+ and E.D. Fletcher*

* Philips Research Laboratories, Redhill, Surrey, England.
+ Cathodeon Crystals Ltd, Linton, Cambridge, England.
- now at Xtal do Brasil, CEP 20,000, Rio de Janeiro, Brasil.

Summary

Plane parallel 10 MHz fundamental mode bulk wave resonators were fabricated by identical methods from 10 blocks of quartz with different infra-red Q's such that 1.3 million < Q_{IR} < 2.6 million. It was found that for Q_{IR} < 1.8 million, the equivalent series resistance R_1 increased by an average of about 50%. Measurements at power levels between 5 nW and 1 mW showed that over the Q_{IR} range 1.8 to 1.3 million the change in R_1 and in frequency also doubled on average. Over the entire range of Q_{IR}, the spread of the temperature coefficients at 25°C also increased at Q_{IR} fell.

The data presented give direct experimental confirmation to the previously formulated hypothesis that the production of devices to tight tolerances requires quartz with a high Q_{IR}.

Even devices with relatively low Q values (in the range 10,000 to 50,000) show a dependence of the device Q on the intrinsic material Q_{IR}. We explain this in terms of the machining properties of the quartz which lead to more badly damaged surfaces for material with lower values of Q_{IR}.

Introduction

Previous work has suggested that it is not desirable to use quartz with a low infra-red Q (Q_{IR}) to make devices. On a statistical basis as Q_{IR} falls, the crystals become less pure, more dislocated and less homogeneous[1]. Bye and Cosier[2] have shown inhomogeneity increases the equivalent series resistance (R_1) of the devices. Also as Q_{IR} falls, the resistance to thermal (and presumably mechanical) shock also falls[3]. This weakening of the material may well account for the observed decreased yield and possibly also drive level dependent effects[4]. Similarly, low Q_{IR} crystals have a greater spread of lattice constants and on average appreciably different ones from high Q_{IR} crystals[5]. This must affect the angle of cut and therefore the frequency temperature characteristics. Toyo workers[6] have shown that Q_{IR} < 1.5 million increases R_1 in overtone models. The present paper shows that for fundamental mode crystals an effect is obvious for Q_{IR} < 1.8 million and that there are also appreciable effects on the frequency temperature curves.

Experimental

The quartz for the experiments came from 5 crystals from three suppliers. Each crystal was cut in half to remove the seed and the infra-red Q was determined by the previous described method[1,7]. The Q_{IR} values quoted are the means of those determined at 3585 and 3410 cm^{-1}. From previous work it is believed that the Q_{IR} value in each half crystal should not differ anywhere in the half crystals by more than ± 0.2 million from the values quoted. The half crystals were aligned by X-rays and cut with a slurry saw.

The devices were made using straightforward processing methods commonly used in the industry for devices not requiring a tight specification.

The devices were constructed on 8 mm diameter plane parallel circular blanks. After a final lap with a 5 μm aluminium oxide abrasive grit, the blanks were etched in ammonium bi-fluoride 0.6 f^2 (i.e. removing about 0.5 μm from each face). Silver electrodes 3.5 mm diameter were evaporated onto the blanks to a thickness which decreased the frequency by about 0.7 f^2 (≈ 80 kHz for the 10.7 MHz devices, i.e. a mass loading of about 0.8% of the mass of quartz under the electrodes). The devices were encapsulated in dry nitrogen (-55°C dew point) in small resistance welded cases with wire leads (type RC or HC - 49/U). Ten batches of devices were made, one from each half crystal. The smallest batch tested contained 14 devices.

The resonant frequency (f) and the equivalent series resistance (R_1) and their variations (Δf and ΔR_1) with power dissipated in the devices were measured in an apparatus designed by Ir H.J.M.W. Dohmen of the Elcoma Division of Philips, Eindhoven. This apparatus consists of a phase zero transmission measuring system[8] with the device in a π-network between two coupled 64 dB attenuators which were changed in 2 dB steps so that while the power to the crystal changed, the total attenuation in the circuit was constant. The power level was changed from 5 nW to 5 mW. Measurements were made with the power increasing and decreasing. The values of R_1 and changes in R_1 (ΔR_1) and Δf quoted here are the means of the values found with the power increasing and decreasing. The maximum divergences hR_1 and hf between the two values were also recorded. In order to remove frequency and

resistance changes due to intrinsic nonlinearities and thermal effects only the changes and hysteresis in R_1 and f occurring up to a power level of 100 μW have been used. The precisions of the various measurements were 0.5% for R_1, 0.2% for ΔR_1 and hR_1 and 1 Hz for Δf.

Results for the ESR variations of a bad resonator are shown in Figure 1 and the corresponding frequency changes are shown in Figure 2. These illustrate the parameters R_1, ΔR_1, hR_1, Δf and hf.

The frequency temperature characteristics were measured using a Saunders type 2000 test system. The data are presented here as the temperature coefficient at 25°C. The precision of individual datum points is roughly 0.1 ppm/°C.

C_0 was measured with an appropriate bridge and C_1 and L_1 were measured in a phase-zero measurement system using the frequency change observed for an added series capacitor.

Results and Discussion

It was found that the values of C_0, C_1 and L_1 were respectively about 6 pF, 0.03 pF and 8 mH so that the device quality factors were in the range 10 to 50 thousand i.e. much less than the value expected for the material with the lowest Q_{IR}.

Figure 3 gives the values of R_1 found in two batches of crystals in histogram form. These data are typical of the high Q_{IR} (> 1.8 million) and low Q_{IR} (< 1.8 million) populations. It can be seen that the high Q_{IR} material gives a symmetrical but probably not Gaussian distribution while the low Q_{IR} material results are asymmetrically distributed over a much wider range. Figure 4 gives quartile, mean and median data for all the batches tested. This figure makes it clear that for Q_{IR} > 1.8 million, R_1 rises as Q_{IR} falls and the spread and asymmetry also increase. Figures 5 and 6 give similar data for ΔR_1 and Δf. Since the 90% confidence limits are about 0.3, 0.4 and 0.5 times the interquartile ranges respectively for the means, medians and quartiles, there is little doubt that the results are statistically significant.

The hysteresis parameters hR_1 and hf have similar distribution to ΔR_1 and Δf and the data can be represented by

$$hR_1 = (0.9 \pm 0.3)\Delta R_1 \quad \ldots (1)$$

$$hf = (0.9 \pm 0.6)\Delta f \quad \ldots (2)$$

Here the variations quoted are standard deviations of the individual ratios of the changes. Most of the variation is due to experimental error.

In the same way, it is possible to relate Δf (Hz) and $\Delta R(\Omega)$ to find that

$$\Delta f = (9.5 \pm 6.0)\Delta R \quad \ldots (3)$$

Again the error quoted is the standard deviation of the ratios of the individual data pairs. From the point of view of making stable devices, the changes Δf and hf are of paramount importance. Here they have been measured over a very large power range. It would be more realistic to consider them over smaller ranges. Over a one decade power range the changes are reduced typically by factors of about 6 and 3 respectively.

We know that unstable surfaces can be a prime cause of the drive level dependence of R_1 and f, especially at low drive levels, and that residual damage in the quartz surface is one contributory factor[4]. It seems reasonable therefore to suggest that the observed dependence of R_1, ΔR_1, f and Δf on the quality of the quartz is a consequence of the nature of the surfaces of the devices.

Figure 7 shows how the depth of damage in a lapped surface is related to the particle size of the abrasive for two qualities of quartz. The results were obtained by progressively etching the surfaces and observing the X-ray rocking curves.

Figure 8 shows how the thermal shock needed to crack a bar of quartz (as grown) is related to Q_{IR}. Thermal shocks occur on a microscopic scale during machining processes.

The above evidence supports the view that the machining properties of quartz and the properties of the resultant surfaces are dependent on Q_{IR}.

It is of course possible to compensate in many respects for the problems introduced by poor quality quartz by improved processing methods. The manufacturer must decide whether this is more cost effective than specifying quartz of a higher quality.

The ability to measure crystal properties over a wide range of power levels can be very important when the efficiency of processing methods is being assessed.

Figure 9 gives the spread of temperature coefficients at 25°C (measured by the standard deviation) as a function of Q_{IR}. The mean temperature coefficient of all the devices tested was -0.005 ppm/°C. The target was, of course, zero. However, if we take the six sample batches with Q_{IR} < 1.8 million, their mean and its standard deviation was -0.032 ± 0.042 and the 4 batches with Q_{IR} > 1.8 million gave +0.035 ± 0.062. The difference between these mean values could be exceeded by chance about one time in 5. However, the known variation of the axial ratio with Q leads us to expect a change of only about 0.02 ppm/°C. The right hand ordinate of Figure 9 is marked with the equivalent cutting angle error. (The calculated frequency temperature curves give a change of -0.082 ppm/°C for every arc minute increase in the angle of cut(θ)). The cutting and lapping methods used to fabricate these devices introduce a spread of θ values of between 2 and 3 arc minutes. Thus the spread at large Q_{IR} values is probably accounted for by a combination of cutting errors and

measurement errors. There is, however, a significant excess spread at low Q_{IR} values. The data given above about the mean temperature coefficients also suggest that the angle of cut to give zero temperature coefficient may change with Q_{IR}.

Conclusions

The work described here gives direct experimental evidence that device parameters depend on Q_{IR} even for relatively low Q devices. Low Q_{IR} values are associated with increases in the value and spread of the equivalent series resistance (R_1) and with changes with drive level of both R_1 and the resonant frequency. Low values of Q_{IR} also increase the variations in frequency temperature behaviour. For devices fabricated in the same way as the ones measured here, the effects become apparent for $Q_{IR} < 1.8$ million. Other fabrication techniques may give different limiting values but this study supports the general hypothesis that when devices to tight tolerances are needed, high Q_{IR} material should be used.

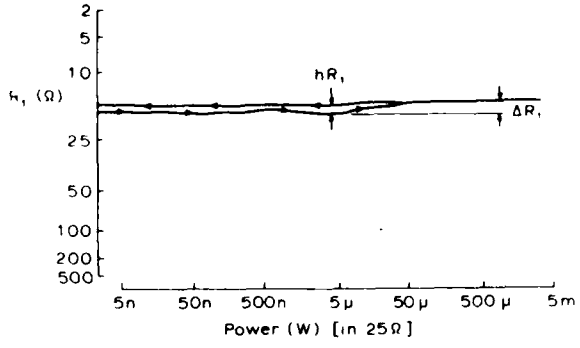


Fig 1 R_1 as a function of drive level (bad crystal)

References

1. J.C. Brice and A.M. Cole: "The Characterization of Synthetic Quartz by Using Infra-Red Absorption". Proc. 1978 Annual Frequency Control Symposium, 1-10.

2. K.L. Bye and R.S. Cosier: "An X-ray Double Crystal Topographic Assessment of Defects in Quartz Resonators". J. Mater. Sci. Vol. 14 (1979), 800-810.

3. J.C. Brice and C.C. Li: "The Breaking of α-Quartz by Thermal Shock". J. Mater. Sci. in press.

4. J.E. Knowles: "On the Origin of the Second Level of Drive Effect in Quartz Oscillators". Proc. 1975 Annual Frequency Control Symposium, 230-236.

5. J.C. Brice: "The Lattice Constants of α-Quartz". J. Mater. Sci. Vol. 15 (1980), 161-167.

6. Toyo Technical Paper No. 77-03 (1977): "Evaluation of Synthetic Quartz by Infra-Red Absorption Technique". Toyo Commercial Equipment Company, Koza-Gun, Japan.

7. J.C. Brice and A.M. Cole: "Infra-Red Absorption in α-Quartz". J. Phys. D. Vol. 12 (1979), 459-463.

8. IEC Publication 444.

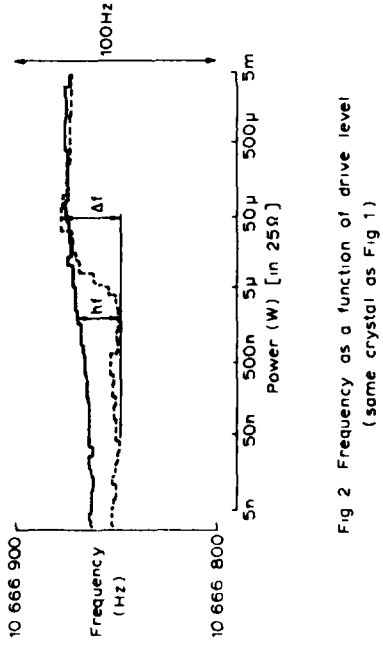


Fig 2 Frequency as a function of drive level (same crystal as Fig 1)

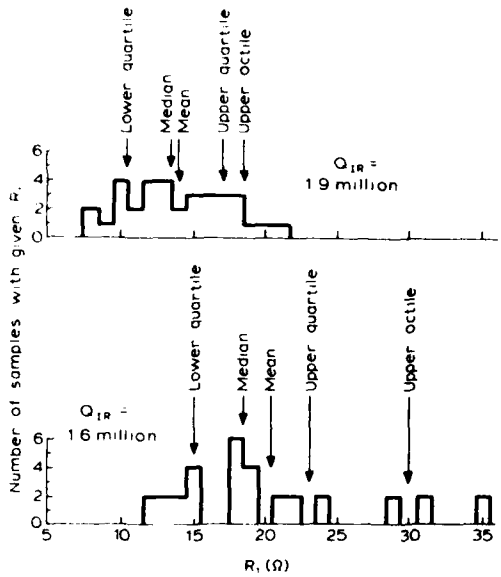


Fig 3 Distribution of R_1 for devices from crystals with Q_{IR} of 1.6 and 1.9 million

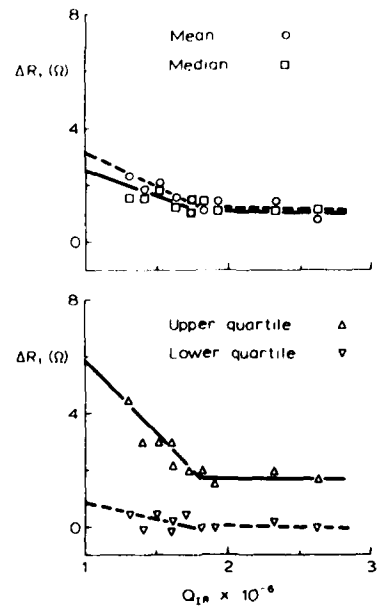


Fig 5 The changes ΔR_1 in R_1 over the range 5nW to 100 μW as a function of Q_{IR}

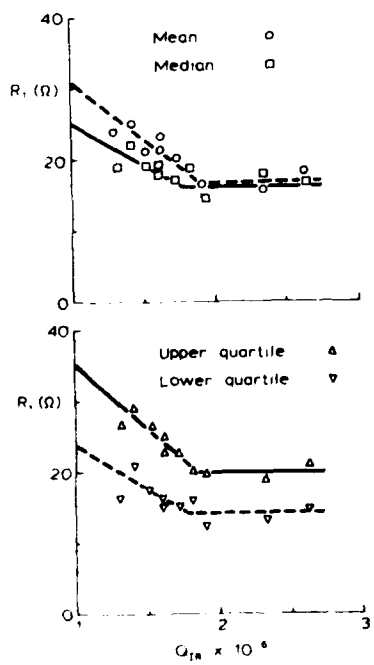


Fig 4 Mean median and quartiles of the R_1 distributions as a function of Q_{IR} for R_1 measured at 1mW

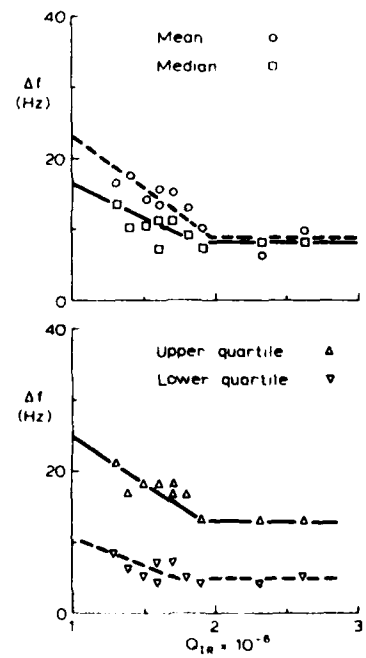


Fig 6 The changes Δf in f measured over the range 5nW to 100 μW as a function of Q_{IR}

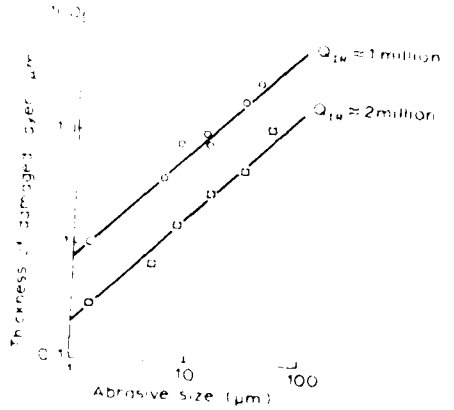


Fig 7 Depth of damage in a quartz surface as a function of particle size

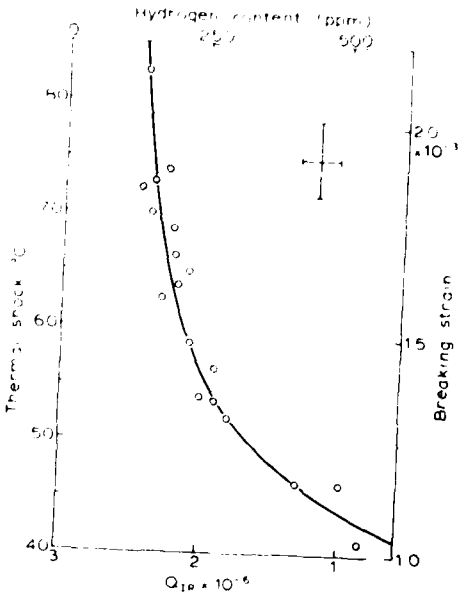


Fig 8 Thermal shock to fracture as a function of Q_{IR}

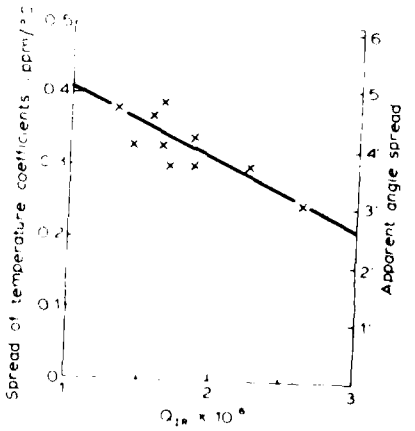


Fig 9 The spread of temperature coefficients (measured as the standard deviation) as a function of Q_{IR}

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703 May 1981

POINT DEFECTS IN CULTURED QUARTZ:

RECENT ACOUSTIC LOSS, INFRARED, AND MAGNETIC RESONANCE RESULTS

J.J. Martin, L.E. Halliburton, and R.B. Bossoli

Department of Physics
Oklahoma State University
Stillwater, Oklahoma 74078

Summary

Acoustic loss measurements on a 5 MHz 5th overtone resonator blank shows no loss peaks that can be attributed to the Al-Li^+ center, even though the blank was fabricated from a high-aluminum-content Premium Q bar and had been Li-swept, i.e., Li was swept into the blank. The Al-Na^+ center loss peak was removed by the Li sweep. Irradiation of unswept, Li-swept, and H_2-swept blanks at room temperature introduces loss peaks at 25 K, 100 K and 145 K. In H_2-swept material these peaks are approximately a factor of five smaller than the same peaks in similar unswept or Li-swept blanks. Since these three peaks anneal between 250 and 300°C, the same as the Al-hole center electron spin resonance spectrum, we conclude that the three peaks are caused by the Al-hole center. In unswept blanks, the Al-Na^+ loss peak is removed by radiation at room temperature but recoveres upon annealing to temperatures above 350°C. The decay of the Al-OH^- center correlates with the recovery of the Al-Na^+ center in unswept material.

Introduction

Alpha-quartz is used in a wide variety of precision electronic devices where aging and radiation-induced instabilities are undesirable. It is now well-known that quartz-controlled oscillators may exhibit transient and steady-state frequency shifts when exposed to ionizing radiation.[1-3] Early results obtained by King[4] and other investigators[5-7] suggested that these effects were correlated with the presence of impurities.

Substitutional Al^{3+} is present in all quartz[8] and requires charge compensation. Examples of such charge compensators are interstitial Li^+ or Na^+ ions, or holes or protons at an oxygen ion adjacent to the aluminum. The proton forms an OH^- molecule which is infrared active.[9,10] The Al-Na^+ defect is responsible for the acoustic loss peak observed near 50 K in 5 MHz 5th overtone AT-cut crystals.[11] Irradiation at room temperature destroys the Al-Na^+ centers[4,12] and this is responsible for much of the steady state frequency offset. Recent work at Oklahoma State University has shown that the alkali ions become mobile under irradiation only if the temperature is greater than 200 K.[12-14] Following a room temperature irradiation, either a hole which can be observed by ESR techniques or a proton is found trapped on an oxygen adjacent to the Al^{3+}. The interstitial alkali ions are usually in the relatively large c-axis channels and at high temperatures can move along the channel under an applied electric field. King,[4] and later Kats[10] and Fraser,[11] used this technique to "sweep" hydrogen and specific alkalis into the sample. Sweeping hydrogen in to replace the alkalis has been shown to improve the radiation hardness of quartz oscillators.[7]

The identification of both growth and radiation-induced defects which affect the performance of quartz resonators is an important part of our project. Recently, using sweeping, IR absorption, and acoustic loss measurements Martin and Doherty[15] reported that the Al-OH^- center does not have an acoustic loss peak at temperatures below 370 K. They also reported that irradiation of both unswept and H_2-swept Premium Q quartz resonator blanks produced acoustic loss peaks at 25 K and 100 K and a broad loss peak between 125 K and 165 K. King and Sander[1] had earlier reported the two higher temperature peaks and had suggested that they were caused by the Al-hole center. The 25 K peak had also been observed earlier and was attributed to changes in the interaction between the resonant vibration of the blank and the thermal phonons.[16] We report here an isochronal anneal study of these three loss peaks which shows that all three peaks are associated with the Al-hole center. We also compare the acoustic loss spectrum of unswept and Li-swept Premium Q material.

Experimental Procedure

All samples for this study were cut from an unswept pure Z growth Sawyer Premium Q bar of cultured quartz that has been given an in-house designation PQ-E. Samples from this bar have been extensively studied at Oklahoma State University using ESR,[9,13] IR,[9,15] and acoustic loss[15] technique. All of these investigations show that the bar is of high quality but that it contains somewhat more aluminum (10-15 ppm) than the average Premium Q material (5-8 ppm). Consequently, aluminum-related effects are more readily observed.

5 MHz 5th overtone AT-cut plano convex resonator blanks of the Warner[17] design were fabricated for this study by K&W Mfg., Prague, OK. The acous-

In Proc. 35th Ann. Freq. Control Symp., USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

tic loss, Q^{-1}, of the resonator blanks was measured by the log decrement method from 5 to 300 K. The measurements were made in a variable temperature helium Dewar with the blanks mounted in a gap holder similar to the one described by Fraser.[16] The blank was driven for 10 ms at its series resonant frequency and then allowed to freely decay. The decaying rf signal was detected with a superheterodyne detector and displayed on a variable persistence storage oscilloscope. The exponential decay times were measured using a window detector to gate a digital timer.

Electrolysis, or sweeping, was carried out at temperatures of 470-480°C in a controlled atmosphere system. This system allowed the use of H_2, D_2, or gettered N_2 atmospheres, or an oil-diffusion-pumped vacuum of approximately 2×10^{-6} Torr. For the acoustic loss studies, the AT-cut resonator blanks were directly swept. Vapor-deposited Au electrodes were used. For the Li sweeping runs, LiCl was vapor deposited on the sample surface and then an Au electrode was deposited over the LiCl layer. The Li electrolysis was carried out in a vacuum better than 5×10^{-6} Torr.

Irradiations were carried out at room temperature using 1.7 MeV electrons from a Van de Graaff accelerator. Typical doses were approximately 2000 J/cm^3 ($\sim 10^8$ rads). This dose is not expected to produce significant knock-on damage, but it has been shown to saturate the Aℓ-related defects. The production of Aℓ-OH^- centers by the irradiation was measured by a liquid nitrogen temperature polarized infrared scan on the resonator blank.

Results and Discussion

Figure 1 compares the current-versus-time curves for typical H_2 and Li sweeping runs. The

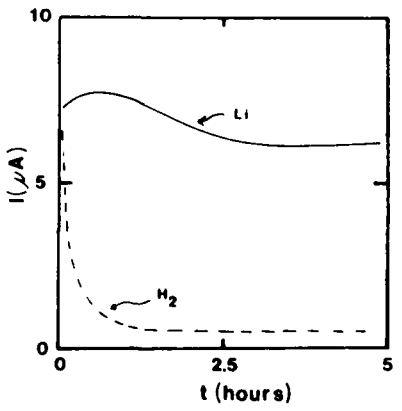


Figure 1. Current-versus-time curves are shown for lithium and hydrogen electrolysis of PQ-E resonator blanks.

component of the applied field parallel to the c axis is 35 V/cm for the Li sweep and 1500 V/cm for the H_2 sweep. In the H_2 case, the large initial current and subsequent decay is most likely caused by the alkalis being removed from the sample and the low final current is most likely protons being swept through the lattice. Since Li ions are being supplied from the LiCl film at the positive electrode, the current remains nearly constant in the Li sweeping runs. Infrared absorption scans show that Aℓ-OH^- centers are present in the H_2-swept sample but not in the Li-swept sample. Figure 2 compares the acoustic loss, Q^{-1}, versus

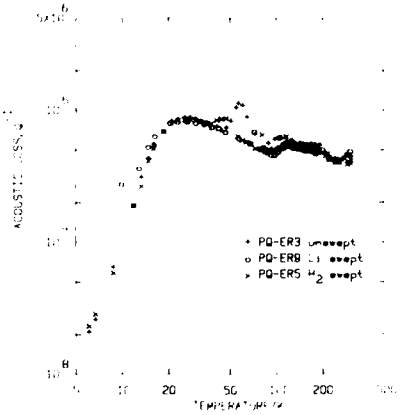


Figure 2. Acoustic loss, Q^{-1}, versus temperature spectra are shown for unswept, hydrogen- and lithium-swept PQ-E resonator blanks.

temperature spectra for unswept, H_2-swept, and Li-swept resonators fabricated from the PQ-E bar. The Aℓ-Na^+ loss peak near 50 K is absent in both of the swept resonators. We do not see any loss peaks caused by the Aℓ-Li^+ center. Evidently the coupling of this center to the AT thickness shear mode is much weaker than that of the Aℓ-Na^+ center. An earlier comparison[15] of H_2- and D_2-swept PQ-E resonators showed that the Aℓ-OH^- center does not have loss peaks at temperatures below 370 K.

Recently, Martin and Doherty[15] reported that irradiation of unswept and H_2-swept PQ-E resonator blanks at room temperature produced loss peaks at 25 K and 100 K and a broad loss between 125 and 165 K. The peaks were much larger in the unswept material than in the H_2-swept blank. The room temperature irradiation also removed the Aℓ-Na^+ loss peak which was present initially in the unswept blank. King and Sander[1] have previously observed the peak at 100 K and the broad loss between 125 K and 165 K. They attributed these loss peaks to the Aℓ-hole center. Martin and Doherty suggested that the 25 K peak is also due to the Aℓ-hole center. Figure 3 shows the Q^{-1} vs temperature spectrum for the Li-swept resonator blank in the as-swept condition and after a room temperature irradiation. The irradiation-induced peaks are in excellent agreement with the earlier results on an

unswept blank.

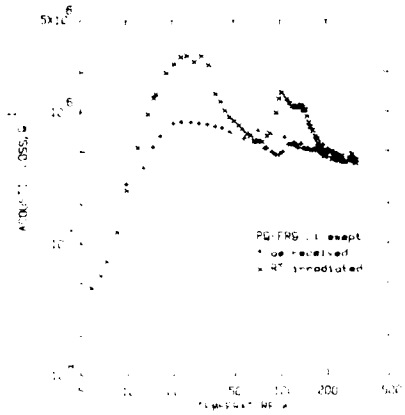


Figure 3. The acoustic loss spectrum of the lithium-swept resonator blank shows radiation-induced loss peaks at 25 K, 100 K, and 135 K. The same peaks are observed in unswept blanks. In hydrogen-swept blanks, the three peaks are significantly reduced.

...fect-related acoustic loss peaks can be described by

$$Q_i^{-1} = D\Delta/(1 + \omega^2\tau^2) \qquad (1)$$

where Q_i^{-1} is the loss above the intrinsic background, D is the strength factor, ω is the angular frequency, and τ is the relaxation time for reorientation of the defect. The relaxation is usually thermally activated as the defect must go over an energy barrier to reach the equivalent sites.[18] Thus,

$$\tau = \tau_0 \exp(E/kT) \qquad (2)$$

where E is the barrier height and τ_0 contains the number of equivalent orientations and the attack rate. We have fit Eq. 1, with the relaxation time given by Eq. 2, to the Al-Na^+ center loss peak and to the three radiation-induced peaks. For purposes of this calculation the broad loss between 125 K and 145 K is described by a single peak centered at 135 K. The calculated parameters are listed in Table I.

Table I. Loss Peak Parameters

T(K)	E(meV)	τ (sec)	Defect
54	59	2.5×10^{-13}	Al-Na^+
25	8	8.1×10^{-14}	Al-hole
100	9	1.0×10^{-13}	Al-hole
135	11	2.1×10^{-13}	Al-hole

Our activation energy and relaxation time for the Al-Na^+ center loss peak are in excellent agreement with those determined by Stevels and Volger[13] from dielectric loss measurements. Stevels and Volger have also reported a radiation-induced dielectric loss peak with E = 7.5 meV and $\tau_0 = 5 \times 10^{-7}$ sec. This activation energy is in reasonable agreement with our 25 K peak, but the relaxation time is much longer. Taylor and Farnell[19] have also made dielectric loss measurements on irradiated quartz; they found a loss peak near E = 7.5 meV in agreement with Stevels and Volger and an additional peak at low temperature with E = 1.2 meV and $\tau_0 = 6.2 \times 10^{-5}$ s.

The thermal anneal behavior of the Al-hole and Al-OH^- centers in unswept quartz has been studied by Markes and Halliburton[17] and by Sibley et al.,[14] respectively. The Al-hole center, as observed by ESR techniques, anneals out slightly below 300°C while the Al-OH^- center anneals out between 350°C and 400°C. These results strongly suggest that the alkali does not return to the aluminum site until the 350°C to 400°C range is reached. If the three radiation-induced loss peaks are due to the Al-hole center, they should show the same annealing pattern as the Al-hole center ESR spectrum. We have carried out an isochronal anneal study on the unswept blank and found that the 25 K, 100 K, and 135 K loss peaks all anneal out at 270°C, as shown in Figure 4.

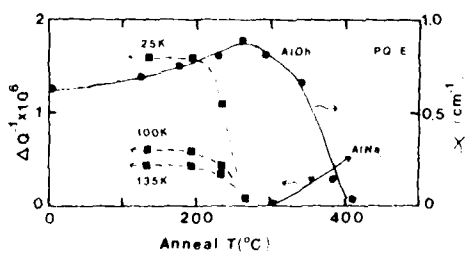


Figure 4. The 25 K, 100 K, and 135 K radiation-induced loss peaks anneal out together at the same temperature as the Al-hole center. The Al-OH^- center anneals out while the Al-Na^+ loss peak returns above ...

Thus, all three ... peaks are most likely caused by the Al-hole center. When the anneal is continued to higher temperatures, the Al-Na^+ center loss peak recovers between ... and ... as shown in Figure 4. The thermal anneal of the Al-OH^- center, as measured by infrared absorption, on a room-temperature-irradiated unswept optical sample taken from the ...-E bar is also shown in Figure 4. The Al-OH^- strength grows slightly when the Al-hole centers decay and then ... at when the Na^+ ions return to the Al sites.

Conclusions

When a Premium Q resonator blank was swept with Li, the 54 K Al-Na^+ center peak was removed from the acoustic loss versus temperature spectrum but no new peak that could be attributed to the Al-Li^+ center was observed. Therefore, we conclude that the Al-Li^+ center shows much less coupling to the AT thickness shear mode than the Al-Na^+ center. Both unswept and Li-swept Premium Q blanks show large acoustic loss peaks at 25 K, 100 K, and 145 K when they are irradiated at room temperature. These peaks are also present in H_2-swept material, in which case they are approximately 20% of the height observed in the unswept quartz. Annealing studies show that all three of these peaks are due to the Al-hole center and that it anneals out in the 250-300°C range. The Al-Na^+ center loss peak which is removed by a room temperature irradiation returns for annealing temperatures above 350°C. The Al-OH^- center in unswept material anneals out at that same temperature.

References

1. J.C. King and H.H. Sander, IEEE Trans. Nucl. Sci. NS-19, 23 (1972).

2. F. Pellegrini, F. Euler, A. Kahan, T.M. Flanagan, and T.F. Wrobel, IEEE Trans. Nucl. Sci. NS-25, 1267 (1978).

3. T.J. Young, D.F. Koehler, and R.A. Adams, Proceedings of the 32nd Annual Symposium on Frequency Control, 34 (1978). Copies available from the Electronic Industries Association, 2001 Eye Street, N.W., Washington, D.C. 20006.

4. J.C. King, Bell System Technical J. 38, 573 (1959).

5. R.A. Toll and S.L. Ridgway, IEEE Trans. Nucl. Sci. NS-13, 130 (December, 1966).

6. T.M. Flanagan and T.F. Wrobel, IEEE Trans. Nucl. Sci. NS-16, 130 (December, 1969).

7. B.F. Capone, A. Kahan, R.N. Brown, and J.R. Buckmelter, IEEE Trans. Nucl. Sci. NS-17, 217 (December, 1970).

8. T.M. Flanagan, IEEE Trans. Nucl. Sci. NS-21, 296 (December, 1974).

9. L.E. Halliburton, N. Koumvakalis, M.E. Markes, and J.J. Martin, J. Appl. Phys. (in press).

10. A. Kats, Philips Res. Repts. 17, 133 (1962).

11. D.B. Fraser, J. Appl. Phys. 35, 2913 (1964).

12. S.P. Doherty, J.J. Martin, A.F. Armington, and R.N. Brown, J. Appl. Phys. 51, 4164 (1980).

13. M.E. Markes and L.E. Halliburton, J. Appl. Phys. 50, 8172 (1979).

14. W.A. Sibley, J.J. Martin, M.C. Wintersgill, and J.D. Brown, J. Appl. Phys. 50, 5449 (1979).

15. J.J. Martin and S.P. Doherty, Proceedings of the 34th Annual Symposium on Frequency Control, U.S. Army Electronics Command, Fort Monmouth, NJ, pp. 81-84 (1980). Copies available from Electronic Industries Association, 2001 Eye Street, N.W., Washington, D.C. 20006.

16. D.B. Fraser, Physical Acoustics, (W.P. Mason, ed.) Vol. V, Chap. 2, Academic Press, New York (1968).

17. A.W. Warner, Bell System Technical J. 40, 1193 (1960).

18. J.M. Stevels and J. Volger, Philips Res. Repts. 17, 284 (1962).

19. A.L. Taylor and G.W. Farnell, Can. J. Phys. 42, 595 (1964).

Acknowledgements

This work was supported by the U.S. Air Force, monitored by A. Armington, RADC, Hanscom AFB, MA 01731.

Conclusions

SAW filter technology is mature with low insertion loss, good out of band rejection, very linear phase and a wide frequency range from 20MHz to 2 GHz. The devices are being utilized in many systems applications from T.V. tuners to pulse compression radars to cable television repeaters.

One challenge still left in surface acoustic wave filter technology is that of finding a high coupling, zero temperature coefficient material that may be produced in large quantities. The material also should have no beam steering problems in the propagation direction. Of course other challenges also exist such as higher frequency (>2 GHz) devices.

Acknowledgements

The authors would like to thank the many engineers and technicians that have contributed to the development of SAW filter technology over the past 10 years. Especially Gay Spence, for being involved in the latest designs and Charlie Shaffer and Bob Vess for working the difficult fabrication problems require special thanks. Thanks also to June Angelino for typing the manuscript.

References

1. B. R. Potter, "L-band Low Loss SAW Filters," 1979 IEEE Ultrasonics Symposium Proceedings, 79CH1482-9SU, pp 533-536.

2. B. R. Potter and C. S. Hartmann, "Surface Acoustic Wave Slanted Device Technology," IEEE Trans. on Son. and Ultras., Vol. SU-26, No. 6, Nov. 1979, pp 411-418.

3. B. R. Potter and C. S. Hartmann, "Low Loss Surface Acoustic Wave Filters," IEEE transactions on Parts, Hybrids and Packaging, December, 1977 pp 348-353.

4. K. Yamanouchi, F. Nyffeler, and K. Shibayama, "Low Insertion Loss Acoustic Surface Wave Filter using Group-Type Unidirectional Interdigital Transducers". 1975 Ultrasonics Symp. Proc., IEEE Cat. No. 75CH0994-4SU, pp. 317-321.

5. D. B. Mac Donald, C. F. Shaffer, T. G. Blocker III and R. C. Vail, "Development of a two-step E-beam Lithography Process for Submicron Surface Acoustic Wave (SAW) Device Fabrication", Optical Engineering, Vol. 18, 59 (1979).

6. R. M. Hays and C. S. Hartmann, "Surface-Acoustic-Wave Devices for Communications", Proc. IEEE, Vol. 64 No. 5, pp 652-670, May 1976.

7. B. R. Potter and T. L. Shoquist, "Multipassband Low Loss SAW Filters", 1977 IEEE Ultrasonics Symp. Proc., Cat. No. 77CH12641-SU, Oct, 1977, pp 736-739.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

RADIATION-INDUCED CONDUCTIVITY AND HIGH TEMPERATURE Q CHANGES IN QUARTZ RESONATORS

D. R. Koehler
Quartz Crystal Oscillator & RTG Development Division, 2531

Sandia National Laboratories
Albuquerque, NM 87185

Summary

While high temperature electrolysis has proven beneficial as a technique to remove interstitial impurities from quartz, reliable indices to measure the efficacy of such a processing step are still under development. The present work is directed toward providing such an index. Two techniques have been investigated--one involves measurement of the radiation-induced conductivity in quartz along the optic axis, and the second involves measurement of high temperature Q changes. Both effects originate when impurity charge compensators are released from their traps, in the first case resulting in an associated increase in ionic conduction and in the second case resulting in increased acoustic losses.

Radiation-induced conductivity measurements have been carried out with a 200 kV, 14 mA X-ray machine producing approximately 5 rads/sec at the sample. With electric fields of the order of 10^4 V/cm, the noise level in the current measuring system is equivalent to an ionic current generated by quartz impurities in the 1 ppb range. The accuracy of the high temperature (300-800 K) Q^{-1} measurement technique is limited by the uncertainties associated with quantitative correlation of the high temperature acoustic losses with the concentration of impurity centers. A number of resonators constructed of quartz material of different impurity contents have been tested, and both the radiation-induced conductivity and the high temperature Q^{-1} results compared with earlier radiation-induced frequency and resonator resistance changes.[1] A postirradiation-induced conductivity index and a high temperature Q index show excellent correlation with the earlier pulsed irradiation-induced dynamic resonator motional resistance changes, and it is therefore concluded that either measurement can be employed to serve as an acceptance criterion for radiation hardness.

Introduction

High temperature electrolysis, as a technique to remove interstitial impurity cations from quartz, has been used successfully for some years.[2] As far as quartz resonator sensitivity to radiation is concerned, electrolysis (sweeping) has proved beneficial because the elastic modulus and acoustic loss changes caused by ionizing radiation have been minimized or, in many cases, reduced to negligible levels. The electrolysis process, however, has not been sufficiently reproducible that the degree of cation removal can be guaranteed. Some index is necessary, therefore, to assure that the quartz is indeed relatively devoid of mobile cations. During the sweeping process, the current through the apparatus has not proved to be an accurate indicator, apparently because parasitic currents, other than the impurity cation current in the crystal, are also measured. Markes and Halliburton[3] have recently suggested an ESR technique, where hole-compensated Al centers are detected at various stages in an irradiation and temperature cycling sequence as a quality assurance methodology.

In earlier studies by Hughes[4] of radiation-induced conductivity (RIC) in quartz, observations were made of both electronic and ionic charge carriers wherein the ionic carriers were characterized as impurity cations freed indirectly by the radiation and moved by the electric field along the optic axis of the crystal. The reduced magnitude of these ion currents in swept Z-cut quartz and the absence of such currents in X- and Y-cuts, stemming from the anisotropy of quartz and the consequent inability of ions to move perceptibly except along the Z-axis, led to the suggestion of utilizing prompt radiation-induced conductivity, following pulsed irradiation, as an indicator of the presence or absence of ionic impurities. This suggestion has stimulated the RIC part of the present effort which is aimed at developing a quartz purity index appropriate from a radiation hardness perspective.

The very large increases in acoustic losses observed in natural and synthetic quartz at high temperatures had been studied as early as 1953 by Cook and Breckenridge[5] and the work up until 1968 discussed by Fraser.[6] These resistance effects can be explained by a mechanical relaxation phenomenon where the associated thermally activated high temperature loss is attributed to the stress induced motion of impurity ions along the Z-axis channels. The activation energy is the sum of the energy to free the cations from their trapping sites and the motional energy to move along the optic axis potential barriers. Electrolysis has been shown to reduce this loss peak in both natural and synthetic material,[2] thus making possible high temperature frequency control applications of quartz crystals. It should be recognized, then, that the high temperature acoustic loss increase is another phenomenological manifestation of the freeing of impurity ions. At high temperatures, the energy to release the

ions from their traps is provided thermally, whereas in the case of radiation-induced conductivity. the energy is provided by the ionizing radiation. In these radiation-induced conductivity processes or in radiation-induced acoustic loss processes, since the energy to release the cations from their traps is provided by the radiation, the observed activation energy is simply the motional energy.[7] If the absence of the high temperature loss peak in quartz can, therefore, be interpreted as an indication of the absence of the Al-associated impurity ions, then the magnitude of this high temperature loss is another measure of the purity of the material.

Experiments

For the radiation-induced conductivity measurements, a 200 kV, 14 mA X-ray machine was used to irradiate the quartz crystal samples which were in the form of finished resonators. The resonators were placed in a vacuum chamber to avoid air ionization currents, and doubly shielded cabling was used from the high voltage supply and current meter to the feedthroughs on the vacuum chamber, thereby minimizing stray currents throughout the experimental setup. A block diagram of the experimental apparatus is shown in Figure 1. The currents were recorded on a strip chart recorder to provide rise and decay time information after radiation or electric field changes at the sample. All of the measurements reported here were made at room temperature. The use of a 200 kV X-ray source assures a fairly uniform radiation deposition and in the present experimental configuration resulted in a fluence of approximately 3-5 rads/sec at the sample. No voltage magnitude or polarity dependence of the background current was observed up to 2000 volts. Furthermore, there appeared to be no measurable X-ray-induced current since the background was constant [~ $3(10)^{-11}$ amperes/cm^2] at all voltages and X-ray fluences.

The high temperature resonator-resistance (Q^{-1}) measurements were done in an absorption network with a gain-phase meter measuring the voltage and phase of the signals at the input and output of the network. A diagram of the apparatus is shown in Figure 2. Temperature measurements were taken with a thermocouple in close proximity to the crystal resonator in the furnace chamber. Resonator fabrication technology limited the highest temperature (~ 400°C) at which the resonators could be tested.

For the most part, the quartz crystal resonator specimens were 5 MHz, fifth overtone AT-cuts of Premium Q* quartz material. The Y-cut units and the natural quartz specimens are obvious exceptions. The swept quartz samples were fabricated from material that was vacuum electrolyzed at Sandia National Laboratories Albuquerque. These resonators were used in an earlier radiation study program where frequency transient data was taken after exposure to pulsed gamma irradiation (pulse width approximately 70 nanoseconds) at 10^4 rads. In that program, both frequency and resonator resistance (Q^{-1}) changes were measured. Unswept or incompletely swept quartz manifested itself as showing transient and permanent resistance (Q^{-1}) changes, with associated frequency changes, while well-swept material showed only frequency transients which were characterized and modeled as thermal in origin.[8,9] In the present studies, correlation was, therefore, looked for relative to this earlier frequency and resistance change (Q^{-1}) characterization, as opposed to the less definitive "electrolyzed or nonelectrolyzed" characterization. Although all of the unswept material showed transient resistance changes, not all of the swept material displayed the absence of resistance changes. The Y-cut samples were not part of the earlier pulsed irradiation testing, and, therefore, no information is available as to ionic purity or resistance-change effects.

*Premium Q quartz produced by Sawyer Research Products, Inc., Eastlake, Ohio.

Results and Discussion

Radiation-Induced Conductivity

The steady state radiation-induced current that results for a constant dose of X-rays is

$$I(\text{amps}) = \alpha e f_1 f_2 A \phi d \qquad (1)$$

where

α = concentration of carriers per rad,

e = electronic charge,

f_1 = number of impurity ions released per carrier ≈ 1,

f_2 = field dependence of ion collection

$$= \frac{\mu\tau E}{d}\left[1 + \frac{\mu\tau E}{d}\exp(-d/\mu\tau E) - \frac{\mu\tau E}{d}\right] \qquad (2)$$

[assuming a single carrier and that the carrier concentration can be represented in the exponentially time-decreasing form $\exp(-t/\tau)$],

A = area of the electrodes in cm^2,

ϕ = radiation rate in rads/sec,

d = sample thickness in cm,

μ = carrier mobility in cm^2/V-sec,

τ = carrier lifetime in seconds, and

E = applied field in V/cm.

Hughes[4] has determined that $\alpha \approx 1\text{-}4(10)^{12}$ carriers/cm^3/rad, therefore, at 5 rads/sec, we are freeing $5\text{-}20(10)^{12}$ ions/cm^3/sec, which is a small

fraction of the total number of impurities present in most quartz. For example, $20(10)^{12}$ ions/cm^3 is equivalent to $7.5(10)^{-4}$ ppm of impurity ions. If the effects of impurities reduction due to electrolysis or high purity growth are to be detectable as a reduction in the RIC, however, this extremely low impurity level must be present in the quartz. In other words, the radiation-induced impurity ion release rate, $\alpha f_1 \phi$, will show a decrease only when this rate is greater than the impurity concentration. When this occurs, the impurity concentration is the RIC limiting factor and the number of impurity ions released per carrier, f_1, will be less than unity. Of course, higher impurity levels could be detected as a reduction in the RIC if higher irradiation rates were used. This conclusion assumes the mobility and the lifetime are unchanged by sweeping, but if the trap levels and numbers change with sweeping time (as, for example, if hydrogen ions or holes compensate the Al sites), then both mobility and lifetime could change, and lower radiation induced currents could still ensue. Lifetime and mobility changes affect the field dependence factor, f_2, in Equation (1). Aside from the work of Hughes, and more recently Nowick and Jain,[10] little has been done in the measurement of ion mobilities and lifetimes in quartz.

Figure 3 displays a typical time record of the current in a resonator during an irradiation sequence. The magnitudes of the steady state RIC and the postradiation conductivity vary, of course, from resonator to resonator. Figure 4 is a plot of the steady state radiation-induced currents for the various crystals tested. Note the units of the ordinate where the dependence on radiation rate, applied electric field, and electrode area has been normalized. This is particularly necessary due to differences in resonator thicknesses and electrode areas. The observed currents for the electronic grade, Y-cut, 2 MHz resonators could not be distinguished above background. This result is in agreement with Hughes' work and consistent with the model of ionic conductivity in quartz wherein the motion is restricted to the open Z-axis channels.

The figure shows the results for the two groups of resonators according to the earlier mentioned characterization, namely resonators showing thermal transient effects only (TT units) and resonators showing resistance changes (RT units) under pulsed irradiation at 10^4 rads; the Y-cut units were not part of the earlier radiation-effects characterization study. For purposes of comparison with expected electron currents in quartz, the values of electron radiation-induced conductivities from Hughes[4] work is displayed in the figure also. Radiation conditioning, that is, preirradiation at high dose levels, has been shown[1] earlier to improve the response of resonators to radiation, and in this work it causes an apparent reduction in the RIC; the lowest observed conductivities for the unswept resonators occurred in those units that had been radiation preconditioned.

While the RT units displayed the highest conductivities observed and while the TT units displayed the lowest conductivities observed, there is a significant overlap on the conductivity scale of the two distributions. It would be, therefore, difficult to specify an acceptance criterion using this steady state RIC index. Furthermore, the fact that the well-swept quartz resonators did not show any appreciable conductivity difference from the unswept or partially swept resonators suggests that the impurity concentration in the well-swept units was still greater than the radiation-induced impurity ion release rate.

Although the steady state RIC results do not lead to an acceptable index, the postirradiation transient RIC results, shown in Figure 5, seem to clearly establish this parameter as a viable index. The ordinate in Figure 5 is the conductivity measured in the first second after termination of the x-irradiation. A conductivity of 10^{-18} amperes/cm^2/volt/cm/rad represents the detection limit of the experimental apparatus. Because the postirradiation conductivity decay times varied significantly between resonators (from seconds to hours), the normalization to total radiation exposure is not completely satisfactory since irradiations took place over several minutes. Postirradiation conductivity measurements following a pulsed irradiation, with a pulse width less than a second, would be superior in this respect. In relating to the marked difference in results between the steady state RIC and the postirradiation RIC, it would appear that the greater sensitivity of the postirradiation conductivity to the ion lifetime constitutes an important difference. If this is the responsible difference factor, then the ion lifetimes in quartz are being significantly reduced during electrolysis. For crystal unit A, a pulsed irradiation resistance transient was observed, but it was sufficiently fast and small that no effect was detectable in the frequency transient. It is concluded that this quartz material does possess some residual ionic impurity content, as confirmed by the postirradiation conductivity measurement.

High Temperature Resistance

Resonator resistances as a function of temperature are shown in Figures 6 and 7. Because of a lower temperature fabrication technology, the Y-cut resonators were not able to be tested. Here again the correlation follows the RT and TT characterization, and the results with the RT units are shown in Figure 6 while the results for the TT units are shown in Figure 7. The magnitude of the Q change for the RT units was sufficiently large at the higher temperatures that a zero phase condition did not ensue. This can be shown to be the case whenever $R_1 > 1/(2\ \omega_p C_S)$. Under this circumstance, however, one can use the peak phase point to calculate the resistance from the equation

$$R_1 = \left[(\tan^2 \theta + 1)^{1/2} - \tan \theta \right] / 2\omega_p C_S \qquad (3)$$

where

$\tan \theta = \sin \phi / (\cos \phi - V_B/V_A)$,

V_A = voltage into the absorption network,

V_B = voltage at the resonator,

ϕ = phase of V_B relative to V_A,

ω_p = angular frequency at maximum phase ϕ (peak phase point), and

C_S = total capacitance at the resonator node at which V_B is measured.

The degree of improvement in high temperature performance, that is, the absence of any marked resistance increase for the electrolyzed (well-swept) quartz resonators is apparent and should be compared with earlier results of Fraser[11] and King and Fraser[12] as displayed in Figures 8 and 9. Although not as prominent as in Fraser's results, there appears to be a similar secondary peak at approximately 300°-320°C for the RT units. Here, as in the RIC results, crystal unit A displays evidence of a residual impurity content in that there is a measurable high temperature increase in resonator resistance.

Perhaps a better high temperature comparison point is nearer 500°C than 400°C because it would be nearer the loss peak; the restrictions of the present experiments (i.e., the use of finished resonators) limited us, however, to the lower temperature. We have formed an index ratio, $(R_{400°C} - R_{25°C})/100$, for the tested resonators which is shown in Figure 10. The denominator, or reference value of 100 ohms in this ratio, is considered to represent the quartz internal frictional loss at 5 MHz. The motional capacitance of these resonators is approximately 10^{-4} pF. The clear separation of the resistance-ratio distributions for the two groups (RT and TT) indicates the viability of this measurement index as an acceptance criterion.

Conclusions

Based on the radiation effects model of cations freed from substitutional Al sites in quartz, two indices for determining the ion impurity levels have been suggested. The measurement techniques have been investigated and demonstrated to be viable radiation-hardness indicators. The postradiation-induced conductivity index and the high temperature Q index show excellent correlation with dynamic resonator motional resistance changes under irradiation. It is, therefore, concluded that either measurement could be employed to serve as an acceptance criterion for radiation hardness. Earlier[1] resonator radiation performance variability, from blanks taken out of the same quartz bar, suggests the need to utilize the acceptance criterion with individual resonators. Furthermore, to preclude the loss of production costs incurred in packaging the resonators, it is recommended that testing, with appropriate fixturing, be performed on the crystal blanks prior to resonator packaging.

Acknowledgment

The author wishes to acknowledge the invaluable assistance of Charles Joerg with the high temperature resonator Q measurements and of John Jewell with the x-irradiation conductivity experiments.

References

1. Young, T. J., Koehler, D. R., and Adams, R. A., "Radiation-Induced Frequency and Resistance Changes in Electrolyzed High Purity Quartz," Proceedings of the 32nd Annual Frequency Control Symposium, pp. 34-42 (1978).

2. King, J. C., "Electrolysis of Synthetic Quartz: Effect Upon Resonator Performance," Proceedings of the I.E.E. 109, pp. 1-7 (1962).

3. Markes, M. E. and Halliburton, L. E., "Defects in Synthetic Quartz: Radiation-Induced Mobility of Interstitial Ions," Journal of Appl. Phys. 50, pp. 8172-8180 (1979).

4. Hughes, R. C., "Electronic and Ionic Charge Carriers in Irradiated Single Crystal and Fused Quartz," Radiation Effects 26, pp. 225-235 (1975).

5. Cook, R. K. and Breckenridge, R. G., "Anelasticity of Quartz," Phys. Rev. 92, pp. 1419-1423 (1953).

6. Fraser, D. B., "Impurities and Anelasticity in Crystalline Quartz," in Physical Acoustics, Vol. V, W. P. Mason, ed, (New York: Academic Press, 1968), pp. 59-110.

7. King, J. C. and Sander, H. H., "Transient Changes in Quartz Resonators Following Exposure to Pulse Ionization," Radiation Effects 26, pp 203-212 (1975).

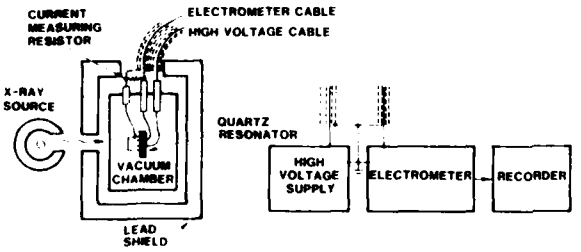
8. Koehler, D. R., Young, T. J., and Adams, R. A., "Radiation-Induced Transient Thermal Effects in 5 MHz AT-Cut Quartz Resonators," Ultrasonics Symposium Proceedings 77 CH 1264 ISU, pp. 877-881 (1977).

9. Koehler, D. R., "Radiation-Induced Frequency Transients in AT-, BT-, and SC-Cut Quartz Resonators," Proceedings of the 33rd Annual Frequency Control Symposium, pp. 118-121 (1979).

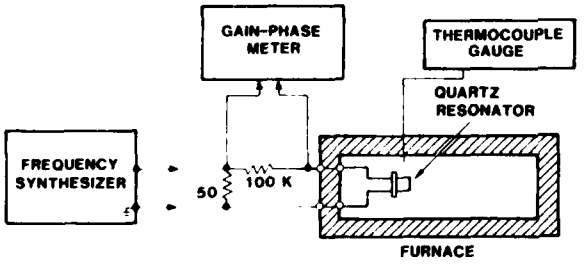
10. Nowick, A. S. and Jain, H., "Electrical Conductivity and Dielectric Loss of Quartz Crystals Before and After Irradiation," Proceedings of the 34th Annual Frequency Control Symposium, pp. 9-13 (1980).

11. Fraser, D. B., First Interim Report, DA36-039 SC-85364 (Whippany, New Jersey: Bell Telephone Lab., Inc., 1960).

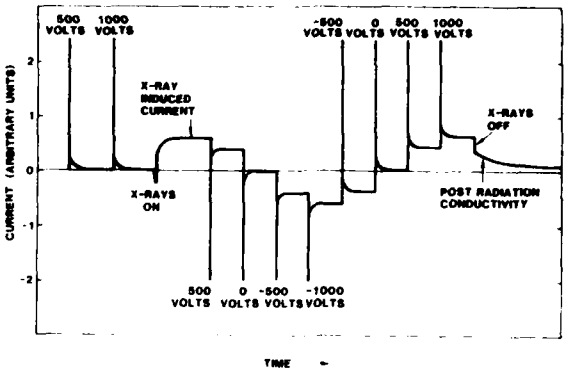
12. King, J. C. and Fraser, D. B., "Performance of Quartz Resonators Near the Alpha-Beta Inversion Point," Proceedings of the 15th Annual Frequency Control Symposium, pp. 2-21 (1961).



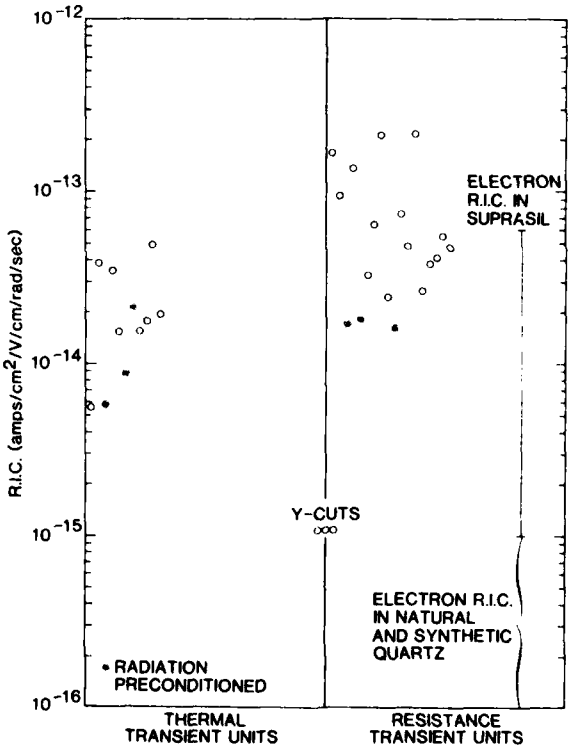
1. Diagram of the experimental apparatus for measuring radiation-induced conductivity.



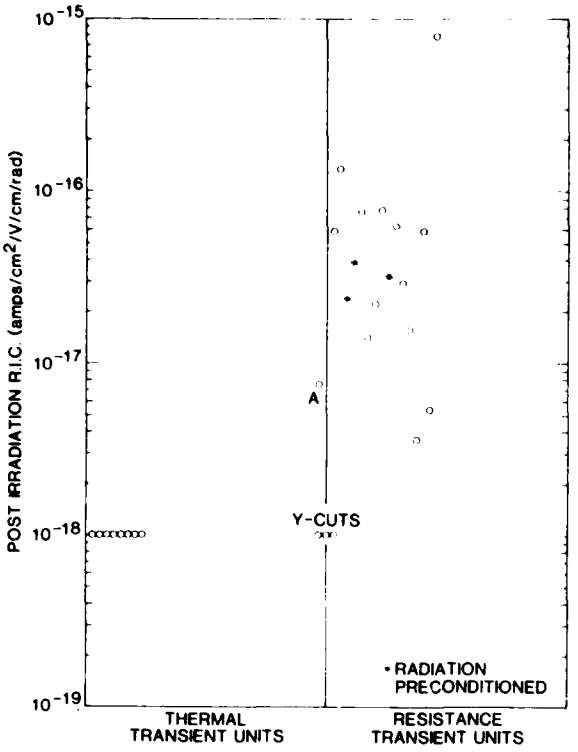
2. Diagram of the experimental apparatus for measuring high temperature resonator resistance.



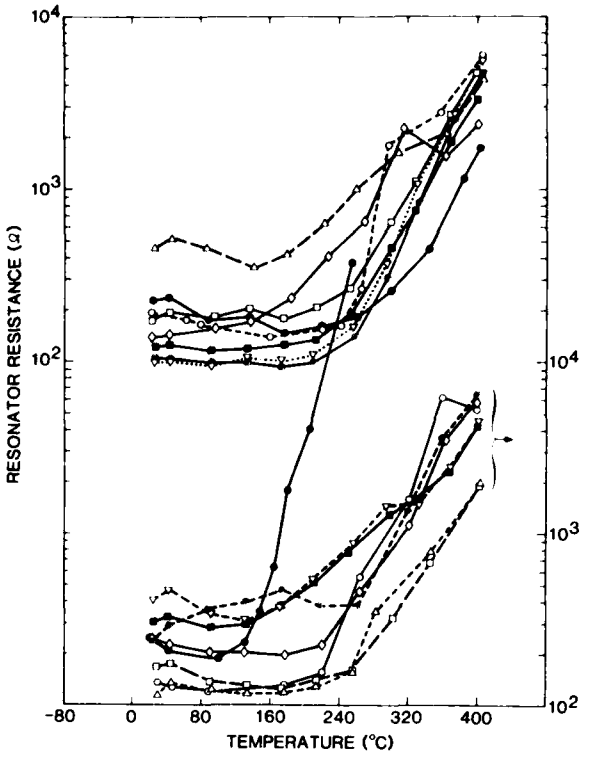
3. Resonator current during x-irradiation. The magnitude of the current during irradiation varied from resonator to resonator. The post-irradiation conductivity was not measurable in well-swept quartz.



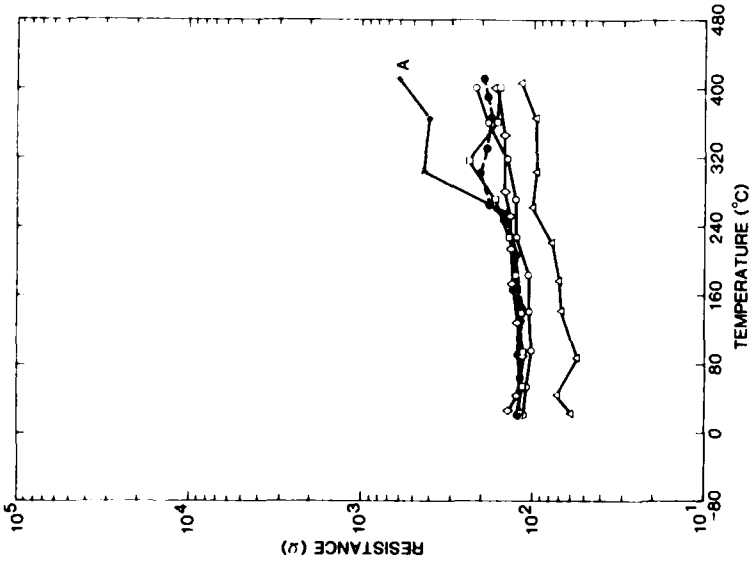
4. Radiation-induced conductivity for well-swept versus unswept and partially swept resonators. Only the crystal sample number is connoted by position along the abscissa. The Y-cut samples were not part of the earlier pulsed irradiation testing and, therefore, no information is available as to ionic purity or resistance change effects.



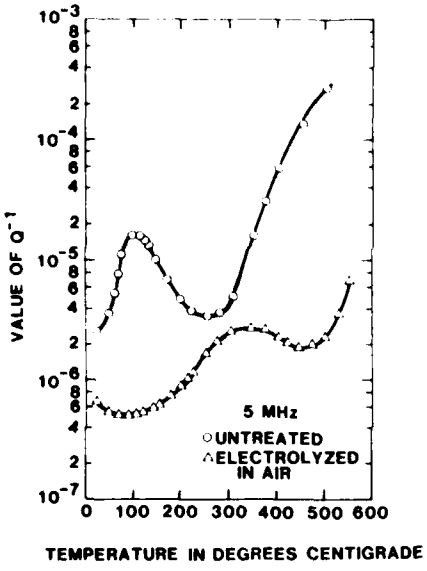
5. Postirradiation conductivity for well-swept versus unswept and partially swept resonators. Only the crystal sample number is connoted by position along the abscissa.



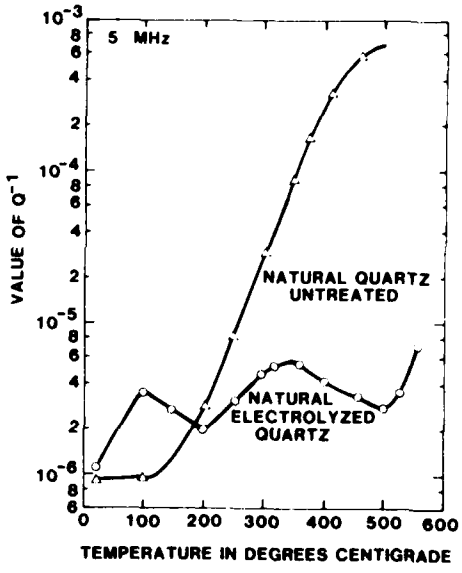
6. Temperature dependence of resonator resistance for those resonators which displayed resistance transients (RT) under pulse irradiation. These samples were of either unswept quartz or partially swept quartz. Separation of the data into the upper and lower groups is for visual clarity. The closed circle data in the bottom group is for a natural quartz resonator.



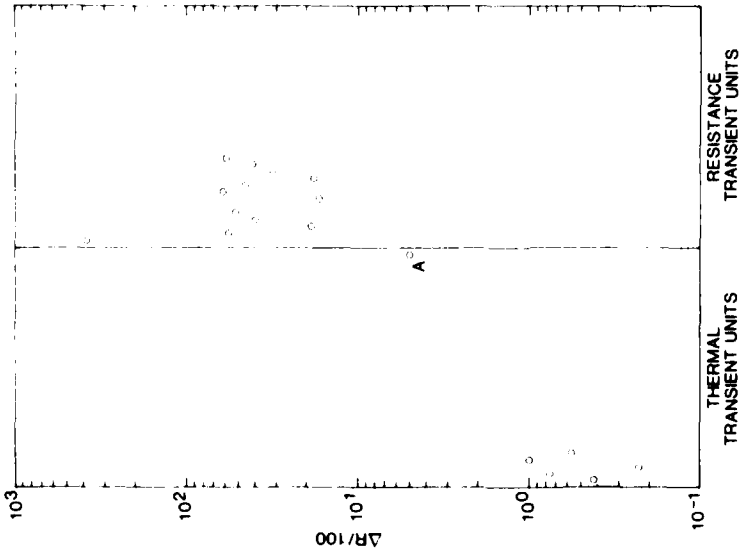
7. Temperature dependence of resonator resistance for thermal transient (TT) samples, that is, of well-swept quartz.



8. Internal friction in untreated and electrolyzed fast grown Z-growth synthetic quartz.



9. Effect of electrolysis on the acoustic absorption in natural quartz at elevated temperatures.



10. High temperature Q-1 index for well-swept quartz versus partially swept and unswept quartz. Only the crystal sample number is connoted by position along the abscissa.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

HIGH TEMPERATURE RESONANCE LOSS AND INFRARED CHARACTERIZATION OF QUARTZ

Herbert G. Lipson, Alfred Kahan, Richard N. Brown and Ferdinand K. Euler

Rome Air Development Center
Deputy for Electronic Technology
Hanscom AFB, MA 01731

Summary

We report the results of electrical resistance at series resonance as a measure of acoustic loss between 150 and 350 °C for resonators fabricated from high grade synthetic and electrolyzed (swept) quartz. An exponential increase of resonator resistance with temperature is observed for all as-grown quartz. This increase is attributed to either lithium or sodium impurity migration. The swept companions of these materials do not show the exponential rise and no essential difference is found between air and vacuum swept material.

Strong Al-OH^- related bands are observed by infrared spectroscopy in both air and vacuum swept samples. From short-term vacuum sweeping experiments we find that a reduction in as-grown OH^--band strengths occurs almost immediately after the onset of sweeping, but the Al-OH^- bands do not form until after a few hours of sweeping time. In air swept samples, with hydrogen supplied by the water vapor in the atmosphere, Al-OH^- bands appear almost immediately and as-grown OH^--bands remain essentially unchanged.

Key words. Quartz, Resonators, Acoustic Loss, Infrared Spectra, Sweeping, Impurity Bands, Defect Centers.

Introduction

It is desirable to have sensitive experimental characterization techniques available to monitor impurity concentration and intrinsic defects in high quality quartz, their transformation and dissociation as a function of electrolysis (sweeping) or irradiation, and to be able to correlate these results with observed device performance. It is also important to develop comparative material evaluation techniques to be used for manufacturing quality control of high precision resonators.

The impurity and intrinsic defect structure of quartz and their characterization by infrared spectroscopy, anelastic loss as a function of temperature, $Q^{-1}(T)$, and electron spin resonance has been reviewed by Halliburton[1] and by Nowick.[2] The aluminum-metal impurity center, Al-M^+, shows a loss peak near 50 K for Al-Na^+, but the low temperature loss pattern of Al-Li^+, Al-K^+, Al-OH^-, and aluminum-hole, Al-e^+, are not clearly identified. Al-OH^-, as well as other hydrogen related bands, are well characterized by low temperature near infrared absorption spectroscopy, whereas Al-e^+ is best studied by ESR. The dominant impurity center in high grade natural and synthetic quartz, Al-Li^+, is not observable by ESR, nor has it been detected by optical techniques unless associated with OH^- ions.[3]

Previous investigations by King[4] and by Fraser[5,6] have shown that $Q^{-1}(T)$ exhibits a high temperature loss behavior which increases exponentially with temperature above 200 °C. The exponential increase reflects the stress induced alkali ion migration, and it is at least three orders of magnitude more sensitive to impurity concentration than the alkali motion between equivalent sites, giving rise to the low temperature Q^{-1} peaks. In addition, a loss peak observed for swept natural and synthetic quartz near 350 °C has been attributed by Fraser to diffusion of alkali ions.

Sweeping synthetic α-quartz is a well-established method for improving quartz resonator performance, but many questions still remain as to the defect mechanisms contributing to this process. In samples swept in an air atmosphere, the dominant mechanism is the replacement of compensating alkali metal ions M^+ (M^+ = Li^+, Na^+, or K^+) by hydrogen. This ion bonds to an oxygen to form OH^- and it provides the compensation for the substitutional aluminum at the silicon sites. Sweeping quartz in vacuum also results in improved material and device properties.[7,8] It is expected that the lack of available hydrogen during vacuum sweeping would inhibit the formation of Al-OH^-, diminish the strengths of as-grown hydrogen related bands, and form Al-e^+ as the prominent defect.

In this paper we discuss characterization techniques for high grade synthetic quartz, and how to differentiate between Al-Li^+ and other

$Al\text{-}M^+$ centers. In addition, a method is needed to distinguish between vacuum and air swept quartz and to test the completeness of sweeping. We report high temperature electrical resistance results for resonators fabricated from state-of-the-art quartz, both as-grown and swept by different techniques, as well as ESR and infrared spectra of these materials.

Experimental Procedures

The Electronic Grade material used in these investigations was obtained from Motorola. Premium-Q and High-Q quartz were grown at Sawyer Research Products (SARP) by the same process, except that for Premium-Q, lithium salt was added to the mineralizer. The aluminum impurity content of High-Q quartz, 2×10^{17} Al/cm^3, as determined from infrared and electron spin resonance spectroscopy, is approximately one order of magnitude greater than that of our Premium-Q samples, 2×10^{16} Al/cm^3.

Sweeping experiments were performed in air, nitrogen, and vacuum ambients. The experimental procedures for sweeping were described in a previous publication.[9] The alkali ion diffused materials were prepared by first sweeping the as-grown Premium-Q quartz in a nitrogen atmosphere, coating the surface with the appropriate alkali carbonate, and electrodiffusing it in a nitrogen atmosphere. Precision 5 MHz, 5th overtone, AT-cut resonators were fabricated from as-grown, swept, and ion diffused materials by Frequency Electronics, Inc. (FEI).

For vacuum sweeping, the furnace and sample holder were enclosed in a bell jar and pumped to approximately 5×10^{-5} Torr. For short-term sweeping experiments the sample was clamped between platinum foils, and for long-term sweeping platinum layers were also sputtered on the sample. The sample was gradually brought up to 500 °C over a two day period and swept at this temperature for a prescribed time with an electric field of 1000 V/cm applied along the crystallographic z-axis. The electric field was then turned off, and the sample cooled to room temperature over a period of two days.

Most of the samples used for comparing effects of sweeping in air and in vacuum by infrared spectroscopy were prepared from the same as-grown quartz bar. Sample preparation and infrared transmission procedures were described in a previous publication.[9] For these measurements transmissions were obtained between 3100-3700 cm^{-1} at 85 K using unpolarized radiation with the electric vector $E \perp z$.

The high temperature electrical resistance measurements on resonators were performed with a frequency synthesizer, vector impedance meter, and x-y recorder. The resonator was enclosed in a small furnace, and the temperature was set with a Eurotherm controller capable of 0.2 °C stability. Between 25 and 200 °C, the temperature was increased at a typical rate of 0.5 °C/min. At higher temperatures, owing to the rapid change of frequency as a function of temperature, a much slower rate was required to obtain a measureable resonance. The minimum impedance, corresponding to the resonance peak, was recorded at approximately 1 °C intervals. Due to the lead capacitance, the off-resonance circuit impedance Z_o is approximately 2600 ohms. Resonance resistance R_1 and measured minimum impedance Z_m differ by less than 10% up to 1000 ohms. For higher Z_m values, the resistance is calculated from

$$R_1 = (Z_m^{-2} - Z_o^{-2})^{-1/2}.$$

Results and Discussion

High Temperature Resistance

Figure 1 shows the electrical resistance R_1 at series resonance between 150 and 350 °C measured on resonators fabricated from as-grown and dried nitrogen atmosphere swept electronic grade quartz. The data is consistent with results obtained independently by King[4] and by Fraser[5,6] on natural and electronic grade quartz. These investigators have shown that the exponential rise above 200 °C is associated with alkali impurities, lithium in natural crystals and sodium in synthetic material. The electronic grade quartz used in our investigations was grown with a lithium additive, and resonators fabricated from this material do not show the sodium related 50 K loss peak. We therefore assign the exponential high temperature loss to the lithium impurity.

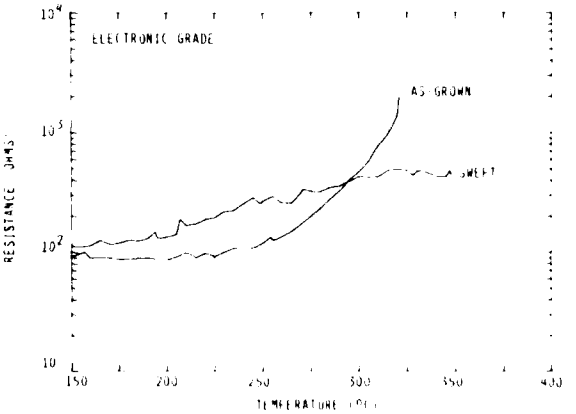


Figure 1. High temperature electrical resistance of resonators fabricated from as-grown and nitrogen swept Electronic Grade quartz.

Sweeping natural or electronic grade quartz in air reduces or removes the exponential loss, and introduces a peak, centered around 350 °C. This reduction in loss corresponds to the removal of the sodium or lithium ions, while the peak has been attributed to the diffusion of the alkali impurity. The upper temperature range of our measurements is limited to 350 °C by the physical survivability of the high precision, 5 MHz, 5th overtone, AT-cut resonators. In our measurements, the 350 °C peak is in evidence but it is not completely defined. The investigations of King and of Fraser, carried out on quartz disks in a ceramic gap holder, were extended to higher temperatures and clearly show this peak.

Currently, the most widely utilized material for high precision resonator applications is SARP grown and swept Premium-Q quartz. These crystals are grown with a lithium additive. The low and room temperature acoustic loss properties of this material, which shows no identifiable loss peaks between 4 and 400 K, have been reported extensively.[10,11] Figure 2 shows the high temperature resonance resistance of as-grown, vaccum, and air atmosphere swept Premium-Q quartz. In the as-grown state, it is similar to lithium additive grown electronic grade material, with the exponential rise above 200 °C attributed to the motion of the lithium impurity.

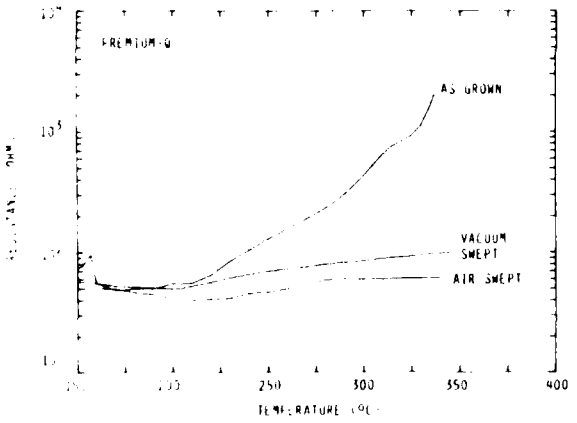


Figure 2. High temperature electrical resistance of resonators fabricated from as-grown, vacuum swept, and air swept Premium-Q quartz.

Sweeping Premium-Q quartz either in air or in vacuum removes the lithium impurity and correspondingly eliminates the exponential rise. However, in neither case do we observe a 350 °C loss peak. The suggestion that the mechanism for this peak is alkali ion diffusion is not substantiated by our data. Infrared data (Figure 5) indicates that Al-OH^- as well as other OH^--related bands are present in both air and vacuum swept samples. The absence of the 350°C peak in these materials tends to rule out OH^--ion as the active defect. Additional experiments are needed to clarify the mechanism, or identify the impurity causing this loss.

Figure 3 shows the high temperature electrical resistance of resonators fabricated from as-grown and air swept High-Q quartz. Similar to all other as-grown or natural quartz, High-Q quartz exhibits an exponential rise in resistance. As established by King[12], the addition of a lithium salt during crystal growth limits the incorporation of sodium into the crystal. As expected, resonators fabricated from this High-Q material grown without a lithium salt additive show a large Al-Na^+ loss peak, 2.5×10^{-5}, at 50 K.[13] The activation energies for the high temperature exponential associated with Li, Na, and K were calculated by Fraser[5] as 0.72, 0.77, and 0.92 eV, respectively. The closeness of these values makes it difficult to assign the high temperature loss to a particular alkali ion. The combination of high and low temperature measurements can differentiate between sodium and lithium, and indicates that in High-Q quartz the active compensating ion is sodium.

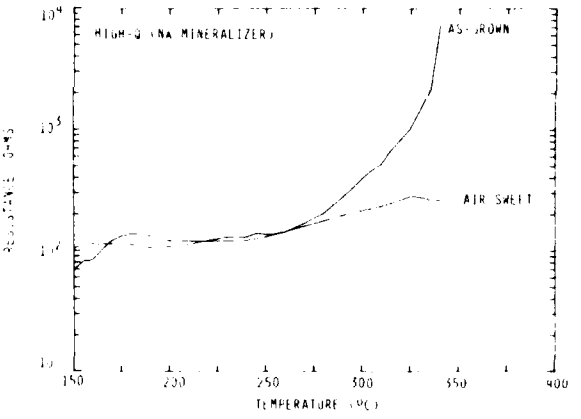


Figure 3. High temperature electrical resistance of resonators fabricated from as-grown and air swept High-Q quartz.

Figure 3 also shows that sweeping the High-Q crystal reduces the exponential loss. The air swept resonator shows a very gradual rise to about 325 °C, with a possible indication of a peak at this temperature. The data has to be extended to higher temperatures before we can reach any conclusion regarding this peak.

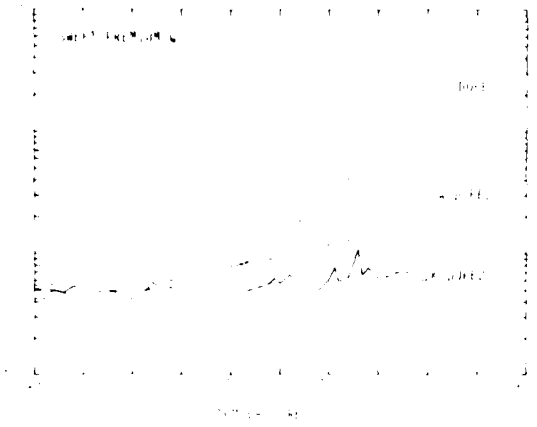


Figure 4. High temperature electrical resistance of resonators fabricated from nitrogen swept Premium-Q quartz, subsequently electrodiffused with Li-, Na-, and K-ions.

Figure 4 shows the high temperature electrical resistance of resonators fabricated from Premium-Q crystals which were swept in nitrogen and then electrodiffused with various alkali impurities. The slope of the lithium diffused resonator is similar to that of as-grown Premium-Q, but the curve is shifted towards higher temperatures indicating a lower lithium impurity concentration. For the sodium doped material, the 50 K loss peak, 1.2×10^{-5}, is approximately one half that of the as-grown High-Q crystal. This reduced sodium impurity concentration shifts the exponential rise beyond the 350 °C limit of our experimental data.

The electrical resistance of the K-ion diffused resonator shows only very slight changes up to 350 °C. This behavior has been confirmed on two additional K-ion diffused resonators. Also, low temperature acoustic loss data indicates that these resonators do not have the 210 K peak[5] which may be associated with $Al\text{-}K^{+}$. It is well known that it is very difficult to sweep potassium into synthetic quartz, owing to the potassium ionic size relative to the open channels of the quartz structure. It is possible that in our experiments either the impurity was not introduced, or that its concentration is too low for the exponential rise to be observed within the temperature range of these experiments. (The infrared spectra of the alkali diffused materials will be discussed in the next section).

We have also made preliminary acoustic loss measurements on resonator disks in an evacuated ceramic gap holder set-up similar to the type described by Fraser.[6] With this equipment we have extended the temperature range of our measurements to above 500 °C. The as-grown Premium-Q disk was was removed from the resonator and measured in this set-up. The exponential loss behavior is seen to continue to above 500 °C. Extended high temperature measurements on other disks reported on in this paper are in progress.

Infrared Spectra.

The infrared bands characteristic of as-grown and swept Premium-Q quartz are shown i the 85 K absorption spectra of Figure 5. The principal OH^{-}-bands of as-grown synthetic quartz are designated as s_1, s_2, s_3, and s_4, while those resulting from the replacement of alkali ions by hydrogen near the substitutional aluminum are labeled e_1 and e_2. The e-bands are always introduced in air swept samples with the compensating ion provided by the water vapor in the air. At the same time, no significant changes occur in the s-bands. In vacuum swept material, contrary to our expectations, we do observe a strong e_2-band, and at the same time the s-bands diminish only slightly. This s- and e-band behavior has been confirmed on other samples swept in vacuum for periods of 1 to 4 days.

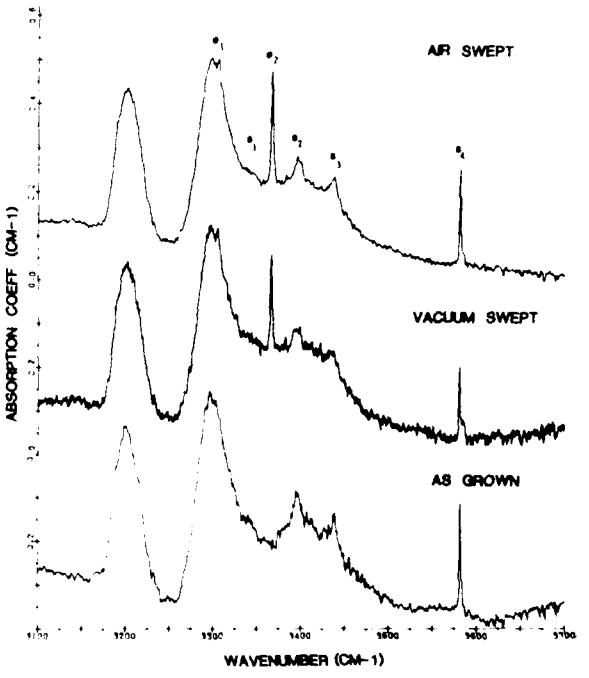


Figure 5. Absorption spectra of as-grown, vacuum swept, and air swept Premium-Q quartz measured at 85 K with $E \perp z$. $Al\text{-}OH^{-}$ impurity bands are designated e_1 and e_2, and as-grown hydroxide related bands as s_1, s_2, s_3, and s_4.

The data shown in Figure 5 demonstrates the qualitative changes in the infrared bands for as-grown, air, and vacuum swept crystals. However, it is difficult to establish a quantitative relationship between the strength of these bands before and after sweeping, owing to variation in s-band strength throughout the as-grown crystal. The variation in the $s_4(3581\ cm^{-1})$band along the y-axis of a Premium-Q bar is illustrated in Figure 6. This type of variation also occurs along the z-axis. It is not uncommon that over a 2 cm sample the s_4-band will vary by a factor of two or more.

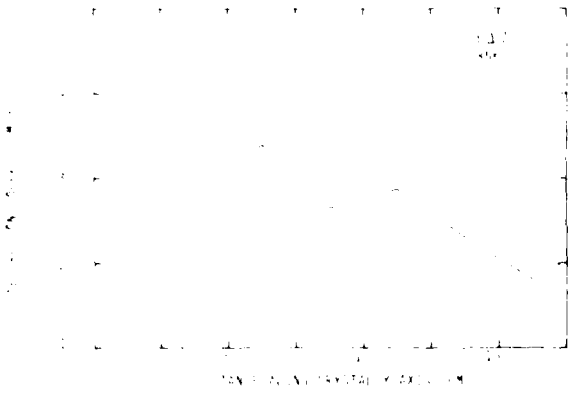


Figure 6. Variation of $s_4(3581\ cm^{-1})$ absorption band strength along the crystal y-axis for an as-grown Premium-Q quartz bar.

In discussing the high temperature electrical resistance data of resonators (Figure 4) fabricated from Na-, Li-, and K-ion diffused material, it was suggested that the potassium diffusion may have been unsuccessful. Further evidence for this is indicated from infrared mesurements on material from which the resonators were fabricated. The e_2-band strength for Na- and Li-ion diffused samples is reduced to approximately one half of its initial value. This decrease in the $Al\text{-}OH^-$ band strength implies the replacement of some hydrogen by alkali ions to form $Al\text{-}M^+$. For the K-ion diffused sample the e_2-band strength showed a much smaller percentage reduction by the diffusion process, indicating very small or negligible concentration of potassium impurity.

Kats[3] and Dodd and Fraser[14] have reported Li-OH and Na-OH related infrared bands for natural and electronic grade quartz. For Premium-Q grade quartz we do not observe these bands, neither in the as-grown nor in the alkali ion diffused state.

Vacuum Swept Material

The existence of the e-bands in vacuum swept material prompted us to extend our investigations on the formation of these bands. We have performed a series of short term vacuum sweeping experiments on a number of samples. Figure 7 shows the $e_2(3366\ cm^{-1})$ and $s_4(3581\ cm^{-1})$ bands of two samples with different initial s-band strengths as a function of sweeping time up to 12 hours. The as-grown s_4-band strength in both samples shows a small initial decrease almost immediately after onset of sweeping, and then reaches saturation. Even though most of the s_4-band strength change occurs before 3 hours of sweeping, the e_2-band of sample #2 does not appear until after this time period, while no band is present in sample #1 even after 12 hours of vacuum sweeping. Thus there is no direct correlation between the formation of the e_2-band and the decrease in the s_4-band. Possible explanations of this phenomenon may invoke a model in which the hydrogen moves to infrared inactive sites, or simply that the hydrogen concentration in the beginning of sweeping is too low to be observed in the infrared.

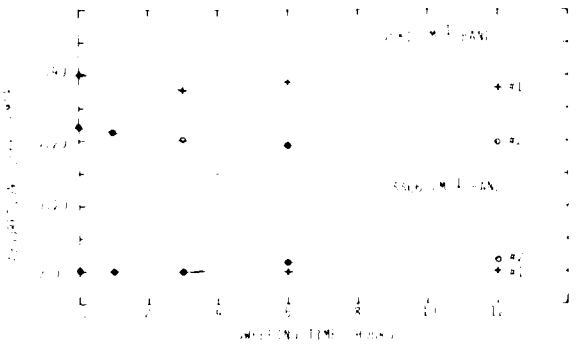


Figure 7. Changes in $s_4(3581\ cm^{-1})$ and $e_2(3366\ cm^{-1})$ absorption band strength with sweeping time for vacuum swept Premium-Q quartz samples, measured at 85 K with $E \perp z$.

Table I shows changes in e_2- and s_4-band strengths (measured at 85 K with $E \perp z$) for samples swept for different time periods in vacuum, air, and nitrogen atmospheres. Sample #3 had a relatively large e_2-band strength of 0.29 cm^{-1} after sweeping in vacuum for 24 hours. During the same time period, the s_4-band strength decreased to about 85% of its initial value. After 96 hours of sweeping, the e_2-band showed little further change, but the s_4-band decreased to about 70% of its initial value. For sample #4, swept in air, the e_2-band is observed almost immediately and reaches close to its full strength after 12 hours of sweeping. Strong e_2-bands are introduced at all sweeping temperatures in the nitrogen swept samples. Also, the rate of increase of e_2-band as a function

Table I. Influence of Sweeping Parameter on Infrared Band Strength (Premium-Q Quartz, 1000 V/cm, 500 C).

Sample #	Total Time (hours)	Sweeping Method	Band Strength Change cm^{-1} e_2 (3366 cm^{-1})	s_4 (3581 cm^{-1})
1	12	vacuum	0.00	-0.04
2	12	vacuum	0.04	-0.04
3	24	vacuum	0.29	-0.05
3	48		0.27	-0.08
3	96		0.26	-0.10
4	12	air	0.22	0.00
5	24	nitrogen	0.42	0.03
5	96		0.31	-0.07
6	168	nitrogen (400 °C)	0.46	0.06
7	168	nitrogen (550 °C)	0.33	-0.05

of sweeping time appears to vary from sample to sample for all the sweeping processes. All the vacuum swept samples show a decrease in as-grown s_4-band strength, but air or nitrogen swept samples exhibit changes in either direction.

It was anticipated that sample #1, vacuum swept for 12 hours, showing decreased as-grown hydrogen related bands and no Al-OH^- bands, might have a large concentration of Al-e^+. ESR measurements on this sample do not exhibit any resonances associated with the Al-e^+ center.[15] This raises the general question whether it is feasible to introduce Al-e^+ in α-quartz by the sweeping process. To the best of our knowledge, all the investigations on aluminum-hole centers in synthetic quartz reported on in the literature were produced either by irradiation or by vacuum sweeping above the phase transition temperature, 573 °C. Once Al-e^+ is introduced in the hexagonal phase it remains stable when the quartz is converted to the trigonal α-phase, but data taken on such a sample may not be representitive of α-quartz. In addition, the crystal twins when raised and lowered above and below the phase transition temperature and this makes it useless for resonator applications.

In summary, we were unsuccessful in obtaining clear-cut signatures of vacuum swept α-quartz with the experimental techniques of low or high temperature acoustic loss, infrared spectroscopy, and ESR. We are further investigating these effects to elucidate the defect mechanisms involved.

Acknowledgement

The authors would like to thank Mr. John O'Connor of RADC/ESM for performing some of the sweeping experiments.

References

1. L.E.Halliburton, M.E.Markes, J.J.Martin, 34th ASFC, 1(1980).

2. A.S.Nowick, H.Jain, 34th ASFC, 9(1980).

3. A.Kats, Philips Research Rep., 17, 133(1962).

4. J.C.King, Proc. IEEE, B109, Suppl. 22, 1(1962).

5. D.B.Fraser, J. Applied Phys., 35, 2913(1964).

6. D.B.Fraser, Physical Acoustics, W.P.Mason, ed., (Academic Press, NY), Vol. V, 59(1968).

7. J.R.Vig, J.W.LeBus, R.L.Filler, 31st ASFC, 131(1977).

8. T.J.Young, D.R.Koehler, R.A.Adams, 32nd ASFC, 34(1978).

9. H.G.Lipson, F.Euler, A.F.Armington, 32nd ASFC, 11(1978).

10. F.Euler, P.A.Ligor, A.Kahan, P.Pellegrini, T.M.Flanagan, T.F.Wrobel, 32nd ASFC, 24(1978).

11. H.G.Lipson, F.Euler, P.A.Ligor, 33rd ASFC, 122(1979).

12. J.C.King, A.A.Ballman, R.A.Laudise, J. Phys. Chem. Solids, 23, 1019(1962).

13. F.Euler, H.G.Lipson, P.A.Ligor, 34th ASFC, 72(1980).

14. D.M.Dodd, D.B.Fraser, J. Phys. Chem. Solids, 26, 673(1965).

15. The ESR measurements on this and other samples were performed by L.Halliburton at Oklahoma State University.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

QUARTZ CRYSTAL OSCILLATOR AT CRYOGENIC TEMPERATURE *

Bokuji Komiyama

Radio Research Laboratories
Ministry of Posts and Telecommunications
Koganei, Tokyo 184, Japan

Summary

This paper presents the experimental results on the properties of quartz crystal resonators at cryogenic temperature and their application to highly stable oscillators.

Properties of quartz resonators have been measured for two different types of crystals: 5 MHz resonators of the BVA_2 design (electrodeless design) and 5 MHz, 5th overtone, AT cut resonators of a conventional design. The Q is increased about one order of magnitude at 4.2K. No significant increase in Q factor below 4.2K has been obtained in either type of resonators. The temperature coefficient near 2K is greatly reduced compared with the value at 4.2K for one sample of both types of crystals. The BVA_2 sample shows the temperature coefficient of 3×10^{-10}/K at 2K. The amplitude frequency effect taken near 2K for the same crystals shows that the frequency versus dissipated power in crystals is very steep above 10^{-7}W.

Some results of the oscillator system stabilized with the cryogenic temperature crystal resonator are given.

Introduction

Because of their excellent short term stability, low volume and weight, low cost and low power consumption, quartz crystal oscillators are indispensable componentz for precision oscillator systems.

Atomic frequency standards generally also employ quartz oscillators as the slave oscillator so that their short-term stability is that of the crystal oscillator. In order to reach high frequency stability in as short a time as possible, high stability quartz crystal oscillators are required. Therefore the quest for improved quartz crystal oscillator stability still remains extremely important and pertinent.

To improve the problem of both short-term and long-term stability in quartz oscillators, the passive dual crystal system, in which a voltage controlled oscillator (VCO) having good short-term stability is locked to a passive quartz resonator, is suggested and the frequency stability of $\sigma_y=6\times10^{-14}$ has been reported.(1) Final long-term stability of this system depends almost entirely on the physical stability of the passive crystal unit.

One approach to the problem of long-term stability is to maintain a crystal unit at cryogenic temperature. Some of the advantages of using cryogenic temperature are improvement of aging and Q cactor, and the simplification of temperature regulation.(2-4)

In this paper, the application of the cryogenic temperature crystal resonator to the passive dual crystal system is studied.

Properties of quartz crystal resonators

The Q factor (unloaded Q) of quartz resonators has been measured for two different types of quartz resonators; 5 MHz resonators of the BVA_2 design (electrodeless design)(5) and 5 MHz, 5th overtone, AT cut resonators of a conventional design.

Highest Q of 10 samples of the BVA_2 resonators is 4×10^7 at 4.2K and highest Q of 4 samples of the conventional resonators is 8.6×10^6 at 4.2K. No significant increase in Q factor below 4.2K(6) has been obtained in either types of resonators. The frequency variation between 300K and 4.2K is approximately 6 KHz for both resonators.

The variations of frequency with temparature below 4.2K for one sample of both types of crystal is shown in Fig. 1. In both cases, the temperature coefficient near 2K is greatly reduced compared with 1.2×10^{-8}/K at 4.2K. The BVA_2 resonator, which is synthetic Russian quartz AT cut, shows the temperature coefficient of 3×10^{-10}/K at 2K, which is less than that of superconducting cavities(4) at 1.3K.

The amplitude frequency effect is an important characteristic of precision crystal units. This effect is a limiting factor for drive level in crystal units. Fig. 2 shows the amplitude frequency effect taken near 2K for the same crystals as shown

*This work is being performed at the Frequency and Time Standards Division, National Bureau of Standards, Boulder, CO. 80303.

in Fig. 1. Q factor and coupling factor, β, to the 50 ohme transmission line are also given in Fig. 2. It is observed that the frequency versus drive power is very steep above 10^{-7} watts at cryogenic temperatures. The BVA_2 sample shows a "turn-over" point versus drive power.

The amplitude frequency effect of quartz crtstal resonators are due to two different physical mechanisms.[7] The first is the thermal effect and the second is non-linearity in the stress strain relation. The specific heat of quartz is extremely decreased at cryogenic temperature. At 10K, specific heat of quartz is 7×10^{-4} joule/g·K.[8] If the Debye approximation is assumed below 10K, the specific heat at 2K is equal to 5.6×10^{-6} joule/g·K. Both the low specific heat and small weight of a crystal unit imply the possibility that the thermal effect may become more dominant at cryogenic temperature. The frequency amplitude effect, however, taken at 4.2K for the same crystals where the temperature coefficient is almost ten times of that at 2K, did not show any noticeable change to the results shown in Fig. 2.

Passive Dual Crystal System and its Stability

The passive dual crystal system has been used for frequency stabilization with cryogenic temperature crystal units. This system shown in Fig. 3 is similar to the one presented in [1] except for the synthesizer.

The same sample of the BVA_2 design shown in Fig. 1 and Fig. 2 was chosen for this system because of its small temperature coefficient. S, the sensitivity of the discriminator, is proportional to $Q\sqrt{P_d}$, where P_d is dissipated power in a crystal resonator. Larger S can be realized also in the sample of the BVA_2 design.

P_d is set to -51 dBm which is turn-over point. This drive level is much lower than that at room temperature, typcially P_d=1μw, $Q=2\times10^6$. However, $Q\sqrt{P_d}$ of the sample at 2.0K is more than twice of that of the typical crystal unit at room temperature.

The oven was regulated at 2.0K. The performance of the temperature controller was monitored during experiments by a germanium thermometer. The temperature of the oven was maintained within 9.1×10^{-7}K. The example of the record is shown in Fig. 4.

The measured fractional frequency stability of the system is shown in Fig. 5. In this system, the residual FM noise of a synthesizer is low pass filtered. The stability of the system, however, is almost unaffected by the residual noise of the synthesizer as shown in Fig. 5. The measured noise floor is 8.2×10^{-13}. This value is much worse than expected from the static properties and the achieved temperature regulation.

Study of Frequency Fluctuation

The spectral density of frequency fluctuations of the system output is given by

$$S_y(f,\text{system}) = S_y(f,\text{VCO}) \left|\frac{1}{1+G(f)}\right|^2$$

G(f) is the open loop frequency response function. As Vx, the output voltage of the discriminator, is proportional to the frequency fluctuation of VCO, the spectral density of V_x was measured to check the system performance. Results are shown in Fig.6. The second order loop filter reduces the frequency fluctuation by 12 dB/octave at fourier frequency below 4 Hz[10]. It will be observed that sufficient loop gain is obtained for the long-term stability. The measured sensitivity of the discriminator is 39.5 V/Hz. The calculated spectral density from the data of Fig. 5 is also shown in Fig. 6.

Measured thermal noise of the electronics at the output of the discriminator is 3.5×10^{-5} V/Hz at fourier frequency between 0.03 Hz and 10 Hz. The power spectrum of frequency fluctuations due to thermal noise is given by

$$S_y(f) = 3.2 \times 10^{-26}/\text{Hz}$$

Under this condition, the rms fractional frequency fluctuations due to the noise in the electronics is given by

$$\sigma_y(\tau) = 1.3 \times 10^{-13}\tau^{-1/2}$$

Instability of the phase due to the temperature coefficient of the transmission line conducting the rf power to the crystal causes a frequency drift. The measured frequency offset is 1.5×10^{-2} Hz/m, which causes a fractional drift rate of approximately 5×10^{-14}/°C for one meter copper rigid coaxial cable.

The drift of the d-c amplifier in the filter circuit also causes frequency fluctuations. The voltage drift for the device used in this system is less than 4.5 μV/°C, which causes a fractional frequency rate of approximately 2×10^{-14}/°C.

The above effects show that the noise floor of frequency fluctuations in the current system would reach a few parts per 10^{14}.

The measured noise floor appears to be caused entirely by the frequency fluctuation of the passive crystal unit itself. One of the most likely reasons for this rather poor performance is due to the neglect of the dynamic temperature frequency behavior of the crystal.[11] Warner[2] shows that for a 10 MHz, AT cut crystal, the temperature gradients have a greater effect than is indicated by the temperature coefficient at 4.2K.

Conclusion

The temperature coefficient of quartz crystal resonator is extremely low near 2K. A 5 MHz resonator of the BVA_2 design show a temperature coefficient of 3×10^{-10}/K below 2K. Although the drive level of a crystal unit at cryogenic temperature must be held at a low value, the increased Q makes it possible to maintain the sensitivity of the discriminator at a higher level than at room temperature.

The passive dual crystal system stabilized with a crystal resonator of the BVA, design at 2K yielded the noise floor of frequency stability as low as 8.2×10^{-13}. Studies of frequency fluctuations due to electronics of this system show that this noise floor is entirely limited by the frequency fluctuation of the crystal unit under the test.

Acknowledgements

This work has been performed at the Frequency & Time Standards Division, National Bureau of Standards. The author wishes to thank S. R. Stein for giving the opportunity to perform this work and for helpful discussion. He acknowledges the help of many useful discussions and suggestions from C. M. Manney and F. L. Walls. Thanks are also due to Prof. R. J. Besson for the offers of several quartz crystals.

References

(1) S. R. Stein, C. M. Manney, F.L. Walls, J. E. Gray, R. J. Besson, "A system approach to high performance oscillators", Proc. 32nd Ann. Symp. on Frequency Control, 527-300, 1978.

(2) A. W. Warner, "Ultra-precise quartz crystal frequency standards", IRE Trans.on Instrumentation 17, 185-188, Dec. 1958.

(3) G. Mossuz and J. J. Gagnepain, "Quartz crystal oscillator at very low temperature", Cryogenics 16, No. 11, 652-656, Nov. 1976.

(4) S. R. Stein, "Application of superconductivity to precision oscillators", Proc. 29th Ann.Symp. on Freq. Control, 321-327, 1976.

(5) R.J.Besson, "A new electrodeless resonator design", Proc. 31st Ann. Symp.on Freq. control, 3-7, 1977.

(6) A. G. Smagin, "A 1-MHz quartz resonator with a Q factor of $4.2 \cdot 10^9$ at a temperature of 2 °K", Prib. Tekh Eksp No. 6 (1974) 143.

(7) J. J. Gagnepain, J. C. Poncot and C. Pegeot, "Amplitude-frequency behavior of doubly rotated quartz resonators", Proc. 31st Ann. Symp. on Freq. Control, 17-22, 1977.

(8) V. J. Johnson, "Compendium of the properties of materials at low temperature", WADD Tech. Report 60/56.

(9) G. K. White, "Experimental techniques in low-temperature physics", Oxford University Press, 1968.

(10) F. L. Walls and S. R. Stein, "Servo techniques in oscillators and measurement systems", NBS Tech. Note 692.

(11) A. Ballato and J. R. Vig, "Static and dynamic frequency - temperature behavior of singly and doubly rotated, oven-controlled quartz resonators", Proc. 32nd Ann. Symp. on Freq. Control, 180-188, 1978.

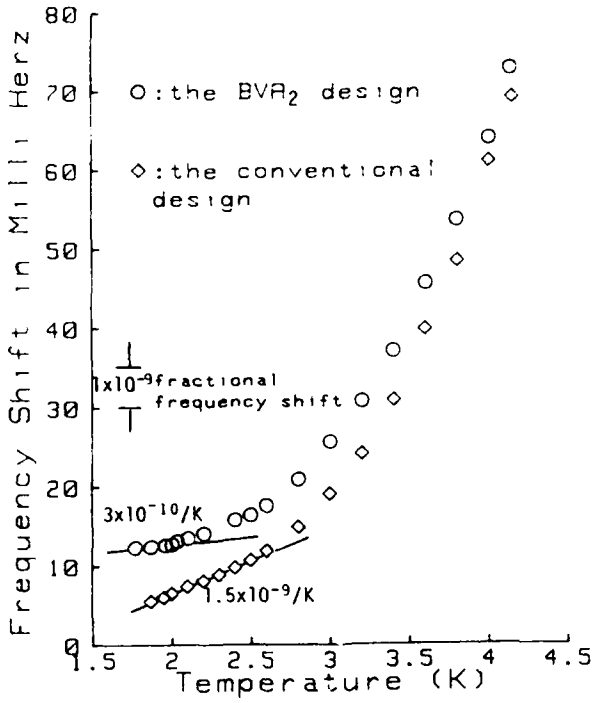


Fig. 1. Frequency variation as a function of temperature below 4.2K.

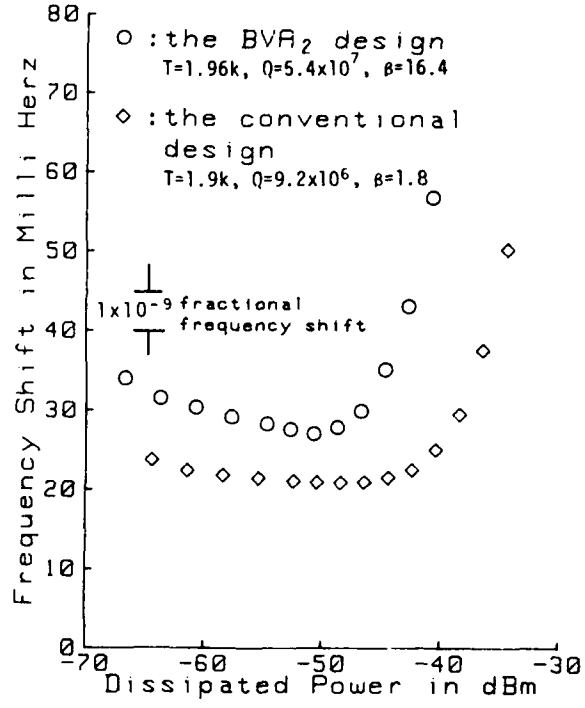


Fig. 2. Frequency vs. dissipated power in crystal resonators near 2K for the same samples as shown in Fig. 1.

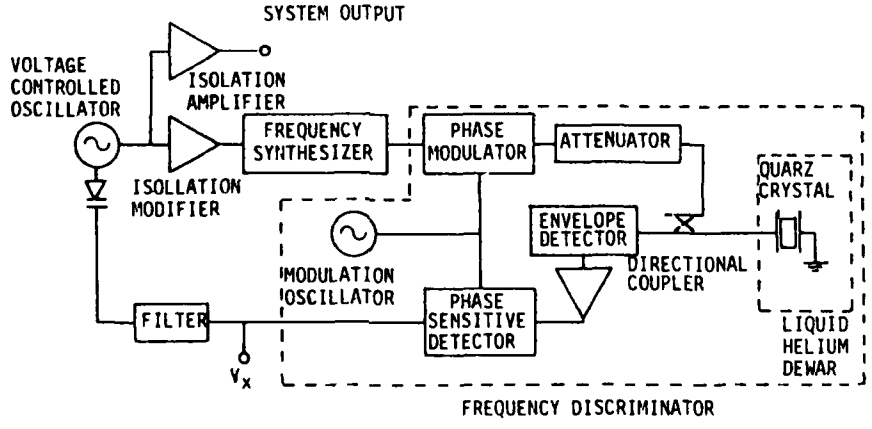


Fig. 3. Block diagram of the passive dual crystal system, in which a VCO having good short-term stability is locked to a crystal resonator frequency discriminator.

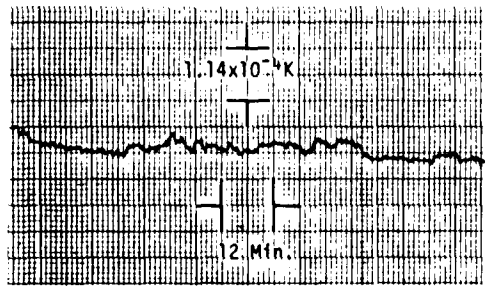


Fig. 4. Temperature variations in the oven.

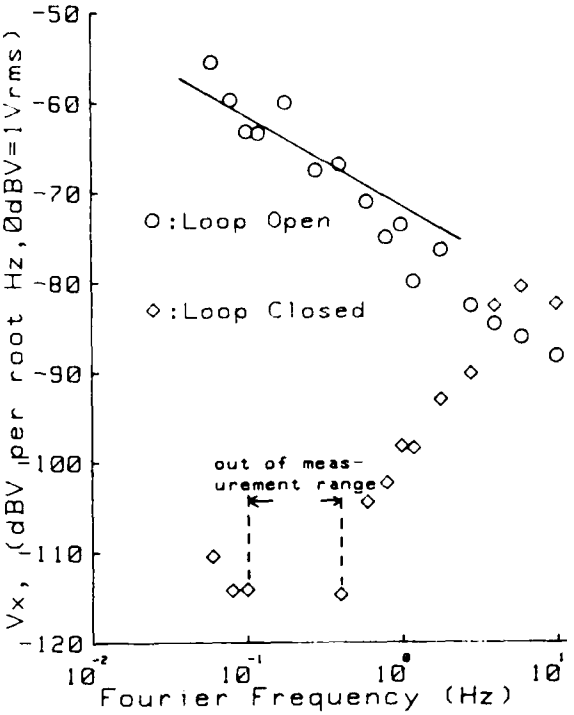


Fig. 6. Spectral density of output voltage of the discriminator. Solid line shows the flicker spectral density calculated from the flicker floor of Fig.5 for free running crystal oscillator.

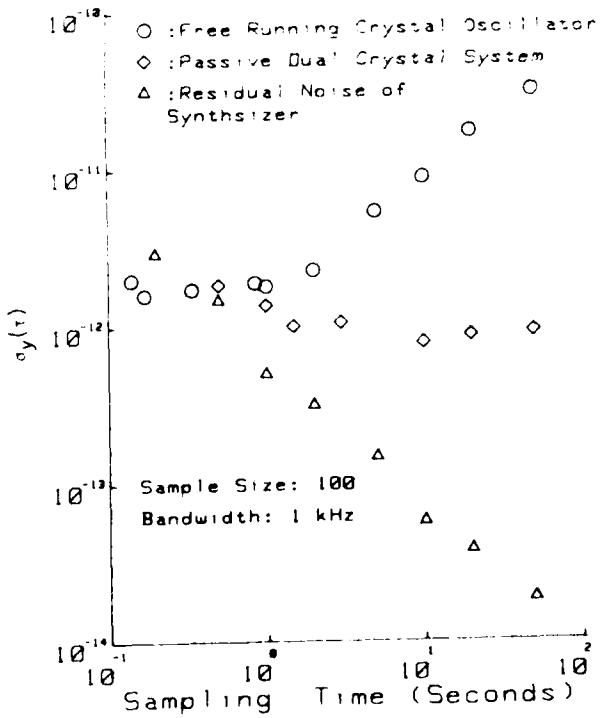


Fig. 5. Fractional frequency stability $\sigma_y(\tau)$.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, 07703, May 1981

QUARTZ RESONATOR THERMAL TRANSIENT DUE TO CRYSTAL SUPPORT

D. Janiaud*, R. Besson**, J.J. Gagnepain*** and M. Valdois*

* Office National d'Etudes et de Recherches Aérospatiales
92320 Chatillon - France

** Ecole Nationale Supérieure de Mécanique et des Microtechniques
Besançon - France

*** Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté-Besançon
25000 Besançon - France

Summary

The dynamic thermal behavior of a quartz resonator is caused by nonuniform temperature distributions in the structure of the resonator, which are time dependant.

These temperature gradients occur in the crystal plate and in the supports. Both effects contribute to the total dynamic thermal behavior of the resonator.

A model of the static and dynamic thermal behaviors of the supports is presented. It is applied to regular commercially available resonators. The dynamic temperature coefficient $\tilde{a}_s$ of the supports is derived. Its value is evaluated for AT and SC-cuts and compared to experimental figures.

Introduction

Dynamic temperature effects in quartz crystals alone, i.e. when the influence of the mounting is not included, have been presented in previous AFCS by J. Théobald[1] and B. Sinha[2]. These theoretical studies have defined a dynamic thermal sensitivity coefficient $\tilde{a}$ such as $\frac{\Delta F}{F}_{dynamic} = \tilde{a}\frac{dT}{dt}$, in agreement with the phenomenological model established by A. Ballato[3]. The present work gives a more complete description of the resonator by introducing the influence of the support on the thermal sensitivity.

When a quartz crystal resonator is submitted to a temperature change, supports contribute to both aspects (static and dynamic) of the frequency change. The contribution of supports to the static effect is introduced by the different thermal expansion coefficients of the resonator's elements. Concerning the dynamic effect, the contribution of supports is related to the spatial thermal gradient which characterize the thermal transient. These two effects create in-plane diametric forces acting on the crystal at the fixation points of the mounting.

Theoretical formulae of supports contribution to the frequency change is then the superposition of two terms. The first one represents the static effect, the second one represents the dynamic effect and is of the form $\tilde{a}_s\frac{dT}{dt}$. The dynamic coefficient $\tilde{a}$ of the whole resonator is the sum of the coefficient $\tilde{a}_Q$ of the crystal and the coefficient $\tilde{a}_s$ of the support.

Diametric forces acting on the crystal plate

The resonator is constituted by a circular quartz plate, of diameter D and thickness e, and by a glass-metal header on which two stiff metal supports are mounted, as shown on fig. 1.

Each support is composed of a folded metal plate (width a, thickness b and length ℓ) which is welded on a cylindrical rod (length h and section S). In the initial state the crystal and the mounting are in thermal equilibrium, and it is considered that no forces are applied to the crystal by the mounting.

When an external thermal perturbation occurs the thermal equilibrium is modified and each of the components of the resonator is submitted to a temperature change following the notation

$\theta_H(t)$ for the header which is supposed to have a uniform temperature distribution

$\theta_R(y,t)$ for a cross section, of coordinate y, of the cylindrical rod

$\theta_J(t)$ for the junction between rod and plate

$\theta_\ell(x,t)$ for the cross section, of coordinate x, of the plate

$\theta_E(t)$ for the electrodes, which are deposited on the main faces of the quartz crystal and with are considered as isothermal planes

$\theta_Q(x_2,t)$ for the cross section, of coordinate x_2, of the crystal

Finally $\overline{\theta}_Q(t)$ is the mean temperature of the crystal.

... temperature gradients induce two forces F, F, which diametrically act on the crystal at the fixation points.

The influence of such in-plane diametric forces on the frequency of a thickness shear quartz resonator was already studied [4-5]. The relative frequency deviation versus the applied forces can be written

$$\frac{\Delta f(t)}{f} = K F(t) \qquad (1)$$

where the force-sensitivity coefficient K depends on the geometry and orientation of the crystal, and on the direction of the in-plane forces. For a circular quartz crystal with diameter D = 15 mm and thickness e = 1,05 mm, and for forces applied along the X' axis, the coefficient K is

$$K_{AT} = 1,3 \times 10^{-9} N^{-1} \quad \text{for AT-cut}$$

$$K_{BT} = 2,3 \times 10^{-9} N^{-1} \quad \text{for BT-cut}$$

The calculation of the F(t) forces can be ... initially considering that the supports are flexible enough and so the different elements of the resonator can expand almost independantly ; the possibilities of the displacements at the ... are allowed by the flexure of the metal plates.

Therefore the displacement at the end of the metal plate is (...

$$\Delta l(t) = \ldots [\Delta_H(t) + \Delta_E(t)] - \Delta_Q(t) \ldots$$

...

... expresses the distance between the ... in the header

... expresses the length of the metal ...

... stands between the metal plates and the ...

... expresses the quartz crystal dia...

These different expressions are given by the equations

$$\Delta_H(t) = l_H \alpha_H \bar{\theta}_H(t) \qquad (2)$$

$$\Delta_E(t) = l_E \alpha_E \bar{\theta}_E(t) \qquad (4)$$

$$\Delta_Q(t) = l_Q \alpha_{33} \bar{\theta}_Q(t) \qquad (5)$$

... α_H, α_E and α_{33} respectively are the thermal expansion coefficients of the header, metal plates and quartz. $\bar{\theta}_H(t)$ and $\bar{\theta}_E(t)$ are the mean temperatures of metal plates and quartz.

The ... component of the diametric ... is a function of the end-... by the classical flexure equation

$$F(t) \sin\delta = \frac{12EI}{\ell^3} f(t) \qquad (6)$$

... the Young modulus of the metal plate and I the inertia momentum of a cross section ($I = ab^3/12$).

By using eq. (2), (3), (4), (5) and (6) the diametric force F(t) is calculated as a function of the temperature of the header and of the mean temperatures of the metal plates and crystal.

$$F(t) = \frac{Eab^3}{2\ell^3 \sin^2\delta} \left[d\alpha_H \theta_H(t) + 2l\alpha_E \bar{\theta}_E(t) \cos\delta - D\alpha_{33} \bar{\theta}_Q(t) \right] \qquad (7)$$

Temperature distribution in the supports

$\theta_Q(t)$ is choosen as reference temperature. To express $\theta_H(t)$ and $\bar{\theta}_E(t)$ versus $\bar{\theta}_Q(t)$ it becomes necessary to study heat transfer in the cylindrical rod and in the metal plates.

These calculations are similar to those of ref. [3] and lead, as a first step, to temperature in any point of the plate

$$\theta_E(x,t) = \frac{2\kappa_E \pi}{\ell^2} \sum_{n=1}^{\infty} e^{-\kappa_E \frac{n^2\pi^2}{\ell^2} t} \, n \sin\frac{n\pi x}{\ell} \, I(t) \qquad (8)$$

with

$$I(t) = \int_0^t e^{\kappa_E \frac{n^2\pi^2}{\ell^2}\lambda} \left[\theta_J(\lambda) - (-1)^n \theta_E(\lambda)\right] d\lambda$$

where κ_E is the thermal diffusivity constant of plate material.

I(t) is obtained by integration ignoring the higher order derivatives of $\theta_J(t)$ and $\theta_E(t)$. Temperature is assumed to be slowly variable with time and one obtains

$$I(t) = \frac{\ell^2}{\kappa_E n^2\pi^2} e^{\kappa_E \frac{n^2\pi^2}{\ell^2} t} \left[\theta_J(t) - (-1)^n \theta_E(t)\right] - \frac{\ell^2}{\kappa_E n^2\pi^2} \left[\dot{\theta}_J(t) - (-1)^n \dot{\theta}_E(t)\right] \qquad (9)$$

Using (9) in expression (8) yields

$$\theta_E(x,t) = \frac{2}{\pi} \left[\theta_J(t) \sum_{n=1}^{\infty} \frac{1}{n} \sin\frac{n\pi x}{\ell} - \theta_E(t) \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \sin\frac{n\pi x}{\ell} - \frac{\ell^2}{\kappa_E \pi^2} \dot{\theta}_J(t) \sum_{n=1}^{\infty} \frac{1}{n^3} \sin\frac{n\pi x}{\ell} + \frac{\ell^2}{\kappa_E \pi^2} \dot{\theta}_E(t) \sum_{n=1}^{\infty} \frac{(-1)^n}{n^3} \sin\frac{n\pi x}{\ell} \right] \qquad (10)$$

These series were evaluated by use of ref. [8]. This yields

$$\theta(x,t) = \left(1-\frac{x}{\ell}\right)\theta_J(t) + \frac{x}{\ell}\theta_E(t) - \frac{\ell^2}{6\kappa_\ell}\frac{x}{\ell}\left(1-\frac{x}{\ell}\right)\left[\left(2-\frac{x}{\ell}\right)\dot{\theta}_J(t) + \left(1+\frac{x}{\ell}\right)\dot{\theta}_E(t)\right] \quad (11)$$

Since heat transfer in the cylindrical rod is similar to the previous case, the variation of temperature $\theta_R(y,t)$ in a cross-section of coordinate y in the cylindrical rod can be written

$$\theta(y,t) = \left(1-\frac{y}{h}\right)\theta_H(t) + \frac{y}{h}\theta_J(t) - \frac{h^2}{6\kappa_r}\frac{y}{h}\left(1-\frac{y}{h}\right)\left[\left(2-\frac{y}{h}\right)\dot{\theta}_H(t) + \left(1+\frac{y}{h}\right)\dot{\theta}_J(t)\right] \quad (12)$$

where κ_r is the thermal diffusivity coefficient of rod material.

Using eq. (11) and eq. (12) in addition with heat conservation law at the rod-plate junction yields a relation between $\theta_J(t)$ and $\theta_H(t)$:

$$\theta_J(t) = \theta_H(t) - \frac{h^2}{2\kappa_\ell}\dot{\theta}_H(t) \quad (13)$$

For simplification, the thermal diffusivities of material in rod and plate have been considered equal. In addition the term $\frac{ab}{S}\frac{h}{\ell}$ of the order of 10^{-2} has been ignored by respect to unity.

From eq. (11) and eq. (13) $\theta(x,t)$ is obtained :

$$\theta(x,t) = \left(1-\frac{x}{\ell}\right)\theta_H(t) + \frac{x}{\ell}\theta_E(t) - \frac{\ell^2}{6\kappa_\ell}\left(1-\frac{x}{\ell}\right)\left\{\left[\frac{x}{\ell}\left(2-\frac{x}{\ell}\right) + \frac{3h^2}{\ell^2}\right]\dot{\theta}_B(t) + \frac{x}{\ell}\left(1+\frac{x}{\ell}\right)\dot{\theta}_E(t)\right\} \quad (14)$$

Temperature variation inside the crystal is obtained from ref. [1] :

$$\theta_Q(x_2,t) = \theta_E(t) + \frac{1}{2\kappa_Q}\left(x_2^2 - \frac{e^2}{4}\right)\dot{\theta}_E(t) \quad (15)$$

where κ_Q is the thermal diffusivity coefficient of quartz in x_2 direction.

Heat conservation laws and eq. (14) with eq. (15) yield :

$$\theta_H(t) = \theta_E(t) + \dot{\theta}_E(t)\frac{\ell^2}{2\kappa_\ell}\frac{C_Q}{C_\ell} \quad (16)$$

where C_Q and C_ℓ are the specific heat capacities of quartz plate and metal plate.

$\theta_Q(t)$ in the crystal is obtained from eq. (15) :

$$\bar{\theta}_Q(t) = \frac{1}{e}\int_{-e/2}^{e/2}\theta(x_2,t)\,dx_2 = \theta_E(t) - \frac{e^2}{12\kappa_Q}\dot{\theta}_E(t) \quad (17)$$

$\theta_H(t)$ is given by eq. (16) and eq. (17) :

$$\theta_H(t) = \bar{\theta}_Q(t) + \dot{\bar{\theta}}_Q(t)\frac{\ell^2}{2\kappa_\ell}\frac{C_Q}{C_\ell} \quad (18)$$

$\bar{\theta}_\ell(t) = \frac{1}{\ell}\int_0^\ell \theta(x,t)\,dx$ is obtained through eq. (14), eq. (17) and eq. (18)

$$\bar{\theta}_\ell(t) = \bar{\theta}_Q(t) + \dot{\bar{\theta}}_Q(t)\frac{\ell^2}{4\kappa_\ell}\frac{C_Q}{C_\ell} \quad (19)$$

Analytical expression of frequency deviation

Eq. (18) and eq. (19) are introduced in eq. (3)

$$\frac{\Delta F(t)}{F} = \frac{Eab^3}{2\ell^3\sin^2\delta}\left[(d\alpha_H + 2\ell\alpha_\ell\cos\delta - D\alpha_{11})\bar{\theta}_Q + \frac{\ell^2}{2\kappa_\ell}\frac{C_Q}{C_\ell}(d\alpha_H + \ell\alpha_\ell\cos\delta)\dot{\bar{\theta}}_Q(t)\right] \quad (20)$$

Under those conditions, frequency deviation can be expressed from eq. (1) and eq. (20)

$$\frac{\Delta F(t)}{F} = a_S\bar{\theta}_Q(t) + \tilde{a}_S\dot{\bar{\theta}}_Q(t) \quad (21)$$

where a_S and $\tilde{a}_S$ are the static and dynamic thermal coefficients related to the only effect of fixation type.

These coefficients are given by :

$$a_S = k\frac{Eab^3}{2\ell^3\sin^2\delta}(d\alpha_H + 2\ell\alpha_\ell\cos\delta - D\alpha_{11}) \quad (22)$$

$$\tilde{a}_S = k\frac{Eab^3}{4\ell\kappa_\ell\sin^2\delta}\frac{C_Q}{C_\ell}(d\alpha_H + \ell\alpha_\ell\cos\delta) \quad (23)$$

It can be seen that a_S is proportional to the differential expansion of quartz plate with respect to the support configuration, and that $\tilde{a}_S$ is only related to expansion of support configuration. As a consequence a resonator whose supports have no thermal expansion yields a dynamical behaviour without influence of the supports.

Numerical evaluation has been made for a regular 5 MHz fifth overtone resonator of AT and SC cuts :

AT-cut
$$a_S = -5.4 \times 10^{-9}\ /°C$$
$$\tilde{a}_S = +5.5 \times 10^{-7}\ s/°C$$

SC-cut
$$a_S = -8.1 \times 10^{-10}\ /°C$$
$$\tilde{a}_S = +8.3 \times 10^{-8}\ s/°C$$

Experimental results

The total dynamic temperature coefficient, $\alpha_{quartz}^{support}$, was measured on 5 MHz AT quartz resonators mounted with supports of the type described on fig. 1. The resonator was submitted to sinusoidal temperature cyclings by using a programmable temperature controlled oven. The total temperature variation was 8°C and the period of a cycle 60 mn. The corresponding frequency-temperature pattern is shown on fig. 3, and the dynamic temperature coefficient is $\alpha = 5.10^{-8}$/°C.

Then the two supports of the resonator were grooved in order to reduce the support width by an amount of 60% at a point located at one third of the total length from the bottom, as shown in the detail of fig. 4.

The α coefficient was measured again. The corresponding cycling curve is given on fig. 3 and it was found that $\alpha = 0$.

The grooving of the support has two effects. It reduces the stiffness of the metal plate and it restricts the diffusion of heat from the header to the crystal. The first effect has a small influence, but the second one rises up the thermal gradients and therefore increases the α coefficient. The two fixation points of the crystal plate are located on a diameter which makes an angle of 45° with respect to the Z' axis of the crystal. This position corresponds to a maximum of the force-sensitivity, which makes the influence of the dynamic thermal behavior of the mounting much more important as it appears from the comparison between the two curves of fig. 3.

Conclusion

The contribution of the supports to the first order static temperature coefficient of quartz resonators is of the order of 10^{-7}/°C for SC-cut resonators and a few 10^{-8}/°C for AT-cut resonators. Such an order of magnitude still is small when compared to the maximum slopes of the frequency-temperature characteristics and therefore the static influence of the supports mainly appears as a shift of the turn over point by an amount of .1°C.

For the dynamic thermal behavior, the results are to be compared to the experimental values already measured, which correspond to -1.3×10^{-5} s/°C for AT-cut and to 3×10^{-7} s/°C for SC-cut. If the influence of the supports can be neglected for AT-cut, it is not the case for SC-cut and both dynamic effects in the supports and in the crystal must be taken into account. Even a compensation between both effects could be possible by adjusting the position of the fixation points and the geometrical parameters of the supports.

Acknowledgements

The authors would like to thank G. Marianneau and R. Prétot of CNRS for providing the experimental data of the SC-cut resonator with modified supports.

References

(1) G. Théobald and al, "Dynamic thermal behavior of quartz resonators", 32nd AFCS, 1978.

(2) B.K. Sinha, H.F. Tiersten, "Transient thermally induced frequency excursion in doubly rotated quartz thickness mode resonators", 34th AFCS, 1980.

(3) A. Ballato, J.R. Vig, "Advances in the stability of high precision crystal resonators", PTTI, 1979.

(4) J.F. Perdigues, "Etude de l'influence des forces sur la fréquence d'un résonateur à quartz", Thèse de Docteur-Ingénieur, Besançon (avril 1975).

(5) D. Janiaud, L. Nissim, J.J. Gagnepain, "Analytical calculation of initial stress effects in anisotropic crystals : application to quartz resonators", 32nd AFCS, 1978.

(6) R. Holland, "Nonuniformly heated anisotropic plates : I. Mechanical distorsion and relaxation", IEEE Trans. Sonics and Ultrasonics, vol. SU-21, n°3 (July 1974).

(7) H.S. Carslaw, J.C. Jaeger, "Conduction of heat in solids", Oxford University Press, London (1959).

(8) H.B. Dwight, "Tables of integrals and other mathematical data", Fourth Edition, Mac Millan Publishing Co, Inc., New York (1961).

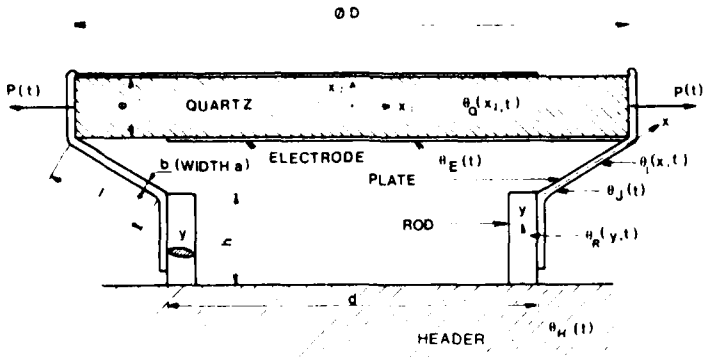


fig 1 CONFIGURATION OF THE SUPPORTS OF THE RESONATOR

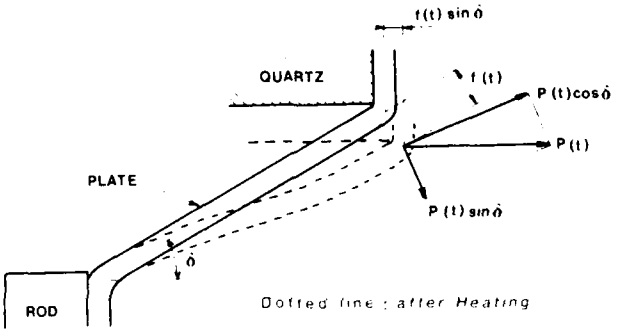


fig 2 FLEXURE OF THE SUPPORT

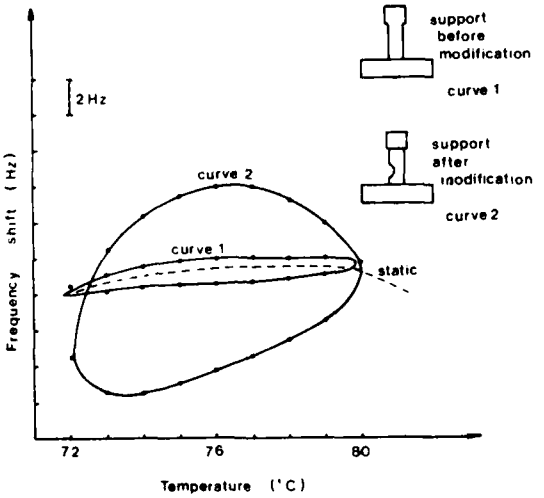


fig 3 DYNAMIC THERMAL BEHAVIOR BEFORE AND AFTER MODIFICATION OF THE SUPPORTS

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703. May 1981

COMMERCIAL SATELLITE NAVIGATION USING SAW OSCILLATOR

B. Y. Lao, N. J. Schneier, D. A. Rowe, R. E. Dietterle
Magnavox Advanced Systems and Products Company
Torrance, California 90503

J. S. Schoenwald, E. J. Staples, J. Wise
Rockwell International Corporation
Thousand Oaks, California 91360

ABSTRACT

A 375 MHz SAW resonator controlled oscillator is developed for application in the Transit satellite marine navigation system. The SAW oscillator, in a two-cubic-inch hybrid package, contains a heater, a voltage regulator and a divider and is a direct replacement for a bulk wave oscillator and its multiplier chain. Short term stability of 2E-10 and aging of 3E-8/day were achieved at 75 °C. Comparison tests showed that the navigation system accuracy with the SAW oscillator was equivalent to a bulk oscillator.

INTRODUCTION

The application of a surface-acoustic-wave controlled oscillator in a commercial transit satellite marine navigator and the test results will be presented.

The MX1102 marine navigation system operates in the UHF band at 400 MHz. The Transit satellites circle the earth in 107-minute polar orbits at an altitute of 600 nautical miles.[1] The orbits do not rotate with the earth so that every point on the earth's surface passes under each of the six orbits approximately twice a day. The signal transmitted by the satellite contains orbital information regarding the position of the satellite and the GMT. Updating by ground stations keeps the information accurate. The navigation receiver detects the signal and obtains the location of the satellite as a function of time. Using this information, the reciever calculates its location in a least-square solution which best fits the Doppler shifts observed in the satellite signal.[2] In addition to location fixes thus obtained, the navigator also provides dead reckoning between fixes based on inputs of the ship's speed and heading.

In the detection process, the RF signal from the satellite is first down-converted to 25MHz by the use of a mixer and a local reference signal at 375MHz. The Doppler shift is generally small (8KHz/5min), and the stability of the local oscillator directly determines the system fix accuracy. Therefore a high stability LO is required. Fig. 1 is a block diagram of the RF section of the navigation reciever. The local oscillator signal is normally generated by multiplying a precision 5MHz quartz bulk-wave oscillator output to 375MHz. This approach suffers from the large volume occupied by the crystal oscillator and the associated multiplying chain and its relatively high cost.

These problems can be solved by replacing the crystal oscillator and the multiplier chain by a more compact SAW oscillator with a fundamental frequency at 375MHz. This technique leads to potential cost savings due to the ability to batch process SAW resonators. The critical requirements for the local oscillator are stabilities of 1E-10 for 1 to 100 sec averaging time, 7E-10 for 1000 sec and 2E-9 per day aging. The oscillator frequency is required to be within 1 ppm of 375MHz because of the limited pass band of the IF stages. Other requirements are to supress spurious and harmonic outputs to less than 60dB below the desired output; operation in an ambient temperature of -20 to 65 °C; a frequency load stability of -1E-9 for a 10% change in the 50 ohm output load and a frequency voltage stability of -5E-9 for a 10% change in the supply voltage.

SAW RESONATOR

For this purpose, SAW resonators were fabricated on ST-quartz cut to provide a turning point of 75 °C. They are single port devices using shallow groove reflective gratings and are evacuated and sealed in TO-5 cans.

The single port design was chosen over a two-port design because of its lower gain requirement which means fewer oscillator components and lower cost. The resonator was designed to operate in its series resonance mode with a series impedance of about 50 to 100 ohms.

During the fabrication process, over 200 metalized patterns (40Å of Cr and 1000Å of Al) are photolithographically formed on a 2"x2" quartz substrate (x-propagating, 34° rotated for 75°C

turnover). After spot-check probing of random devices a second masking with photoresist is used to protect the transducers during CF_4 plasma etching to form grooves in the reflector portion of the device. Following etching, the resist is removed and the wafer is spot probed again. The wafer is then diced into chips which are mounted on TO-5 headers using Abelbond 71-1 high temperature polyimide adhesive. A low power CF_4 plasma etch is used to place the mounted crystal on frequency just prior to sealing. Bake-outs before and after sealing are performed at 125 °C for 24 hours. Fig. 2 shows the typical polar Smith chart response of the S_{11} scattering parameter of a sealed SAW resonator.

OSCILLATOR

Initial attempts to replace the crystal oscillator with a SAW oscillator were hampered by turning point shifts, excessive voltage and load dependence. These problems were traced to circuit component variations which are relatively unimportant for crystal oscillators but become significant for SAW oscillators because of the lower Q. High resonator drive levels on the order of -5dBm into 50 ohms were also found to degrade the device over extended periods. Metal migration and subsequent crystalite formation are suspected to be the cause of this degradation.

These problems were solved by stablizing the oscillator with an AGC circuit which limits the power into the SAW resonator to -15dBm. Short term stabilities of 1 to 2E-10 and aging rates of 3E-8 per day at 75 °C were achieved.[3]

Fig. 3 shows a typical stability curve after one week aging. The long term drift, which is the limiting factor for stability at 1000 sec sampling time, generally improves by a factor of 4 after a one-month aging. A block diagram of the oscillator circuit is shown in Fig. 4.

HIBRID PACKAGING

In order to take the full advantage of the reduction in size that a SAW resonator can offer, a hybrid oscillator package was built that contains the oscillator, buffer amplifiers, the temperature controller, the heater, the voltage regulator, provision for frequency adjustment, and a frequency divider all within a 1.8 x 1.8 x 0.65 inch (4.6 x 4.6 x 1.65 cm) package.[4] The assembled package is shown in Fig. 5. The heater is thermally coupled to the oscillator circuit by attaching it on the back side of the circuit board. The design rules for the oscillator board were intentionally conservative with a minimum line width of .020 inches. Further size reduction is therefore possible. Chip resistors and PC boards were used rather than the standard film resistors and ceramic substrates for turn-around time considerations and low heat transfer requirements. The hybrid construction represents a volume reduction by a factor of 20 over the current production system. This is illustrated in Fig. 6, which shows the crystal oscillator and the RF board. The middle-right portion of the board is the multiplier circuitry. The g-sensitivity of the hybrid oscillator was measured through phase noise measurement at 100 Hz. The corresponding fractional frequency change was found to be 5E-9/g in two axes and 3E-9/g in one axis. The g-sensitivity at this level is comparable to AT-cut quartz bulk-wave oscillators. The g-sensitivity may be reduced further by making the circuit assembly more mechanically rigid.

COMPARISON TEST IN SYSTEM PERFORMANCE

The bulk-wave crystal oscillator and its multiplier chain in the MX1102 navigator were then replaced by the SAW oscillator with its simpler divider circuitry. Satellite tracking and fixes were obtained before and after the replacement. The error of the fixes was calculated for both arrangements. The system accuracy for the location fixes with the SAW oscillator was as good as the bulk wave crystal oscillator. The standard deviation for both arrangements was on the order of .1 nautical miles at night for October 1980 when the sun spot activity was relatively high. Fig. 7 is a plot of the position of the fixes obtained with the crystal oscillator and the SAW oscillator in relation to the receiver antenna location. The average deviation of the fixes for both arrangements is shown in the lower left corner. The confidence level is indicated by the error bars. Thus, within experimental error, the SAW oscillator and the crystal oscillator provide comparable accuracy.

CONCLUSION

The use of hybrid SAW oscillators has been demonstrated in commercial satellite navigation receivers that require precision local oscillators with a short term stability of 1E-10. The SAW oscillator yields the same system accuracy as the bulk-wave crystal oscillator which it replaces.

REFERENCE

1. T. A. Stansell, Navigation, 25, 1 (1978).

2. W. H. Guier and G. C. Weiffenbach, Proc. IRE, 48, 4 (1960).

3. J. S. Schoenwald, A. B. Harker, W. W. Ho, J. Wise and E. J. Staples, 1980 Ultrasonics Sym. Proc., IEEE Cat. No. 80CH1602-2, 193 (1980).

4. S. J. Dolochycki, E. J. Staples, J. Wise, J. S. Schoenwald and T. C. Lim, Proc. 33 Annual Symp. on Freq. Control, 374 (1979).

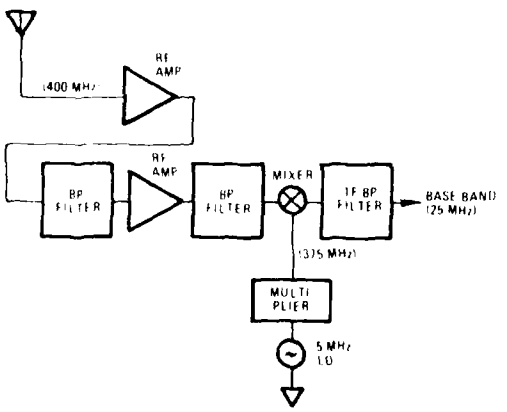


Figure 1. Block diagram of the RF section of the navigation receiver

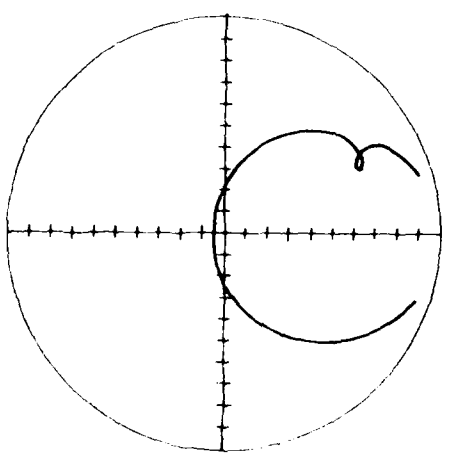


Figure 2. Typical plot of reflection coefficient of SAW resonator

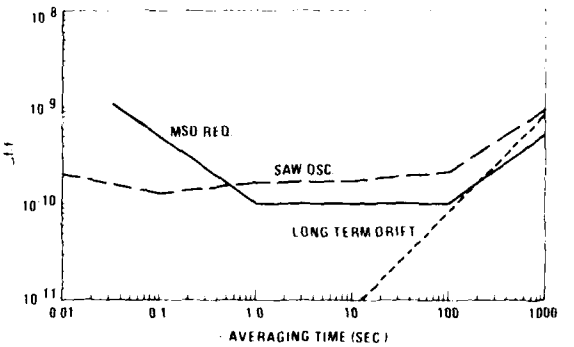


Figure 3. Stability of SAW oscillator after one week aging

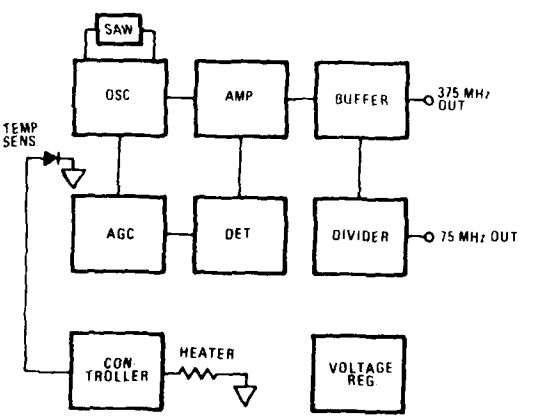


Figure 4. Block diagram of oscillator with control circuitry

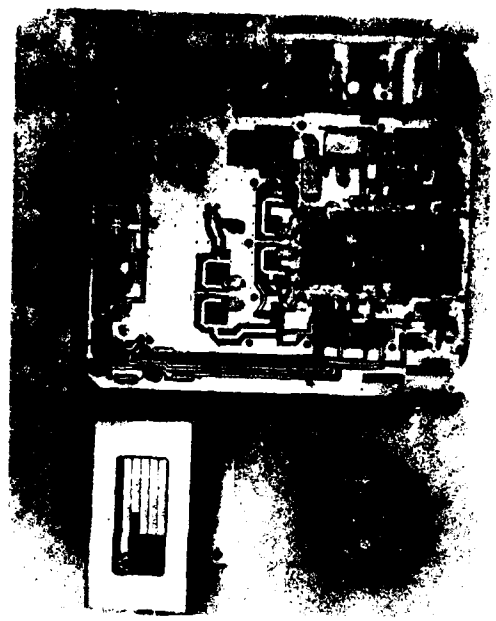


Figure 6. Size comparison of hybrid SAW oscillator and crystal equivalent

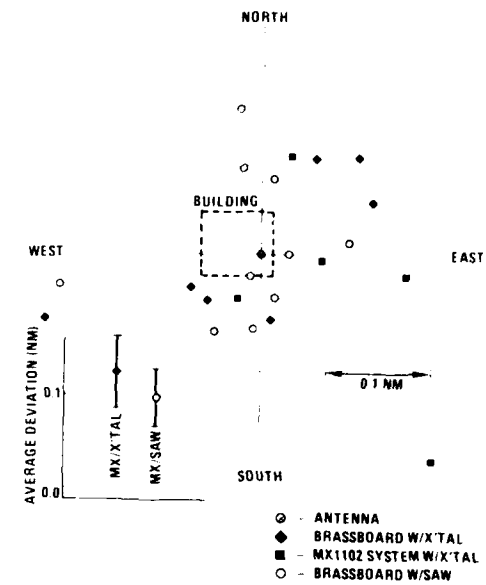


Figure 7. System fix accuracy comparison of SAW and crystal oscillators

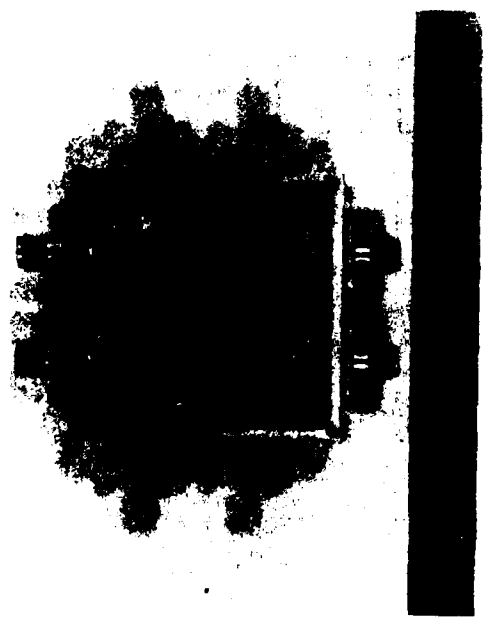


Figure 5. Assembled hybrid SAW oscillator

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ, 07703, May 1981

UHF VOLTAGE-CONTROLLED NARROW-BANDWIDTH SAW FILTERS

Jeannine HENAFF, Michel FELDMANN and Michel CAREL

CNET Paris-B, Département DTS/MAE
92220 BAGNEUX, FRANCE

Abstract

Tunable SAW filters with 30-kilohertz bandwidth in the UHF range are described in view of future vehicular communication systems around 900 MHz. Presently, different implementations around 200 MHz using $LiTaO_3$ and $AlPO_4$ are compared and stability with automatic gain control is discussed.

Introduction

Agile electrically tunable filters are of great interest in signal processing for spectrum analysis, e.g. in radars. Important civil applications are arising from the development of vehicular communication systems.

A new tunable filter has been recently experimented on $LiTaO_3$ [1] and Berlinite [2]. This paper compares the results on these materials in view of future 900 MHz vehicular system. In addition, an automatic gain control has been inserted into the device to improve the stability.

Principle of electrically tunable filter

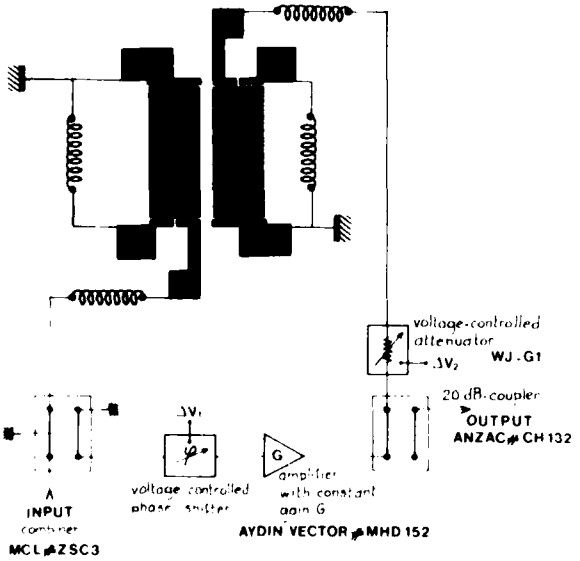


Fig. 1 : Principle of electrically tunable filters.

The electrically tunable filter is built as a tunable SAW oscillator [3,4] (cf. Fig.1). A SAW delay line is inserted in a closed loop together with an amplifier and a phase shifter. If the gain of the amplifier increases to compensate for the loss, the device is indeed an oscillator operating at a frequency ω_o, defined by $\varphi + \omega_o T = 2k\pi$, where φ is the auxiliary phase shift, T is the delay of the SAW line and k is an integer. From the well known design rules of such an oscillator, the transducer lengths are chosen in order to have a single-mode operation. If, next, the gain of the amplifier decreases, leading to a linear closed-loop gain r less than unity, the transfer function from the input to the output is roughly

$$S_{21} \simeq \frac{1}{1 + 2jQ\left|(\omega - \omega_o)/\omega_o\right|}$$

where $Q \simeq \omega_o T/2(1 - r) = k\pi/(1 - r)$ and $\omega_o = (2k\pi - \varphi)/T$. Here k is the effective number of wavelengths in the delay line accounting for both electrical and mechanical reflections.

Such a tunable filter has been implemented first, around 200 MHz on a YZ lithium tantalate substrate, and second around 170 MHz on a YX berlinite substrate, which is a zero frequency temperature coefficient configuration.

The results concerning the two experiments are described in Ref. [1], and [2]. The frequency responses are plotted in Fig. 2 and 3 for lithium tantalate and in Fig. 4 for berlinite, for different values of the loop gain (Fig. 2a and 4a) and for different values of the phase shift (Fig. 2b and 4b). The frequency responses are reproducible and stable and prove the feasibility of this principle of electrically tunable narrow-bandwidth filters.

Table I sums up some characteristics of the first lithium tantalate filter and extrapolates these results to a center frequency equal to 900 MHz. For berlinite the same characteristics are summarized in Table II. It is seen that Q-factors up to 20000 are achievable and the goal of Q = 30000 at 900 MHz is well within reach.

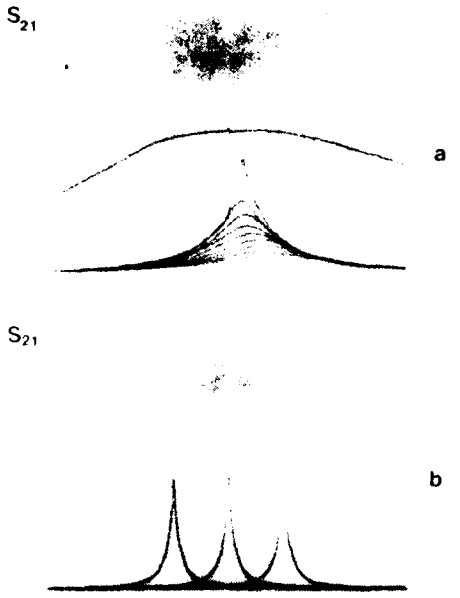


Fig. 2 : Frequency responses of the tantalate filter
a) for different values of the loop gain (step 0.1 dB)
horizontal scale : 50 kHz/c
vertical scale : 10 dB/c

b) for different values of the phase shift ($\pm$ 0.1v and o)
horizontal scale : 200 kHz/c
vertical scale : 10 dB/c

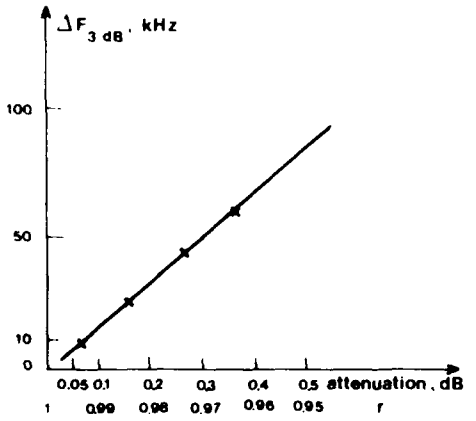


Fig. 3 : 3 dB-bandwidth of the tantalate filter versus the loop gain
step attenuation of 0.1 dB or linear gain r

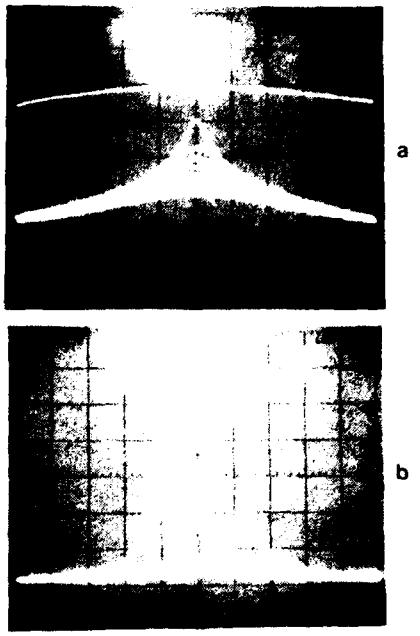


Fig. 4 : Frequency responses of the berlinite filter
a) for different values of the loop gain (step 0.1 dB)
horizontal scale : 50 kHz/c
vertical scale : 10 dB/c

b) for different values of the phase shift ($\pm$ 0.1v and o)
horizontal scale : 200 kHz/c
vertical scale : 10 dB/c

Table I : Some characteristics of electrically tunable lithium tantalate filters

	First tantalate filter	Projected filter
Centre Frequency F_c, MHz	200	900
Open loop ± 1dB bandwidth Δf, MHz	2.5	30
Channel 3dB bandwidth δf, kHz	25	30
Q	8000	30000
r	0.982	0.995
L/l	30	40
Number of channels	100	1000

Table II : Some characteristics of electrically tunable berlinite filters

	First berlinite filter	Projected filter
Centre frequency F_c, MHz	169	900
Open loop ± 1 dB bandwidth Δf, MHz	1.9	30
Channel 3dB bandwidth Δf, kHz	21	30
Q	8000	30000
r	0.99	0.996
L/λ	30	40
Number of channels	90	1000

Stabilization of the filter

In order to get a more stable operation an automatic gain control (AGC) can be provided, as shown in Figure 5. This AGC is conveniently realized by designing a second device identical to the filter except that a slight offset in the gains G and G' of the amplifiers is maintained. This auxiliary device is then operating as a SAW oscillator ($r' = 1$) at the center frequency of the tunable filter ($r < 1$). The AGC is applied to both devices in order to limit the output level of the oscillator. Finally, the gain r in the filter loop is kept at a constant value, whatever the center frequency and the temperature.

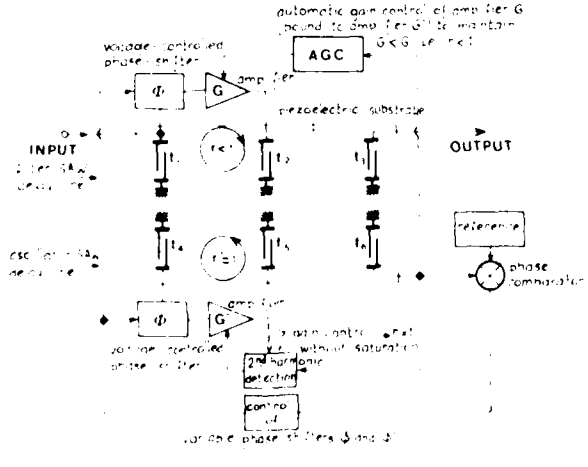


Fig. 5 : Schematic of the stabilized filter

For example Figure 6 describes the first experiment on a lithium tantalate filter stabilized between -10°C and +50°C, and exhibiting a stability of 10^{-4} over the complete range of operation. Better results are expected for narrower fractional bandwidth around 900 MHz.

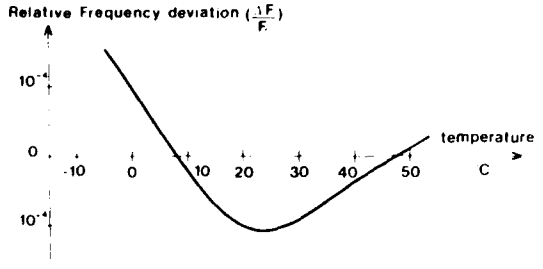


Fig. 6 : Experimental stability of the lithium tantalate tunable filter versus temperature.

Since the goal is rather an overall stability of 10^{-5}, this nevertheless indicates the large interest of berlinite due to its zero frequency temperature coefficient.

Conclusion

We have described an electrically tunable narrow bandwidth SAW filter designed for future vehicular communication systems around 900 MHz. Present results suggest that a berlinite stabilized filter should be necessary to meet the stability requirements in the complete temperature range. This is currently under investigation.

References

1 HENAFF, J., and FELDMANN, M. : "Electrically tunable narrow-bandwidth SAW filter". 1980 IEEE Ultrasonics Symposium, Boston USA, Cat. 80 CH 1602-2, pp.332-335.

2 HENAFF, J., FELDMANN, M., and CAREL, M. : "Voltage-controlled SAW filter on berlinite". Electron. Lett., vol.17, n°2, pp.86-87 (1981).

3 CRABB, J., LEWIS, M.F., and MAINES, J.D. : "Surface acoustic wave oscillators : mode selection and frequency modulation". Electron. Lett. 1973, 9, pp.195-197.

4 HENAFF, J. : "Oscillateurs à ondes élastiques de surface". Onde électrique, 1976, 56, pp.189-196.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

SAW FILTER TECHNOLOGY AND APPLICATIONS

B. R. Potter and D. B. MacDonald

Texas Instruments

Summary

Surface Acoustic Wave (SAW) filter technology has been employed in a number of systems applications. These filters range in frequency from about 30 MHz to as high as 2 GHz. Loss as low as .6 dB has been achieved at 35 MHz and 4.5 dB at 1233 MHz. Narrowband filters utilizing St-Quartz as a substrate will be discussed along with medium and wideband devices constructed on lithium niobate. A variety of transducer technology will be presented which includes narrowbandwidth three phase unidirectional transducers, medium bandwidth apodized three phase unidirectional transducers and very flat frequency response wideband slanted array transducers.

Key Words. SAW filters, group-type transducers, SAW fabrication, E-beam lithography, slanted transducers.

Introduction

Low loss surface acoustic wave (SAW) filters are being utilized effectively in many electronic systems applications. These applications include both military and commercial electronic systems. Millions of television receivers employ SAW filters. Military applications include matched filters for pulse compression radar, delay lines and resonators for highly stable oscillators, front end filters for communication systems such as the global positioning system (GPS) receivers[1], medium bandwidth (3% to 15%) filters for filter bank application in electronic warfare (EW) receivers, and wideband filters for ultraflat delay lines[2] required in some EW applications.

Capabilities of SAW Bandpass Filters and Transducer Technology

Table 1 is a presentation of the electrical capabilities of SAW filters. Some parameters may be possible only at the exclusion of others. For example, a very wideband SAW device cannot have low insertion loss.

Table 1

Capabilities of SAW Bandpass Filters

PARAMETER	PRESENT	PROJECTED
CENTER FREQUENCY	10 MHz TO 2 GHz	2.5 GHz
BANDWIDTH	20 KHz TO .7 F_0	.8 F_0
INSERTION LOSS	2 dB	1 dB
MINIMUM SHAPE FACTOR	1.2	1.2
MINIMUM TRANSITION BANDWIDTH	50 KHz	20 KHz
STRIPline REJECTION	60 dB	70 dB
ULTIMATE REJECTION	90 dB	100 dB
DEVIATION FROM LINEAR PHASE	± 1°	±.5°
AMPLITUDE RIPPLE	.01 dB	.01 dB

The transducer technology employed with a SAW filter will depend on the application, frequency and bandwidth. There are three basic transducer types: bidirectional, unidirectional and chirped. Actually the chirped transducer can also be bidirectional or unidirectional but is definitely a transducer type. The transducers may be arranged in the configurations shown in Figure 1. Figure 1a is the conventional bi-di-

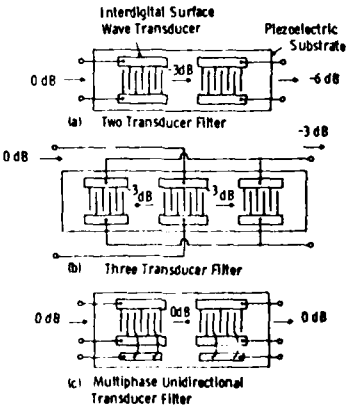


Figure 1. Transducer Configurations

rectional transducer filter that has at least 6dB insertion loss. Figure 1b is a three bi-directional transducer arrangement that has at least 3dB insertion loss and triple transit signals are cancelled at the device center frequency. The outside two transducers may also be unidirectional as will be seen in one of the later examples. Figure 1c is a three phase unidirectional transducer configuration. At high frequencies above about 600 MHz the three phase unidirectional transducer cannot be fabricated because of processing limitations.

The group-type unidirectional transducer may be utilized for frequencies above 600 MHz to build low loss filters. A schematic layout for the group-type transducer is shown in Figure 2. The electrodes are arranged in two groups[1] that are placed 90 degrees out of phase from each other on the substrate. Another 90 degrees of phase shift is applied at the input to the transducer.

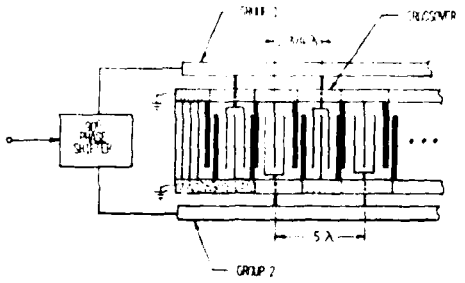


Figure 2. Group-Type Transducer Schematic Layout

Narrow Bandwidth SAW Filters

The SAW transducer technology utilized for each systems application depends on the device insertion loss, bandwidth and center frequency requirements of the system that will employ the device. If a narrow bandwidth (.2 to 3%) SAW filter is required and the frequency is less than about 600 MHz, then three phase unidirectional transducer technology[3] may be utilized effectively and a high quality low loss (<3.0dB) filter will result. If the frequency is above about 600 MHz and the system requirement is still for a narrow bandwidth filter, then the group-type unidirectional transducer[1,4] offers a technique for achieving low insertion loss and good out-of-frequency band rejection. Of course, if low loss and smooth phase response are not required then bidirectional transducers may be employed in any of these applications at somewhat reduced cost.

Medium Bandwidth SAW Filters

Some systems require medium bandwidth filters with low shape factors , low insertion loss, and good out of band rejection. Medium bandwidth SAW filters constructed on ST-cut quartz crystals will have excess insertion loss depending on the bandwidth of the device. Consequently, most medium bandwidth SAW filters are fabricated on lithium niobate or lithium tantalate. A typical design procedure for these filters would be to utilize apodize weighted three phase unidirectional transducers and employ a multistrip coupler to suppress bulk mode spurious responses. At frequencies above 600 MHz the group-type unidirectional may be used for both apodized and withdrawal weighted transducers with good results. Another technique that may be used with excellent results for medium bandwidth devices is to use an apodized bidirectional transducer in between two unidirectional transducers. The bi-directional transducer will allow for good sidelobe suppression and low-loss may be obtained because of the two unidirectional transducers.

Widebandwidth Filters and Delay Lines

Widebandwidth SAW filters have only found application to any extent in matched filters for pulse compression radar. The reason for this is obvious since excellent low loss widebandwidth filters can be constructed using lumped element resistors, inductors, and capacitors Since acoustic delay times are relatively long even for small SAW devices, widebandwidth ultraflat delay lines are being employed in a number of systems applications. Bandwidths as large as 70% have designed and tested. SAW slanted transducer technology[2] is employed in designing these devices and several microseconds of delay is achievable over a 100% bandwidth.

An integral part of SAW filter technology, along with design considerations, is filter fabrication. Lower frequency photomasks have been constructed with pattern generators and the devices are fabricated with highly developed photoresist and wet etch techniques. Fabrication of devices at L-band frequencies requires special mask making and fabrication procedures. The next section addresses fabrication techniques.

L-Band SAW Device Fabrication

Interdigital electrode line width required for fabrication of L-band SAW filters necessitates the use of electron beam pattern generators. Typical line widths are in the range of 1 micron down to 0.325 micron. All device designs reported here were generated using the two step E-beam process reported previously with minor variations.[5] Software designs were converted to card output and filed onto disk using an interpretative pattern defining language. Transducer placement is facilitated using a run description which is filed on the same disk. At run time, the disk is taken to the E-beam pattern generator. A typical mask takes approximately 10 minutes to run.

At this point, the process breaks out into two distinct paths. The first path is characterized by single level devices similar to the L-band

SAC delay line shown in Figure 3. In addition to high resolution requirements, these devices depend critically upon transducer placement accuracy which determines their static delay. Transducer separation is controlled using a laser interferometer stage with electronic beam offset to within ± 1/2 micro inch. This corresponds to ± 3.64 and ± 4.04 picoseconds for YZ-$LiNbO_3$ and ST-quartz respectively.



Figure 3. Transducer Layout for an L-band 1.36GHz, 1.6usec Delay Line

The second path is characterized by three level unidirectional filters. Figure 4 shows a first level high resolution mask which contains the critical interdigitial geometries. This particular example is a 1575 MHz narrow band filter with one apodized weighted and one withdrawal weighted transducer. The additional problem of level to level registration can be seen from Figure 5 which is the second level photomask corresponding to the first level in Figure 4.



Figure 4. Typical First Level L-Band Group-Type Unidirectional Transducer

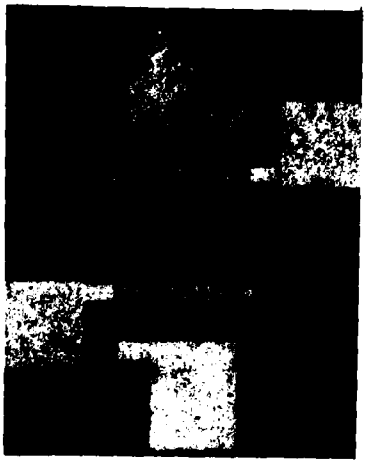


Figure 5. Typical Second Level L-Band Group-Type Unidirectional Transducer

The key problem is alignment of the small metal vias over every third finger without shorting to adjacent electrode stripes. Allowance for etching undercut must be provided at this step. A third level mask with a conducting bar is used to connect every third finger with a metal conducting bar. The third level is the least critical in regard to alignment registration and resolution. The problem of alignment becomes more difficult as the operating frequency of SAW filters increases into the L-band. For critical level to level requirements, improved registration is obtained by using a master mask to replicate reference alignment marks on three edges of a 2 1/2 inch chrome plate. These marks are then used to calibrate the E-beam stage control mechanism.

Levels one and two are typically dark image masks made using 2200°of TI-309 negative E-beam resist on glass plates with 750 to 800 A of chrome metalization. Level three is a clear image mask to facilitate the plating process. It is formed using 2400A of PMMA positive E-beam resist. Resist lithography is transmitted into chrome by wet etching techniques for devices with minimum geometries above 1 micron. Plasma etching in chloroform for approximately two minutes or ion milling is used for submicron geometries down to 0.3 micron.

The masks are then contact printed to $LiNbO_3$ or quartz substrates which have been metallized with 500A of aluminum. Positive Hunt MPR 2:1 photoresist is used to define the critical high resolution interdigital fingers on the first level. Luminum wet etching is used to define the final transducer pattern on its peizoelectric substrate.

Saw Filter Device Results and Applications

Table 2 is a list of the devices to be discussed in this section. The devices are divided up into the three basic groups described before. The devices listed here are not all that have been constructed but only representative of the state-of-the-art in each category.

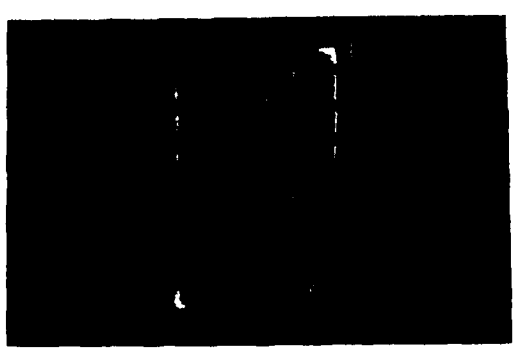
TABLE 2

SAW FILTER DEVICE RESULTS

	CENTER FREQ. MHz	BANDWIDTH MHz, %		INSERTION LOSS dB	SIDELOBE LEVEL DB
NARROW	287	6	2.1	10	-70
	253	1	.4	3	-45
	321	1.7	.5	2	-55
	1233	15	1.2	4.7	-18
	1575	18	1.1	15	-55
MEDIUM	253	21	8.3	9	-60
	165.5	5	3.0	2	-30
WIDE	150	100	67	27	-40
	1500	1000	67	27	-40
	1320	700	53	30	-40

Narrow Bandwidth SAW Filters

287 MHz Mariner Jupiter Device[6]. As shown in Figure 6 the Advanced Transponder System device utilized a commercially available package (Isotronics) and a three bidirectional transducer design. The transducers were withdrawal weighted and the packaged devices had excellent RF feed through characteristics. 70dB near-in sidelobes and 100dB far out sidelobes is the best ever achieved at Texas Instruments. There is phase distortion in the passband due to triple transit signals.



Three-Transducer withdrawal weighted SAW Filter

Figure 6a. Package.

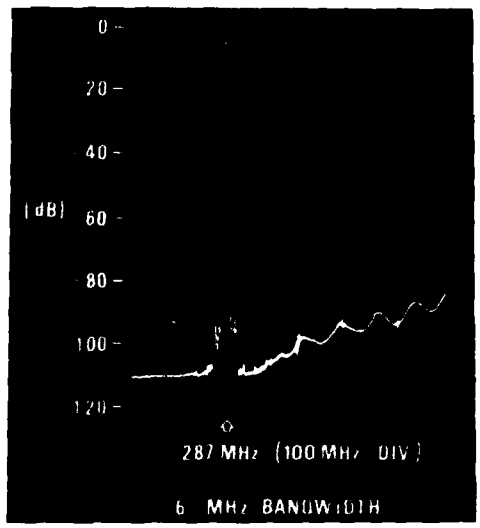


Figure 6b. Frequency Response.

253 MHz Dual Passband Fleet Sat Com Front End Filter[7]. The frequency response for this device is shown in Figure 7. Cosine modulation of the basic transducer time function results in a dual passband frequency response. The device combines low insertion loss of 3dB with good out of band rejection of -40dB. Two withdrawal weighted transducers that utilized Dolph-Chebyshev weighting were employed in the design of this device. The filter was installed in a Motorola Fleet Sat Com receiver on board a Navy Ship between the receiving antenna and the RF front end of the system. Extraneous signals from other radios aboard ship are then rejected by the SAW filter.

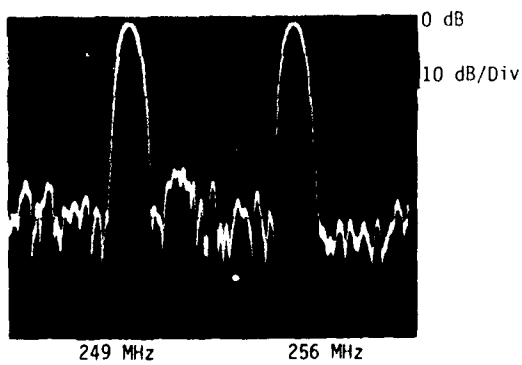


Figure 7. Dual Passband SAW Low Loss Filter Frequency Response

321 MHz Low Loss SAW Filter[3]. This filter is typical of narrow bandwidth SAW filters that combine both low insertion loss and good out-of-band rejection. Two three-phase unidirectional withdrawal weighted transducers were designed for low insertion loss (2dB) and low frequency sidelobes (-55dB). Required matching network is four components per transducer to both match and phase the device. The frequency response of this device is shown in Figure 8.

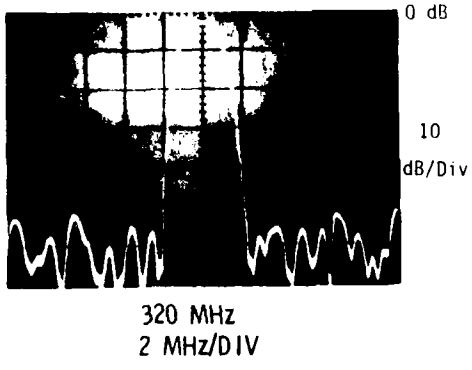


Figure 8. Single Passband SAW Low Loss Filter Frequency Response

L-band Filters for the GPS Receiver[1]. The GPS manpack receiver requires two SAW filters: one at 1227.6 MHz and one at 1575.42 MHz. For temperature stability most of these devices have been constructed on St-cut quartz. However, a prototype device fabricated on lithium niobate had 4.7dB insertion loss and the passband response is shown in Figure 9. Later results on St-Cut quartz utilized both apodized and withdrawal weighted group-type unidirectional transducers. As shown in Figure 10 good sidelobe rejection is achieved over a wide frequency band. The insertion loss is higher than expected because the bandwidth of the device is

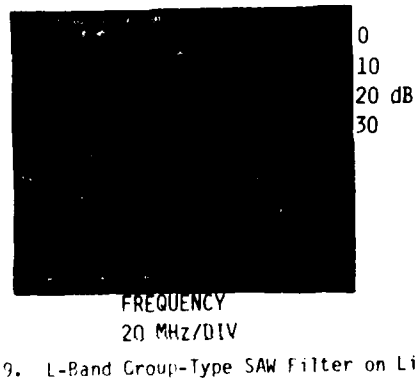


Figure 9. L-Band Group-Type SAW Filter on Lithium Niobate

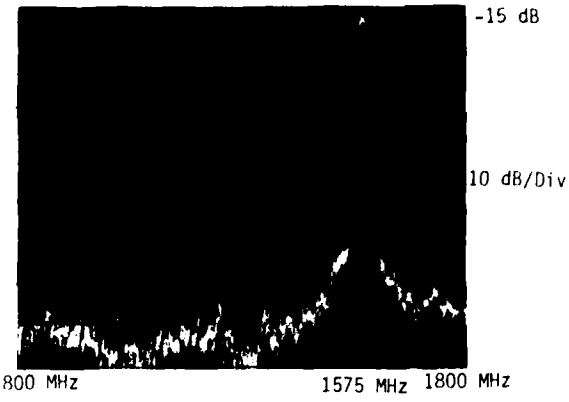


Figure 10. L-Band Group-Type SAW Filter on St-Cut Quartz

more than one percent and energy is being dissipated in parasitic resistance. The phase of these filters varies about ± 3° over the 1 dB bandwidth and the delay variation is less than 10 ns.

Medium Bandwidth SAW Filters

SAW Filters for EW Filter Banks. The filter shown in Figure 11 is utilized in a filter bank for an EW system. Frequency sidelobe requirements are -60dB and spurious time signals when the device is impulsed are down -60dB. Both of these requirements are met by a device that utilizes two, three phase apodized weighted transducers and a multistrip coupler. The bulk modes that are so strong in Y-Z lithium niobate are not transferred by the multistrip coupler and consequently, frequency spurious signals are down greater than -60dB. The transducers are tuned using a four element matching and phasing network. One phase is grounded and a 60° phase shift is created between the other two ports using a pi-phase shift network.

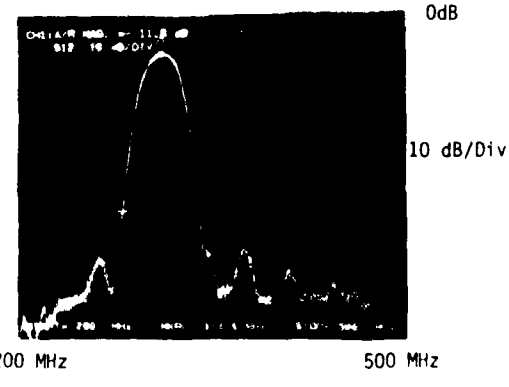
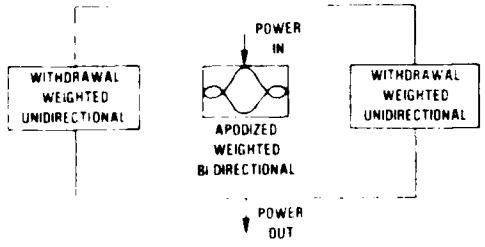
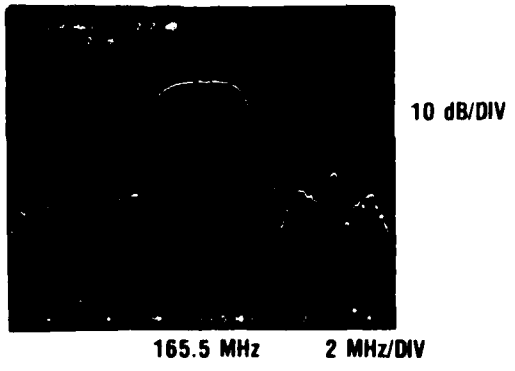


Figure 11. EW Filter Bank SAW Device with -60dB Time Spurious

Low-Loss Medium Bandwidth SAW Filter at 165.5 MHz. The previous filter bank device had 8-10dB insertion loss because of the two apodized transducers and the multistrip coupler arrangement. If low loss is required at medium bandwidth and moderate sidelobe response, then a three transducer configuration as shown in Figure 12 is appropriate. This device was constructed on 128° rotated Y-Cut lithium niobate for the bulk mode suppression characteristic. Insertion loss was 2.2dB and sidelobes of -30dB were required. The 128° cut has better temperature characteristics (76ppm/oc) as compared to the Y-Z lithium niobate (92ppm/oc). Coupling is higher (.059) and the velocity is faster (3884 M/SEC). Beam Steering, however, is a problem and resulted in some excess insertion loss in this particular device. The material is pyro-electric and can break if not cooled slowly during processing.



a. Transducer Layout



b. Frequency Response

Figure 12. Three Transducer Low Loss SAW Filter on 128° Rotated Y-Cut Lithium Niobate

Wideband Width SAW Filters

150 MHz Slanted Correlator. This device was a prototype for the L-band slanted correlator for a microscan receiver program. The device frequency response is shown in Figure 13. It had a passband flat to within .5dB and insertion loss less than 27 dB over a 66% bandwidth. If the slanted transducers are configured like Figure 3 a widebandwidth SAW delay line results.

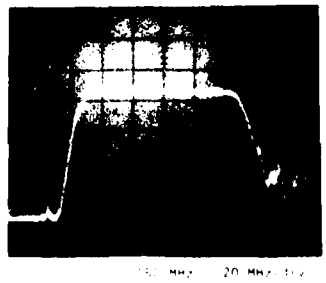


Figure 13. 150 MHz Slanted Correlator Frequency Response

1350 MHz Slanted Transducer Delay Line. This device is utilized in an EW application where ultra-flat delay lines are required over a wide frequency Band. The slanted transducer allows a device to be designed without matching networks as long as compensation for bond lead parasitics, diffraction effects and surface wave attenuation is included in the transducer design. The device has about -30dB insertion loss and is flat to within 1dB over a 700 MHz bandwidth. The group delay is constant at 1.6usec across the entire frequency band. A photograph of the device was shown in Figure 3 and the frequency response is in Figure 14.

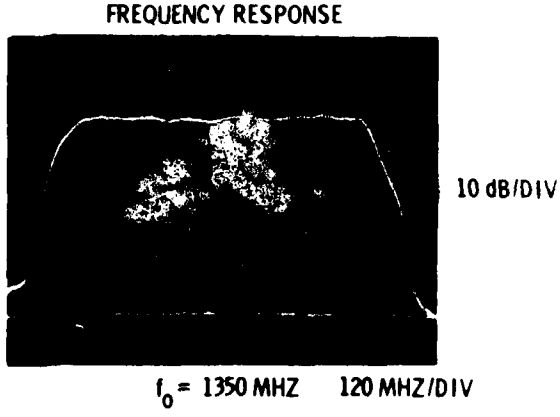


Figure 14. Wideband SAW Delay Line at 1.35 GHz

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

THE STATUS OF MAGNETOSTATIC WAVE DEVICES

J. M. Owens, C. V. Smith, Jr. & R. L. Carter
Electrical Engineering Dept.
University of Texas at Arlington
Arlington, Texas USA 76019

Abstract

The continual demand for increased performance in modern communication and radar systems in terms of increased bandwidths and higher operating frequencies has led to investigation of novel techniques and technologies for analog signal processing. In particular, surface acoustic waves (SAW) have been extensively exploited with great success to this end, but systems requiring bandwidths greater than 500 MHz and center frequencies greater than 1 GHz have pushed SAW devices near the practical physical limit of the technology. A novel technology promising increased bandwidths at higher frequencies is based on magnetostatic waves (MSW) propagating in epitaxial ferrite films such as Yttrium Iron Garnet (YIG). These waves can be exploited in devices offering instantaneous bandwidths up to 2.2 GHz at microwave center frequencies from 0.5 to 20 GHz.

This MSW signal processing technology, based on transversal filtering concepts used in SAW, has been under extensive investigation for the past 8 years. This paper will first discuss physical properties and limits of magnetostatic waves, and consider similarities and differences with SAW. Next, the state of development of MSW technology in the United States and abroad will be summarized. Finally, some significant remaining problems for device application of MSW will be discussed followed by some projections for MSW technology limits.

Introduction

Magnetostatic waves are slow, dispersive, magnetic field dominated electromagnetic waves which propagate in magnetically biased ferrite materials at microwave frequencies (1-20 GHz in Yttrium Iron Garnet). A key to the recent interest in MSW devices has been the development of high quality low linewidth (low loss), large area, Yttrium Iron Garnet ($Y_3Fe_5O_{12}$) grown on Gadolinium Gallium Garnet ($Gd_3Ga_5O_{12}$) substrates as epitaxial films. As an example, magnetostatic waves in EPI-YIG offer lower propagation loss than SAW on lithium niobate at 3 GHz and losses are less than 3 dB/msec at 10 GHz.

Magnetostatic wave propagation in thin films has been extensively considered, and three major propagating modes with the propagation direction in the film plane have been used in device applications. These three modes are determined by the relative orientation of bias field, propagation direction and the crystalline orientation of the ferrite slab normal. The modes are dispersive and characterized by a limited propagation passband width and magnetic bias field tunability. This tunability in principle can permit propagation of a particular wave length at any particular frequency in a range of frequencies. First consider the Magnetostatic Surface Wave (MSSW) mode in which the bias field, $\bar{H}$, is perpendicular to the direction of the wave propagation and in the plane of the film. This mode has highly anisotropic propagation in the plane of the film. This mode has highly anisotropic propagation in the film plane and the mode energy is confined to the "top" surface for forward propagation and to the "bottom" surface for reverse propagation as indicated in Figure 1. A second

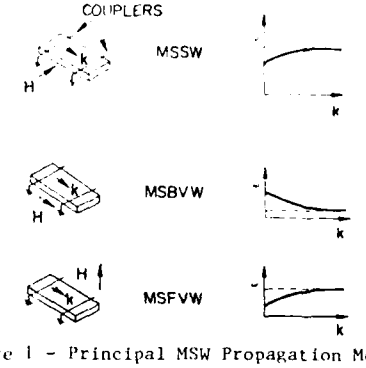


Figure 1 - Principal MSW Propagation Modes

mode is the Magnetostatic Forward Volume Wave (MSFVW) in which the bias field, H, is perpendicular to the film plane and is characterized by nearly isotropic propagation in the plane of the slab with mode energy distributions resembling those found rectangular metal pipe wave-guides. The lowest order mode is usually most easily excited. The third mode, the Magnetostatic Backward Volume Wave (MSBVW) exists when the direction of the bias field and propagation are the same in the film plane. This mode has opposite phase and group velocity directions, is highly anisotropic and is multimoded as in the MSFVW case.

All modes typically have velocities in the 3-300 km/sec range and wavelengths from 1 μm to 1 mm for 10 μm thick epitaxial films; thus, transducer and periodic structure dimensions maybe chosen to permit easy fabrication by standard photolithographic techniques. Since practically any wavelength can be achieved at any frequency, devices designed to work at specific wavelength will also work at any frequency provided a suitable bias field can be obtained and the ferrite is uniformly satur-

ated. Propagation delays of ten to several hundred nanosecond per cm are typical.

The objectives of this paper are to summarize the state of MSW research both in the U.S.A. and abroad, to compare MSW devices with SAW counterparts and to outline primary areas requiring further work. Finally, the future for MSW signal processing will be evaluated.

MSW Propagation and Dispersion Control

The propagation characteristics of magnetostatics waves have been extensively investigated both experimentally and theoretically since the initial study by Damon and Eshbach[1] in 1961. Dispersion control by ground plane proximity has received much attention. Bongianni[2] at Rockwell International first showed that MSSW dispersion characteristics could be significantly altered to produce both non-dispersive and linearly dispersive regions by placing a ground plane in close proximity to the YIG layer. Subsequent work by Tsai et al[3] of the University of Texas at Arlington, and Daniel et al[4] at Westinghouse have shown that a similar situation exists for MSFVW and MSBVW. Adams et al[5] of Westinghouse has shown experimentally using MSFVW that constant delay devices with a 100 nsec/cm delay and a bandwidth of 400 MHz are possible at 8 GHz and that linearly dispersive delays of greater than 1 GHz bandwidth at 10 GHz and a time-bandwidth product of 230 can be realized with an insertion loss of 30 db. Glass et al[6] of Rockwell International has demonstrated that dispersion control can be accomplished through the use of layered magnetic structures. These structures consist of 2 or more magnetic layers of slightly different magnetization spaced by a dielectric layer (either the substrate or an epitaxially grown nonmagnetic layer). This technique produces non-dispersive regions with 200 MHz bandwidths at S-band and 1° delay flatness is achievable with lower propagation loss than is available with metal layered structures. Morgenthaler[7] of Massachusetts Institute of Technology (MIT) has shown dispersion can be accurately controlled by bias field gradients.

All of these studies underscore significantly that MSW's are basically dispersive in nature, with a nonlinear dispersion characteristic and a finite passband width. These studies demonstrated that the MSW dispersion can be closely controlled over a limited bandwidth.

Tranducers

MSW transducers are basically the magnetic analog of the SAW interdigital transducer. Coupling is direct from RF currents to MSW magnetic field components and shorted "fine wire" couplers have been used since experimentation with MSW was first initiated. Ganguly and Webb[8] of the Naval Research Labs first theoretically analyzed the simple "fine wire" MSW coupler using a Poynting Vector analysis, Wu et al.[9] of The University of Texas at Arlington analyzed single wire and simple arrays for MSSW and MSFVW and Parekh et al.[10] of New York State University studied the MSBVW case. Both groups used an array factor model for the simple arrays considered. Experimental work by Wu has shown that very low loss

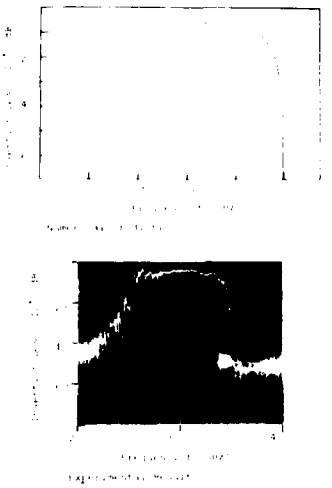


Figure 2 - Simple MSSW Delay Line, Path Length 3mm, Couplers 50μm wide 5μm thick shorted Aluminum Microstrips.

(2 dB) unmatched delay lines are possible using MSSW's (see Figure 2). This is possible due to the anisotropic propagation characteristics of MSSW and the near 50Ω radiation resistance of a shorted "narrow" microstrip. This model assumes independent non-interacting radiators and works well with small arrays but its accuracy decreases with increasing array size. The reason for this inaccuracy lies in the fact that coupling in the MSW case is strong and V_{ij}/V_{ii} for metal on YIG can be 2 or greater. Owens et al[11] of The University of Texas at Arlington have developed a 3-port model for the MSFVW transducer arrays. This model shows greatly increased accuracy for larger arrays (figure 3), but still requires further refinements to yield the high accuracy seen in SAW transducer models. MSW

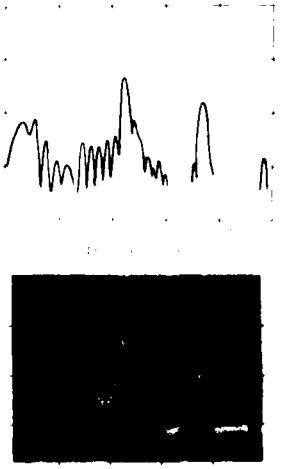


Figure 3 - Theoretical and Experimental Response of a Uniform 13-bar MSFVW Transducer Array Filter using the 3-port model. YIG: 2μm; Alumina: 250μm; Bar Width: 35μm; Bar Length: 3mm; Spacing: 300μm.

devices also suffer from a further problem related to the strongly dispersive nature of the waves. Transducer filters basically rely on spatial filtering techniques and MSW wavelengths within the passband cover a range from infinity at uniform resonance to ~ 1 μm. This implies that spatial harmonic operation presents a significant problem in this type of filter.

Inherently the electromagnetic excitation of MSSW allows directional transduction and this has resulted in very low (2 dB) insertion loss delay lines being reported by The University of Texas at Arlington. In contrast, MSFVW are reciprocal like SAW and with normal microstrip transducers bidirectional transduction is obtained. Vollucet[12] of Thomson-CSF, France, reported at the 1980 IEEE Intermag Conference on the realization of a unidirectional transducer (UDT) for MSFVW. The transducer configuration is similar to that reported in much earlier SAW work by Collins et al[13], in using two basic transducer (microstrip couplers) separated by D on the YIG film (D equal to λ/4 or 5 λ/4) and driven with equal amplitude electromagnetic signals π/2 out of phase. The Thomson-CSF results for a delay line with 2 UDT's operating at D = 200 μm (λ= 800 μm) are an insertion loss of 10 dB at 2.12 GHz for a 10 dB bandwidth of 340 Mhz.

TERMINATIONS

As in the surface acoustic wave case, MSW delay lines require terminations to remove wave reflection at sample ends. The first work by Collins et al[14] provided a significant reduction of end reflections through the use of lossy propagation regions induced by overlaying thin metal films and lossy ferrites. The best terminations seen thus far involve the use of soft magnetic materials[15] or small bias magnets[16] to change the field in the region of the sample ends and effectively increase the propagation delay and thus the loss in the end regions. These terminations are still not as good as silicone rubber in the SAW case, but reflection are not a limiting factor in device design at this point.

PERIODIC REFLECTING STRUCTURES

Several significant SAW devices based on normal and oblique reflection of surface waves from periodic structures have been developed, in particular resonators and R.A.C.'s. The analogs of these devices exist in MSW technology. Initial studies of periodic reflecting structures by Sykes et al[17] Owens et al,[18] have shown that periodic etched grooves and metal arrays (bars and dots) can provide periodic impedance variations in the propagation medium and thus filtering. Either reflector works well in the MSSW case, but etched groove arrays can result in mode conversion in the MSFVW case[19] reducing its usefulness. Studies of oblique reflection of MSFVW by metal arrays[20] yielded the analog of SAW-RAC's at microwave frequencies. Workers at UTA have also demonstrated that strain induced anisotropy created by ion implantation produces an excellent, controllable periodic structure. These reflective array devices show significant advantages over their SAW counterparts in that for etched groove arrays the effective wave impedance is proportional to the film thickness allowing more design flexibility.

RESONATORS

Initial studies by Sykes et al[21] of The University of Edinburgh, U.K. demonstrates that etched groove periodic structures could be used to yield a single port resonator using MSSW. Work by Owens et al,[22] has shown that 2-port Fabry-Perot MSSW resonators with loaded Q's of greater 800 at 3 GHz are possible with octave tunability (Figure 4). These devices had less than 12 dB insertion loss with spurious down another 14 dB.

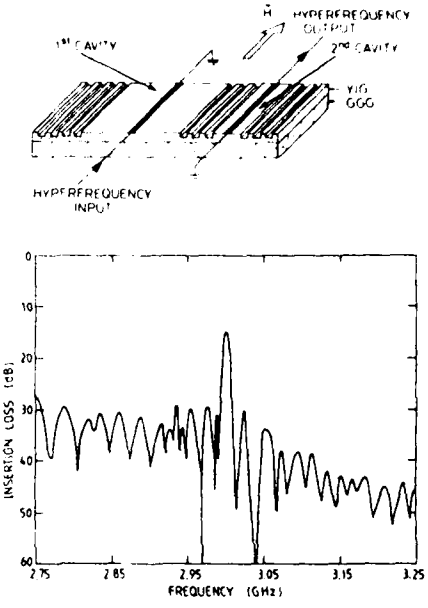


Figure 4 - Two Cavity MSSW Resonator
(a) Schematic
(b) Amplitude versus Frequency Response

In 1979 Castera of Thomson-CSF, France reported results on 2-port S-band single cavity magnetostatic volume wave (MSVW) resonators[23]. Ion beam milled periodic etched groove gratings were used in both magnetostatic forward volume wave (MSFVW) and magnetostatic backward volume wave (MSBVW) configurations. Best results for MSFVW were obtained with an 8 μm wavelength of 300 μm, 56 periods in each grating and a spacing between reflectors of 3 mm. An insertion loss of 12 dB and a loaded Q of 550 at 2.2 GHz was obtained with a single resolved mode; and the inherent coupling of MSFVW with resonant exchange dominated spin wave modes was largely avoided. With MSBVW the results for the same reflector arrays were insertion loss of 24 dB and loaded Q of 870 at 4.35 GHz. Both the MSFVW and MSBVW resonators were tunable by bias field adjustment over an octave. In contrast to 'ring-around' MSSW resonators, these resonators were found disadvantageously to have larger off-resonance transmission due to the reciprocal nature of MSVW. However, these devices allow operation up to +10 dBm c compared with the typical -10 dBm saturation level of MSSW resonators.

Castera, Vallaet and Hartemann[24] reported initial results on cascaded 2-port MSSW resonators. For a single cavity resonator they obtained at 3 GHz, an insertion loss of 12 dB, rejection of 11 dB and a loaded Q of 600. To obtain better rejection of the off-resonance transmission they have investigated a new configuration consisting of two cavities (Figure 4), which is similar to that of a single mode laser. The first resonator is long with two high reflection coefficient arrays. Thus, it is multimoded but has high Q. The second resonator is single-moded but has a lower Q. In cascade a mode of the first resonator is filtered by the second giving overall single-mode resonator with high Q (Figure 4). At 3 GHz the insertion loss is 15 dB, the off-resonance rejection is 15 dB and the loaded Q is 500.

DISPERSIVE FILTERS

Work on linearly dispersive filters have centered on two techniques. First, control of the existing dispersion characteristic by layering, and second, reflective array devices. Adam et al[5] have achieved a MSSW device with 1 GHz bandwidth at 10 GHz and a time-bandwidth product of 230, with a linearity of ± 5 nsec over the bandwidth (Figure 5). Insertion loss was 30-45 db. Also, a backward wave down chirp device has been built with 300 MHz bandwidth and a time bandwidth product of 70. Using reflective array technology Owens et al[20] have achieved a 250 MHz bandwidth device at 3 GHz with a time-bandwidth product of 70, and a linearity of better than ± 2 nsec (Figure 6). While these devices have large instantaneous bandwidths the phase accuracy is still far from that seen SAW RAC's. Initial studies of ion implanted structures indicate that they show promise for controlled RAC devices.

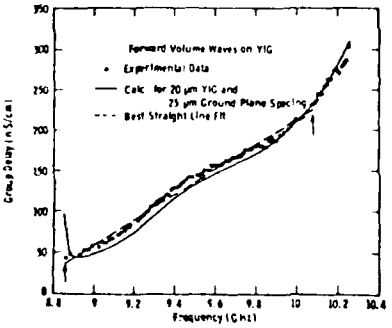


Figure 5 - Group Delay versus Frequency at X-band for a Forward Volume Wave Dispersive Delay Line using a 19.8 μm thick YIG film. Solid curve shows the calculated variation, broken line is best-fit straight line between arrows, and dots show measured group delay.

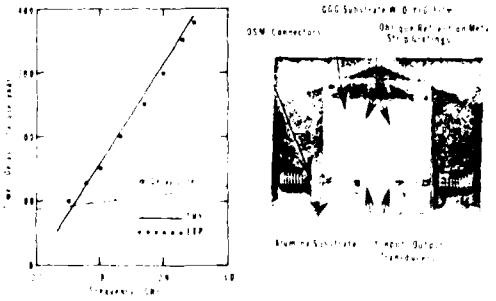


Figure 6 - Quadratic Phase Reflective-Array pulse compression. Comparison between theory and experiment.

DIRECTIONAL COUPLERS

Sasaki and Mikoshiba from Tohoku University in Japan have recently reported[25] theoretical results on the propagation characteristics of MSSW in a layered system essentially consisting of a YIG film, dielectric and a second YIG film. As expected, two modes of propagation occur; a quasi-symmetric mode with wavenumber k_s and a quasi-antisymmetric mode with wavenumber k_a. These degenerate when the dielectric spacing between the YIG films is sufficiently large. Calculations are made in the frequency range 2.5 to 3.5 GHz for 20 μm thick YIG films and a dielectric thickness 250 μm. An MSSW launched on a single YIG film will couple equally with both modes in the axial overlap region of the two YIG films. Power transfer is then periodic with distance. Because the MSSW is strongly dispersive, the directional coupling between the two YIG films is frequency dependent. The coupling length for complete power transfer at 2.9 GHz is 2.24 mm. Thus, this device behaves both as a directional coupler and a frequency filter.

Castera and Hartemann of Thomson-CSF, France, have recently reported[26] the results of detailed studies of a MSSW directional coupler, first reported by Owens at the 1978 Ultrasonics Symposium, using the familiar SAW multistrip coupler (MSC) configuration originally due to Marshall and Paige[27]. The MSC consists of 45 aluminum microstrips 10 μm wide, each separated by 10 μm (L=0.89 mm). For a 19 μm thick YIG film

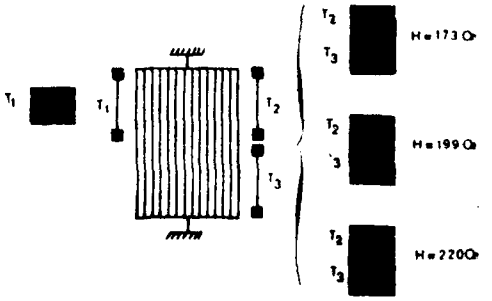


Figure 7 - Transmission Response of a Multistrip Coupler for 3 values of the applied magnetic field: 173 Oe, 199 Oe, and 220 Oe corresponding respectively to No Power Transfer, Half Power Transfer, and Full Power Transfer between the two tracks.

virtually perfect power transfer was observed at 2.528 GHz for a magnetic bias field of 220 Oe, which is closely consistent with the theoretical calculation. Results for no power transfer, 1/2 power transfer and full power transfer are shown in Figure 7. These are obtained at 2.528 GHz purely by adjusting the magnetic bias field. Significantly and unlike the SAW case, the power may be switched at a given frequency to the coupled port rather than the straight-forward port by a bias field change of 50 Oe. Again the directional coupling has frequency filtering characteristics due to the dispersion of the MSSW.

MSSW OSCILLATORS

The first reported work on MSSW oscillators was by Miller & Brown[28] from the GEC Hirst Research Centre, London. The configuration used was essentially a S-band MSSW delay line in the feedback loop of a solid-state amplifier having gain adjustable to 50 dB over its bandwidth of 2.8 to 3.3 GHz. The epitaxial YIG film was 30 μm thick and the 50 μm wide microstrip couplers were separated by 7mm. To encourage single mode oscillation the MSSW delay line passband was narrowed by flipping the GGG substrate so that it contacted the microstrip couplers and hence decoupled the MSW from the electromagnetic ports. A 10 dB insertion loss delay line was obtained with a 3 dB bandwidth of 25 MHz at a particular bias field. The group delay was 33 nsec and the mean k-vector, 500 m^{-1}. Oscillations were obtained covering the 500 MHz amplifier bandwidth by tuning the bias field. However, unexplained discontinuous jumps of 25 MHz occurred approximately every 100 MHz. The medium term stability of the oscillator was poor by SAW standards being .2 MHz/°C.

Castera[29] of Thomson-CSF has reported advances on the Miller & Brown studies. He built a S-band oscillator using multibar microstrip transducers, originally researched at The University of Texas at Arlington. With 6-bar parallel transducers, each 100 μm wide and spaced by 419 μm, a MSSW delay line with a 3 dB bandwidth of [illegible] MHz at 3.2 GHz was obtained. This transducer technique has the disadvantage that high order MSSW responses occur. Bias field tunable oscillation was obtained from 2.7 to 3.8 GHz but again, like Miller & Brown, discontinuous mode jumps (in this case of 10 MHz) occurred. Castera correctly attributed these to the difference between the variation of the oscillator [illegible] versus magnetic field. To overcome the mode jumping problem, Castera then followed [illegible] oscillator principles and built an oscillator [illegible] an MSSW delay line having two output transducers, which are summed at the amplifier input. In this configuration an oscillator, continuously tunable by the bias field, can be realized when the allowable modes for both loops correspond. Experiments were performed with a 10 μm thick YIG film with 10 cm and 15 cm long MSSW paths to the output transducers. Respective delays and mode spacings were 77 nsec and 13 MHz, and 115 nsec and 8.69 MHz, giving modes of oscillation spaced at 26 MHz. By this technique Castera obtained continuous tuning of oscillation in a linear manner with bias field over several hundreds of MHz around 2.1 GHz. However, this oscillator, because of the single microstrip transducers used, exhibited a comb of frequencies spaced by 26 MHz. Thus, judging from European research, MSSW oscillators have not yet been achieved with single mode continuous tuning, low medium term temperature stability, and no measurements are available on the phase noise governing the short term stability.

Carter et al.[30] of The University of Texas at Arlington will report work on an octave tunable resonator based oscillator at the 1981 MTT Symposium. In this work a two port single cavity resonator was placed in the feedback loop of a broad band amplifier to give continuous single mode tuning from 1.8 - 4 GHz. Noise performance measurements indicate that this oscillator is superior to many bulk sphere oscillators with potential significant improvement.

MSW SYSTEM COMPONENTS

Several system oriented MSW devices have been developed to date. However, these devices are intended to demonstrate feasibility not application in any particular system. Adams et al.[31] of Westinghouse has developed a 4 tap programmable MSW taped delay line operating at 9 GHz. The device had a 22 dB insertion loss and when used to as a correlator for a 4-bit Barker code (bit duration 20 nsec) achieved a -10 dB peak to side-lobe correlation peak. Further, the Westinghouse workers have also developed a unique device, the frequency selective signal to noise enhancer, or power expander. This device provides low loss for high level signals and high loss for low level signals. This characteristic is achieved over 1 GHz bandwidths with a differential loss between a +10 dBm signal and a -6 dBm signal of >20 dB. Adams et al.[31] has also achieved temperature compensation of a narrow band MSW filter by utilizing a compensated magnet structure with stability better than 1.2×10^{-5}/°C. This represents a significant step toward system application of these devices.

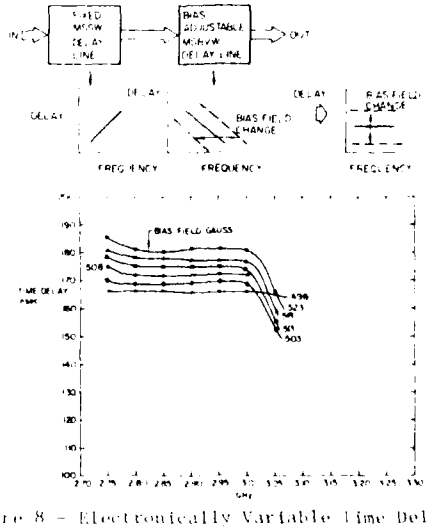


Figure 8 - Electronically Variable Time Delay Device

(a) Schematic
(b) Time Delay versus Frequency, showing bias field tuning

Acknowledgements

This research has been supported by AFOSR under Grant AFOSR-80-0264.

REFERENCES

1. R.W. Damon and J.R. Eshbach, J. Phys. Chem. Solids, 19, 308-320, 1961.

2. W.L. Bongianni, J. Appl. Phys., 93, pp. 2541, 1972.

3. M.C. Tsai, H.J. Wu, J.M. Owens and C.V. Smith Jr., A.I.P. Conf. Proc. #34, pp. 280, 1976.

4. M.R. Daniel, J.D. Adam and T.W. O'Keeffe, Proc. 1979 IEEE Ultrasonics Symp., CH 1482-9/7910000-0800, pp. 806-809.

5. J.D. Adams, T.W. O'Keeffe and M.R. Daniel, Proc. 1980 SPIE Conf., Vol. 111, pp. 241.

6. H.L. Glass and L.R. Atkins to be presented 1980 Ultrasonics Symposium, pp. 5.

7. F.R. Morgenthaller, "Synthesis of Magnetostatic Waves and Modes Using Nonuniform Bias Fields", 1980 Ultrasonics Symposium, IEEE Cat. No. 80CH 1602-2, pp. 532, November 1980.

8. A.K. Ganguly and D.C. Web, IEEE Trans. MTT 23, #12, 998, 1006, 1975.

9. H.J. Wu, C.V. Smith Jr., and J.M. Owens, J. Appl. Phys. 50 (3), pp. 2455-2457, 1977

10. J.P. Parekh and H.S. Tuan, IEEE Trans Mag 16, #5, pp. 1165-1168, 1980.

11. J.M. Owens, R.L. Carter, C.V. Smith, Jr., and G. Hasnain, "A 3-Port Model for Magnetostatic Wave Transducer," 1980 Ultrasonics Symposium, IEEE Cat. No. 80CH1602-2, pp. 538-542, November 1980.

12. G. Vollnet, IEEE Trans Mag 16, 1980.

13. J.H. Collins et al Proc. IEEE 57, pp. 833, 1969.

14. J.H. Collins, D.M. Hastie, J.M. Owens and C.V. Smith, Jr., J. Appl. Phys., pp. 49 (3), Part II pp. 1800-1802, 1978.

15. E. Faber, Electronics Letters, 16, pp. 452, 1980

16. V.L. Taylor, J.C. Sethares and C.V. Smith Jr. "MSW Terminations," 1980 Ultrasonics Symposium, IEEE Cat. No. 80CH1602-2, pp. 562-566, November 1980.

17. C.G. Sykes et al, Appl Phys. Letters, 29, pp. 388, 1976.

18. J.M.Owens et al, IEEE Trans Mag 14, #5, pp. 388, 1976.

19. R.L. Carter, C.V. Smith Jr., and J.M. Owens, IEEE Trans Mag 16, #5, pp. 1159-1162, 1980.

20. J.M. Owens, C.V. Smith Jr., and T.J. Mears II, Int. Microwave Symp. Proc. IEEE #79 CH1439-9-MTTS, pp. 154-156.

21. Sykes et al, Appl. Phys. Letters, 29, p. 388, 1976.

22. J.M. Owens, C.V. Smith Jr., and E.P. Snapka, 1978 INt. Microwave Symp., Proc. IEEE 78CH1355-7 MTT, pp. 440-442, 1978.

23. J.P. Castera, 1979 Int. Microwave Symp. Proc. #79 CH1439-9-MTTS, pp. 157.

24. J.P. Castera, G. Vollnet and P. Hartemann, "New Configurations for Magnetostatic Wave Devices 1980 Ultrasonics Symposium IEEE Cat. No. 80CH 1602-2, pp. 514, November 1980.

25. T. Sasaki and M. Mikoshiba, Electronics Letters. 15, pp. 1772, 1979.

26. T. Marshal and E.G. S.,Paige, Proc. IEEE, 57, pp. 839, 1969.

27. J.P. Castera and P. Hartemann, Electronics Letters, 16, pp. 195, 1980.

28. G. Miller and T. Brown, Electronics Letters 12, pp. 209, 1976.

29. J.P. Castera, IEEE Trans Mag 14, pp 826, 1978.

30. R.L. Carter et al to be presented 1981 MTT-S International Microwave Symposium.

31. J.D. Adams, 1979 Int. Microwave Symp., Proc. 79 CH1439-9 MTTS, pp. 160-161.

32. J.H. Sethares, J.M. Owens and C.V. Smith Jr., "MSW Nondispersive Electronically Tunable Time Delay Elements," Electronic Let 16, pp. 825-826, 1980.

33. K.W. Reed, J.M. Owens and C.V. Smith Jr., and R.L. Carter, 1980 Int. Microwave Symp. Digest, 80 CH1545-3 MTT, pp. 40-42.

Owens et al[32] of The University of Texas at Arlington have developed an electronically variable time delay device utilizing cascaded MSBVW and an MSSW delay lines (Figure 8). By altering the bias field of one device the time delay is varied over 15 nsec with a 200 MHz constant delay bandwidth at 3 GHz. Reed et al.[33] also of UTA have utilized a MSSW delay line with transducer weighting to achieve an active compression loop with 600 MHz bandwidth at 3 GHz and time-bandwidth product of 60 and time sidelobes of -24 dB (Figure 9).

Adams et al.[34] of Westinghouse will report on the first application of a linearly dispersive MSW delay line to a microwave compressive receiver. This system has yielded 12 MHz resolution over a 1 GHz bandwidth.

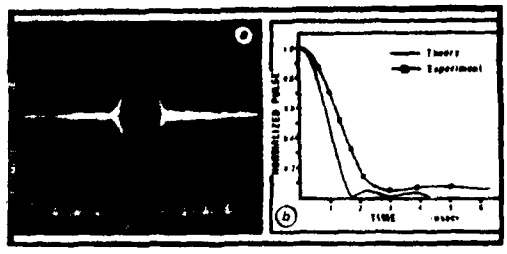


Figure 9 - Recompressed Pulse from Amplitude Weighted MSSW Delay Line in Active Pulse-Compression Loop.

TABLE I MSW PERFORMANCE

Parameter	Tranducer Array		Reflective Array		Resonator	
	Current	Future	Current	Future	Current	Future
Center Frequency (GHz) Tunable	1-20	1-40	1-20	1-40	1-20	1-40
Bandwidth (MHz)	30-1000	10-1500	30-500	10-1800	1-50	5-50
Time Dispersion/ Delay	.5 μsec	2 μsec	.5 μsec	5 μsec	NA	NA
Time x Bandwidth	30	<200	60	>1000	NA	NA
AM Ripple dB	2	<1	4	<1	NA	NA
Phase Error DEG (RMS)	NA	NA	20	5	NA	NA
Temperature Stability	NA	NA	NA	NA	1: 6×10^{-6}	1: $10^{-6}/^{o}C$
Insertion dB Loss	10	<3	30	20	12	6

Present and Potential MSW Device Status

While SAW devices have been under intensive study for the past 12 years, MSW devices have only been under limited active study for about 8 years. Thus, only the basic proof of principles have been reported. Also various technical problems, including higher frequency of operation and dispersion to name a few, makes MSW more complex than SAW in general. Still progress has been significant and the promise is great. Table I shows the present status and future projections for MSW devices. Note that not all parameters are obtainable from a single device. Table II summarizes possible MSW applications.

SUMMARY AND CONCLUSIONS

Surface acoustic wave devices in the VHF, UHF region have little competition as broadband high performance signal processors. Likewise, MSW technology appears to have no competition in the microwave region if its full potential can be realized. While problems remain in the MSW area little effort has been expended to date as compared with SAW and further research should yield much progress.

TABLE II

PRIME APPLICATIONS OF MSW DEVICES

DEVICE	
Resonator	Narrow band filters, Oscillators
Band pass Filters	ECM, Radar, Communication satellite repeaters
Non-dispersive delay lines	Fusing, MTI Radar, Communication path length equalizers, Altimetry recirculating storage loops
Tapped delay lines	Fourier Transformation, Clutter-reference radar, ECM deception, PSK matched filters
Dispersive Delay Lines	Radar pulse compression Variable delay for target simulation and phase array radars, Fourier transforms, Compressive receivers, Group delay equalization

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

DIRECT FREQUENCY CRYSTAL OSCILLATORS

Louis BIDART
&
Jacques CHAUVIN

Compagnie d'Electronique et
de Piezo - Electricité-78500
SARTROUVILLE - FRANCE

Summary

The realization of quartz crystals or Lithium Tantalate resonators in bulk wave is now possible between 500 MHz and 3 GHz.

The new technology of ion etching allows one to obtain resonators working in fundamental mode up to 600 MHz, or on the third overtone mode up to 1,2 GHz, that means to build oscillators in direct frequency, in opposition to multiplied frequencies.

Different cuts have been experimented : AT, BT doubly rotated. The high Q -factor obtained, i.e., 80 000 at 500 MHz, allows the conception of high spectral density and excellent stability oscillators.

Within some months, such high frequency oscillators will be used in aircraft and telecommunication applications.

Introduction

Following a relative stagnation in the field of piezoelectric devices a sudden reversal has been observed recently.

It is believed or was led to believe that a quartz crystal is an obsolete object must revise his opinion and recognize that today bulk mode piezoelectric devices provide the most satisfactory solution for filtering systems and for frequency sources.

Of course, for certain particular conditions, other more efficient solutions may be found ; however they will relinquish the industrial suppleness that is necessary for wide-range applications.

Recent technological advances have enhanced the application of piezoelectricity in three principal directions :

- obtaining very high frequencies - between 400 MHz and 3 GHz - accompanied by an improvement in the frequency noise spectrum of the oscillators
- an increase in the traditional stability of quartz oscillators which in the future may attain better than 10^{-11}/day thereby realizing standards comparable to Rubidium or even Cesium time-bases
- the use of new piezoelectric materials such as Lithium Tantalate or Berlinate.

Although the two last points may seem important they will be referred to occasionally in this paper ; however our main concern is the first point: a widening of the frequency range for modern piezoelectric devices.

Limiting Frequencies

Regardless of an AT or BT cut of quartz or of double rotation, including Lithium Niobate and Lithium Tantalate, the determining factor concerning the resonant frequency - for bulk modes - is the thickness of the resonator. Thus, there is a frequency limit due to the mechanical fragility of the substates as they become very thin. This limit is largely due to the mechanical process used to treat and to finish the resonators.

At CEPE, the industrial limitations (several thousands of componants per month) are as follows:

- AT cut : 50 MHz with a thickness of 33 µ
- BT cut : 75 MHz with a thickness of 34 µ
- SC cut : 50 MHz with a thickness of 37 µ
- Lithium Tantalate : 60 MHz with a thickness of 35 µ

These values are valid only for the fundamental mode. If the finishing of the resonators is adequate then operation on overtones allows the multiplication of the fundamental frequency by an integer n, which corresponds approximately to the rank of the overtone. It is possible for n to attain 13 for the AT cut and 7 for the SC and BT cuts whereas one may easily obtain a 21 st overtone with Lithium Tantalate.

However, the higher the overtone the more difficult it is to adjust the frequency ; thus in practice it is preferable to limit one's objectives to the following :

- 7 th overtone for the AT cut
- 5 th overtone for the BT and SC cuts
- 9 or 11 th overtones for Lithium Tantalate.

In the case of filters, VCXO's or ovened oscillators, adequate frequency control may be achieved by using the fundamental mode or at most the 3 rd overtone.

Thus a first limitation resides in the realization of a resonator functionning in the fundamental mode.

By means of an Helium ion beam the crystal blanks may be brought to the following frequencies:

- 420 MHz in the fundamental mode and 2 GHz for the 5 th overtone for the AT cut
- 600 MHz in the fundamental mode and 3 Ghz for the 5 th overtone for the BT cut
- 450 MHz in the fundamental and 2,5 GHz for the 5th overtone for the SC cut
- 400 MHz in fundamental and nearly 4 GHz in overtones 9 and 11 for Lithium Tantalate.

Working with high level overtones poses certain problems among which figure the state of the crystal surface and the parallelism of the faces ; furthermore, the ratio $\frac{Co}{C}$ which determines the frequency shift of the crystal is altered unfavourably due to the extremely small dimensions of the electrodes.

For crystals produced in the traditional manner, $\frac{Co}{C} \# 300$ whereas for an Helium ion treatment, $\frac{Co}{C} \# 1800$.

Unloaded quality factor Qo=

The unloaded Q of a crystal defines its quality and determines the possible uses of the crystal.

This characteristic depends on the nature of the material as well as the technology that is applied. The constant q.F is very useful in this wise. This means that for any frequency and for a given crystal and a particular cut, Qo F is a constant as may be verified by consulting the adjoined tables.

Loaded Q : Qc

If the unloaded Q of the crystal defines its quality, the loaded Q determines the possibilities of utilization. The loaded Q is not only responsible for operating conditions but must respect certain restrictions due to the associated electronic circuits. For an oscillator, the loaded quality factor will depend on the following points:

- an appropriate electronic circuit
- proper operating temperature
- suitable mechanical environmental conditions
- frequency adjustments or imposed feedback circuits.

In general, one admits that the unloaded quality factor is reduced by three when the resonator is under load, however this factor may vary between 1,5 and 10 according to its utilization ; furthermore this notion is preponderant whenever the spectral quality of an oscillator is concerned.

Series resistance

The unloaded quality factor (Qo) depends principaly on the self_inductance of the resonator which may vary between 0,15 mH for a fundamental mode BT quartz at 600 MHz to 5 mH for a BT quartz operating in the 5th overtone at 500 MHz. At the same time, the resistance may vary from 25 Ω to 400 Ω.

One realizes immediately that the associated electronic circuits will be radicaly different.

One may thus conclude, as a first approximation that the search for a maximal unloaded quality factor (Qo) must make a compromise to achieve a minimum series resistance accompanied by acceptable electronic frequency adjustments.

Drive Level

The drive level of a resonator is important not only for ensuring a high spectral quality but also for obtaining a minimal ageing rate.

A priori, the spectral quality of an oscillator depends on the following points :

- a drive level sufficient to ensure a carrier level superior to the thermal noise
- appropriate electronic components
- the quality factor

It is not possible to consider spectral quality without examining the possible repercussions on ageing, which is a notion that is not entirely clear and that should be examined more closely ; on the other hand, one must make a strict distinction between the drive level as a function of the various cuts and the nature of the piezoelectric material that is employed.

The AT cut rapidly presents isochronomism faults of about - 15 dBm whereas a BT cut remains at 0 dBm.

Experiments have shown that Lithium Tantalate allows more than 5 watts for filters and more than 500 mW for oscillators.

Sensitivity to accelerations

The sensitivity of the frequency of an oscillator to gravitation remains a moot question when the desired precision exceeds 10^{-10}.

An AT quartz realized according to classical methods yields a sensitivity of 1 to $2.10^{-9}/g$ whereas BT cuts give regularly $1.10^{-10}/g$.

It is said but poorly shown that the SC cut of quartz yields a sensitivity of 5 to $1.10^{-10}/g$.

In the sense that sensitivity to accelerations depends in part to the mounting of the crystal, one sees the advantages provided by an ionic treatment since it ensures a relatively neutral crown around the resonator.

This affirmation is corroborated by results obtained on a resonator machined by ionic means under vibrations between 0 and 10.000 Hz. For equal volumes, one should expect to find vibration resonances between 1000 and 3000 Hz whereas, in our case, nothing occurs before 10.000 Hz.

Ageing

The notion of ageing of a piezoelectric resonator - i.e., the comportment of the frequency-time characteristic - is vague and qualitative for the moment ; however the empirical results may be used to advantage in the choice of a model.

As concerns ageing, among the apparent or hidden causes, one may retain the following :

- the nature of the piezoelectric material
- operating conditions such as
  - . the type of the oscillator
  - . the drive level
  - . the temperature range
  - . mechanical constraints
  - . continuous or intermitted operation
  - . the various technologies that may be used.

Ageing contributes to frequency stability in general and serves as a compromise among many stability notions, an only aspect of which is long-term stability.

Table (3) gives the results obtained presently for various resonators realized by means of ionic machining methods.

Measurements

The measurements consist in an analysis of the transfer function of a passive quadripole incorporating a quartz crystal.

The determination of the S-parameters in the vicinity of resonance allows one to specify the characteristics of the resonator. For the moment, the ultimate frequency is 1,3 GHz.

QUARTZ OSCILLATORS

Various electronic structures for oscillators may be taken into consideration.

The primary criterium is the series resistance of the resonator. For 100 Ω or less, starting of the oscillator is facilited as well as the frequency setting ; however, the unloaded quality factor (Qo)does not achieve the maximum value.

On the other hand, a high value of the loaded quality factor leads to difficulties concerning frequency tuning possibilities.

As a result, one must choose between two or three types of oscillators according as one desires spectral purety or proper ageing also taking into account instantaneous or retarded operation.

Finally, of the frequencies considered, the basic electrical circuit must be easily transposed into an integrated network.

Electrical Structures

- A common-base circuit has been used successfully on several oscillators between 400 and 600 MHz for AT or BT cut resonators in the 3rd overtone. Ageing is of the order of 2.10^{-7} for three months. This type of oscillator admits almost all resonators ; however, a priori, it is limited in its frequency range.

- A second "négative resistance" circuit utilizes the negative impedance presented by a common-collector HF transistor. This resistance depresses the resonator and allows oscillations to begin.

This circuit leads naturally to a hybrid configuration for high frequencies, however the level of oscillation may be difficult to control.

Results

The charts that follow resume the results already obtained as well as those to come.

These performances in the laboratory may be improved or adjusted according to specific industrial applications.

Many characteristics may be favorized at the expense of others ; we have tried to reach a compromise taking into account four types of oscillators corresponding to different applications.

- The BP3 oscillator is designed for use in avionics. The spectral purity and the frequency stability are not optimal, however the starting time must be very short.

- The AP3 oscillator is meant for wide diffusion equipment such as the radio-telephone and microwave links. It consists of a frequency source which may be piloted by division by a reference oscillator. Its consists of portable equipment of reduced size and power dissipation (50 mW) with intermitant operation and instantaneous start-up.

- Model SP3 has its principal application in missiles. Starting is instantaneous ; furthermore, these oscillators are conditioned in order to resist to severe storage conditions.

- The LP3 oscillator is characterized by a low -noise level and high output power. It is frequency controllable and, in a latter version, could be transformed into a VCXO.

- The BP5 oscillator is a ground-based version of the BP3. It functions in C.W.

Advantages and drawbacks of resonators machined by ionic beams (R.U.I.)

A new technological innovation using an ion beam for finishing the crystal blanks presents certain advantages. At the start, it was a question of obtaining bulkwave piezoelectrical crystal filters operating in the fundamental mode up to 500 MHz. Now, keeping in mind the initial objective, this new technique provides extended possibilities in the domain of crystal oscillators :

- the frequency may be extended to 3 or 4 GHz
- the resonator is very small and is easily ovened
- vibrations present no particular problems
- sensitivity to acceleration is reduced
- the frequency temperature characteristic is not altered and remains that of the original substrate
- ageing and retrace are not affected ; furthermore, they will profit from modern technologies such as plasma mettalization and ion implantation
- one may easily accentuate one of the following characteristics : frequency stability, spectral purity, output power. The frequency sources may be naturally stable or else locked to an other source compensated analogue or digital means in temperature (TCXO) or voltage frequency control (VCXO).

The main drawback is that one can no longer rely on classical realization techniques which implies the use of integrated circuits as much as possible.

Concerning fabrication cost, it is relatively still high for filters comprising a large number of resonators, however, it is acceptable for a frequency source.

In conclusion, the possibilities that are offered are sufficiently attractive to justify general research project on this subject, which is actually under way.

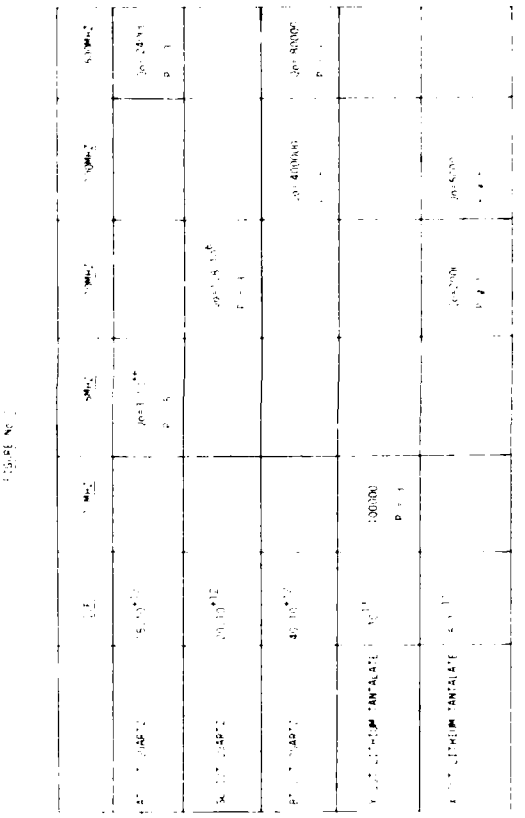


Fig. 1.

Bibliography

1.- R.N. Castellano and J.L. Hokanson " A survey of Ion Beam Milling Techniques for Piezoelectric device Fabrication "
Proc. 29th Annual Symposium on Frequency Control, pp. 128 - 134, 1975.

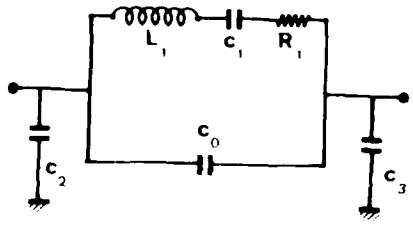
2.- M. Bertó " Acoustic-Bulk-wave resonators and filters operating in the fundamental mode at frequencies greater than 100 MHz "
Electronics Letters, VOL. 13, 1977, pp. 248.

3.- F.F. Schmitt " High-Q BT-cut resonators in flat configuration "
Proc. of the 28th Symposium on Frequency Control, 1974, pp. 67.

4.- P. Hartemann " Acoustic surface wave resonator using ion-implanted gratings "
1975 Ultrasonics Symposium Proceedings, IEE Cat. 75 CHO 994-4SU, 1975, pp. 303.

VHF Quartz Crystal Resonators

Equivalent Circuit





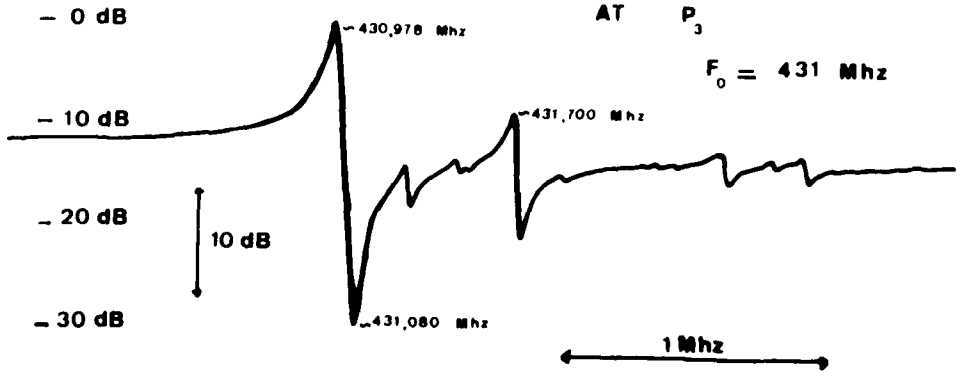
Measurements (Typical Values)

AT P_3	BT P_3
F_0 = 450 Mhz Q = 24000	F_0 = 465 Mhz Q = 76000
L_1 = 2,2 mH	L_1 = 3,04 mH
R_1 = 250 Ω	R_1 = 116 Ω
C_1 = 5,6 10^{-5} pF	C_1 = 3,8 10^{-5} pF
C_0 = 0,6 pF	C_0 = 0,7 pF
C_2 = 0,4 pF	C_2 = 0,8 pF
C_3 = 0,4 pF	C_3 = 0,8 pF
Q.F = 11 . 10^{12}	Q.F = 35 . 10^{12}
Q.F = 15 . 10^{12}	Q.F = 40 . 10^{12}

Transmission Measurements

AT P_3

F_0 = 431 Mhz



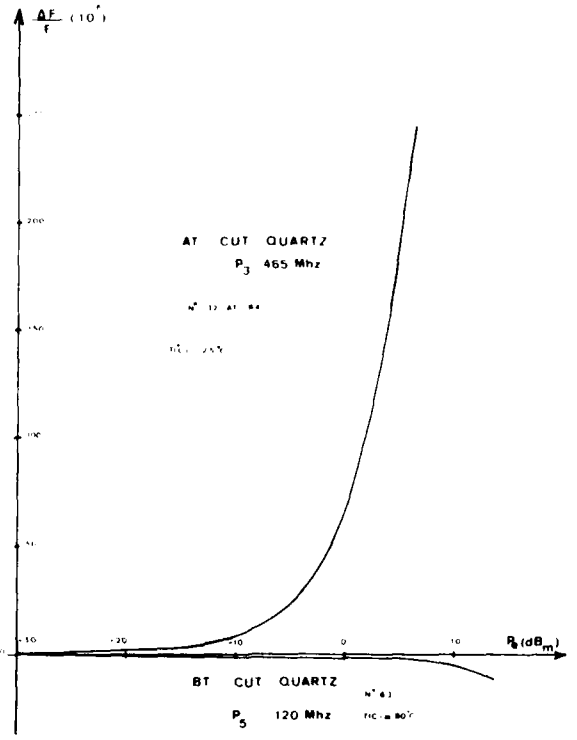


Fig. 2.

FIGURE No 3

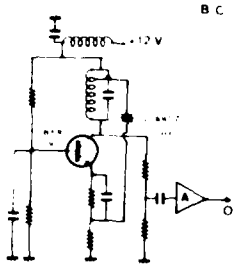
	Qo	L	r	Co/c	Excitation Level	Δf/g	Annual Ageing rate	Frequency
AT P1	67000	3 mH	30 Ω	1800	80 µw	3.10^{-9}	3.10^{-6}	100 MHZ
AT P3	24000	2,2 mH	250 Ω	30000	100 µw	1.10^{-9}	1.10^{-6}	450 MHZ
AT P5	25000	4 mH	550 Ω	60000	150 µw	1.10^{-9}	1.10^{-6}	500 MHZ
BT P1	30000	0.150 mH	20 Ω	3000	200 µw	5.10^{-10}	1.10^{-6}	575 MHZ
BT P3	80000	2,5 mH	90 Ω	35000	500 µw	1.10^{-10}	1.10^{-7}	470 MHZ
BT P5	36000	5,4 mH	420 Ω	70000	500 µw	1.10^{-10}	1.10^{-8}	475 MHZ
LITA03 P1	300	0.032 mH	150 Ω	100	50 µw		10.10^{-6}	240 MHZ
LITA03 P3	750	0.032 mH	150 Ω	750	100 µw		5.10^{-6}	720 MHZ

Fig. 3.

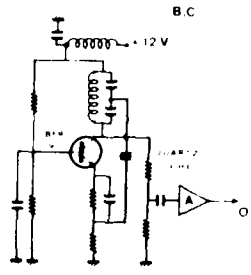
VHF Quartz Crystal Oscillators

1) Reaction

Hartley Oscillator



Colpitts Oscillator



2) Negative Dynamic Resistance

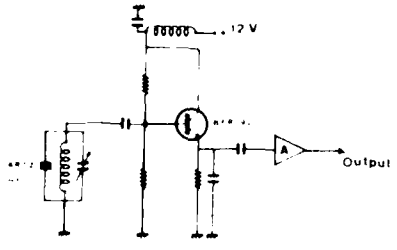


Fig. 5.

Principal Characteristics of the Possible Oscillators

AP_3 Oscillator

frequency f_0 = 500 mHz

AT cut resonator P_3 Q > 20000

stability

— temperature (-20C +70C) $\pm 5.10^{-6}$

— long term $\pm 3.10^{-6}$/year

frequency noise L(fm)

— $f_0/2Q$ < 50 kHz

— ground level ≃ -150 dB_c/Hz

BP_3 Oscillator

frequency f_0 = 500 mHz

BT cut resonator P_3 Q > 60000

stability

— temperature (-40C +85C) $\pm 1.10^{-9}$

— long term $\pm 5.10^{-7}$/year

frequency noise L(fm)

— $f_0/2Q$ < 15 kHz

— ground level ≃ -160 dB_c/Hz

SP_3 Oscillator

frequency 500 mHz

SC cut resonator P_3 Q > 30000

stability

— temperature (+10C +80C) $\pm 2.10^{-6}$

— temperature (-40C +80C) $\pm 20.10^{-6}$

— long term 1.10^{-6}/year

frequency noise L(fm)

— $f_0/2Q$ < 20 kHz

— ground level ≃ -160 dB_c/Hz

LP_3 Oscillator

frequency 500 mHz

$LiTaO_3$ resonator P_3 Q > 1000

stability

— temperature (-40C +80C) $\pm 20.10^{-6}$

— long term $\pm 5.10^{-6}$/year

frequency noise L(fm)

— $f_0/2Q$ < 200 kHz

— ground level ≃ -170 dB_c/Hz

BP_5 Oscillator

frequency 1,5 GHz

BT cut resonator P_5/P_3 Q > 20000

stability

— temperature (+10C +50C) $\pm 1.10^{-7}$ ovened at 60°c

— long term $\pm 2.10^{-6}$/year

frequency noise L(fm)

— $f_0/2Q$ < 30 kHz

— ground level ≃ -160 dB_c/Hz

Fig. 6.

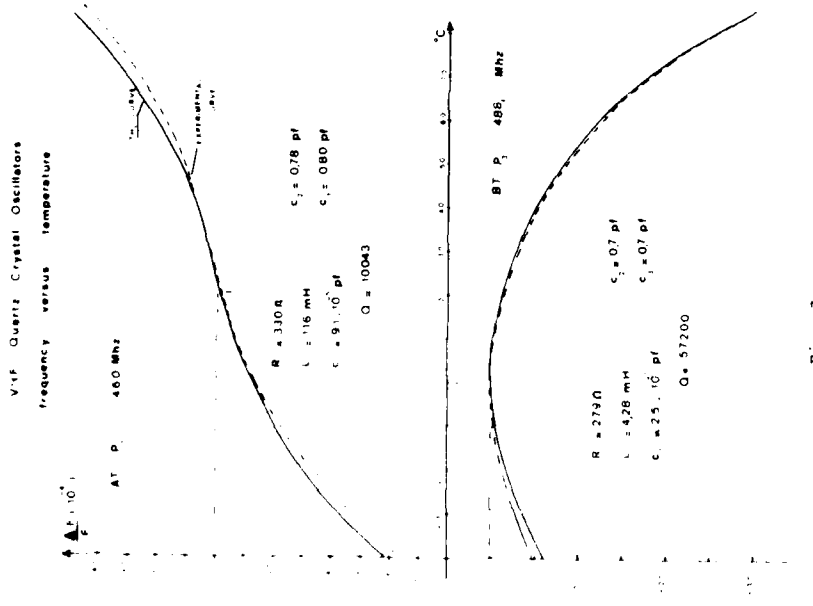
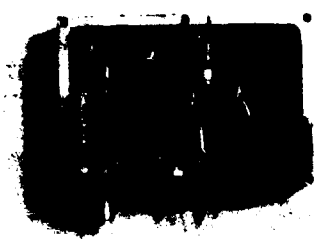
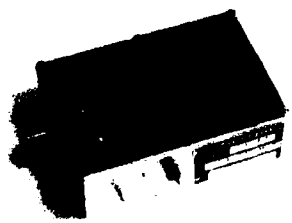


Fig. 7.

AT P_3 Quartz Cristal Oscillator

400 Mhz 500 Mhz





Bred Board Model N_2^o



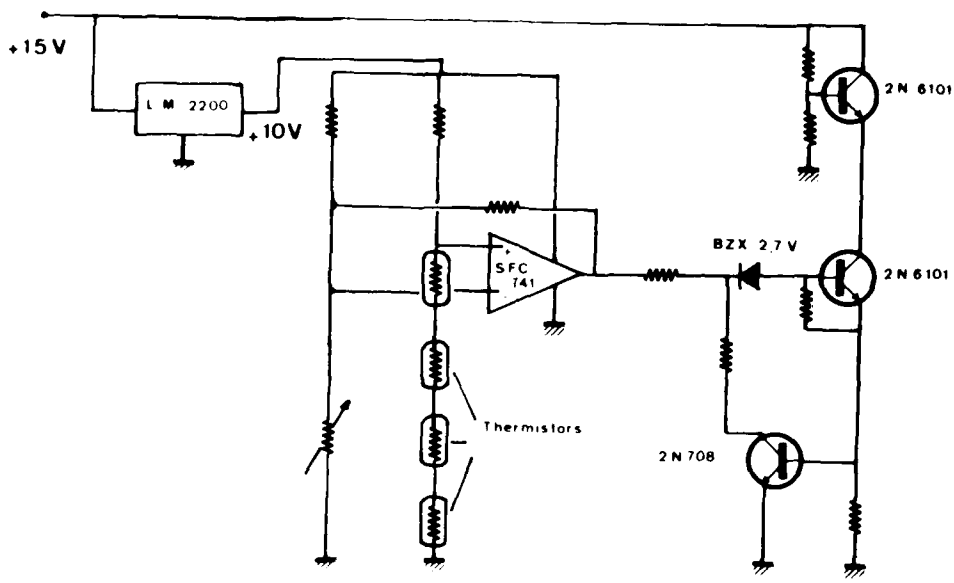
Oscillateur HARTLEY 500Mhz

VHF Oscillator with discrete components

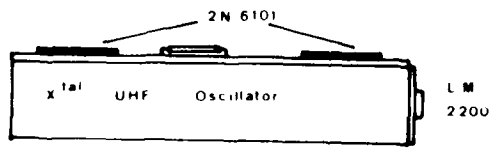
Fig. 8.

BT P_3 Thermostat

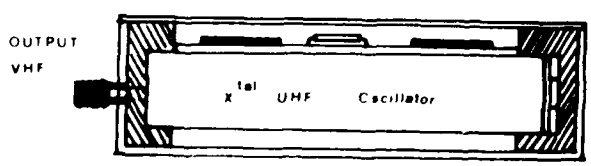
Principal Scheme



Mecanical Description



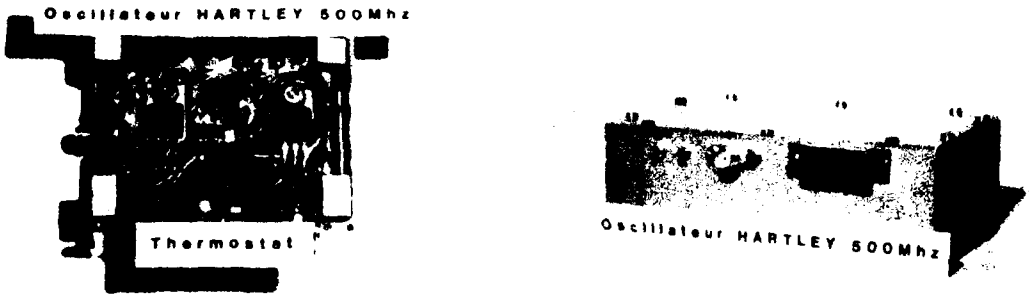
STYLE CASE



EXTERNAL DIMENSIONS : 75 x 60 x 30 (mm)

Fig. 9.

VHF BT P_3 Quartz Crystal Oscillator (ovened)



Ageing

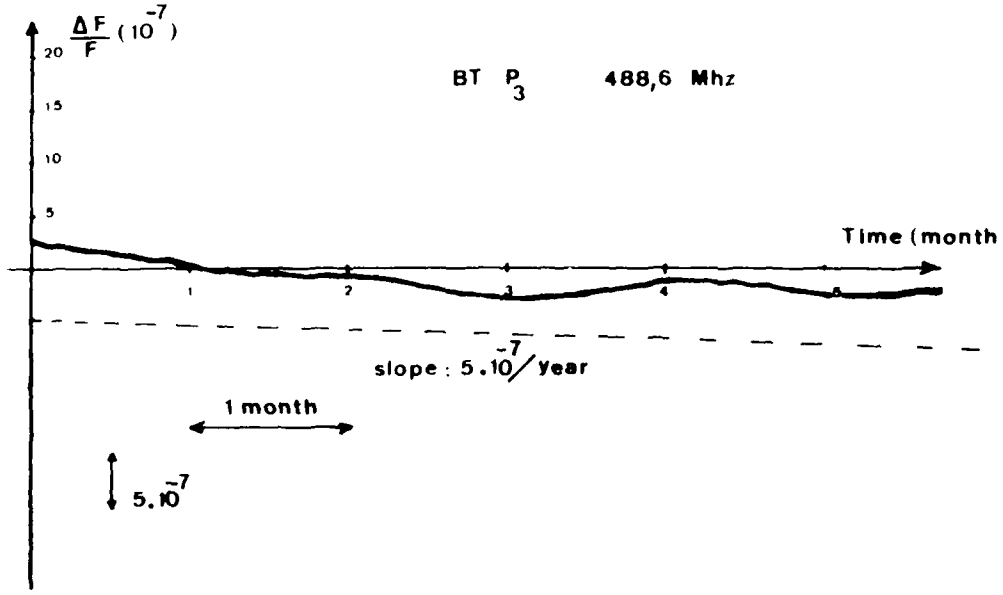
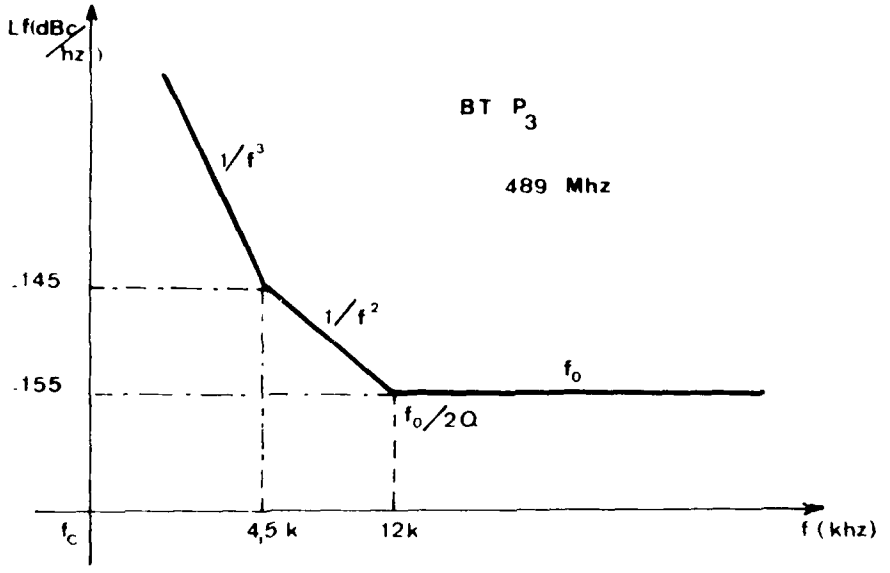


Fig. 10.

Phase Noise Measurements

Experimental Values



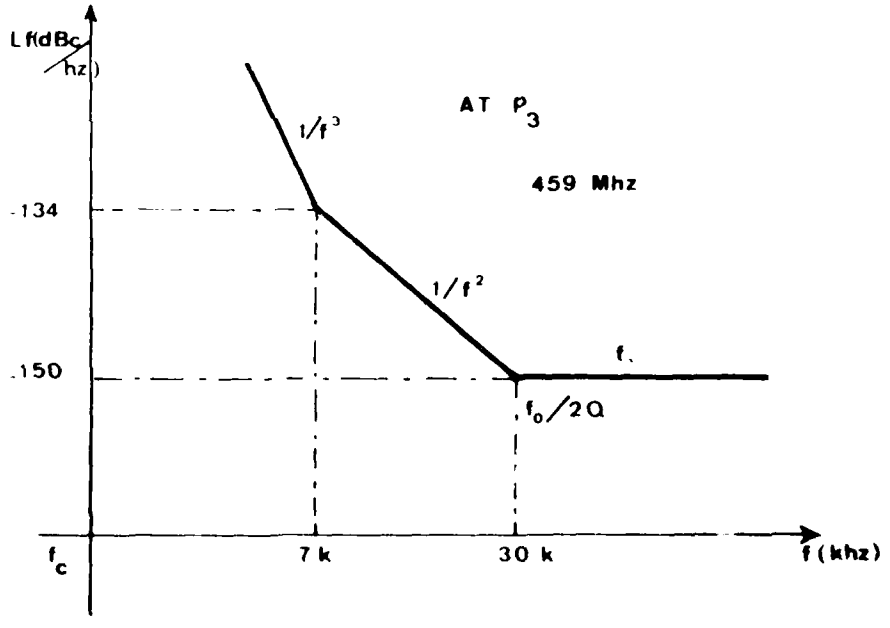
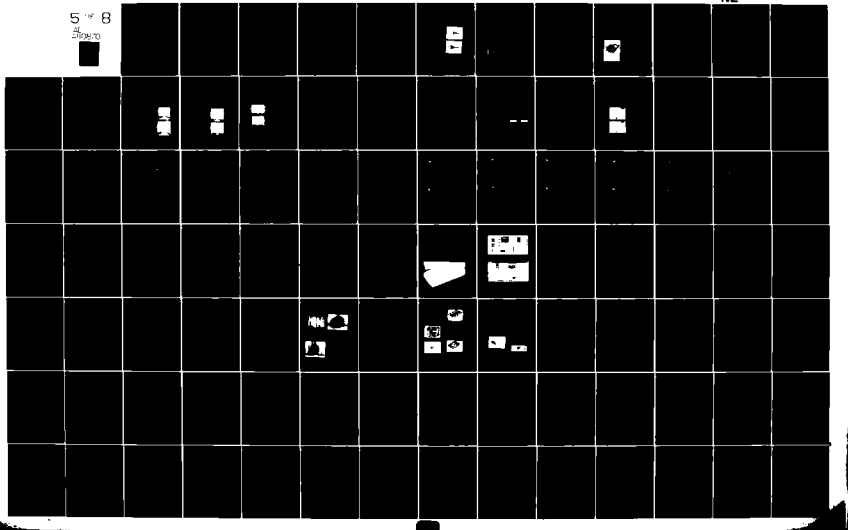


Fig. 11.

AD-A110 870 ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND AD--ETC F/G 9/5
PROCEEDINGS OF THE 35TH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL, --ETC(U)
1981

UNCLASSIFIED

NL



Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

THE PROPAGATION CHARACTERISTICS OF SURFACE ACOUSTIC WAVES ON SINGLY AND DOUBLY ROTATED CUTS OF QUARTZ

D.F. Williams and F.Y. Cho
Motorola Government Electronics Division

A. Ballato and T. Lukaszek
U. S. Army Electronics Research & Development Command

ABSTRACT

The temperature variation of surface acoustic wave (SAW) propagation characteristics on doubly rotated cuts of alpha quartz are discussed. Extensive numerical calculations were used previously to identify families of doubly rotated cuts on which surface acoustic waves exhibit a greater delay time temperature stability than previously reported. Second order temperature coefficients of frequency (TCF) as low as -1.3×10^{-8} and zero first order temperature coefficients of frequency were calculated for these families of cuts. In this paper, experimental measurements of the temperature coefficients of the frequency of selected members of these families of doubly rotated cuts of quartz are presented. The experimental measurements of the temperature coefficients of frequency are compared with the previously calculated values. Phase-temperature plots, coupling coefficients and velocities for these cuts are also presented.

In addition, numerical calculations of the temperature dependence of the surface acoustic wave power flow angle and of the anisotropy parameter γ, which is related to the diffraction effects of the surface acoustic wave, are presented. Power flow variations of 3 degrees and variations of the anisotropy parameter γ of 10 percent (5 percent for the ST-cut) over the temperature range -50°C to 150°C are typical.

Experimental measurements of the temperature variation of the surface acoustic wave power flow angle (PFA) are shown to be in close agreement with the calculated variation. The effect of the temperature dependence of the power flow angle and the anisotropy parameter γ on surface wave filters, reflective array devices, and delay lines are discussed both for doubly rotated and the more popular singly rotated orientations. Techniques for utilizing the temperature dependence of the surface acoustic wave power flow angle to produce temperature stable delay lines are presented. Choice of optimum doubly rotated cuts of quartz for this purpose is discussed, and experimental results are presented.

INTRODUCTION

Recently, much attention has been given to doubly rotated orientations of quartz suitable for Surface Acoustic Wave (SAW) propagation. Previous computation[1] has pointed to the existence of orientations with improved temperature stability. Additional experimental data presented here confirms these theoretical predictions. The temperature variation of the SAW power flow angle has been reported previously[2]. The power flow angle and anisotropy parameter temperature variation for temperature stable doubly rotated cuts is reported here, and a temperature compensation technique utilizing the phenomena is demonstrated.

ANALYTICAL APPROACH

IRE standard angle definitions (YX wlt) PHI/THETA/PSI for quartz were used throughout the investigation[3]. Using the Finite Difference approach with the available crystal constants, previously calculated results show that the zero TCF^1 surfaces do not intersect with the zero TCF^2 surfaces, based on the interpolated results of the 10° x 10° x 10° resolution. It is not likely that a finer resolution will provide contrary information, because TCF^1 and TCF^2 are relatively slow varying functions. Calculations were performed on a 10° x 10° x 10° grid over the angular ranges $0 \leq PHI \leq 30°$, $0 \leq PSI \leq 180°$, and $-90° \leq THETA \leq 90°$. These initial calculations defined the "angular volumes" of low TCF orientations. Calculations were then performed on a 2.5° X 2.5° x 2.5° grid near promising orientations. Calculations were then performed on a 1° x 1° x 1° grid near the best cuts located by this technique. In this way, the entire angular range was explored and a large computer-based data file built. Maps of first and second order TCF's were generated. Despite the number and density of points at which the first and second order TCF's were calculated, wherever TCF^1 was found to be less than or equal to zero, we found TCF^2 to be less than 0.

Table 1 consists of a summary of the results of using the investigative approach described above. Out of the many areas with low TCF cuts, some were identified in this program and some were previously identified. These orientations have zero TCF^1 calculated by the approach of Sinha and Tiersten with TCF^2 and TCF^3 calculated using the Finite Difference approach. These areas are chosen because of zero TCF^1 and low TCF^2. TCF^3 can be mostly cancelled out by TCF^1 if the propagation direction is slightly rotated away from the zero TCF^1 direction, so that the TCF^2 term will dominate the performance characteristics.

The doubly rotated cuts in the angular region near (YX wlt) 15°/30°/38° have a calculated TCF^2 of less than $1.0 \times 10^{-8}/(C°)^2$. The doubly rotated cuts in the angular regions near (YX wlt) 15/40/40, (YX wlt) 12.5/35/130, (YX wlt) 7/27/135, and (YX wlt) 0/27/138 have calculated TCF^2 with magnitudes less than $1.5 \times 10^{-8}/(C°)^2$. In all of these regions, the low calculated TCF^2 indicates the suitability of these cuts for fabricating highly temperature-stable SAW devices on doubly rotated cuts of quartz.

To ensure the suitability of the cuts described above for SAW applications, the coupling coefficients, SAW velocity, and power flow angles were calculated for these cuts and are summarized in Table 2. Table 3 from reference 1 contains a summary of $\partial TCF^1/\partial\psi$. Table 4 from reference 1 contains a summary of the calculated values of $\partial TCF^1/\partial\phi$ and $\partial TCF^1/\partial\theta$. The large values of $\partial TCF^1/\partial\phi$, $\partial TCF^1/\partial\theta$, and $\partial TCF^1/\partial\psi$ impose strict fabricational tolerances on the SAW cuts and on mask alignment. All of the cuts were found to be suitable for SAW devices, however.

EXPERIMENTAL VERIFICATION

Wafers with the orientations listed in Table 1 were fabricated. Considerable care has been taken to reduce fabrication tolerances of the devices fabricated for this study. For these cuts, ϕ and θ are known to be within ± 4 minutes. ψ is known within ± 15 minutes. Wafers fabricated at Motorola as well as a commercially supplied ST-cut wafer have been tested. A delay line oscillator was used to measure the temperature dependence of the SAW delay time. The experimental apparatus is shown in Figure 1. The complete experimental error is estimated to be within ± 20 ppm.

Table 1. Propagation Characteristics of Crystal Orientations

ANGLES OF ZTCF(1) DEGREES (S AND T'S PROGRAM)			TCF(2)/°C2 (X10-6) FINITE DIFFERENCE PROGRAM	TCF(3)/°C3 (X10-10) FINITE DIFFERENCE PROGRAM
PHI	THETA	PSI		
6	26	136.31	-1.4	
6	27	135.93	-1.3	0.67
6	28	135.59	-1.3	0.57
7	26	135.99	-1.5	
7	27	135.64	-1.4	
7	28	135.27	-1.3	0.65
8	26	135.74	-1.4	0.65
8	27	135.36	-1.4	
8	28	134.97	-1.3	
1	26	137.78	-1.2	0.68
1	27	137.48	-1.2	0.65
1	28	137.17	-1.1	0.67
0	26	138.07	-1.2	0.67
0	27	137.78	-1.1	0.68
0	28	137.49	-1.1	0.62
-1	26	138.37	-1.2	0.60
-1	27	138.09	-1.2	0.62
-1	28	137.80	-1.1	0.73
14	39	40.195	-1.0	0.64
14	40	40.415	-1.0	0.66
14	41	40.64	-1.0	0.75
15	39	39.79	-1.0	0.63
15	40	40	-1.0	0.74
15	41	40.23	-1.0	0.73
16	39	39.4	-1.0	0.68
16	40	39.605	-1.0	0.66
16	41	39.825	-1.1	0.60
7.5	35.0	41.77	-0.95	0.58
10	35	40.82	-0.94	0.58
12.5	30	38.88	-0.93	0.57
12.5	32.5	39.4	-0.93	0.57
15.0	30.0	38.12	-0.93	0.57
15.0	32.5	38.55	-0.93	0.57
17.5	30.0	37.35	-0.94	0.56
20.0	30.0	36.6	-0.97	0.55
22.5	30.0	35.85	-1.0	0.54
25.0	30.0	35.07	-1.0	0.52
27.5	30.0	34.28	-1.1	0.50
7.5	35.0	132.68	-1.1	0.54
10.0	40.0	129.4	-1.1	0.57
12.5	35.0	130.62	-1.0	0.41
12.5	42.5	127.15	-1.1	0.58
15.0	45.0	124.43	-1.1	0.58
17.5	40.0	126.14	-1.1	0.54
20.0	40.0	124.92	-1.0	0.44

Table 2. Propagation Characteristics of Selected Orientations

ANGLES OF ZTCF(1), DEGREES (S AND T'S PROGRAM)			VELOCITY (MSEC)	K2 (X10-3)	POWER FLOW ANGLE (DEGREES)
PHI	THETA	PSI			
6	26	136.31	3288.84	1.12	-0.3
6	27	135.93	3289.89	1.12	-0.2
6	28	135.59	3290.83	1.12	-0.1
7	26	135.99	3303.33	1.12	-0.5
7	27	135.64	3299.70	1.12	-0.4
7	28	135.27	3296.33	1.12	-0.3
8	26	135.74	3310.15	1.12	-0.7
8	27	135.36	3306.11	1.12	-0.6
8	28	134.97	3302.32	1.10	-0.5
1	26	137.78	3268.80	1.10	+0.7
1	27	137.48	3267.44	1.10	+0.9
1	28	137.17	3266.36	1.10	+1.0
0	26	138.07	3264.09	1.12	+0.9
0	27	137.78	3263.09	1.10	+1.1
0	28	137.49	3262.36	1.10	+1.2
-1	26	138.37	3259.65	1.10	+1.1
-1	27	138.09	3259.01	1.10	-1.3
-1	28	137.80	3258.64	1.08	+1.5
14	39	40.195	3299.80	0.96	-7.7
14	40	40.415	3306.67	0.96	-8.1
14	41	40.64	3313.19	0.94	-8.8
15	39	39.79	3301.82	0.96	-7.8
15	40	40.00	3310.14	0.94	-8.3
15	41	40.23	3319.09	0.96	-8.6
16	39	39.4	3305.38	0.96	-8.0
16	40	39.605	3314.03	0.96	-8.4
7.5	35.0	41.77	3262.43	1.00	-4.8
10	35.0	40.82	3264.23	1.01	-5.3
12.5	30.0	38.88	3243.27	1.01	-4.1
12.5	32.5	39.4	3254.49	1.00	-4.9
15.0	30.0	38.12	3244.99	0.98	-4.6
15.0	32.5	38.55	3257.82	0.98	-5.3

Table 3. $\partial TCF^{(1)}/\partial\psi$ for Selected Cuts

Angles of ZTCF(1), Degrees (S and T'S Program)			$\partial TCF^{(1)}/\partial\psi$
PHI	THETA	PSI	
6	26	136.31	+2.7 (PPM/C°)/DEGREE
6	27	135.93	+2.7
6	28	135.59	+2.7
7	26	135.99	+2.7
7	27	135.64	+2.7
7	28	135.27	+2.7
8	26	135.74	+2.7
8	27	135.36	+2.7
8	28	134.97	+2.7
1	26	137.78	+2.8
1	27	137.48	+2.8
1	28	137.17	+2.8
0	26	138.07	+3.0
0	27	137.78	+3.0
0	28	137.49	+3.0

Table 3. $\partial TCF^{(1)}/\partial\psi$ for Selected Cuts (Cont)

Angles of $ZTCF^{(1)}$, Degrees (S and T'S Program)			$\partial TCF^{(1)}/\partial\psi$
PHI	THETA	PSI	
−1	26	138.37	+3.0
−1	27	138.09	+3.0
−1	28	137.08	+3.0
14	39	40.195	−3.5
14	40	40.415	−3.5
14	41	40.64	−3.5
15	39	39.79	−3.5
15	40	40	−3.5
15	41	40.23	−3.5
16	39	39.4	−3.7
16	40	39.605	−3.7
16	41	39.825	−3.7
7.5	35.0	41.77	−3.3
10.0	35.0	40.82	−3.4
12.5	30.0	38.88	−3.3
12.5	32.5	39.4	−3.2
15.0	30.0	38.12	−3.4
15.0	32.5	38.55	−3.6
17.5	30.0	37.35	−3.4
20.0	30.0	36.6	−3.5
22.5	30.0	35.85	−3.5
25.0	30.0	35.07	−3.3
27.5	30.0	34.28	−3.4
7.5	35.0	132.68	+2.7
10.0	40.0	129.40	+2.8
12.5	35.0	130.62	+2.0
12.5	42.5	127.15	+2.5
15.0	45.0	124.43	+2.6
17.5	40.0	126.14	+2.1
20.0	40.0	124.92	+2.6

Table 4. $\partial TCF^{(1)}/\partial\phi$ and $\partial TCF^{(1)}/\partial\theta$ for Selected Cuts

Angles of $ZTCF^{(1)}$, Degrees (S and T's Program)			$\partial TCF^{(1)}/\partial\phi$	$\partial TCF^{(1)}/\partial\theta$
PHI	THETA	PSI		
7	27	135.64	−0.7 (PPM/C°)/ DEGREE	−0.5 (PPM/C°)/ DEGREE
0	27	137.78	−0.8	−0.8
15	40	40.00	+1.5	−0.7
15.0	32.5	38.55	−1.2	0.6
12.5	35.0	130.62	0.95	1.1

The frequency-temperature behavior of the device fabricated on the commercially supplied ST-cut wafer (YX wlt) 0/42.75/0 (angular tolerance is not known) is shown in Figure 2. Both the calculated and experimental results are plotted. Plots of fitted temperature curves for a set of cuts (YX wlt) 7.417/27.833/ψ where ψ =134.0°, 134.2°, 135.0° and 135.2° are shown in Figure 3. These

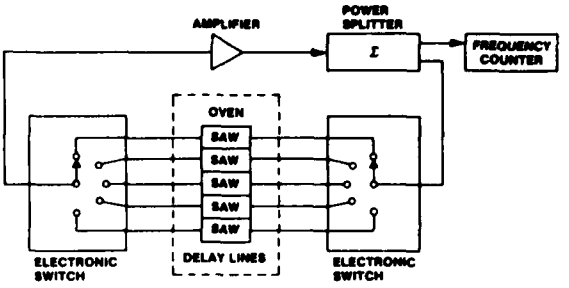


Figure 1. Measurement System

curves were obtained from parts fabricated on the same substrate with a single mask. Each device propagates in a different ψ direction, as determined by the rotation of the patterns on the mask. Each device is rotated by 0.20° with respect to its neighbor. Figure 3 demonstrates the need for a high degree of fabricational accuracy. Figure 3 also illustrates the ability to compensate for fabrication errors by using a mask patterned with devices oriented with a range of ψ angles.

A few representative frequency-temperature measurements are presented in Figures 4 through 7. The stars represent experimental data points. The solid lines are linearly regressed curves used to define the measured first, second, and third order TCF's for these cuts found in Table 5 of reference 1. Figures 4 and 5 are representative of cuts in region (YX wlt) 7/27/135. Cut (YX wlt) 6.57/26.88/134.9 of Figure 4 has a small linear frequency term at room temperature and is well-suited for use at both high and low temperatures. Its second order TCF is in good agreement with the computer calculations and is considerably smaller than that of ST-cut quartz (see Table 5 of reference 1). Cut (YX wlt) 0.633/26.15/137.0 of Figure 5 is representative of cuts in region (YX wlt) 0/27/138 with the small first order TCF. A slight rotation of ψ could be used to set the first order TCF to zero while slightly improving the second order TCF of the device. This cut also represents a substantial improvement over the ST-cut (see Table 5 of reference 1). Results in this area are in excellent agreement with the theory (Table 5 of reference 1) as well.

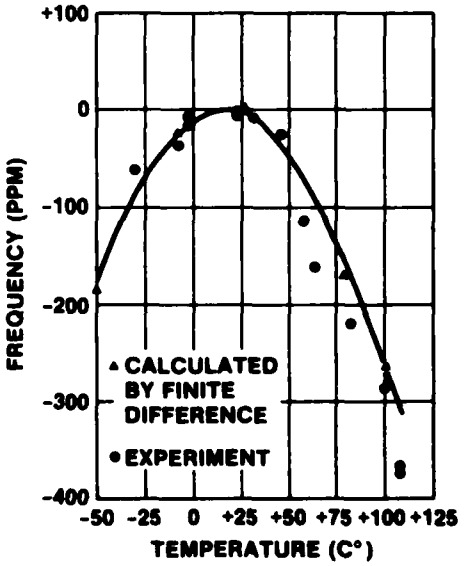


Figure 2. (YX wlt) 0/42.75/0 ST-Cut

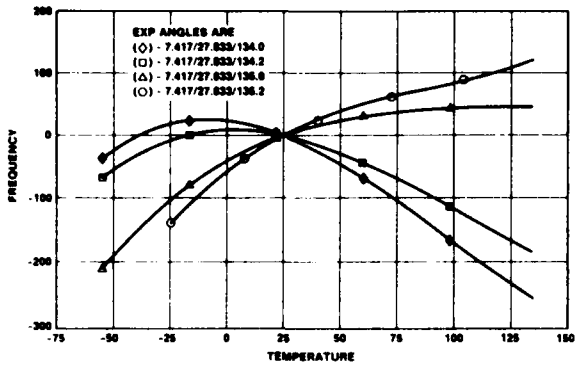


Figure 3. Doubly Rotated Cut With Different ψ Angles

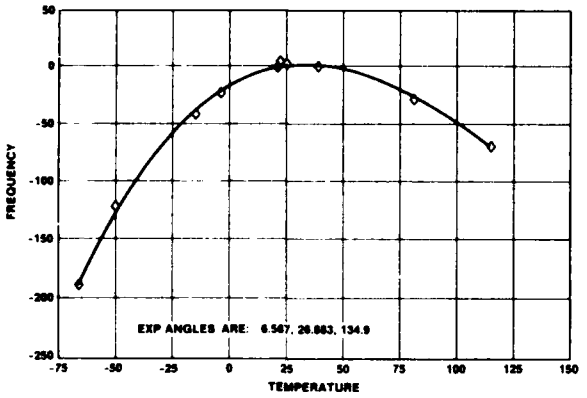


Figure 4. (YX wlt) 6.57/26.88/134.9

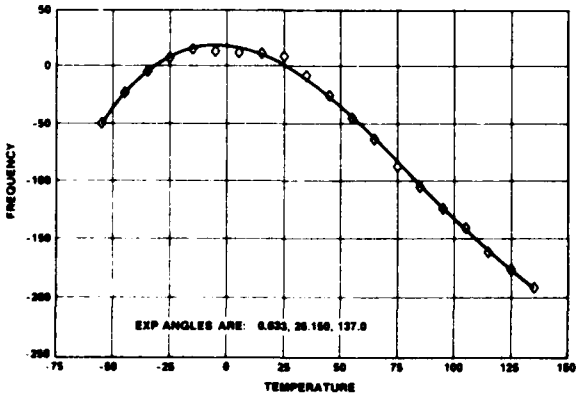


Figure 5. (YX wlt) 0.633/26.15/137.0

Figure 6 shows results typical of the area (YX wlt) 15/40/40. The cut has a measured second order TCF of $-1.4 \times 10^{-8}/(C°)$. The total temperature variation over the range shown is small enough that the effect of third order TCF becomes noticeable, decreasing the total frequency deviation at the higher temperatures. The effect of the third order TCF is also noticeable in the experimental plot of Figure 7. The cut has a measured TCF^2 of -1.5×10^{-8} and a calculated TCF^2 of -1.4×10^{-8}.

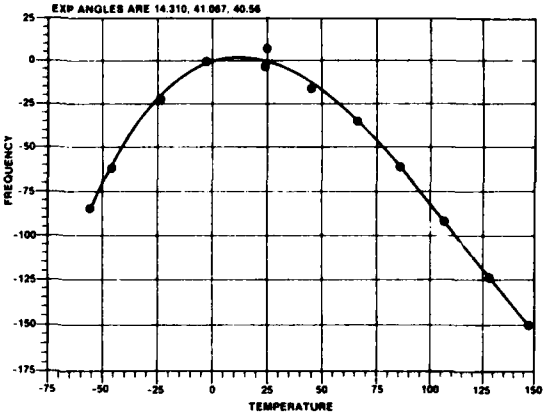


Figure 6. (YX wlt) 14.31/41.067/40.56

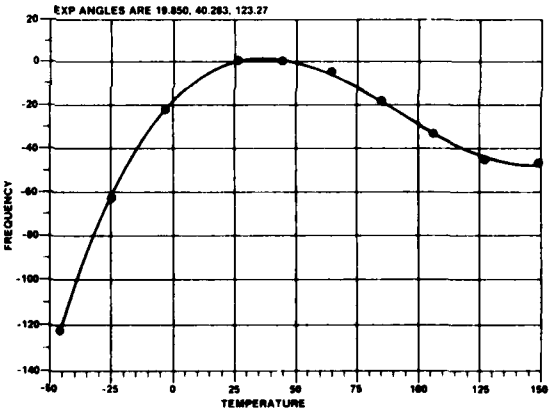


Figure 7. (YX wlt) 19.85/40.283/123.27

Table 5. Comparison of Experimental and Calculated Results

ANGLES			CALCULATED				MEASURED		
PHI	THET	PSI	TCF^{1}††	TCF^{1}†††	TCF^{2}†††	TCF^{3}†††	TCF^{1}	TCF^{2}	TCF^{3}
0	42.75	0†	-0.07×10^{-5}	0.06×10^{-5}	-0.40×10^{-7}	0.11×10^{-10}	-0.1×10^{-5}	-0.37×10^{-7}	-0.17×10^{-10}
8.05	25.9	135.7	-0.01	0.74	-0.15	0.42	0.16	-0.16	0.56
6.57	26.88	134.9	-0.24	0.55	-0.15	0.43	0.025	-0.16	0.47
8.03	26.97	134.6	-0.18	0.60	-0.15	0.46	0.067	-0.13	0.46
7.41	27.83	134.2	-0.26	0.54	-0.15	0.49	-0.08	-0.15	0.63
6.00	26.97	135.8	-0.04	0.75	-0.14	0.46	0.15	-0.13	0.30
5.56	27.83	135.2	-0.15	0.65	-0.14	0.49	0.13	-0.13	0.28
-0.03	26.70	138.9	0.28	1.10	-0.11	0.48	0.43	-0.11	0.49
- 0.13	28.1	137.7	0.06	0.89	-0.11	0.53	0.24	-0.097	0.25
0.633	26.15	137.0	-0.24	0.57	-0.13	0.47	-0.11	-0.15	0.66
-0.967	26.23	138.4	0.01	0.82	-0.12	0.51	0.34	-0.13	0.79
14.337	39.016	40.384	0.11	0.84	-0.11	0.55	-0.098	-0.14	0.65
14.31	41.067	40.56	0.01	0.96	-0.11	0.53	-0.038	-0.14	0.53
19.85	40.283	123.67	-0.26	0.76	-0.14	0.57	0.13	-0.14	0.60
19.85	40.283	123.27	-0.36	0.66	-0.14	0.57	0.03	-0.15	0.74

ANGLES		PSI FOR ZTCF (1)		
PSI	THETA	EXPERIMENT	SINHA AND TIERSTEN	FINITE
8.05	25.9	135.12	135.75	132.97
6.67	26.88	134.81	135.79	132.86
8.03	26.97	134.57	135.27	132.38
7.41	27.83	134.50	135.16	132.20
6.00	26.97	135.24	135.94	133.02
5.56	27.83	134.72	135.76	132.79
-0.03	26.70	137.47	137.97	135.23
-0.13	28.10	137.90	137.50	135.73
0.633	26.15	137.37	137.80	135.10
-0.967	26.23	137.27	138.37	135.67
14.337	39.016	40.10	40.05	42.78
14.31	41.067	40.45	40.53	43.30
19.85	40.283	123.15	124.75	120.59

†WAFER OBTAINED COMMERCIALLY. ANGULAR TOLERANCE IS UNKNOWN.

††CALCULATED USING SINHA AND TIERSTEN'S PROGRAM

†††CALCULATED USING FINITE DIFFERENCE APPROACH

TEMPERATURE VARIATION OF THE SAW POWER FLOW ANGLE

An unexpected rapid fluctuation of the surface acoustic wave power flow angle on doubly rotated cut quartz SAW devices was discovered during the testing of temperature-stable SAW devices. In this section, the phenomenon of a temperature variation in the SAW power flow angle will be discussed.

Calculations[4] of the power flow angle at different temperatures for doubly rotated cuts of quartz have been performed. The cuts are designated by the 1949 IRE standard. Table 6 contains a summary of the power flow angles temperature dependence for several temperature-stable doubly rotated cuts. Figure 8 illustrates the temperature variation of the power flow angle for the doubly rotated cut of quartz (YX wlt) 14.283/39.117/40.6. The important feature of this dependence is the large variation of the power flow angle over the temperature shown. The input transducer on the left generates an acoustic wave which only partially illuminates the output transducer on the right. Figure 9 (a and b) has photographs of the device response with a short gated RF pulse as the input, showing the response at 131°C and 34°C, respectively. The first notch is a result of missing finger pairs. The transducer apertures are 34 mils and 24 mils, the length of the device is 260 mils and the acoustic wavelength is 0.48 mils. The temperature-dependent effects of diffraction on the envelope were found to be negligible. The shortening of the device response is clearly evident from the photographs and is due to the rapidly decreasing power flow angle successively illuminating more of the output transducer as the temperature increases. The power flow angles estimated from these photographs are displayed in Figure 10, alongside the theoretical results.

Table 6. Temperature Dependence of the Power Flow Angle on Doubly Rotated Cuts at Orientations (YX wlt) PHI/THETA/PSI

Orientation			Power Flow Angle		
PHI	THETA	PSI	T = 25°C	T = 150°C	T = -50°C
1 330	28 100	137 692	+1 2	-0 1	+1 8
1 050	28 067	136 554	+2 5	+1 0	+3 2
0 967	26 233	138 449	+1 1	-0 2	+1 8
0 33	26 700	138 859	+0 5	-0 7	+1 1
0 633	26 150	137 016	-1 4	+0 1	+2 1
5 583	27 833	135 194	-0 3	0 9	+1 0
5 583	27 833	134 940	+0 5	+0 1	+1 1
5 583	27 833	134 994	-0 4	-0 8	+1 1
6 000	26 967	135 812	0 1	1 2	+0 5
6 067	25 933	133 099	-1 7	+0 3	+2 4
6 567	26 883	134 925	-0 1	1 0	+0 7
7 410	27 380	134 2	+0 1	1 0	+0 8
8 033	26 967	134 618	0 3	1 4	-0 3
8 05	25 900	135 71	0 7	1 6	-0 1
14 283	39 117	40 627	8 1	6 2	-9 0
15 300	40 683	40 031	8 6	6 8	-9 6
16 117	41 267	37 309	7 2	5 5	-8 1

Table 7 lists the anisotropy parameter γ for various cuts as a function of temperature. The variation of the anisotropy parameter is small (typically 10 percent). No problems arising from the temperature-dependent anisotropy paramters are anticipated.

Table 7. Temperature Dependence of γ

ORIENTATION			ANISOTROPY PARAMETER γ		
PHI	THETA	PSI	T= 25°C	T=150°C	T=-50°C
0	42.5	0	0.375	0.386	0.369
0	27.0	137.78	0.616	0.445	0.669
7	27.0	135.64	0.379	0.234	0.444
12.5	35.0	130.62	0.645	0.507	0.704
14.3	39.1	40.6	0.614	0.586	0.625
15.0	32.5	38.55	0.756	0.705	0.775
15.0	40.0	40.0	0.621	0.585	0.637

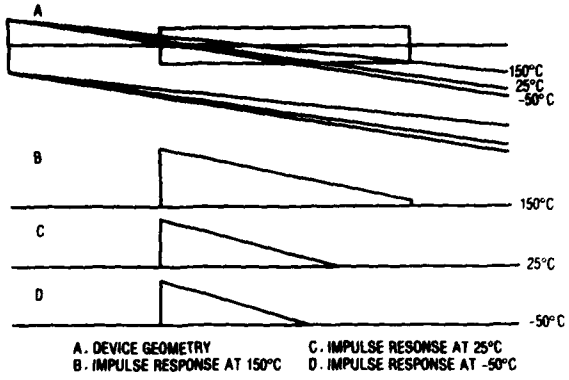
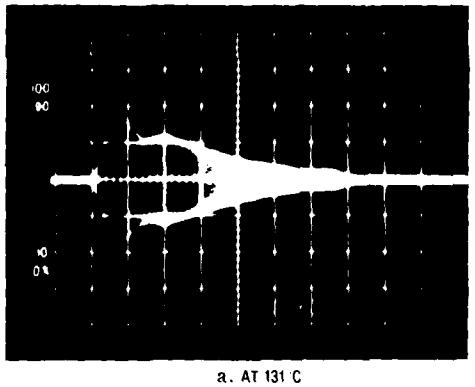
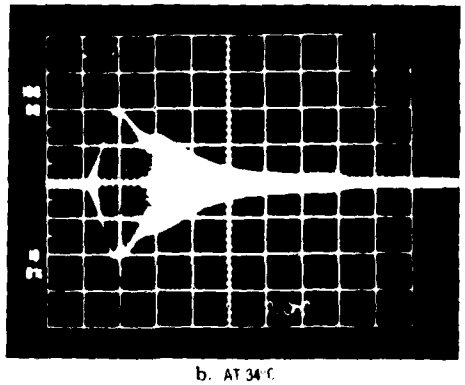


Figure 8. Pictorial Representation of Device Response



a. AT 131°C



b. AT 34°C

Figure 9. Device Response to Short Gated 270.4 MHz Input Pulse

On cuts for which the power flow angle is temperature-dependent, however, when the oscillator frequency is not exactly at the synchronous frequency of the SAW delay line, the phase response of the delay line becomes temperature dependent. The use of doubly rotated cut SAW devices clearly requires designs which can accommodate a large variation in the SAW power flow angle.

In designing filters and reflective array devices on rotated cuts of quartz, additional care must be taken to ensure that the temperature dependent power flow angle does not degrade device response. If improperly designed, device time delay, bandwidth and phase all become temperature dependent when fabricated on a rotated cut. These observations are especially important for

reflective array devices in which rotated cuts of quartz are often used to achieve temperature compensation in two different propagation directions.

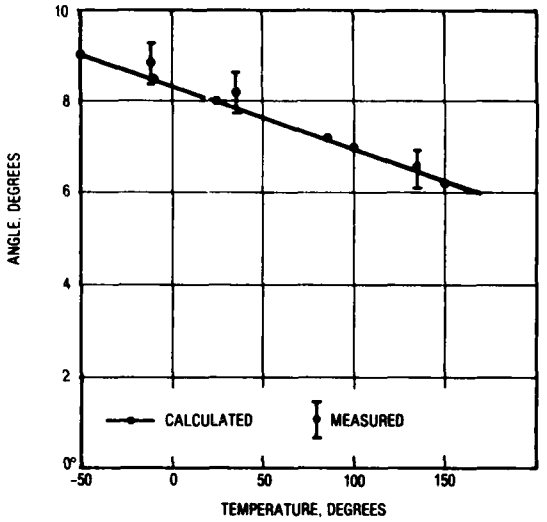


Figure 10. Power Flow Angle Against Temperature for (YX wlt) 14.283/39.117/40.6

Suitable device design can overcome these problems and result in temperature compensated delay lines useful for oscillator applications. The moving acoustic beam may be made to illuminate different parts of the output transducer at different temperatures in such a way as to maintain a steady phase over the entire temperature range. A design using the temperature dependent PFA has been fabricated in which a compensation signal maintains a relatively constant output phase despite the temperature dependent delay time.

Figure 11 shows the device design with the compensators on the left. The compensator output is summed with the output signal from the transducer structure on the right to adjust the temperature dependent phase response. Figure 12 shows the results of using the compensators. With suitable design, the second order curvature could also be compensated for. The advantage of using this compensation mechanism is that it is inherently monolithic.

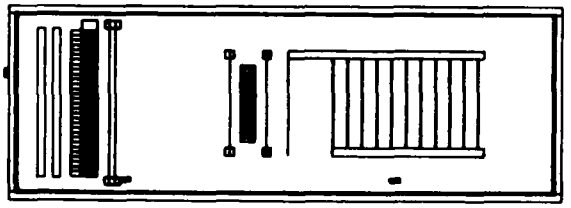


Figure 11. Transducer Design with Compensators on the left. The Compensator Output is 90° Out of Phase with the Main Transducer on the Right. The Two Outputs Are Summed to Give a Phase Response Insensitive to Temperature.

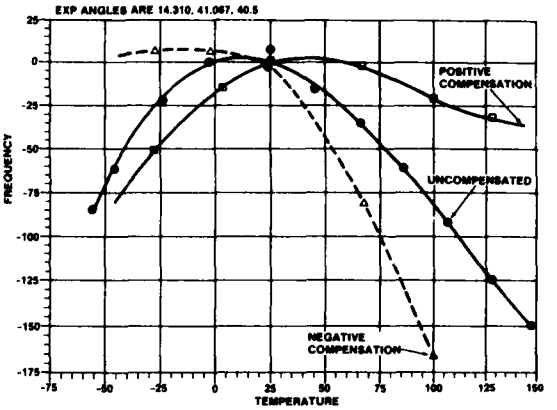


Figure 12. Effect of Utilizing the Compensation Transducers (see Figure 11) to Adjust the Temperature Dependent on the SAW Oscillator

CONCLUSION

Additional experimental results have shown continued agreement with theoretical calculations. Doubly rotated devices with second-order temperature coefficients of frequency as small as -1.3×10^{-8} have been fabricated and tested. These devices are three times as stable as the ST-cut. The analytical evaluation yielded temperature stable families of cuts, one with optimum orientation possessing the lowest second-order temperature coefficient of frequency yet predicted in quartz.

The temperature variation of the power flow angle and the anisotropy parameter on doubly rotated cuts of quartz and its effect on device design was discussed. An oscillator design accommodating large temperature-varying power flow angles was discussed. The temperature variation of the power flow angle was utilized to achieve additional SAW oscillators fabricated on doubly rotated cuts of quartz.

ACKNOWLEDGEMENTS

The authors wish to thank Mr. R. Caputo and Mr. T. Chatam for their valuable contributions to the project.

REFERENCES

[1] D.F. Williams, F.Y. Cho, J.J. Sanchez, "Temperature Stable SAW Devices Using Doubly Rotated Cuts of Quartz," 1980 IEEE Ultrasonics Symposing, pp. 429-433.

[2] D.F. Williams, F.Y. Cho, "Temperature Variation of SAW Power Flow Angle," Electronics Letters, Feb. 19, 1981, Vol. 17, No. 4, pp. 164-165.

[3] "Standards on Piezoelectric Crystals 1949," Pro. IRE 14, Dec. 1949, pp. 1378-1395.

[4] "Higher Order Temperature Coefficients of the Elastic Stiffness and Compliances of Alpha-Quartz," Bechmann, R., Ballato, A., and Lukaszek, T., IRE Trans., 1962, pp. 1812-1822.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

ABSOLUTE AND DIFFERENTIAL AGING OF SAW RESONATOR PAIRS*

J. S. Schoenwald, J. Wise and E. J. Staples

Rockwell International Corporation
Thousand Oaks, California 91360

Summary

This paper examines the absolute and differential aging of SAW resonators fabricated as pairs. Low absolute aging rates of individual SAW resonators have been observed after hermetic sealing and temperature stabilization inside 70°C pre-aging ovens. Typical drift rates fall below 0.5 ppm/day within 48 hours and 0.1-0.5 ppm/wk after 3 weeks. The differential drift rate between pairs was less than 0.25 ppm/day after 48 hours and typically 0.1-0.2 ppm/wk after 3 weeks.

The effect of drive level was studied for devices fabricated using evaporated chromium (50Å)/aluminum (1000Å), rf sputtered aluminum (1000Å) and evaporated aluminum + 4.5°% copper alloy. Significant performance degradation occurs at 75°C for power levels above +10 dBm delivered to the crystal at resonance. Comparative studies of frequency dependence on drive level show no significant differences among the three metallization systems described.

Key words (for information retrieval)
Surface Acoustic Wave (SAW) Resonators, Aging, Drive Level, Fabrication, Electromigration.

Introduction

Systems requiring long term stability and insensitivity to undesirable effects provide strong motivation for investigating the common mode rejection characteristics of appropriately configured pairs of surface acoustic wave (SAW) resonators.

Low absolute aging of individual SAW resonators depends on device material composition, process controls and cleanliness, packaging environment and hermeticity. Extended pre-aging appears to be necessary to reduce long term drift below 1 ppm/yr with typical aging studies reported in the literature now approaching the two year mark.[1-3] An objective of this study was to isolate as much as possible those factors contributing to the long term instability. By a study of resonator pairs in a common enclosure, effects such as packaging could be eliminated or deduced by evaluating differential aging.

*Work supported by Defense Advanced Research Projects Agency, Contract No. MDA 903-81-C-0081.

This paper will examine the absolute and differential aging of SAW resonators fabricated as pairs and sealed under vacuum in a common TO-5 enclosure. The design of the fabrication process and control of the process is described. Also, the effects of power level at which the devices are driven and the damage encountered under sustained conditions of high power and elevated temperature are discussed.

Fabrication Process Design and Control

Reliable device performance requires a fabrication process that is reproducible. Analysis procedures were instituted at appropriate steps in the process to monitor the condition of the SAW substrate as it proceeded from wafer to chip and finally to packaged environment. Figure 1 is a flow chart of the fabrication process. As pilot wafers progress through the fabrication process they are subjected to a variety of surface and bulk analytical measurements to determine the presence of contaminants, interdiffusion of deposited metals, or surface chemical reactions that may have significant consequences for device long term stability after the package is sealed.[4] This required vacuum baking prior to sealing, to pump off moisture and accelerate short term surface reactions. Also a small plasma etch chamber for frequency calibration was installed directly in the dry nitrogen glove box containing the resistance weld sealer. After vacuum sealing the packages were bombed with helium and leak tested. Typically better than 90% of all sealed devices indicated leak rates of less than 1×10^{-10} std cc/sec air equivalent.

Pre-Aging System

An oven was designed for monitoring the aging of 36 SAW resonators simultaneously. By using interchangeable PC boards, either 18 TO-5 packages with resonator pairs or 36 packages containing single resonators could be monitored. Resonator groups of six were connected to a single pole six throw (SP6T) coaxial switch. The six switches were connected to a seventh in a switching matrix arrangement. Figure 2 shows photograph of the oven, containing the PC board into which the TO-5 packages were inserted and the coaxial switches.

The resonators were periodically interrogated through the switches by a network analyzer, frequency synthesizer and counter, all of which were controlled by a desktop computer. The resonant

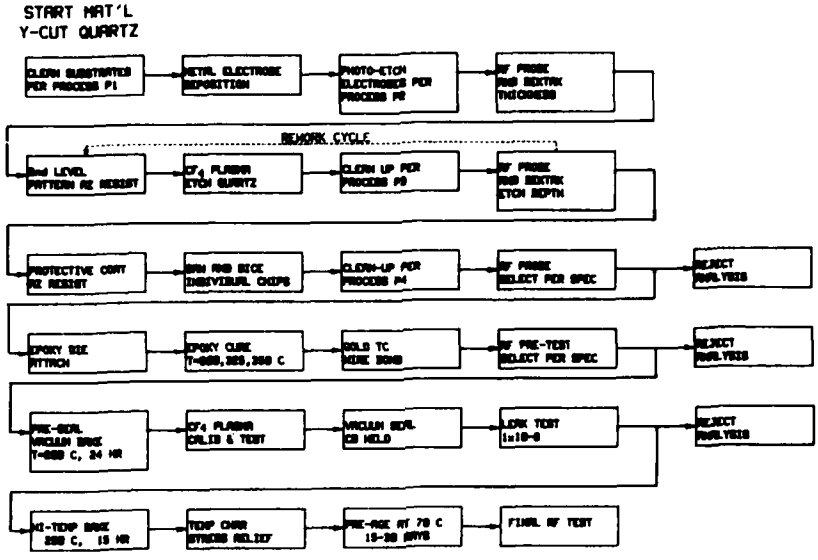


Figure 1 Fabrication process flow chart.



Figure 2 Photograph of oven showing SP6T coaxial switches and PC board for TO-5 crystal enclosures.

frequencies were sequentially measured, the oven temperature recorded, and the time logged from the computer's real time clock. In this manner, the frequency/temperature drift was tracked with time. Software computed the differential drift between packaged pairs of SAW resonators.

Absolute and Differential Pre-Aging Results

Pre-aging studies were conducted on two paired resonator configurations and two types of electrode metalizations:

1. Pairs: two separate chips in one package; evaporated Cr/Al electrodes.
2. Dual resonators: two resonators on a common chip; evaporated Cr/Al electrodes.
3. Dual resonators: same as #2 but with evaporated Al + 4.5% Cu alloy electrodes.

Figures 3. 4 and 5 show the absolute and differential aging for each of the three cases indicated. In all cases, frequency drift follows a logarithmic time dependence characteristic of chemisorptive reactions. Small oscillations in the aging were primarily due to fluctuations in oven temperature that occurred with a 24 hour cycle. These frequency variations result from an offset between the oven operating temperature (70°C) and the turning point of the crystals (85°C). In the case of paired Cr/Al crystals the absolute aging rates typically showed between 5 and 20 ppm frequency increases within the first

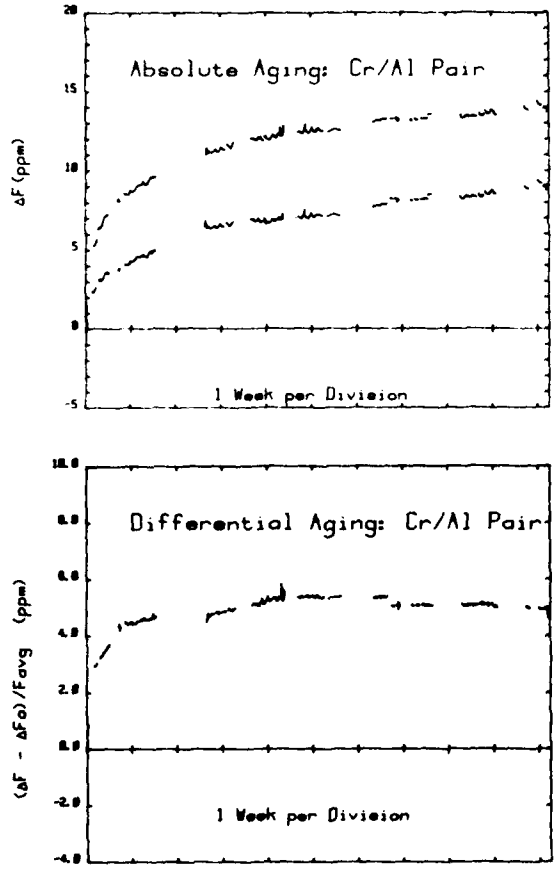


Figure 3 Absolute and differential aging for pairs of Cr/Al resonators sealed in a common package.

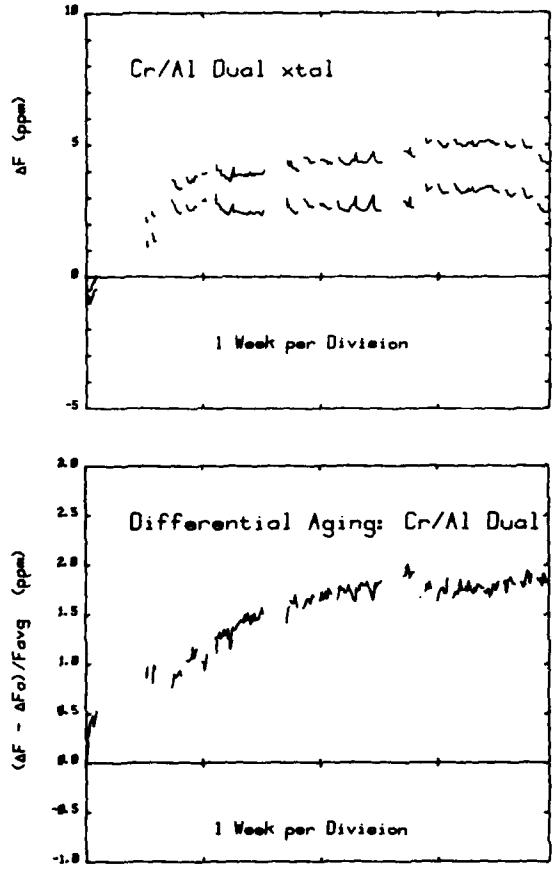


Figure 4 Absolute and differential aging for Cr/Al dual resonators.

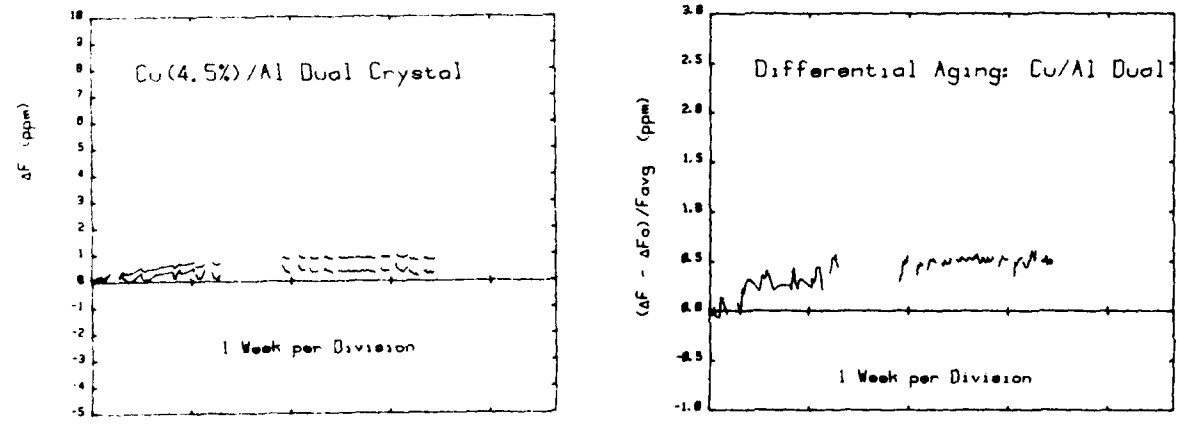


Figure 5 Absolute and differential aging for Cu doped Al dual resonators.

30 days. typical drift rates fell below 0.5 ppm/day within 48 hours and 0.1-0.5 ppm/wk after three weeks. Typical differential drift rates fell below 0.25 ppm/day within 48 hours and typically 0.1-0.2 ppm/wk after three weeks (5-10 ppm/year). Fabricating the devices as dual resonators on the same crystal, Fig. 4, improved the differential aging rates but absolute aging remained high. Replacing the Cr/Al metallization with copper alloyed aluminum remarkably improved the absolute as well as differential aging, Fig. 5. The total pre-aging amounted to less than 2 ppm and drift rates less than 0.1 ppm/mo were obtained after only 3 weeks pre-aging. As expected the differential aging showed a similar improvement. A more accurate assessment of these rates requires either individual oscillator measurements or longer pre-aging times. The former is currently under investigation using a dual oscillator circuit.

Drive Level Studies

The effect of rf power delivered to the resonators from a 50 ohm source at the resonant frequency of each device was investigated and compared for three different types of electrodes: (1) evaporated chrome, 50Å, and aluminum, 1000Å, (2) rf sputtered aluminum and (3) evaporated 4.5% copper-aluminum alloy, 1000Å. It has been reported that excessive power levels in SAW resonators result in drive level dependent frequency shifting and degraded performance with time.[5,6]

Figure 6 shows the frequency dependence on drive level for eight resonators from each of the metallization groups described above. From the results of the three types of metallized devices studied it does not appear that one type provides better frequency stability under varying drive levels than any other. This may indicate that the dependence of the device frequency is controlled primarily by thermal heating due to power dissipation in the bulk of the quartz, and not by any property of the metallization.

Degradation at High Drive Levels

The equivalent circuit series resonant resistance was typically 45-75 ohms for the 375 MHz SAW resonators studied in this work and a 50 ohm resonator will dissipate all the power delivered to it from a 50 ohm frequency source. At +10 dBm drive level, in excess of 1 Kw/cc is being dissipated in the skin depth of the central region of the electrode pattern. In view of the fact that the Q of these devices ranged from 12,000 to 20,000, and the energy stored in the device is proportional to the product of the Q and the energy dissipated, large stored energy densities are achieved. It comes as no surprise, then, that electromigration effects due to large piezoelectrically induced potentials, severe mechanical distortions and thermal stresses are likely to have a degenerative effect on the structure of the device, and particularly the metal electrodes.

SAW resonators were driven for periods exceeding 15 hours at drive levels of +10 dBm.

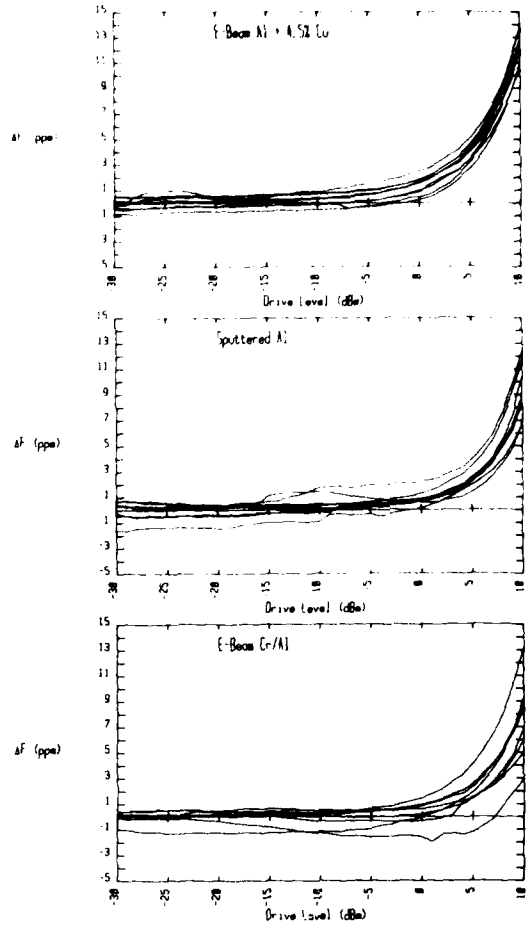
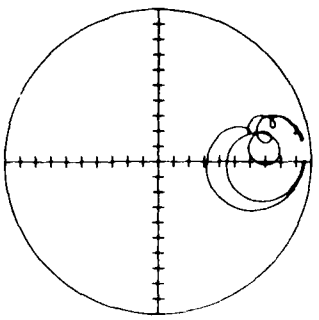


Figure 6 Frequency dependence of drive level for Cr/Al, Cu-doped aluminum, and sputtered aluminum metallized resonators.

Device degradation was found to be closely related to the temperature at which the drive level measurements were made. At room temperature little or no change in the device characteristics occurred. At 75°C, however, significant damage was observed after prolonged active aging at +10 dBm. Figure 7 shows the return loss magnitude and polar response for a SAW resonator fabricated with a Cr/Al electrode pattern before and after 15 hours of active aging at 75°C. The appearance of spurious resonances, a shift in the resonant frequency and an increase in the series resonant resistance indicates damage has occurred either in the acoustic cavity, the interdigital transducer, or both.

Scanning electron micrographs of the transducer region of the SAW resonator, taken afterward confirmed that damage had occurred. Small hillocks, typical of electromigration damage, were



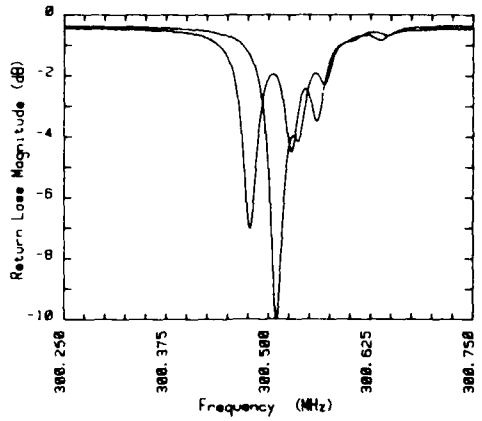


Figure 7 Return loss degradation for +10 dBm drive level at 75°C.

found.[6] In addition, "drag" type effects were noticeable at the edges of the electrodes, as if aluminum had been physically brushed or scraped in the direction of acoustic motion. Neither of these artifacts were present in undamaged devices.

Conclusions

The absolute and differential aging of SAW resonators was investigated. A fabrication process which included techniques for monitoring process conditions resulted in devices with low absolute and differential drift rates after a short period of pre-aging. Absolute aging rates less than 0.1 ppm/mo after only 3 weeks pre-aging at 70°C were demonstrated. Low differential aging rates were obtained using dual resonator crystals. Drive level studies on devices fabricated with different types of aluminum metallizations have not conclusively shown that one type of metallization is superior to another based upon drive level. For the purpose of long term frequency stability however, copper alloyed aluminum appears to be superior to chrome/aluminum metals. Damage due to elevated drive levels for extended periods was found to be more sensitive to operating temperature than the drive level up to +10 dBm.

Acknowledgement

The authors would like to thank A. B. Harker for his assistance in surface chemistry analysis.

References

1. T. E. Parker, "Analysis of Aging Data on SAW Oscillators, Proceedings, 34th Annual Symposium on Frequency Control, U.S. Army Electronics Command, Ft. Monmouth, N.J., pp. 292-297 (1980).
2. M. Gilden, G. K. Montress, and R. A. Wagner, "Long-Term Aging and Mechanical Stability of 1.4 GHz SAW Oscillators," 1980 Ultrasonics Symposium Proc., IEEE Cat. No. 80CH1602-2, pp. 184-187.
3. W. R. Shreve, "Active Aging of SAW Resonators," 1980 Ultrasonics Symposium Proc., IEEE Cat. No. 80CH1602-2, pp. 188-192.
4. J. S. Schoenwald, A. B. Harker, W. W. Ho, J. Wise, and E. J. Staples, "Surface Chemistry Related to SAW resonator Aging," 1980 Ultrasonics Symposium Proc., IEEE Cat. No. 80CH1602-2, pp. 193-199.
5. J. I. Latham, W. R. Shreve, N. J. Tolar, and P. B. Ghate, "Improved Metallization for Surface Acoustic Wave Devices," thin Solid Films, 64 (1979), pp. 9-15.
6. W. J. Tanski, M. Block, and A. Vulcan, "High Performance SAW Resonator Filters for Satellite Use," 1980 Ultrasonics Symposium Proc., IEEE Cat. No. 80CH1602-2, pp. 148-152.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

ELEMENTS OF SAW RESONATOR FABRICATION AND PERFORMANCE

William J. Tanski

Sperry Research Center
Sudbury, MA 01776

ABSTRACT

The application of surface acoustic wave (SAW) resonators[1,2] in systems requires that certain fabrication problems be solved and that various aspects of device performance be well understood. Acceptable device yields require that a post-fabrication frequency adjustment (trimming) procedure be available for the recessed-transducer/etched-groove-reflector resonator configuration which we use. Here we discuss a new trimming technique which involves modification of only the transducer configuration by selective etching of the quartz substrate in a fluorocarbon plasma. Details of the procedure and experimental results are presented for resonators, and we show that the procedure is effective on delay lines as well. Resonator frequencies have been set to an accuracy of five parts-per-million using this technique.

The response of UHF resonator filters to variations in input power level has been studied and data on the power limits, frequency-power effect, and damage mechanisms are presented. Also, the 1/f phase noise (flicker noise) inherent in 500 MHz SAW resonators has been measured and is discussed.

Introduction

SAW resonators have been developed to the extent that the systems application[3] of this class of devices is feasible. However, in order improve the manufacturing yield a post-fabrication frequency trimming procedure is required to compensate for fabrication variances. In addition, the device power handling capacity must be known, the nature of overpower damage mechanisms understood, and when used in an oscillator the flicker noise contribution of a SAW resonator should be anticipated. In this report we discuss a frequency trimming technique which we have recently developed for use on resonators with the recessed-aluminum-transducer/etched-groove-reflector[3] configuration. We have found that one may trim the frequency of a resonator downward by changing the physical configuration of only the transducers. The method, to be discussed in detail, utilizes selective etching of the quartz by means of a fluorcarbon-oxygen plasma to increase the transducer electrode step height. An increase in transducer reflectivity is the dominant frequency shift mechanism.

We have found experimentally that UHF SAW resonators have a relatively high power handling capacity, on the order of +10 d8m and greater. However, for oscillator and filter applications we should know the upper limit on power input, the variation of frequency with power level, and the nature of the overpower damage mechanisms. Here we present data on the effects of device input power level. This data, through specific to a given design, gives us a clear idea of the nature of the power level response.

For oscillator design, the flicker noise (1/f) contribution of a SAW resonator should be anticipated. Recent workers[4] have reported that no resonator flicker noise was detected in a resonator stabilized oscillator. Our experimental results are at variance with their[4] observations and we present flicker noise data[5] for two devices.

Frequency Trimming

Narrowband surface acoustic wave resonators[6,7] and delay-line type filters[8] will generally require post-fabrication adjustment (trimming) of the resonant or the synchronous frequency. Most narrowband SAW devices are fabricated on quartz, for temperature stability, with aluminum being used for the interdigital transducer electrodes. Previously published trimming techniques, such as the implantation of high energy ions,[9] application of overlay films,[10] or alteration of the reflection grating depth by etching through a metal film mask,[11-14] either increase the long term aging rate of the device or have the disadvantages associated with metal reflectors. Here we discuss a frequency trimming technique in which we change the transducer electrode step height, with a resultant change in the SAW velocity and transducer reflectivity, by selectively etching the quartz substrate using a tetrafluoromethane-oxygen ($CF_4 + O_2$) plasma. Previous workers[11-14] have failed to recognize the useful effects of etching the transducers, and for this reason we discuss theory which increases our understanding and

techniques which are more general in application. The significant aspect of this work is that we demonstrate the ability to frequency trim SAW devices, particularly resonators, by changing only the transducer configuration. The procedure is useable for all types of narrowband devices on quartz, it utilizes readily available equipment, and the long-term device aging rate is not changed. Also, the procedure is simple to implement and it has been used successfully to trim the frequencies of resonators used in two systems applications[1,2].

The physical effect of selectively etching the quartz, for several representative aluminum-on-quartz device configurations, is shown schematically in Fig. 1. The section views before etching are shown on the left side of Fig. 1 and the configurations after etching are shown on the right side. In each case the aluminum electrodes remain substantially unchanged (etch rate differences on the order of 10:1 are generally obtained) while the quartz is removed in a uniform manner. In Fig. 1a we show a commonly used resonator configuration and it is for resonator trimming that this procedure was originally developed. However, the procedure is effective for all SAW devices on quartz with aluminum transducer electrodes. Two frequently used configurations for SAW delay-line filters (which utilize travelling waves) are shown in Fig. 1b and c.

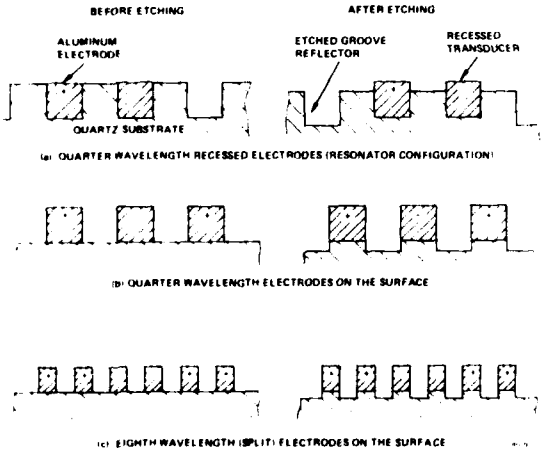


Fig. 1. Section views of several of the SAW device configurations which may be frequency trimmed using selective etching of the quartz. The configuration before etching is shown schematically on the left side and the system after etching is shown on the right for each case.

The etching process may be performed by RF sputtering or plasma assisted etching which processes are discussed in Ref. 15, sections II-1 and V-2, respectively. The gas mixture we use is CF_4 (80 percent) + O_2 (20 percent) by volume, though the percentage of O_2 is not critical, and our trimming is usually performed in an RF sputtering station with a backfill pressure of about 20 microns (of Hg) at a power density in the range of 0.1 W/cm^2. We have successfully performed device trimming in a very simple barrel-type plasma reactor[16] though the trim rates are much lower than for RF sputtering. The etching of the quartz is performed by chemically reactive ions formed by disassociation of the CF_4. The function of the O_2 is to suppress fluorocarbon-polymer formation (which coats the device surface in the absence of O_2), and to enhance etching by the formation of an oxyfluoride (OF) species[17] and by increasing the fluorine-carbon ratio[18]. The frequency trim rate for a given device design is established empirically by subjecting a device to the etching process for a given length of time. For instance, the 500 MHz resonators discussed in Ref. 2 have a trim rate on the order of 7 kHz/sec downward, for the RF sputtering conditions stated above. We generally trim a device to frequency using an iterative procedure so as not to reduce the frequency below that which is desired. Also, the devices are designed and fabricated to have an initial frequency higher than the final frequency desired. The maximum trimming range for resonators being made[1,2] is approximately 500 parts-per-million (ppm). This range is adequate to compensate for the maximum device-to-device frequency variations of about 300 ppm we experience for devices on a given substrate. Our photomask designs and fabrication procedures allow consistent attainment of initial device frequency close enough to the desired frequency such that this trimming procedure may be successfully applied. Devices may be trimmed to within 5 ppm or less. With this trimming procedure we have improved our device yield from less than 5 percent due to frequency inaccuracies, to greater than 70 percent. We will generally lose a number of the devices on a substrate due to other fabrication errors and difficulties.

There are two physical reasons why the frequency is reduced when the quartz is selectively etched. The first is due to the velocity decrease associated with an increase in the energy storage[19] of the larger physical steps. This is the reason for the downward frequency shift of delay-line filters utilizing either quarter-wavelength, recessed, or eighth-wavelength (split) electrodes. We have experimentally demonstrated this downward shift for a variety of devices in our laboratory. For example, the synchronous frequency of a 0.7 percent bandwidth 150 MHz split-electrode delay line filter was trimmed 200 kHz downward by etching for four minutes using the conditions stated earlier. The second reason for the downward frequency shift, in resonators only, is

the increase in transducer reflectivity[20] for recessed (Fig. 1a) or quarter-wavelength-surface (Fig. 1b) electrodes. In resonators on quartz we generally make every effort to reduce the SAW reflectivity of the transducers to a negligible amount to reduce resonance distortion[7]. However, to trim the resonance frequency downward we intentionally increase the transducer reflectivity by increasing the height of the steps. The downward shift in resonance frequency is due to the interactions between the waves reflected from the transducers and those reflected by the reflectors as the transducer reflectivity is increased. For devices with grooved reflectors no change in the groove depth occurs, and the maximum trim range for resonators utilizing such reflectors is set by the amount of resulting resonance distortion which is acceptable and the stopband width of the reflectors. The trim range attainable by etching only the transducers has been found to be completely acceptable. We note that a somewhat larger trim range is possible for devices using metal reflectors because the stopband of such a reflector also shifts down in frequency as the step height is increased. Using our trimming procedure we etch the entire device rendering masking of particular device sections unnecessary, and we can trim resonators which have the desirable characteristic of containing no metal in the reflectors.

For delay line type filters using the configurations shown in Fig. 1 (b) and (c) the increased step height decreases the velocity and the synchronous frequency in a predictable way. It is known that the energy storage effect[19] decreases the velocity as a function of the square of the relative step height $(h/\lambda_0)^2$, where h is the height and λ_0 is the surface wavelength. For a device with initial frequency f_1 and metal step height h_1 (the effect of an aluminum step is approximately the same as that of a step in the quartz), the total step height h_2 required for a final frequency f_2 ($f_2 < f_1$) is given by (1), where

$$h_2 = \lambda_0 \sqrt{\left(\frac{h_1}{\lambda_0}\right)^2 - \frac{(f_1 - f_2)\,\lambda_0}{KV_0}} \qquad (1)$$

V_0 is the free surface velocity and K is approximately[7] 10.5. Using (1) we predict that an increase in step height of 1200 Å is required (from 1700 Å to 2900 Å) in the delay line filter discussed earlier to decrease the frequency by the observed 200 kHz. This is in reasonable agreement with the 1000 Å increase step height actually produced during etching.

In resonators the change in reflectivity of the transducers, as the step height is changed, has a dominant effect on the resonance frequency. Equivalent circuit model calculations we have performed indicate that the frequency decrease due to changes in reflectivity will in general be substantially larger than the decrease due to altering the velocity. Our resonators have utilized the configuration of Fig. 1a (recessed-transducers/grooved-reflectors) and we have always been able to obtain an acceptably large trim range using the above etching procedures.

In order to illustrate the effectiveness of this trimming procedure, we show in Fig. 2 the response of a 500 MHz two-port SAW resonator before and after RF sputter etching as described above. We note the downward shift of 200 kHz (or 400 ppm) with no increase in distortion of the response. This device, described in detail elsewhere[2], was etched for 30 seconds.

This trimming procedure is also being applied in the development of multipole SAW resonator filters. In this type of filter there exists a requirement to routinely set each pole frequency, in isolation, to better that 5 ppm. Recent trimming work on a three-pole filter at 150 MHz demonstrates that we can readily meet this requirement.

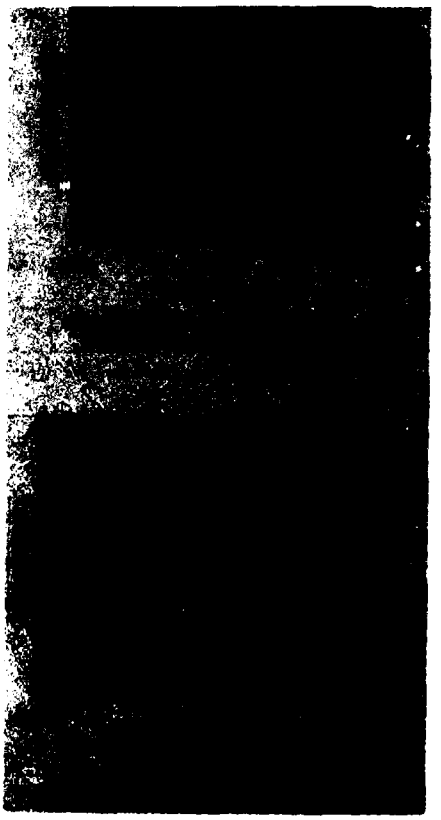


Fig. 2 Frequency response of a 500 MHz two-port SAW resonator before and after trimming using reactive RF sputter etching in a $CF_4 + O_2$ plasma.

Power Level Tests

In order to use SAW resonators most effectively in systems, we must know the maximum safe power input level, and the variation of frequency with drive level. In addition we wish to understand the nature of the over-power failure mechanism and the change in device response under over-power conditions. To these ends we have tested single-pole, two-port SAW resonators[2] which operate at 500 MHz and which are fabricated on the "ST" cut of quartz. The devices were tested by amplifying a sythesized signal to a given power level, passing this signal through the electrically matched resonator filter, and monitoring the output power level. The resonance frequency is defined as the mean of the 3 dB points, and the reference power level is taken at the approximate peak of the filter response. A device was tested with varying input power levels, until a power level was reached at which the resonance frequency changed rapidly (within minutes) with constant input power. These results are summarized in Fig. 3 where we see that the resonance frequency begins to increase at a drive level of about 10 dBm (corresponding to an acoustic power flow of about 7.5 kW/cm^2). The resonance frequency continued to increase rapidly with increasing drive level, but remained constant at a given drive level. However, at about +24 dBm input power the resonance frequency drifted downward by about 20 ppm in the interval of several minutes. The curve of Fig. 3 is thus shown as a dashed line with a steep negative slope, and we take +24 dBm as the threshold for damage. We note that the drive level effect (frequency change with input power) and the damaging power levels both have fairly sharp thresholds. Extended power level tests have been performed at 500 MHz and we have found no change in the filter response following 1200 hours at an input power level of +18 dBm. Also, two-port resonator filters at 840 MHz[1] have been in operating oscillators for over one year at an input power levels of +13 dBm with no degradation in performance. These data show that SAW resonators of the type discussed here[1,2] have a realtively high power handling capacity.

The device of Fig. 3 has a desired operating power level of +13 dBm which is well within the safe drive level range. However, there is an upward frequency shift of about 3 ppm.

The change in device response when the device has been damaged by excessive input power is shown in Fig. 4. The undamaged response is

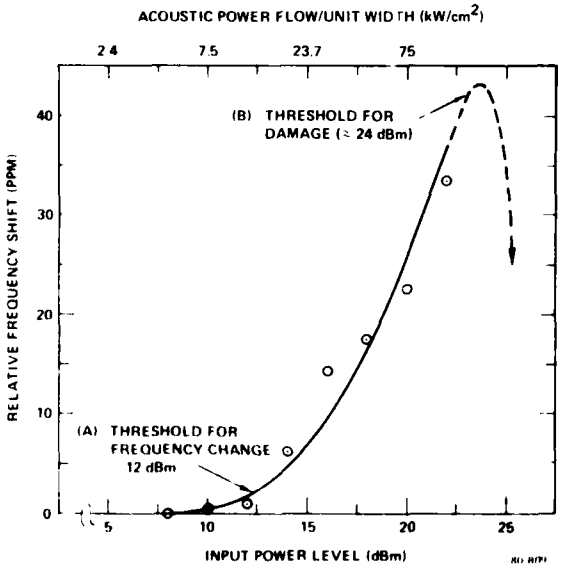


Fig. 3 Variation of resonance frequency with input drive level. The relatively sharp threshold for frequency change (A) and for damage (B) are to be noted.



Fig. 4 Frequency response (electrically matched) of a 500 MHz SAW resonator filter before and after applilcation of power in excess of +26 dBm for about 10 minutes.

shown in Fig. 4(a) where the resonance frequency (500 MHz) and undistorted response are to be noted. In Fig. 4(b) we see the response after application of +26 dBm for about 10 minutes. The resonance frequency has decreased by 240 ppm and a strong third order transverse mode response has developed on the high side of resonance. In Fig. 5 we shown scanning electron micrographs (at 25,000X) of the damaged electodes (aluminum) of the filter of Fig. 4(b). The damaged and undamaged portions are indicated. The damaged areas show the development of many small growths which we attribute to electromigration[21] as well as stress induced migration[22] of the electrode metal. We note from Fig. 5, and from other micrographs, a distribution of damage on the electrode edges which is compatible with the electric field distributions. We have been using a mixture of aluminum plus about 3 percent copper[21] for the device electrodes to increase the threshold power level for damage.

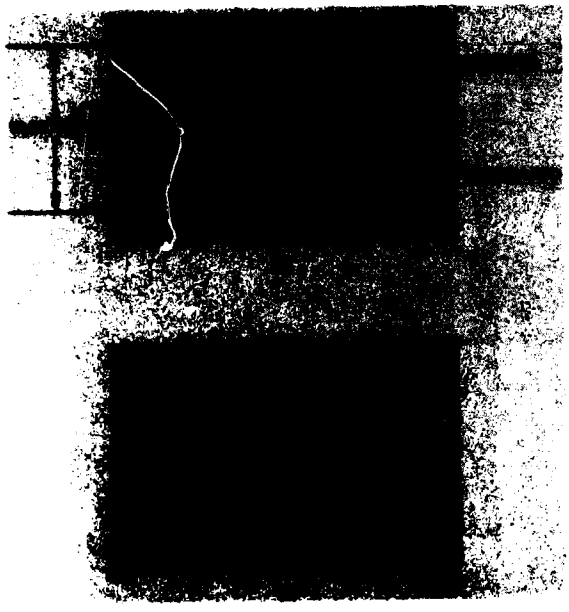


Fig. 5 Scanning electron micrographs (25,000X) showing the damage in the aluminum electrode structure following application of very high power levels. The damage seen is the result of electromigration plus stress induced migration of the metal. The electrodes are 1.6 microns wide and 700 Å thick.

Resonator 1/f Phase Noise

In a recent paper, Parker[23] presented results showing that the close-to-carrier 1/f phase noise (or flicker noise) in a SAW resonator stabilized oscillator was due primarily to the resonator. More recent work by Penavaire[4] et. al. indicated that the resonator in their oscillator contributed little, if any, phase noise. Here we present 1/f phase noise data taken by Parker[5] on resonators made at Sperry Research to show that this device noise is clearly measureable and may be significant in certain oscillator applications. In Fig. 6(a), the two-point fractional frequency deviation, or Allen variance, σ_A, is plotted for two 500 MHz SAW resonator[2] stabilized oscillators versus the frequency counter gate time (τ). These σ_A measurements, which were taken using the closed

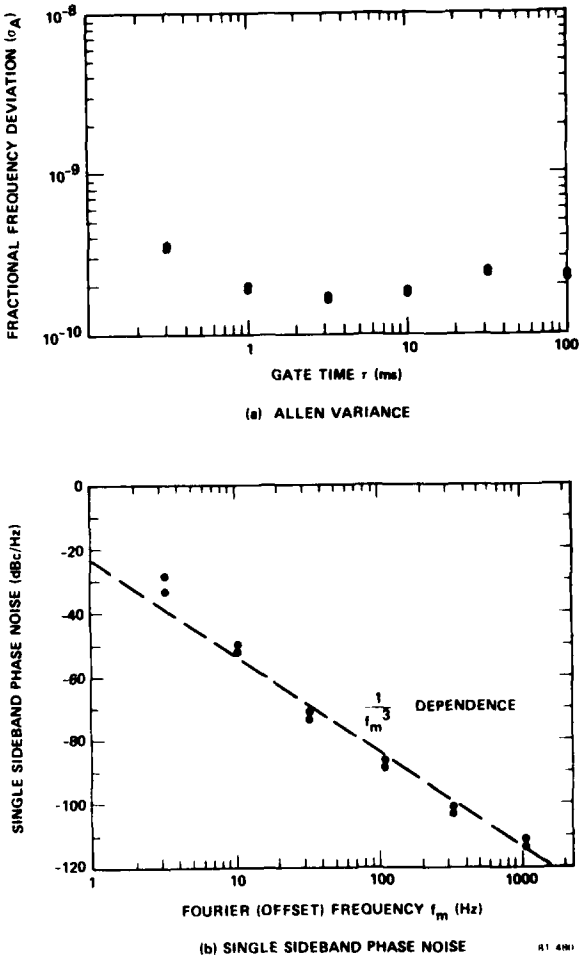


Fig. 6 SAW resonator 1/f phase (flicker) noise data for two 500 MHz devices. The upper section (a) shows the resonator stabilized oscillator two-point fractional frequency deviation (Allen variance) versus frequency counter gate time, and the lower section (b) is the equivalent single sideband phase noise versus off-set frequency with the characteristic $1/f^3$ frequency dependence.

loop circuit (Fig. 1 of Ref. 23), show values which are almost identicle for both devices on this variance scale. The σ_A value is about 2×10^{-10} for $1 < \tau < 100$ msec. The equivalent single sideband phase noise as a function of Fourier (off-set) frequency from the carrier is shown in Fig. 6(b) where the separate device values are distinquishable. The slope of the curve in Fig. 6(b) shows the $1/f^3$ frequency dependence characteristic of device flicker noise in an oscillator.

From the above discussion, we see that flicker noise is indeed present in devices produced at Sperry Research and Raytheon Research. These results are at variance with the data presented by Penavaire et. al. who reports no flicker noise in their devices. It is of interest to determine why no resonator phase noise was detected by these latter workers.

Summary

We have presented a frequency trimming technique for SAW resonator, and also delay-line, filters which utilizes selective etching of the quartz substrate to modify the transducer electrode configuration. In this technique we utilize the velocity slowing and the increase in transducer reflectivity, associated with an increase in transducer electrode step height, to effect a required downward frequency shift. Procedural details and experimental results are presented. The response of SAW resonators to high input power levels has been determined, and we find an increase in resonance frequency with drive level with a fairly sharp threshold. Also, the threshold for damage was found to be rather sharp (at about +24 dBm for the 500 MHz devices tested) and the damage mechanism is shown to be electrode deformation due to electro-and stress induced metal migration. Resonator 1/f phase noise data are presented and these data are compared with the results of other workers, some of whom report no resonator noise of this type. This reported lack of resonator flicker noise[4] invites further investigation.

Acknowledgments

The author wishes to thank Mr. R. A. St. Cyr for fabricating the devices used in this effort, Dr. Thomas Parker of Raytheon Research for obtaining the resonator noise data, and Dr. C. L. Bennett for his support and encouragement in pursuit of this work.

References

1. W. J. Tanski, R. A. St. Cyr, P. Dragonetti and E. Kosco, "A radar system application of an 840 MHz resonator stabilized oscillator," IEEE Trans. Sonics Ultrasonics, SU-28, 146, May 1981.

2. W. J. Tanski, M. Block, and A. Vulcan, "High performance SAW resonator filters for satellite use," Proc. 1980 Ultrasonics Symposium (IEEE Cat. No. 80CH1602-2) pp. 148-152.

3. W. J. Tanski, "UHF SAW resonators and applications," 34th Proceedings of the Annual Symposium on Frequency Control, U.S. Army Electronics Command, Ft. Monmouth, N.J., pp. 278-285.

4. L. Penavaire, D. Seguignes, C. Lardat, J. J. Bummier, J. Y. Chevalier, Y. Besson, "A 120 MHz resonator stabilized oscillator with high spectral purity," Proc. 1980 Ultrasonics Symposium (IEEE Cat. No. 80CH1602-2), pp. 256-259.

5. This data was taken by Dr. Thomas Parker of the Raytheon Research Laboratory, Waltham, MA, using two 500 MHz two-port SAW resonators[2] made at Sperry Research.

6. D. T. Bell and R. C. M. Li, "Surface acoustic wave resonators," Proc. IEEE, 1976, 64, pp. 711-721.

7. W. J. Tanski, "Surface acoustic wave resonators on quartz," IEEE Trans. Sonics and Ultrasonics, 1979, SU-26, pp. 93-104.

8. M. G. Holland and L. T. Claiborne, "Practical surface acoutic wave devices," Proc. IEEE, 1974, 62, pp. 582-611.

9. S. James and I. H. Wilson, "Fine tuning of S.A.W. resonators using argon ion bombardment," Electronics Letter, 1979, 15, pp. 683-684.

10. C. N. Helmick and D. J. White, "Observation of aging and temperature effects on dielectric-coated SAW devices," Proc. 1978 Ultrasonics Symposium (IEEE Cat. No. 78-CH1344-1SU), pp. 580-585.

11. J. Wise, J. Schoenwald, and E. Staples, "Impedance charaterization and design of 2-pole hybrid SAW resonator filters," 1980 Ultrasonics Symposium Proceedings (IEEE Cat. No. 80CH1602-2), pp. 200-203.

12. C. A. Adams and J. A. Kusters, "Deeply etched SAW resonators," 31st Proceedings of the Annual Symposium on Frequency Control, U.S. Army Electronics Command, Ft. Monmouth, N.J., pp. 246-250.

13. S. P. Miller, R. E. Stigal, and W. R. Shreve, "Plasma etched quartz SAW resonators," Proc. 1975 Ultrasonics Symposium (IEEE Cat. No. 75CH0994-4SU), pp. 474-477.

14. W. R. Shreve, "Surface wave resonators and their use in narrowband filters," Proc. 1976 Ultrasonics Symposium (IEEE Cat. No. 76CH1120-5SU), pp. 706-713.

15. J. L. Vossen and W. Kern, Editors, "Thin film processes," Academic Press, New York, 1978.

16. "PLASMOD" etcher/stripper manufactured by TEGAL Corp., Sunnyvale, CA.

17. "Dry plasma ethcing of SiO_2 films," Bulletin 8213-TA1 (LFE Corp., Waltham, MA 02154), April 1975.

18. J. W. Coburn and H. F. Winters, "Plasma etching - A discussion of mechanisms," J. Vac. Sci. Tech., 16, 391 (1979).

19. R. C. M. Li and J. Melngailis, "The influence of stored energy at step discontinuities on the behavior of surface wave gratings," IEEE Trans. Sonic Ultrason. SU-22, 189, (1975).

20. W. H. Haydl, P. Hiesinger, R. S. Smith, B. Dishler, and K. Heber, "Design of quartz and lithium niobate SAW resonators using aluminum metalization," Proc. 30th Annual Symposium on Frequency Control, U.S. Army Electronics Command, Fort Monmouth, N.J., pp. 346-357.

21. J. I. Latham, W. R. Shreve, N. J. Tolan, and P. B. Ghate, "Improved metalization for surface acoustic wave devices," Thin Solid Films, 64 (1979), pp. 9-15.

22. P. B. Ghate, J. C. Blair and C. R. Fuller, "Metalization in microelectronics," Thin Solid Films, 45 (1977), pp. 69-84.

23. T. E. Parker, "1/f phase noise in delay lines and resonators," Proc. 1979 Ultrasonics Symposium (IEEE Cat. No. 79CH1482-9), pp. 878-881.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

GaAs SAW RESONATOR OSCILLATORS WITH ELECTRONIC TUNING

M. Gilden and T. W. Grudkowski
United Technologies Research Center
East Hartford, CT 06108

Abstract

High Q SAW resonators have been fabricated on GaAs and used to control the frequency of oscillators. Since GaAs is both a piezoelectric and semiconducting material, there is the potential of obtaining complete oscillators in monolithic form with frequency tuning by electronic means. SAW resonators were studied parametrically on semi-insulating GaAs substrates. These devices had either groove or metallic electrode reflective gratings. Tunable SAW resonators were fabricated on epitaxial GaAs substrates that used Schottky barrier electrodes as the resonator gratings. The tuning did not materially alter insertion loss. Two-port resonator insertion loss values were typically between 15 and 20 dB and loaded Q values at 180 MHz were as high as 12,000. Although the temperature coefficient of delay for GaAs is 52 ppm/^{o}C, it is possible to temperature compensate GaAs delay lines by the addition of a Au/SiO_2 layer.

1. Introduction

The piezoelectric properties of semiconducting GaAs provide a means for developing high quality SAW resonators with electronic tuning. These GaAs substrates with useful SAW properties are also suitable for other electronic components such as amplifiers. Thus, there is a potential for developing GaAs integrated circuits that would include SAW oscillator as well as other SAW and electronic devices for complete integrated subsystems. An example of the frequency response of a SAW resonator at 180 MHz is shown in Figure 1. It has an insertion loss of 13 dB and a Q of 9300.

The preferred orientation of GaAs substrates for SAW applications is the surface with a <001> normal and propagation in the (110) direction. The SAW properties for this orientation are listed in Table I and are compared, for reference, with those of the commonly used ST cut of quartz. The similarity in properties, except for temperature sensitivity, is evident and indicates the practical value of GaAs.

This work was partially supported by USAERADCOM, Ft. Monmouth, NJ, under Contract DAAK20-79-C-0263.

1981: 35th Annual Frequency Control Symposium

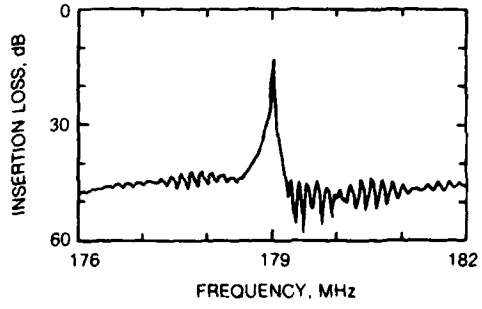


Fig. 1 Frequency Response of a Two Port GaAs Resonator

2. GaAs SAW Resonators and Electronic Tuning

In order to characterize the SAW resonators a series of devices were fabricated on semi-insulating GaAs. The grating reflectors used were formed from either parallel grooves or parallel metallic electrodes. The metallic electrode gratings are later described for use as electronic tuning elements on epitaxial GaAs. A cross section of the SAW resonators is shown in Figure 2. The same photolithographic masks were used for both grooves and metallic lines resonators. Thus, by varying the processing, identical input and output transducer were obtained while the groove grating lines were offset by one-half a period from the metallic

TABLE I
GaAs AND QUARTZ SAW PROPERTIES

	GaAs*	QUARTZ (ST)
VELOCITY (m/sec)	2868	3158
$\Delta V/V$	3.6×10^{-4}	5.8×10^{-4}
ATTENUATION (180 MHz)	1 dB/cm	0.31 dB/cm
LINEAR TEMPERATURE COEFFICIENT OF DELAY (PPM/°C)	+52	0 (AT −26°C)

* SURFACE NORMAL [001], PROPAGATION (110)

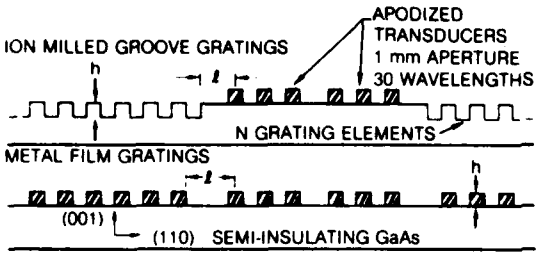


Fig. 2 Cross Sectional Views of GaAs SAW Resonators

grating lines. The sketches of the resonators in Figure 2 also indicate the pertinent dimensions of the devices.

An epitaxial semiconducting layer was provided for electronic tuning[1]. Direct electronic tuning of the resonator frequency was then accomplished by varying the SAW velocity within the Schottky barrier grating by controlling the depletion layer depth. The shorting of the electric fields accompanying the SAW by the free carriers in the undepleted regions alters the SAW velocity but does not materially alter attenuation. Figure 3 shows a section of a tuned grating, using Shottky barrier electrodes, with a depletion region associated with each electrode. As the reverse bias is increased the individual depletion regions expand and then merge into a single region as the tuning continues until the epitaxial region is completely depleted. The reverse bias is applied between the grating resonator electrodes and an ohmic contact electrode that is placed outside the SAW region. Typical doping levels of the GaAs epitaxial layer were about 10^{15} carriers per cubic cm and the typical thicknesses were about 8 μm. At this doping level the undepleted semiconducting region appears as a good conductor at 180 MHz. With this tuning

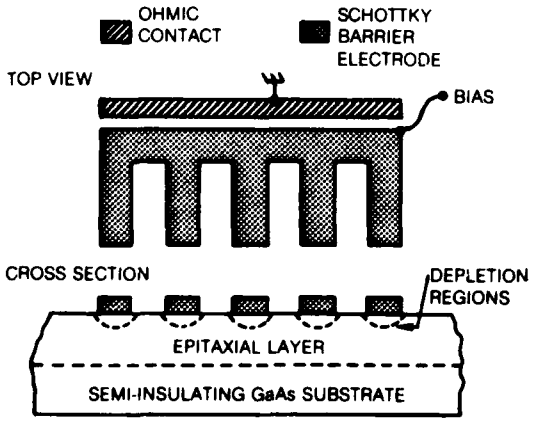


Fig. 3 Tunable Schottky Barrier SAW Grating

technique approximately 700 ppm change in velocity has been observed[2]. The actual tuning range is determined by the epitaxial layer properties, the distribution of the SAW electric field in the layer[1], and the relative length of the tuned portion of the resonator.

The effect of the SAW resonator parameters on the electronic tuning range can be related to the delay times associated with the various sections of the resonators shown in Figure 4. The results of an analysis of the tuning range is given in Eq. (1) which contains only the most significant terms

$$\frac{\Delta f_r}{f_r} = \frac{-\Delta\tau_L + \left(\frac{\Delta f_c}{f_c}\right)\tau_G}{\tau_L + \tau_G} \qquad (1)$$

where

$$\tau_G = 1/\pi\overline{\Delta f} \qquad (2)$$

and where f_c is the center frequency of the response of the grating reflector, v_c is the velocity under the gratings, τ_L is the grating gap delay, τ_G is the equivalent group delay of the grating reflector, and $\overline{\Delta f}$ is the bandwidth of the gratings. The change $\Delta\tau_L$, associated with the gap between the gratings that contains the transducers, is relatively small because the transducers region cannot be tuned electronically along with the gratings. The region around the transducers must be completely depleted because the transducer coupling is greatly reduced by the presence of carriers[3].

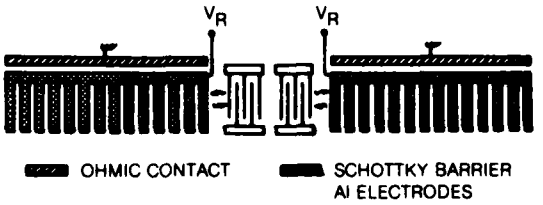


Fig. 4 Sketch of a Complete Tunable GaAs SAW Resonator

From Eq. (1) it is seen that the maximum tuning range is obtained for large values of τ_g. This condition, it can be shown, leads to a relative maximum tuning range equal to the relative change in the SAW velocity in the grating region, Eq. (3).

$$\left.\frac{\Delta f_r}{f_r}\right|_{\tau_G \text{large}} = \frac{\Delta v_c}{v_c} \qquad (3)$$

This condition implies that the grating should be designed with weak reflectors so that the frequency

response of the grating will be relatively narrow in accordance with Eq. (2) which indicates that τ_G is inversely proportional to grating bandwidth[4,5].

3. Experimentally Determined Parameters of GaAs Resonators

Two basic parameters for designing a resonator grating are: 1) the strength of the grating $\Delta Z/Z$, and 2) its center or synchronous frequency f_c. Both are characterized in terms of h/λ, the normalized grating electrode height or groove depth. The parameter $\Delta Z/Z$ is the relative change in characteristic impedance of an equivalent periodically stepped transmission line associated with the grating. From theoretical considerations, a relationship for $\Delta Z/Z$[4,5] is:

$$\left(\frac{\Delta Z}{Z}\right) = \alpha \frac{h}{\lambda} = \left(\frac{\pi^2 \overline{\Delta f}^2}{f_c^2} - \frac{\pi^2}{N^2}\right)^{1/2} \qquad (4a)$$

$$\frac{\Delta Z}{Z} \approx \frac{\pi \overline{\Delta f}}{f_c} \qquad (4b)$$

where the term $\overline{\Delta f}$ is the grating bandwidth, defined as the spacing between the edges of the transmission frequency response for propagation through a single grating, and N is the number of grating lines. The linear coefficient α in Eq. (4) has characteristic values for metallic lines and grooves depending upon the particular substrate material used. The parameter $\Delta Z/Z$ in Eq. (4) is also related to the maximum value of the grating insertion loss response (i.e., transmission through a single grating) at the synchronous frequency.[4]

The synchronous or center frequency of the grating is equal to

$$f_c = v_c/\Lambda \qquad (5)$$

where v_c is the grating velocity and Λ is twice the periodicity of the grating, i.e., the synchronous wavelength is fixed by the geometry of the grating. The grating velocity, v_c, is in turn given approximately by[3]

$$v_c = v_o \left[1 - K_v \left(\frac{h}{\lambda}\right)^2\right] \qquad (6)$$

where v_o is the limiting SAW velocity within the grating as h/λ goes to zero. For grooves this velocity would be that for a free surface while for metal electrodes, it would correspond to an effective value for the case of periodic, massless shorting electrodes.

Measurements of the grating transmission response yield basic parameters required for proper resonator design. The results of measurements made on a series of devices with different values of h were used to obtain values of the coefficient K_v in Eq. (6) for the effective grating velocity for grooves and for metallic gratings. The grating velocity is plotted in Figure 5 as a function of $(h/\lambda)^2$ showing the predicted linear relationships of Eq. (6). The grooves have a stronger effect on the grating velocity. The experimentally determined values for K_v are 4.6 for metal lines and 20.6 for grooves. These values are based upon measurements at 179 MHz. The K_v values may not necessarily be completely independent of frequency as was pointed out for quartz[4]. The grating strength in terms of $\Delta Z/Z$, the incremental change in transmission line impedance at the grating edges,

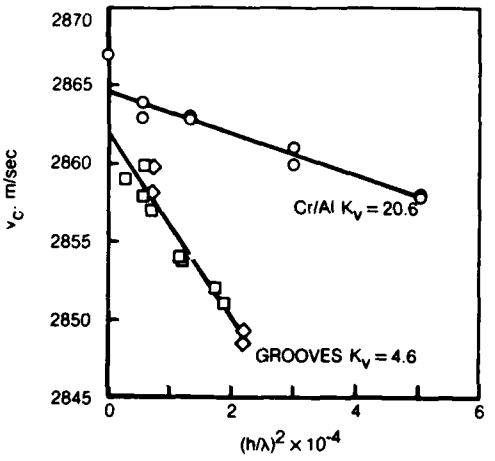


Fig. 5 Synchronous Velocity for GaAs Gratings as a Function of h/λ

is plotted in Figure 6 as a function of h/λ for groove gratings and for metal line gratings. The values of $\Delta Z/Z$ were calculated from grating bandwidth ($\overline{\Delta f}$) measurements using Eq. (4a). The experimental results show the expected linear relationship between $\Delta Z/Z$ and h/λ. The slopes of the curves, comparing grooves and metal lines, do not show as clear a difference as was found for the synchronous velocity. The linear coefficient from our experiment was 1.2 and 1.6 for groove resonators and was 1.1 for metal strip grating devices.

4. Measurements of Electronically Tuned Resonators and Oscillators

Representative frequency response curves for a resonator with zero bias and a -15 volt bias applied to the gratings is shown in Figure 7. The maximum frequency shift observed was 30 kHz, or 170 ppm relative to the resonant frequency of 179 MHz. In

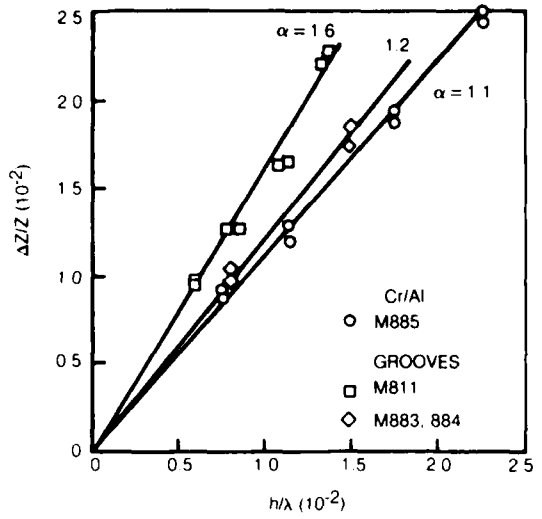


Fig. 6 Dependence of $\Delta Z/Z$ for GaAs Gratings as a Function of h/λ

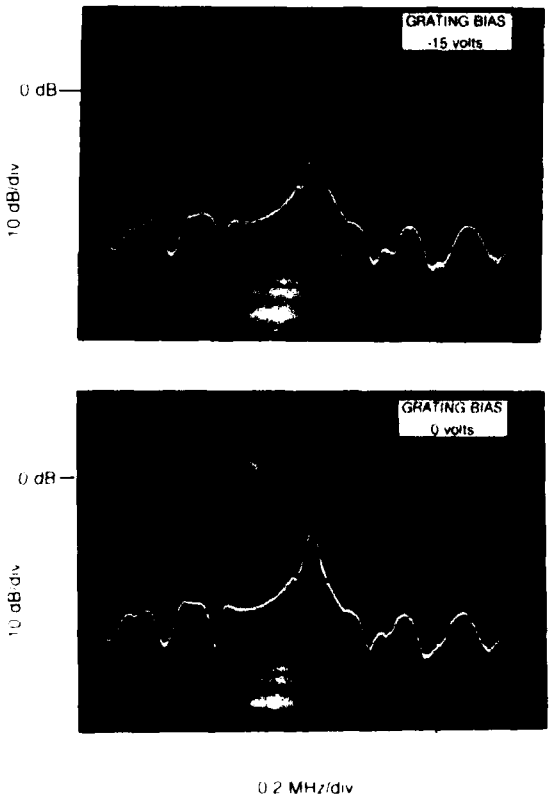


Fig. 7 Frequency Response of Epitaxial GaAs SAW Resonator at 15 and 0 Volts Reverse Tuning Bias

this case the tuned region was only one-third of the total delay path as discussed with respect to Eq. (1). This frequency shift corresponds to about 1.5 times the 3 dB bandwidth. The insertion loss and loaded Q, with transducers operated without series tuning inductors, were 22 dB and 5500, respectively. Over the voltage tuning range, with increasing reverse bias, the insertion loss typically increased by about 0.5 dB. When the transducers were series tuned, the insertion loss was decreased to values as low as 4 dB while the resonator loaded Q was reduced to less than 2000. Voltage tuning by means of an additional intracavity plate electrode phase shifter was typically an order of magnitude less than that observed with grating bias because of the relatively small portion of the acoustic path occupied by the phase shifter.

A circuit schematic for the tunable oscillator is shown in Figure 8. The two port GaAs SAW resonator is used in the feedback loop. The heavy lines trace out the rf feedback path. The remainder of the oscillator rf circuit path contains amplifiers, dc blocking capacitors, and two T-section LC filter stages for fine adjustment of loop transmission phase in order to insure oscillation as close as possible to the resonant frequency of the GaAs SAW resonator. The output of the oscillator is capacitively coupled through C_3. Transducer dc bias is applied through 10k ohm isolation resistors, and bypass capacitors are used to provide direct rf ground for the transducers. The dc bias and ground connections for resonator tuning, using both the Schottky barrier resonator gratings and a plate electrode between input and output transducers, are also indicated. The amplifier bias connections are made through choke inductors for isolation in order to minimize the dc power consumption of the circuit. The SAW resonator was mounted alongside the circuit board and the transducer ports of the GaAs SAW resonator were wirebonded to pads on the circuit board. All other oscillator circuit components were mounted on the printed circuit board.

The frequency tuning range, frequency temperature sensitivity, and short term frequency stability (Allan variance for a one second gating time) were evaluated for a tunable GaAs SAW oscillator. In the oscillator, the small signal amplifier gain available was approximately 28 dB while the tunable GaAs SAW resonator had a minimum insertion loss of about 22 dB. There were provisions on the circuit board to inductively match the transducers if it was desired to use less amplifier gain or fewer amplifiers. In the present circuit the dc bias condition for each amplifier is 6V @ 25 mA, thus the total dc power dissipation of the oscillator electronics is only 300 mW.

The oscillator tuning curve, frequency change

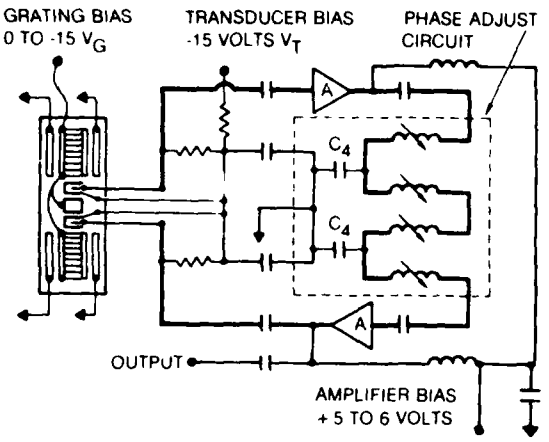


Fig. 8 Diagram of the Tunable GaAs SAW Resonator Oscillator

(ppm) versus dc bias voltage, is shown in Figure 9. The maximum total frequency tuning range for the oscillator was about 80 ppm, for a total reverse grating bias voltage change from 0V to 15V. Most of the frequency tuning occurred for reverse grating bias voltages between 2 to 6 volts. The remaining frequency tuning for reverse bias voltages less than 7 volts was attributable to the residual charge carrier depletion between grating lines. Although GaAs SAW resonators with larger frequency tuning ranges were fabricated, the device used for this oscillator exhibited better insertion loss and phase characteristics.

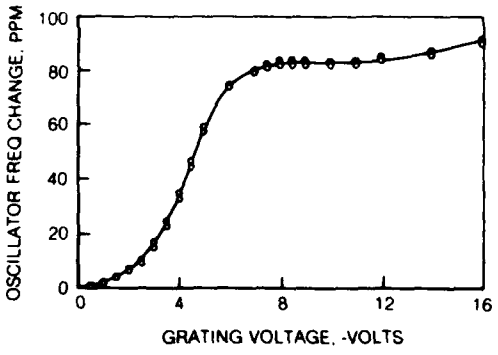


Fig. 9 GaAs SAW Oscillator Tuning Curve with Both Gratings and Plate Biased

The temperature sensitivity of the oscillator was measured for several reverse grating bias voltages between 0 and 15 volts. For a temperature range from 10°C to 45°C the linear temperature coefficients of delay for the various grating bias voltage were essentially equal. The mean value of +52.8 ppm/°C is in very good agreement with the value of +52 ppm/°C measured using long delay path devices.

The oscillator short term stability (Allan variance) for a one second gating time was also measured. Measurements were made for several values of grating reverse bias voltage using sets of 240 consecutive frequency readings. Because the GaAs SAW substrate is not a temperature compensated material, there was frequency drifting during the four minutes required for data taking even though a thermally isolated environment was used. The sensitivity of the oscillator to temperature drifting was estimated to set a lower limit of about 0.25×10^{-8} on the measurements of Allan variance. A measured value of about 1×10^{-8} was obtained for the Allan variances for reverse bias values up to about 12 volts. Near 15 volts the Allan variance increased rapidly to a value of about 5×10^{-8}. This unexpected effect has yet to be explained. In general, noise fluctuations in bias supplies used for the gratings and transducers appear to have contributed to degrading the stability of the oscillator. The best values of Allan variance, 8 to 9×10^{-9}, were measured when both the gratings and tuning plate were held at rf ground potential.

The temperature sensitivity of the surface acoustic waves on GaAs is detrimental both for long term and short term stability. To overcome this limitation, a technique for temperature compensation using an overlay of Au/SiO_2/Ti indicates that a zero value for the temperature coefficient of delay can be obtained[6]. The SiO_2 provides the compensation but the gold layer is required to keep the SAW in a low loss mode.

5. Conclusions

This work has demonstrated that GaAs is a suitable substrate for the fabrication of high Q SAW resonators and voltage tunable resonators for use in oscillator frequency control. A useful temperature compensation technique employing thin film overlays has also been demonstrated[6], but has yet to be applied to resonator design. These GaAs SAW components can be fabricated on the same type of substrate material as is currently being used for GaAs integrated circuit development. It is therefore possible to integrate these SAW components together with electronics on the same monolithic substrate. Such integration would significantly reduce the size and cost of high Q stable VHF/UHF oscillator circuits.

REFERENCES

1. Grudkowski, T. W.: Active Acoustic Waves and Electrons in Gallium Arsenide, M. L. Report 2440, Stanford University, 1974.

2. Grudkowski, T. W., M. Gilden, and G. K. Montress: Integrated Surface Acoustic Wave (SAW) Devices Using Monolithic Gallium Arsenide Circuits. DELET-TR-79-0263-F. Final Report, June 1981, ERADCOM.

3. Grudkowski, T. W., G. K. Montress, M. Gilden, and J. F. Black: GaAs Monolithic SAW Devices for Signal Processing and Frequency Control, 1980 Ultrasonics Symposium Proceedings, pp. 88-97.

4. Tanski, W. J.: Surface Acoustic Wave Resonators on Quartz. IEEE Trans. Sonics and Ultrasonics, SU-26, No. 2, 93-104, March 1979.

5. Cross, P. S.: Properties of Reflective Arrays for Surface Acoustic Resonators. IEEE Trans. of Sonics and Ultrasonics, SU-23, No. 4, July 1976, pp. 255-262.

6. Grudkowski, T. W., M. Gilden: Realization of Temperature-Compensated GaAs Surface Acoustic Wave Delay Lines. Applied Physics Letters, Vol. 38, 15 March 1981, p. 412.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

SHALLOW BULK ACOUSTIC WAVES IN BERLINITE

K. F. Lau, K. H. Yen, R. B. Stokes, R. S. Kagiwada
TRW Electronics and Defense
Redondo Beach, California

and

B. H. T. Chai
Allied Chemical Corporation
Morristown, New Jersey

Summary

This paper examines the shallow bulk acoustic wave (SBAW) properties in berlinite. Single crystals of berlinite were grown hydrothermally by the transport method. The crystal cuts investigated included rotated Y cuts with rotation angle θ near +40.5° and -64°. Theoretical calculations based on published material constants of berlinite indicated that these crystal cuts have zero first order temperature coefficient of delay at near room temperature.

The wave velocity, coupling coefficient, and temperature behavior of the SBAW were characterized by fabricating and evaluating delay lines and oscillators. The temperature behavior was found to differ from that of quartz. For quartz, a parabolic temperature behavior exists for cuts with θ near +40°, while a cubic temperature behavior exists for cuts with θ near -50°. For berlinite, the frequency-temperature curve was observed to be a cubic function for θ = 40.5°, while a parabolic function was observed for θ = -64°. The effect of metallization is to modify the flatness of the frequency-temperature curve. One of our devices showed only 2.5 ppm shift in frequency over the temperature range of 15°C. This is comparable to the stability of SAW in ST cut quartz for the same temperature range.

The investigation confirms the prediction of temperature stable SBAW devices on berlinite. It points out that there are significant third order terms in the temperature coefficient of the material constants. It is likely that by adjusting the substrate angle and metal thickness, even better temperature stability can be achieved with this material.

Introduction

Berlinite (α-$AlPO_4$), an α-quartz isomorph, promises to have a significant impact on frequency control due to its unique combination of acoustic wave properties. The material has a relatively high electromechanical coupling coefficient and at the same time exhibits a temperature-compensated property.[1] Investigations of detailed acoustic wave properties in berlinite have, however, been hampered by the lack of high quality single crystals with adequate sizes. Recent advances in crystal growth have made available crystals with adequate quality and dimensions for further evaluation.

Shallow bulk acoustic waves or surface skimming waves are shear horizontal bulk waves which propagate along the surface of the substrate. Devices based on SBAW in quartz have shown advantages including higher operating frequency, better temperature stability and lower spurious responses. The SBAW device fabricated on -50.5° rotated Y-cut quartz, for example, has demonstrated better temperature stability than SAW on ST quartz.[2-4]

SBAWs in berlinite are expected to possess the same advantages as SBAWs in quartz. In addition, due to their higher coupling, the devices are expected to have much lower insertion losses.

Experimental Details

Crystal Growth

Because of the α-β phase transition at 580°C, berlinite crystals can be grown only by the hydrothermal process. Concentrated phosphoric acid (7.3M) was used as the transport medium. Since the solubility of berlinite in phosphoric acid is retrograde, crystal seeds have to be hung at the warmer part of the autoclave, whereas the nutrient is placed at the cooler part of the autoclave. This is in reverse of the arrangement for quartz crystal growth.

Twinning of berlinite is the major obstacle to overcome in order to grow usable material for device applications. Both the Dauphine (electrical) twins and the Brazil (optical) twins are very common in berlinite. As stated before,[5] Brazil twins are extremely easy to form in berlinite crystals and are most difficult to eliminate. At present, limited success has been achieved in growing twin-free, optically clear berlinite crystals up to 4 cm in length.

*This work was partially supported by the Army Research Office under Contract No. DAAG26-78-C-0043.

Crystal Substrate Fabrication

The berlinite crystal which was to be cut to substrate slices was first mounted on a two-circle goniometer head specially designed to fit to the Rigaku X-Ray Piezo-goniometer machine. Crystal orientation is based on $(10\bar{1}0)$ reflection with $2\theta = 22.31°$. After the orientation was determined, the goniometer head was transferred and mounted on a Varian bread-slicer machine for cutting. It was necessary to cut several crystals at the beginning to make corrections for the transfer of the goniometer head from the X-ray machine to the slicing machine. For single rotated Y-cuts, the accuracy is better than ±0.2° on rotation and ±0.5° on tilting.

The sliced substrates were first fine-lapped with a Hoffman Lapping Machine, and then etched with ammonium bifluoride solution to check twinning. Twinned areas were then trimmed before the substrates were mounted on a LECO planetary lapping machine for polishing. Linde A alumina powder mixed with a 1 to 1 volume ratio of glycerine-water solution was used as the polishing agent.

Device Fabrication

The wave velocity, coupling coefficient and temperature coefficient of delay (TCD) of SBAW in rotated Y-cut berlinite were calculated previously, using material constants measured by Chang and Barsh.[1,2] The result indicated that temperature stable cuts occur in two regions of the rotation angle θ, +40.5° and -64°. The SBAW in the +40.5° substrate corresponds to the SBAW in AT-cut quartz while the -64° corresponds to that of the BT-cut quartz. We have therefore chosen these two orientations for the present investigation. In addition, substrates with θ equal to 39.5°, 40°, 41°, and 41.5° were also selected to allow investigation of orientation dependences of various wave quantities.

SBAW delay lines were fabricated for the evaluation of wave properties. The delay lines were fabricated on crystal blanks with dimensions at least 5 mm x 10 mm. The direction of wave propagation is at 90° off the X-axis. The angle accuracy of the alignment is estimated to be at the order of ±0.5°. Each transducer of the delay line consists of 140 finger pairs with a uniform aperture width of 33 wavelengths. The transducer contained split fingers which generate strong fundamental as well as third harmonic responses. The devices were fabricated with standard lift-off techniques.

Characterization Techniques

The SBAW velocity at room temperature was measured from the delay line synchronous frequency and the spacing of the transducer fingers. This technique measures the SBAW velocity under metallized transducers. The coupling coefficient was calculated from the measured synchronous impedance and static capacitance of the transducer. Equivalent circuit model[2] developed for the SBAW transducer was used for the calculation. The temperature coefficient of delay (TCD) was measured by the oscillator method. The SBAW delay line was incorporated into an oscillator and the output frequency measured as a function of temperature. Only the SBAW delay line was placed in the oven so that all other components of the oscillator circuit remain unaffected by the temperature change. The temperature-frequency curve can therefore be used to deduce the TCD.

Results

+40.5° Rotated Y-Cut Berlinite

Figure 1 shows the frequency response of a SBAW delay line fabricated on 40.5° rotated Y-cut berlinite. The unmatched insertion loss of the device was 22 dB. The response shape was a $(\sin x/x)^2$ function and few spurious signals were observed.

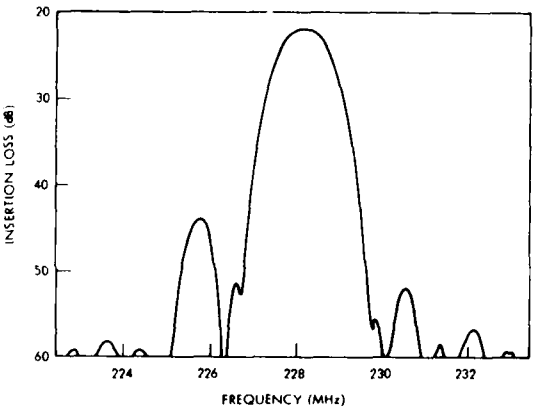


Figure 1. Frequency Response of SBAW Delay Line on 40.5° Rotated Y-Cut Berlinite

The wave velocity as determined by the center frequency of the response was 4310 m/sec. This is compared to a theoretically predicted velocity of 4323 m/sec. Figure 2 summarizes the comparison between theory and experiment of the SBAW velocities for substrates with θ near 40.5°. The agreement is satisfactory. The discrepancy for the 40.5° and 41° devices can be attributed to a number of factors, including metal loading effects, misalignment of transducers and inaccurate crystal angles.

The temperature stability of the SBAW device with θ near 40.5° was estimated theoretically based on published material constants and their temperature coefficients.[1] Since none of the third order coefficients was available, all calculated results showed a parabolic dependence. As shown in Figure 3, the turnover temperature of the θ = 40.5° substrate was calculated to be at room temperature.

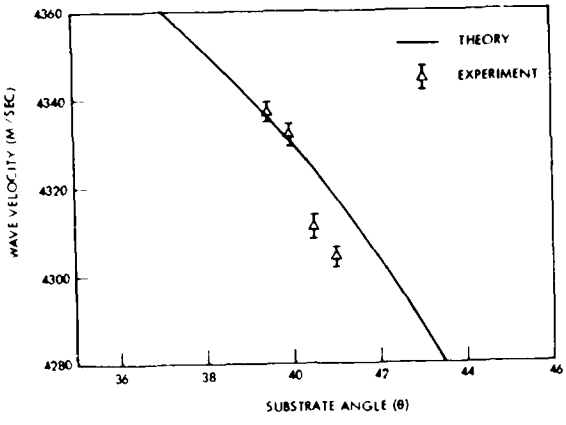


Figure 2. Theoretical and Experimental Data on SBAW Velocity in Berlinite

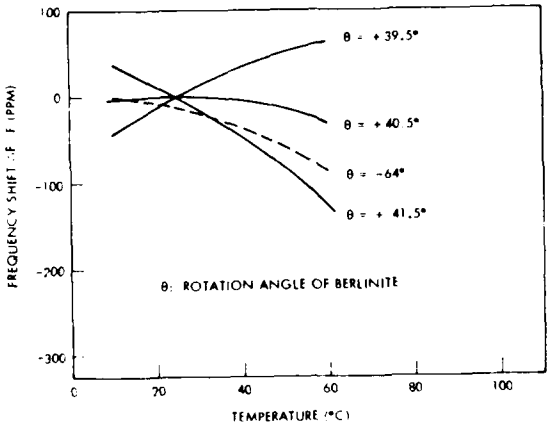


Figure 3. Calculated Frequency-Temperature Curves of Berlinite

The measured temperature behavior of the SBAW with θ near 40.5° is shown in Figure 4. These frequency-temperature curves turned out to be cubic functions. Zero first order coefficient occurred at temperatures between 50 and 60°C. It appeared that with smaller θ the turnover temperature shifted toward lower temperatures. This is in contrast with the calculated result which tends to indicate the opposite. It can be concluded from these measurements that some of the third and higher order terms in the temperature coefficient of material constants are significant and cannot be neglected. Furthermore, the symmetry of the third order coefficient of berlinite is different from that of quartz. In AT-cut quartz, the frequency-temperature curve of SBAW is parabolic.

The shape of the frequency-temperature curves was found to be dependent on the material used in the transducers. By using only Al metallization

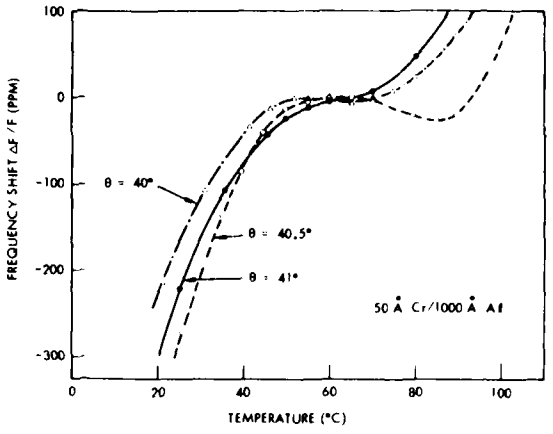


Figure 4. Frequency-Temperature Curves of SBAW Devices on Berlinite with θ Near 40.5°

without a thin Cr layer for adhesion, the frequency temperature curve was significantly changed, as shown in Figure 5. For these devices, the frequency shift is less than 180 ppm over a temperature range of 100°C.

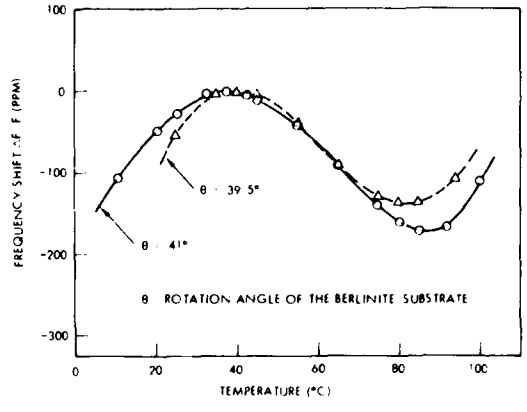


Figure 5. Frequency-Temperature Curves of SBAW Devices Fabricated Without a Cr-Adhesion Layer

The temperature frequency curve was also dependent on the metal thickness. Figure 6 shows a measurement of the thickness dependence of the temperature characteristics of SBAW in 40.5° rotated Y-cut berlinite. The effect of metallization is to modify the flatness of the frequency-temperature curve. The device with thick metallization showed only 2.5 ppm shift in frequency over the temperature range of 15°C. This is comparable to the stability of SAW in ST cut quartz for the same temperature range.

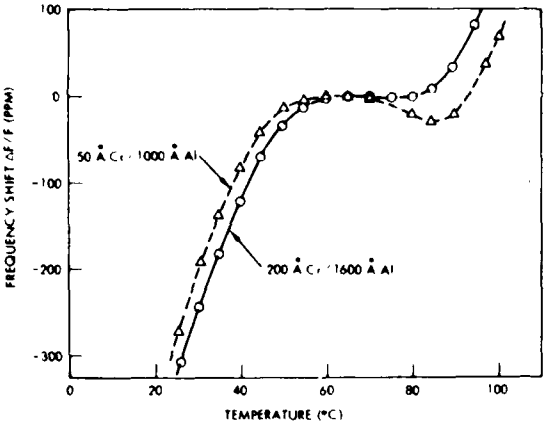


Figure 6. Frequency-Temperature Curves of SBAW Devices on 40.5° Rotated Y-Cut Quartz With Two Different Metal Thicknesses

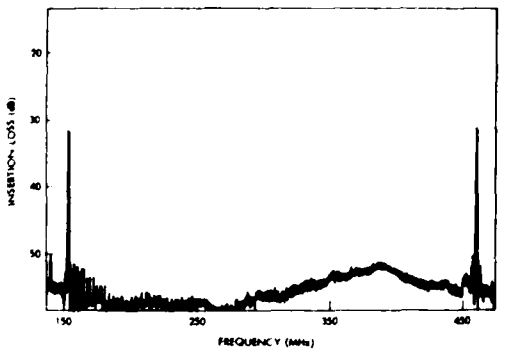


Figure 8. Fundamental and Third Harmonic Responses of SBAW Delay Line Fabricated on -64° Rotated Y-Cut Berlinite

-64° Rotated Y-Cut Berlinite

The frequency response of the SBAW in -64° rotated Y-cut berlinite showed a series of plate mode peaks at frequencies slightly above the SBAW frequency. By roughening the back surfaces of the substrate, these plate modes were eliminated. Figure 7 shows the response of the device with roughened back surface. The wave velocity was measured to be 2893.6 m/sec, while calculation predicted a velocity of 2895.1 m/sec.

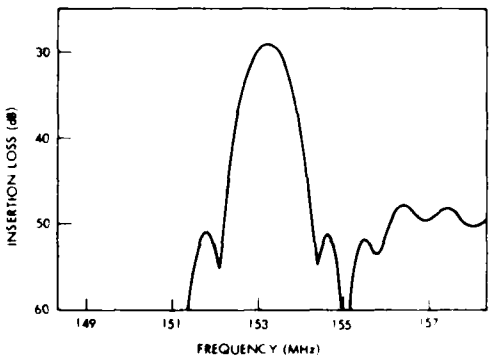


Figure 7. Frequency Response of SBAW Delay Line on -64° Rotated Y-Cut Berlinite

With the plate mode suppressed, the SBAW device on -64° rotated Y-cut berlinite showed no significant spurious responses. This is shown in Figure 8, which also shows the fundamental and third harmonic responses of the delay line.

The frequency-temperature curve of the SBAW device on -64° again showed a marked difference from that of SBAW on quartz, as shown in Figure 9. For SBAW on -50.5° rotated Y-cut quartz, the frequency temperature curve is a cubic function. In berlinite, the slow SBAW showed a parabolic dependence. For quartz, the effect of Cr/Al metallization is to lower the turnover temperature. In berlinite, it appeared to increase the turnover temperature. The 2nd order coefficient of delay for the Cr/Al device was estimated to be approximately 190 x $10^{-9}/°C^2$.

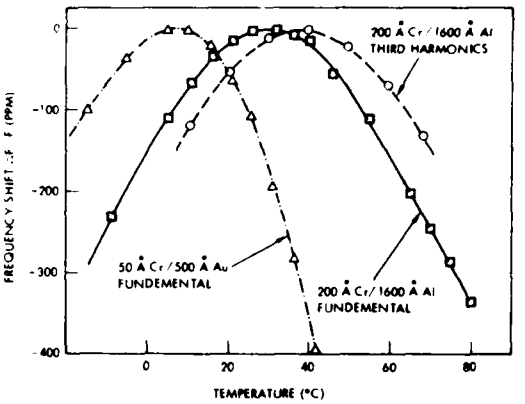


Figure 9. Frequency-Temperature Curves of SBAW Devices on -64° Rotated Y-Cut Berlinite

A device with Cr/Au metallization was also tested. The turnover temperature is lower than the other two devices. Its second order temperature coefficient of delay also appeared larger.

Coupling Coefficient

The coupling coefficients of the berlinite substrate were evaluated, and the results summarized in Table 1. The measurements were not

Table 1. Electromechanical Coupling Coefficient of Berlinie Samples

Crystal Angle	Wave Type	Theory	Experiment
+40.5°	SAW	0.29%	0.15%
+40.5°	SBAW	6.7%*	1.5%*
-64°	SBAW	1.2%	1.0%*

*To obtain k^2, divide the number by the square root of N, where N is the number of finger pairs in the transducer.[2]

entirely consistent. The numbers reported in Table 1 represented the best experimental values. The SBAW measurements indicated a much lower than theoretically predicted coupling for the 40.5° cut, while the coupling is close to theory for the -64° cut.

An independent check was performed by fabricating surface wave devices on a 40.5° substrate and measuring its impedance and static capacitance. This measurement of the k^2 is considered to be more reliable since the equivalent circuit theory of surface wave devices is well understood. The coupling strength was measured to be half of that predicted by theory.

Conclusion

The investigation confirms the prediction of temperature stable SBAW devices on berlinite. It points out that there are significant third order terms in the temperature coefficient of the material constants. Further utilization of this material will depend on the ability to improve the crystal quality. The result of this study indicates that by adjusting the substrate angle and metal thickness, even better temperature stability is likely to be achieved.

Acknowledgement

The authors would like to thank A. M. Kong and S. L. Chen for fabricating the SBAW devices.

References

1. Chang, Y. P., and G. R. Barsh, IEEE Trans. on Sonics and Ultrasonics, SU 23, p. 127 (1976).

2. Yen, K. H., K. F. Lau, and R. S. Kagiwada, "Recent Advances in Shallow Bulk Acoustic Wave Devices," 1979 Ultrasonics Symp. Proceedings, p. 776 (1979).

3. Lau, K. F., K. H. Yen, R. S. Kagiwada, and A. M. Kong," A Temperature Stable 2 GHz SBAW Delay Line Oscillator," Proc. 34th Annual Frequency Control Symposium, U. S. Army Electronics Technology and Device Laboratory, Fort Monmouth, NJ, May 1980.

4. Lau, K. F., K. H. Yen, R. S. Kagiwada, and A. M. Kong, "High Frequency Temperature Stable SBAW Oscillators," 1980 Ultrasonics Symposium Proceedings, 1981.

5. Chai, B. H. T., M. L. Shard, E. Buehler, and M. A. Gilleo, "Experimental Data on the Piezoelectric Properties of Berlinite," 1979 Ultrasonics Symposium Proceedings, p. 577 (1979).

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, 07703, May 1981

DIRECT DIGITAL FREQUENCY SYNTHESIS

Albert L. Bramble

US Army Electronics Technology and Devices Laboratory (ERADCOM)
Fort Monmouth, New Jersey 07703

Abstract

Direct Digital Frequency Synthesis uses a phase accumulator which is supplied by a frequency control digital word. The accumulator addresses a Read Only Memory containing amplitude values of the desired wave shape. The ROM, in turn, supplies commands to a Digital to Analog Converter for conversion to quantitized analog signals. This technique has been relegated to the synthesis of audio and low frequency RF signals by limitations in device technology, specifically in ROMs and D to A converters. Rapid evolution of LSI device technology and incorporation of the direct digital circuit into indirect synthesizer circuits will result in hybrid designs capitalizing on the advantages of both techniques.

The design of the basic direct digital synthesizer and its advantages and limitations in comparison to the indirect phase locked loop is presented. Possible ways of realizing the advantages of both techniques by combining the two are also presented.

Indirect Phase Lock Synthesis

The indirect phase locked loop synthesizer is the workhorse of the synthesizer field and has found considerable success in high volume commercial products as well as military communications equipment. A brief review of a simple phase locked loop is in order to point out the advantages and limitations of this technique. Figure 1 shows a basic phase controlled synthesizer. This simple loop will reach equilibrium only when the output frequency of the variable divider and F_{Ref} are equal. At this point a phase error between the two phase detector inputs will provide a stable D.C. output which tunes the VCO to an output equal to $N \cdot F_{Ref}$. Changing N will cause the output of the variable divider to differ from F_{Ref} and the phase detector output will contain a varying correction voltage with its period a function of the frequency difference. The loop filter is a low pass filter usually designed to have a cutoff of approximately .1 F_{Ref}. The filtered correction voltage tunes the VCO until equilibrium is again reached when the output of the variable divider is equal to F_{Ref}. The minimum output frequency increment is equal to F_{Ref} and is produced when the variable divider is changed by 1. Phase noise and spurious responses may be reduced by lowering the cutoff frequency of the loop filter and thereby decreasing the loop bandwidth (BW). However, this increases response time resulting in switching times greater than the 3/B.W. time obtainable with BW = .1 F_{Ref}. Present emphasis in military communications is to employ frequency hopping to decrease jamming vunerability. At the same time restrictions on the noise floor and spurious output of these systems place conflicting requirements on the frequency synthesizer used in communications systems. H.F. SSB transceivers require frequency increments of 100 hz and extremely low noise floor and spurious outputs. Adding the requirement of frequency hopping places conflicting requirements on the indirect phase locked synthesizer that can require considerable complexity, expense, and volume to meet. Narrow frequency increments dictate a low reference frequency which dictates low frequency cutoff in the loop filter which, in turn, results in long switching times which are not amenable to fast frequency hopping. Increasing the loop bandwidth to allow faster switching raises noise floor and phase jitter. Narrow frequency increments, especially in VHF systems also place critical stability requirements on the VCO since drift correction occurs at relatively wide intervals ($1/F_{Ref}$). The simple loop shown requires a programmable divider capable of operating at the VCO frequency. The counter must be capable of being reset in less than a half cycle of the VCO frequency which could be as high as 400 MHz for some military systems. The constraints for the single loop synthesizer are shown in Figure 2. Multiple loop synthesizers and dual modulus prescaling are two techniques presently used to obtain wide bandwidths, narrow increments as well as low noise and spurious outputs. These techniques are expensive in design time, component cost and size.

Advantages and Disadvantages

Figure 3 list several of the advantages and limitations of the phase locked synthesizer. All of the advantages listed are obtainable but only at the expense of other operating characteristics. A single loop circuit providing low frequency, widely spaced outputs over a limited bandwidth with moderate restrictions on spurious outputs will be relatively simple. All components with the exception of the loop filter could be realized in LSI form. As performance increases so does complexity and increased use of conventional components to realize optimum performance. Fast

hopping, low spurious output and narrow output increments are characteristics which are obtainable singly but with great difficulty together. The disadvantages of the phase locked loop are as shown in Figure 3. The major limitation is relatively slow switching speeds and widely spaced output frequency increments.

Direct Digital Synthesis

The direct digital frequency synthesizer is shown in simplified form in Figure 4. The multiple inputs shown for the Phase Accumulator, Read Only Memory (ROM), and Digital to Analog Converter (D/A), indicate parallel logic inputs. The phase accumulator consists of an N-bit frequency storage register which stores a digital frequency control word followed by an N-bit full adder and data register. The digital input frequency control number is entered in the input storage register. At each clock pulse this data is summed with data previously held in the data register and the new data is stored in the data register. The frequency setting input represents a phase angle step that is added to the value of the previous step each $1/F_C$ seconds to produce a linearly increasing digital value. Periodically the accumulator overflows and begins a second cycle at count zero. The rate of overflow is the output frequency in digital code.

if: K = frequency control word
N = number of accumulator bits
F_C = clock frequency
F_O = output frequency

then $F_O = \frac{K F_C}{2^N}$

The magnitude of the frequency control word, K can occur only in interger values, therefore the minimum frequency increment is $F_C/2^N$. To increase the resolution without increasing the output frequency the number of accumulator bits N must be increased. The resolution may be doubled by increasing the accumulator adder and register by one bit.

The Read Only Memory (ROM) is a sine function table which converts the phase information, in digital form, received from the accumulator into digital phase and amplitude values of a sine wave. The number of words in the ROM will determine the phase quantization error and the number of bits in each word will determine the amplitude quantization error. The digital ROM output is presented to the D/A converter which develops a quantitized analog sine wave. The low pass filter removes the high frequency sampling components and provides a pure sine wave output.

One immediately obvious improvement in the basic circuit is illustrated in Figure 5. Since a sinusoid is a symetrical waveshape it is not necessary to store ampliture/phase points for the full 360^0. The size of the ROM may be decreased by a factor of four by storing only points for one quadrant as shown in Figure 6. The decrease in ROM capacity is slightly offset by the additional logic necessary to generate the complement of the accumulator output. The accumulator is shown recycling each quadrant instead of once each cycle as before. The accumulator output is used "as is" for the first and third quadrants. The bits must be complemented so that the slope of the saw tooth is inverted for the second and fourth quadrant. The digital values of the ROM input and output are represented in Figure 7. The most significant accumulator bit is the Sign bit and is used to control the polarity of the D/A output. The next significant bit is the Quadrant bit and is used to control the complementer logic.

Practical Aspects

The maximum number of phase samples per unit time is fixed by the clock frequency. Therefore as output frequency is increased there are fewer samples per cycle available to accurately define the shape of the output sinusoid. The result is increased distortion as output frequency is increased. As a general rule the maximum output frequency is limited to $F_c/F_{out} = 4$.

In practical applications it is necessary to include a storage register between the ROM and the D/A converter. This register is used to hold the digital data for the D/A during accumulator transitions and to allow all data bits to be presented to the D/A simultaneously to reduce output glitches. This additional register could be eliminated with D/A's that have internal storage. In general, noise generated in the D/A converter increases with frequency while amplitude and phase quantization noise generated as a result of quantization error in the ROM is constant and independent of frequency.

Advantages

The direct digital circuit has advantages and limitations different from those of the indirect phase locked loop. The two different approaches complement each other in several important areas. Advantages of the direct digital synthesizer are summarized in Figure 8. The D/A Converter and the output low pass filter are the only two components of the system not directly amenable to LSI construction. Precision D to A's are usually implemented in film hybrid or modular type construction. Hopping speeds of the direct digital approach are limited only by the delays inherent in the digital circuitry and the response time of the low pass filter.

Switching speeds in the order of 1 usec are possible. Output frequency spacing is limited only by the number of accumulator bits. Doubling the resolution only requires the addition of one adder and register bit.

Disadvantages

The direct digital approach is primarily limited in output frequency range. For most applications the clock frequency should be a minimum of four times the output frequency. ROM access times and D/A switching speeds place a definite limit on the upper frequency range. Spurious output in this approach appears to be limited to approximately 50 to 60 db below the carrier - far above the low levels possible with a phase locked loop.

Hybrid Techniques

The indirect phase locked loop is capable of very high frequency, wide bandwidth operation, but at the expense of switching speed or narrow output frequency increments. The direct digital approach has the capability of very narrow output frequency increments and fast switching speeds but is severly limited to low frequency operation. Both circuit techniques may be implemented, to a large extent, by digital monolithic, custom LSI devices. Let us now explore a few possible hybrid combinations of the two techniques. Figure 9 shows a basic, single, indirect phase locked loop with the fixed reference oscillator replaced by a direct digital synthesizer. The output frequency = $N.F_{Ref}$. This configuration would allow F_{Ref} to be incremented in small steps. However, since the direct digital supplied F_{Ref} is multiplied by N the F_{Ref} increments would have to be very small indeed, especially if the output frequency and N were large. Also, as N is changed the effect of fixed increments available from the direct digital synthesizer would also change. Hence, a logic control circuit would be necessary to program both the variable loop divider and the direct digital circuit to provide the desired output frequency.

The circuit shown in Figure 10 mixes a direct digital low frequency output with a higher frequency VCO signal. A balanced mixer would provide sum and difference outputs. A tunable low pass filter would reject the sum and pass $F_{out} - F_o$. The output of this circuit would then be:

$$F_{out} = N F_{Ref} + F_o$$

This configuration simplifies determination of the output frequency and would allow smaller output frequency increments. The circuit shown would allow use of a high loop reference frequency which would permit a correspondingly wide loop bandwidth which would permit fast switching for rapid frequency hopping. Small, rapidly switched, frequency increments are supplied by the direct digital synthesizer. The obvious drawback here is the close spacing of the mixer sum and difference outputs. A sharp, tunable filter would be required to track the mixer output. The frequency control logic used to provide inputs to the ÷ N counter and the digital synthesizer would also have to supply predetermined tuning information to the mixer filter. Implementing the filter to cover a wide frequency range would be a problem.

A further improvement is shown in Figure 11. The double balanced mixer has been replaced by a single sideband modulator. F_1 is the low frequency input from a direct digital synthesizer. Quadrature outputs from the VCO and direct digital synthesizer would be required for the SSB modulator. The advantage here is deletion of the tunable filter at the expense of added analog circuit complexity for the SSB modulator. The major difficulty would be in developing a wideband SSB modulator capable of sufficiently surpressing the carrier and unwanted sideband.

Conclusion

The basic characteristics of the direct digital synthesizer have been described and the circuit compared to the more familiar indirect phase locked loop synthesizer. The two approaches appear to be complementary and could provide the basis for a hybrid synthesizer capable of wide application and able to be implemented almost entirely in digital LSI form.

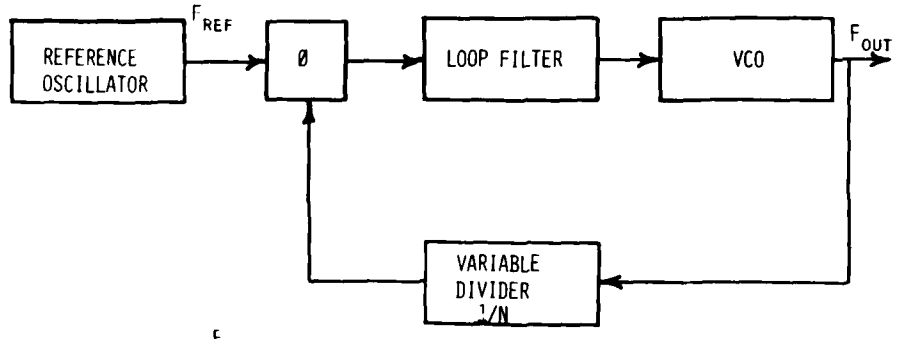
References

1. IEEE Press, "Frequency Synthesis: Techniques and Applications," edited by J. Gorski - Popiel.

2. Hosking, R. H., "Direct Digital Frequency Synthesis."

3. Cooper, H. W., "Why Complicate Frequency Synthesis?", Electronic Design Magazine, July 1974.

PHASE-CONTROLLED INDIRECT SYNTHESIZER



OUTPUT RESOLUTION = F_{REF}
LOOP BANDWIDTH ≈ 1/10 F_{REF}
SWITCHING TIME ≈ 3/B.W.

BASIC PHASE LOCKED LOOP

CONSTRAINTS

- MAX RESOLUTION = REFERENCE FREQUENCY (F_{REF})
- LOOP BANDWIDTH (BW) ≈ 1/10 F_{REF}
- SWITCHING TIME (T_S) ≈ 3/B.W.
- $T_S = 30/F_{REF}$

FOR HOPPING SYSTEM, T_S 1/10 OF FREQUENCY DWELL TIME AT ANY SELECTED FREQUENCY.

- FH ≤ 1/10 T_S
- FH ≤ 1/300 F_{REF}

INDIRECT SYNTHESIS

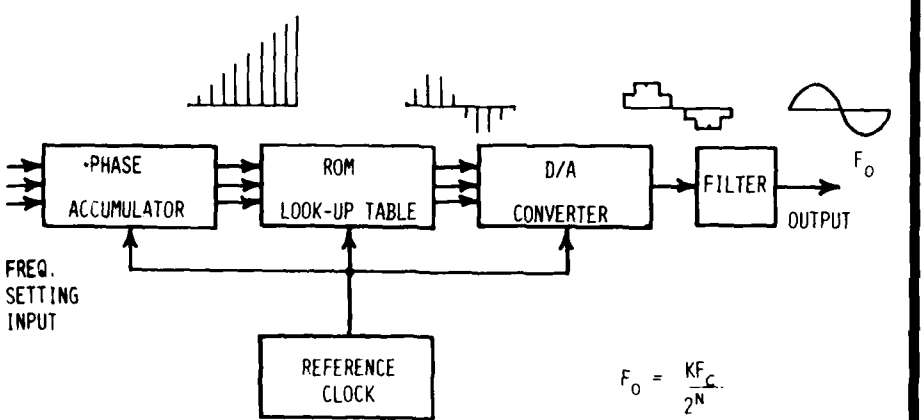
DIGITAL PHASE LOCKED LOOP

ADVANTAGES

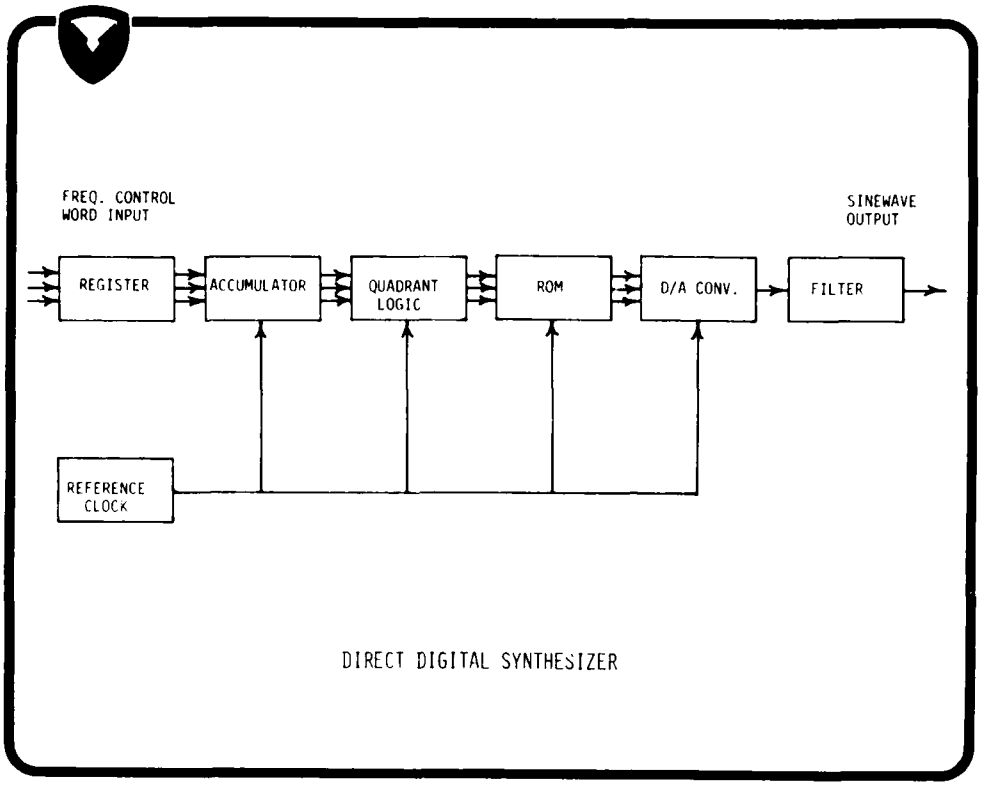
- RELATIVELY SIMPLE, LOW PARTS COUNT.
- WIDE BANDWIDTHS POSSIBLE (OCTAVE OR HIGHER).
- HIGH HOPPING RATES POSSIBLE.
- RELATIVELY CLEAN OUTPUT SIGNAL.

DISADVANTAGES

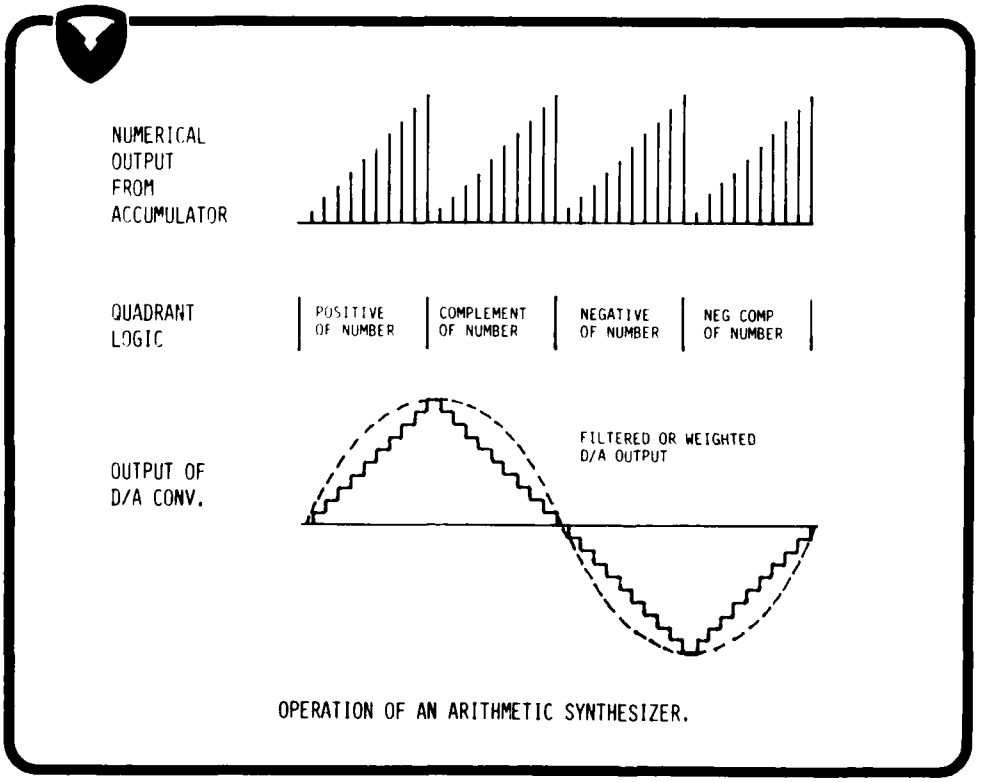
- SOME CIRCUIT COMPONENTS NOT ABLE TO BE IMPLEMENTED IN MONOLITHIC FORM.
- COMPLEX MULTIPLE LOOPS REQUIRED FOR FAST HOPPING.
- CLOSE FREQUENCY SPACING NOT COMPATABLE WITH FAST FREQUENCY HOPPING.



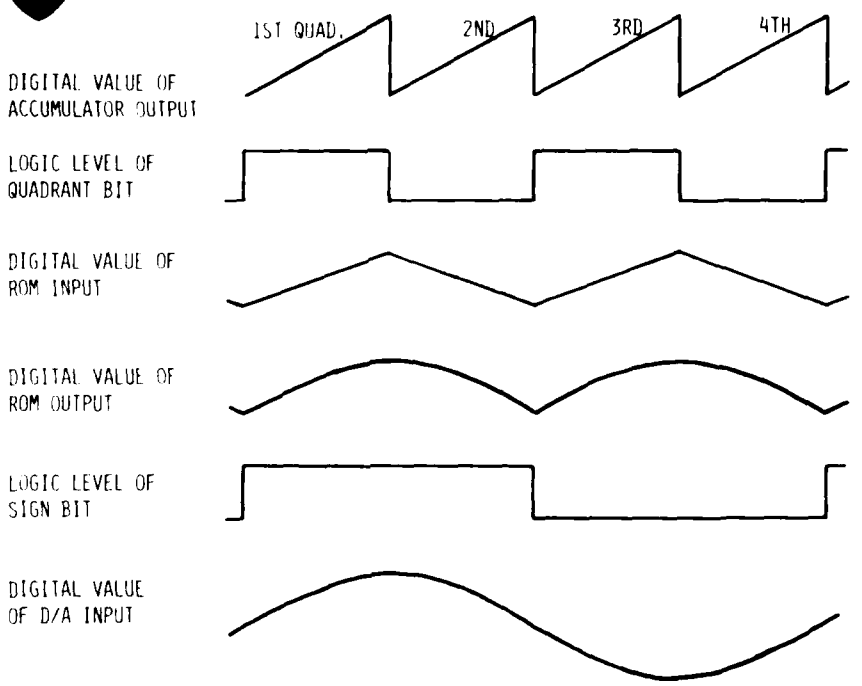
SIMPLIFIED BLOCK DIAGRAM



DIRECT DIGITAL SYNTHESIZER



OPERATION OF AN ARITHMETIC SYNTHESIZER.



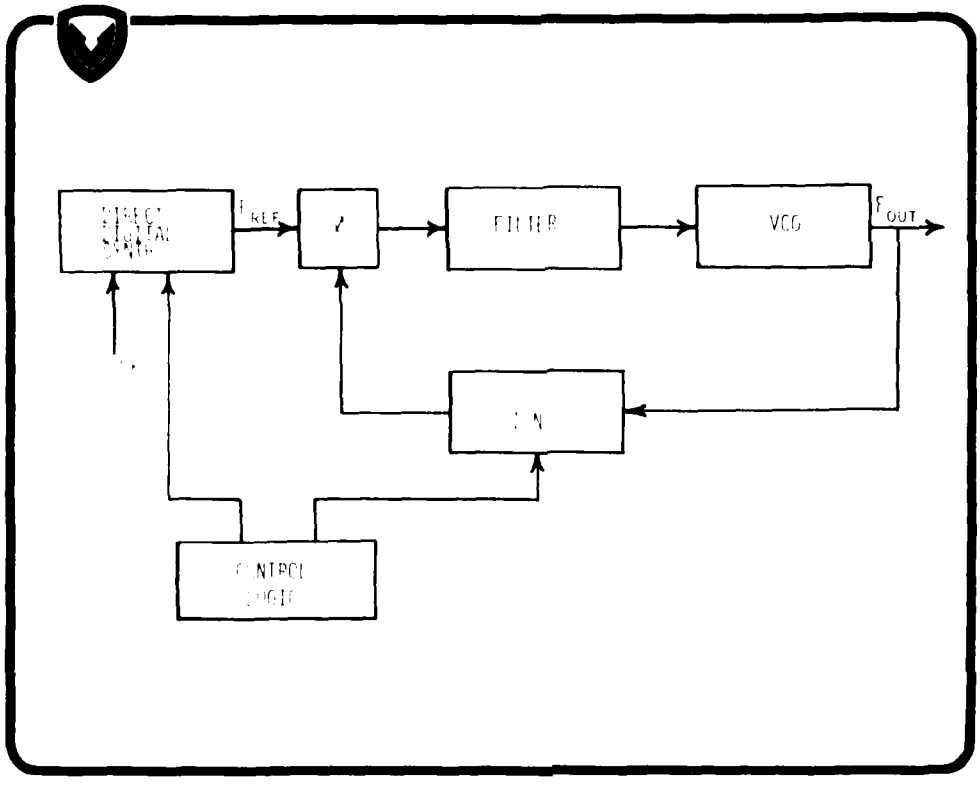
DIRECT SYNTHESIS

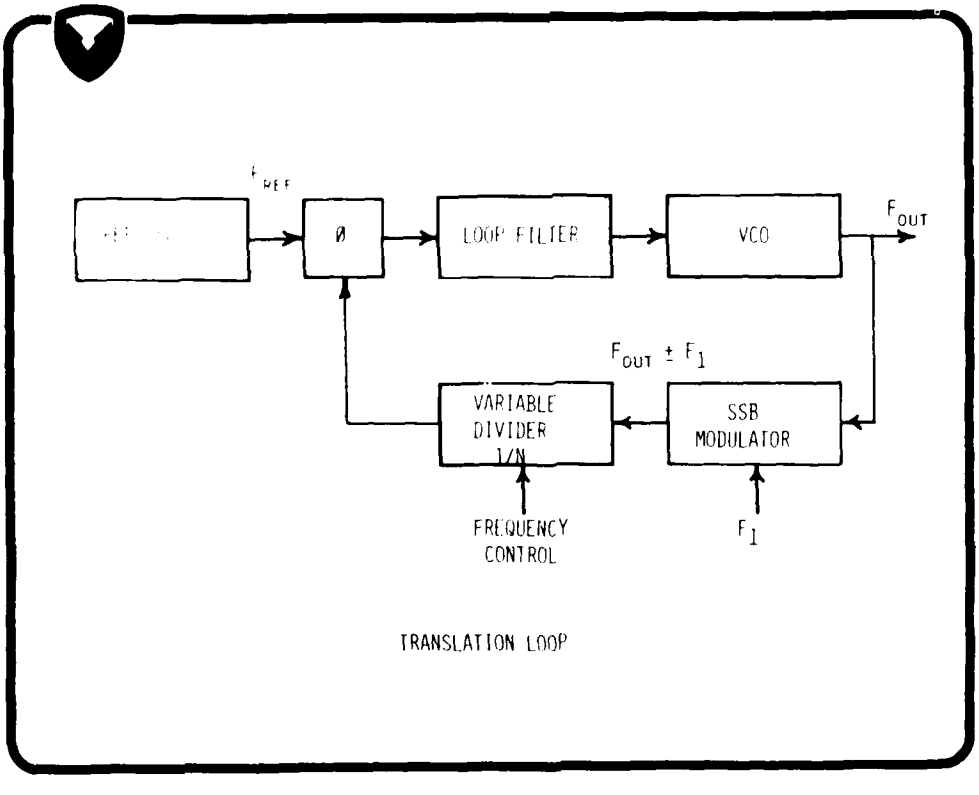
ADVANTAGES

- ALL DIGITAL APPROACH.
- VERY HIGH HOPPING SPEEDS POSSIBLE.
- OUTPUT IS PHASE CONTINUOUS WHEN HOPPING.
- VERY CLOSE FREQUENCY SPACING POSSIBLE.

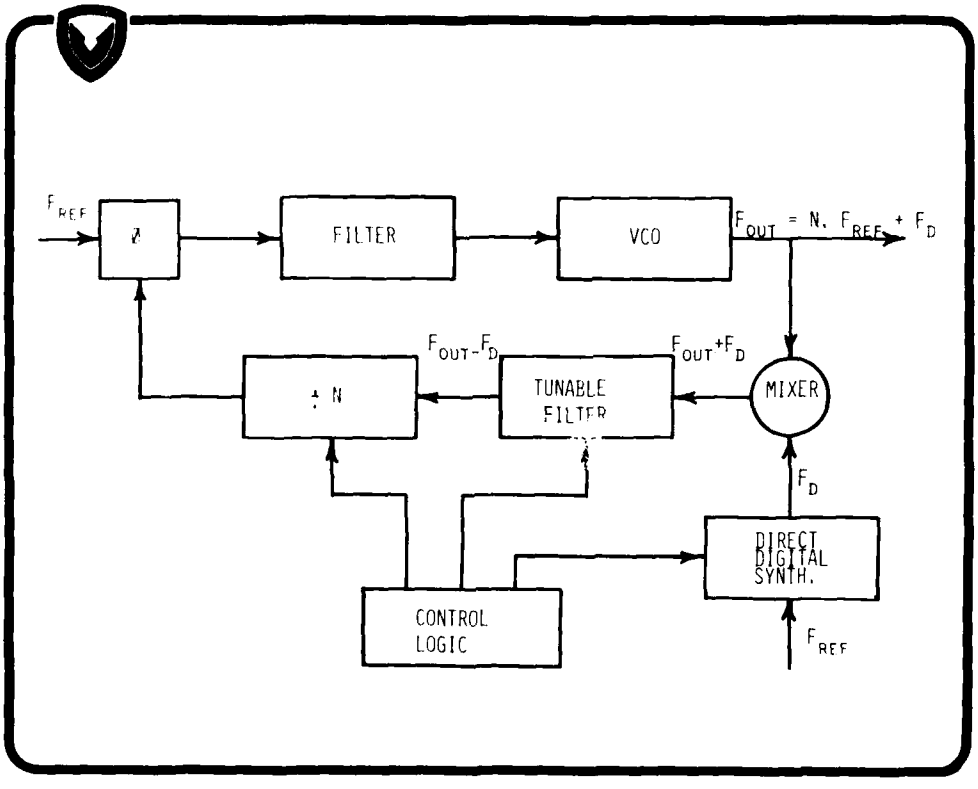
DISADVANTAGES

- HIGH SPEED LOGIC REQUIRED TO APPROACH 10 MHz OUTPUT FREQUENCY.
- D/A CONVERTERS AND ROM ACCESS TIMES ARE A LIMITING FACTOR.
- HIGH LEVELS OF NOISE AND SPURIOUS RESPONSES IN OUTPUT.
- POWER CONSUMPTION IS RELATIVELY HIGH - FUNCTION OF MAXIMUM OUTPUT FREQUENCY.





TRANSLATION LOOP


F_{REF}
÷
FILTER
VCO
$F_{OUT} = N. F_{REF} + F_D$
$F_{OUT}-F_D$
$F_{OUT}+F_D$
÷ N
TUNABLE FILTER
MIXER
F_D
DIRECT DIGITAL SYNTH.
CONTROL LOGIC
F_{REF}

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

A LOW NOISE 500 MHZ FREQUENCY SOURCE

A. Vulcan and M. Bloch
Frequency Electronics, Inc.
New Hyde Park, N. Y. 11040

and

W. Tanski
Sperry Research Center
Sudbury, Massachusetts 01776

Abstract

This paper describes the design of a frequency source for a high reliability, long life application. The unit is totally redundant for reliable failsafe operation and utilizes bulk and surface acoustic wave resonators to insure extremely low phase noise and spurious outputs. Multiple 500 MHz and 400 MHz outputs are generated by direct synthesis from a 25 MHz reference input or an internal 100 MHz TCXO. When using the external reference a 25 MHz input filter and 2 additional crystal filters at 50 MHz and 100 MHz insure that the phase noise floor is reduced to the theoretical minimum before final multiplication. The multiplied output signals at 500 MHz are filtered with two-port SAW resonators. These resonators are designed to operate at input power levels of +15 dBm in a 50 ohm matched system. The filtering scheme also ensures that internally and externally generated modulation sidebands are reduced to negligible levels. Single sideband phase noise is less than -174 dBc/Hz at frequencies spaced greater than 2 MHz from the carrier. Likewise, spurious signals are below -110 dBc at all frequencies.

The frequency source operates in an electrically noisy environment and a high degree of active and passive filtering and package shielding ensures minimum conducted and radiated susceptibility.

Critical system requirements dictate the use of standby redundancy with no single point failures. The packaging concept was developed to minimize subassembly interconnections and provide both physical and electrical independence of the two redundant generators.

Introduction

Achieving a thermal noise floor of -174 dB/hertz at a spacing of 2 MHz from the carrier requires the utilization of bulk and surface wave resonators and a packaging technique that insures internal leakage paths are minimized. The frequency source has 20 outputs, sixteen at 500 MHz and four at 400 MHz. The 400 MHz signals do not require the utilization of SAWS to meet a specified noise floor, and hence, that portion of the system is not discussed in detail. The following paragraph describes the design of the frequency source.

System Design

A functional block diagram of the low noise frequency source is shown in Figure 1. The unit is completely redundant with internal cross strapping occurring at two specific subsystem points. These points were chosen to maximize the reliability of the system and provide an availability factor of 0.99927. The basic frequency reference is an external 25 MHz source or an internal TCXO operating at 100 MHz. The TCXO provides a redundant backup in the event of failure of the external source. The 25 MHz reference frequency is crystal filtered and multiplied to 100 MHz using high efficiency transistor multipliers. A high isolation input cascode amplifier establishes a low system noise figure and reduces internal signal leakage to the input port to a level of -105 dBm. The 25 MHz signal multiplied to 100 MHz, feeds a diode switch along with the TCXO. External power gating selects either the internal or external mode of operation and hence, only one signal is present at the switch output at a given time. A crystal filter using a dual resonator half-lattice design reduces the 100 MHz phase noise at frequencies greater than ±10 kHz from the carrier to an extremely low level. Additionally, spurious signals generated from the multiplication process in the external mode of operation are attenuated by 60 dB. This is important since the output spurious requirements at 500 MHz are stringent. (See Table 1.)

Each crystal filter feeds a low noise amplifier stage providing +17 dBm at its output. The amplifier output feeds cross strap circuitry such that any input simultaneously feeds both output multiplier and amplifier chains. Hence, all modes of operation are available at this point, Internal A, External A, Internal B and External B. The cross strapping, as well as the entire system, is designed so that any single point failure will not compromise output signal integrity. If a failure occurs, switchover is accomplished by the power gating circuitry. The cross straps at 100 MHz feed X4 and X5 multipliers yielding a 400 MHz and 500 MHz signal respectively. The multipliers use microwave stripline transistors in a common emitter configuration to provide stable, low noise multiplication. The 500 MHz multiplier output feeds a SAW resonator assembly consisting of four devices mounted on a single PC board. A 1 X 4

power splitter drives the 4 resonators whose outputs feeds hybrids which are cross-strapped with the outputs from the redundant side. A 50 MHz amplifier boosts the signal level at the input to the SAW resonator assembly to +20 dBm to overcome the SAW assembly insertion loss. The system output level at 500 MHz is +4 dBm. The noise and spurious requirements on the 400 MHz outputs are less stringent and a single transistor stage is used for multiplication with a three pole LC filter providing sufficient rejection of spurious signals. A 500 MHz four pole helical resonator filter is used to suppress the 100 MHz sidebands of the X5 multiplier and provides rejection of 50 dB at +50 MHz around the carrier.

Input power to the frequency source is in the form of redundant +18 Vdc busses for source and side selection. Selection of internal or external frequency source is accomplished by an external command pulse which sets a latching relay to the appropriate position, directing the DC voltage to the TCXO or 25 to 100 multiplier stage. Low noise linear regulators provide an output voltage of +13.5 Vdc to the various circuits and present 80 dB of rejection to low frequency conducted interference on the power lines. The Low Noise Frequency Source must provide spurious free output signals when driven from a DC voltage containing large amounts of corrupting sine waves and transients. The frequency range of the interfering signals are from 30 Hz to 400 MHz. Hence, the judicious use of decoupling and filtering networks is necessary to achieve the desired performance.

Tables 1, 2 and 3 itemize the performance characteristics of the Low Noise Frequency Source. The specified limits are compared with the actual performance data taken from a working engineering model. The specified electrical characteristics are worst case values taking into account a seven year life and the effects of environmental conditions. The spurious output requirements of Table 2 from 5 kHz to 499 MHz from the carrier are met with spurious signals present on the 25 MHz reference input at a level of -111 dBc from 5 kHz to 2 MHz and -68 dBc from 2 MHz to 24 MHz. Additionally, the input phase noise floor at frequencies 100 kHz and up from the carrier is -140 dBc/Hz. A comparison of the phase noise of the 400 MHz and 500 MHz outputs clearly shows the effect of the SAW resonators at frequencies greater than 1 MHz from the carrier. THe 3 dB bandwidth of the SAW's is aproximately 250 kHz and the skirt selectivity provides sufficient noise rejection to permit a level of -174 dBc/Hz to be obtained 6.25 MHz from the carrier. The output to output isolation and output VSWR are a function of the output hybrid design and the integrity of the hybrid source impedance. The SAW resonators are tuned to insure that these requirements are met.

Circuit Design

TCXO

The 100 MHz TCXO is a Butler oscillator using a 5th Overtone TO-5 crystal as the resonator. Table 4 shows the error budget for this internal frequency source. The long term drift error of $\pm 2.5 \times 10^{-6}$ is based upon operational experience of hundreds of similar oscillators used in identical environmental conditions. The buss voltage variation, which is based upon end-of-life parameter variations for the analog regulators and the effect of bus voltage on TCXO frequency, contributes a negligibly small drift parameter to the overall budget. The setting tolerance is based upon initial factory trimming and likewise represents a small offset. The temperature effect of $\pm 4 \times 10^{-6}$ over a range of 0 to 140°F represents the compensated crystal frequency variations. It can be seen that the overall deviation for the life of the unit does not exceed $\pm 6.7 \times 10^{-6}$. This corresponds to ± 3350 Hz, well within the specified range of ± 5000 Hz.

25 to 100 MHz Multiplier

This subassembly consists of a cascode input stage using low noise transistors and emitter degeneration, a full wave diode doubler circuit which rejects the carrier input and unwanted harmonics by 40 dB, a single resonator 50 MHz crystal filter, a common emitter 50 MHz amplifier stage, and a X2 transistor multiplier. The circuits are designed so that the power level at any point does not drop below +3 dBm. This insures that the signal to noise ratio will remain at a reasonable level and succeeding noise figure contributions will have small effect on overall noise performance.

SAW Resonators

Figure 2 is a schematic diagram showing the two quadruple SAW resonator assemblies and their cross-strapping. The 500 MHz input signals at a level of +20 dBm feed 1 X 4 passive power splitters configured as three lumped element 1 X 2 Wilkinson dividers. These splitters have an operational bandwidth in excess of 30%. The four power splitter outputs at a level of +13 dBm feed the SAW resonators via a matching network consisting of a series capacitor and inductance. The unmatched SAW resonators have a natural impedance of approximately 250 ohms and this impedance in conjunction with stray header capacitance in the TO-8 holder is matched to 50 ohms with the series elements. Although a single series inductor would accomplish this function, the capacitor provides variability and permits ease of alignment. The SAW outputs are matched with an identical circuit and feed 1 X 2 power splitters. The eight power splitter outputs from the SAW subassembly feed 4 hybrid

circuits on each board. Figure 2 shows that a failure on either subassembly will not prevent outputting of all 16 500 MHz signals. For example, if resonator FL 1 on subassembly A1 fails, gated power is applied to side B and +20 dBm at 500 MHz appears at the input to the A11 board. This assembly will then feed all 8 hybrids yielding the 16 outputs. Cross-strapping is accomplished via coaxial cables to maintain a 50 ohm impedance throughout the system. The output hybrids are quarter/wave coupled devices which have an operating bandwidth of approximately 20%. Isolation over this bandwidth exceeds 20 dB and the output VSWR is 1.2:1. Each SAW filter assembly consisting of seven 1 X 2 power splitters, four hybrids, and four SAW resonators with their associated matching networks, occupies an area of 3.5 X 3.5 inches.

The SAW resonators were designed by the Sperry Research Center and are described in Reference 1. A sample lot of these devices are undergoing an aging study to ascertain the long term effects of temperature and RF power level on performance characteristics. This study, which is performed on 20 typical resonators will evaluate the effect of stepped stresses on the insertion loss and frequency of the devices. A two year program is projected for the resonators but useful data has already been obtained regarding the safe maximum power levels at which the devices can be used. Based on this data, it is evident that at the +13 dBm levels encountered in the systems, immeasurable insertion loss and resonant frequency changes occur. The accelerated life testing program will yield additional data and the results will be presented i a future paper.

The frequency response of a typical SAW resonator is shown in Figures 3 and 4. Data is plotted from 2 MHz to 1000 MHz in Figure 3 where it is seen that the out-of-band rejection at frequencies below 400 MHz exceeds 60 dB. This high rejection is a function of circuit layout since any parasitic reactances degrade the rejection. Figure 4 shows the SAW resonator frequency characteristics over a band of ±20 MHz around 500 MHz. It is seen that the rejection exceeds 20 dB at spacings greater than 1 MHz from the carrier. This is significant in the present application since the output phase noise approaches the thermal noise floor at ±6.25 MHz from the carrier. The SAW resonators have a matched bandwidth of 250 kHz and an insertion loss of approximately 2.1 dB. Of the 60 devices tested to date, the insertion loss lies between 2.4 and 1.8 dB and the bandwidths between 140 KHz to 300 kHz.

500 MHz Helical Resonator

As shown in Figure 1, a 500 MHz 4-pole helical resonator is placed at the output of the X5 multiplier to provide rejection of unwanted harmonics. The frequency response of this filter is shown in Figure 5. It can be seen that the 400 and 600 MHz multiplier outputs are attenuated by greater than 60 dB. The helical resonator filters have an insertion loss of 1.5 dB and occupy a volume of 3 X 2 X 1 inch. Since the frequency source must operate over stringent shock and vibration conditions, the helical resonators are foamed to reduce the effect of vibration on performance characteristics.

Power Conditioning

The prime input power to the Low Noise Frequency Source is +18 Vdc. This voltage has sine and transient conducted interference which could degrade the overall operating performance of the unit. For this reason RFI filters are used between the various subassemblies and the input voltage is lowpass filtered with a network whose cutoff frequency is 2 kHz. The 18 volt input is regulated to 13.5 volts with linear regulators using specially selected low noise reference Zener diodes. The regulators are short circuit protected and adequately decoupled to insure that input rejection is greater than 80 dB from DC to 30 kHz and 50 dB minimum at 300 kHz. The power conditioning circuits are packaged in separate enclosed compartments which are part of the overall frequency source assembly.

Noise and Spurious Signal Performance

The Low Noise Frequency Source is designed to have low phase noise when used with a 25 MHz external reference having a noise floor of -140 dBc per hertz. Additionally, the spurious requirements which are itemized in Table 2 require that the internally generated sidebands due to the multiplication process are effectively filtered prior to multiplier enhancement. Figure 6 illustrates the spurious rejection and enhancement characteristics of the Low Noise Frequency Source. These spurious requirements are specified over a band of 5 kHz to 499 MHz from the carrier. By following the illustration from left to right, it is seen how spurious signals introduced at the input as well as internally generated spurious signals are rejected by the crystal filters and SAW resonators. It can be seen that the spurious signal level at the input to the A5 amplifier meets the requirements without the use of the SAW filters, the SAW filters being necessary only to meet the noise requirements at spacings greater than 1 MHz from the carrier.

The phase noise performance at various points in the system is graphed in Figure 7. It is seen from the curve that the internal 100 MHz TCXO has better noise than the external reference translated to 100 MHz at frequencies greater than 300 kHz from the carrier. The 100 MHz noise is shaped by the 100 MHz crystal filter to yield Curve II. This noise is essentially at the thermal level due to the selectivity of the filter. The 100 MHz filter output noise is

translated to 500 MHz and degraded 14 dB in the process. The noise at 500 MHz is shaped by the SAW resonators yielding Curve III, the 500 MHz output phase noise. Additional filtering of the 25 MHz reference input signals is obtained from the 25 MHz input filter and 50 MHz filter, although the effect of these devices is not shown in Figures 6 or 7.

In order to actually meet the phase noise and spurious requirements as calculated, the circuit design employed must feature techniques which insure that the requirements are met. This includes the use of low noise transistors, emitter degeneration, transistor biasing for optimum noise figure, the use of optimum source impedance at interstage points for best noise performance, and the use of on board filtering and decoupling networks to minimize interaction between subassemblies.

Mechanical Design

Since the frequency source is a redundant unit, a chassis design is utilized which permits a dual physical configuration with a mirror image on one side of the chassis and a center isolating partition. The overall assembly, which measures 15 X 7 X 2 inches, is shown in Figure 8. The 20 output connectors, 16 at 500 MHz and 4 at 400 MHz are SMA types as are the 25 MHz input connectors. Redundant power connectors are used which also interface fault outputs from the unit to external equipment. Figure 9 is a top view of the Low Noise Frequency Source showing how the various functions are compartmentalized to enhance electrical performance. The RFI filters between the various subassemblies are clearly seen as well as the "dirty box" used for RF filtering the power, command and fault input and output signals. Two particular areas which use low loss dielectric foam to minimize vibration induced noise sidebands are the 100 MHz crystal filter and helical resonators. All other components are staked in place using a structural adhesive with good electrical characteristics. Figure 10 shows the bottom view of the frequency source which is essentially a duplicate of the top. Cross-strapping between the two independent sources is accomplished via coaxial cables passing through the center plate.

The Low Noise Frequency Source is designed to meet all specifications over a 7 year operating lifetime with an ambient temperature range of 0 to 140°F. Normally encountered vibration levels are 0.4 g^2/Hz from 50 to 2000 Hz. The major effect of temperature on the unit is to move the bulk and surface wave resonators in frequency resulting in a reduction of output power and change in output frequency when operating in the internal mode. The overall change in output amplitude is less than 2 dB over the full temperature range. Semiconductor elements which dissipate power in excess of 0.2 watts are heat sunk directly to the main chassis to minimize junction temperature. Junction temperatures are limited to 90°C at the worst case point. Typical junction temperatures are 75°C, with an ambient of 59°C.

Measurement Techniques

Figure 11 shows the test setup used for measuring phase noise and spurious signals. The setup requires the use of two frequency sources whose outputs are mixed to give a DC carrier. The carrier is amplified by 20 or 40 dB and fed to a spectrum analyzer. An elliptic function 100 MHz filter at the output of the mixer attenuates mixer output signals at 400 MHz, 500 MHz, and their harmonics which would overload the amplifier feeding the spectrum analyzer. This filter has 100 dB of attenuation at frequencies above 400 MHz. Low noise amplifiers with 10 dB of gain are used at the output of the frequency sources to override the mixer noise figure which is typically 10 dB (7 dB due to mixer loss and 3 dB amplifier noise figure). These amplifiers must provide linear operation with +14 dBm output power and still have a noise figure below 3.0 dB. The system is set up as shown in the Figure and the phase shifter adjusted to give zero voltage at the mixer output. This insures that the mixer inputs are in quadrature and proper mixing action occurs. The system is calibrated by slightly offsetting the two input frequencies to the preamplifiers and measuring the overall transfer function from input to output. The relative noise and/or spurious level can then be directly converted to dBc. Spurious signals are deliberately injected into one of the frequency sources with a synthesizer to measure the rejection.

Conclusion

This paper describes a Low Noise Signal Source providing VHF outputs whose noise characteristics approach the thermal limit at frequencies spaced greater than 1 MHz from the carrier. The unit is designed for high reliability operation and is redundant. Packaging techniques, electrical design and manufacturing processes are employed which insure that the overall design objective is met.

References

1. Tanski, William J., IEE Ultrasonics Symposium, 1980, Paper G-1, Abstracts.

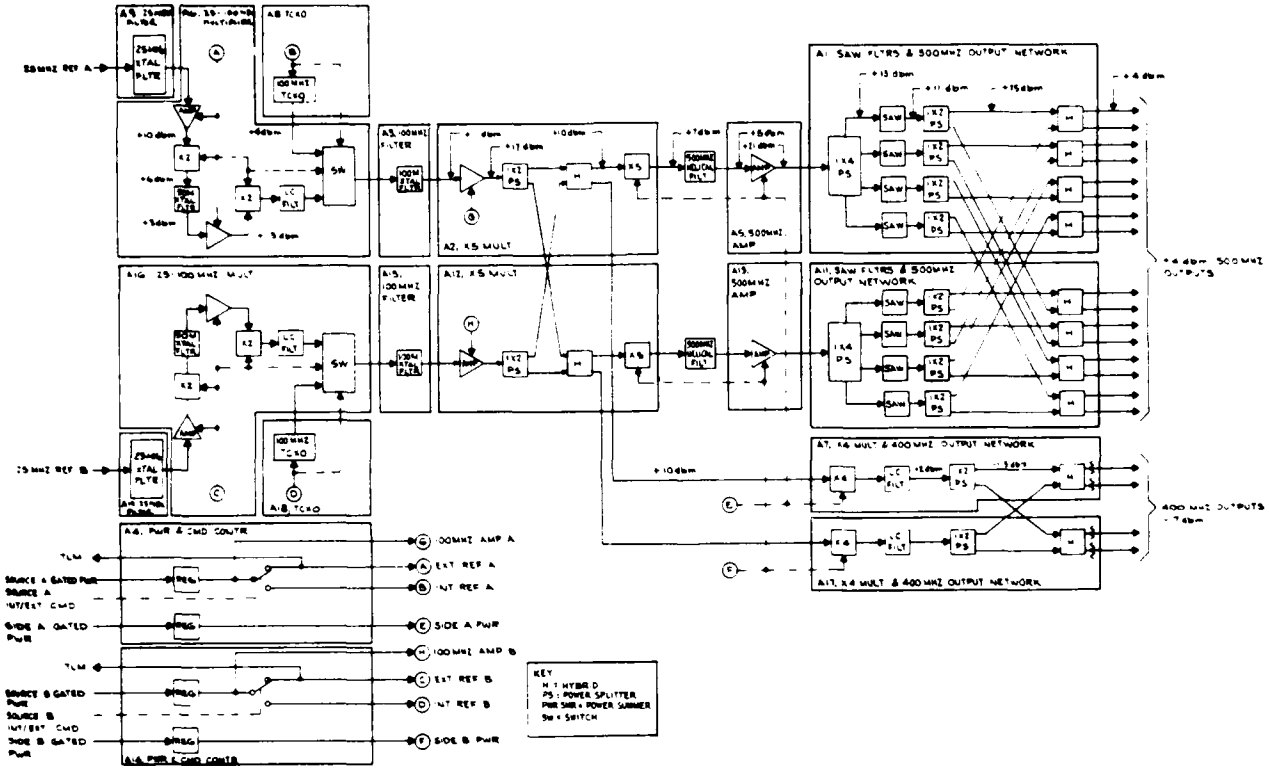


Figure 1. Low Noise 500 MHz Frequency Source, Functional Block Diagram

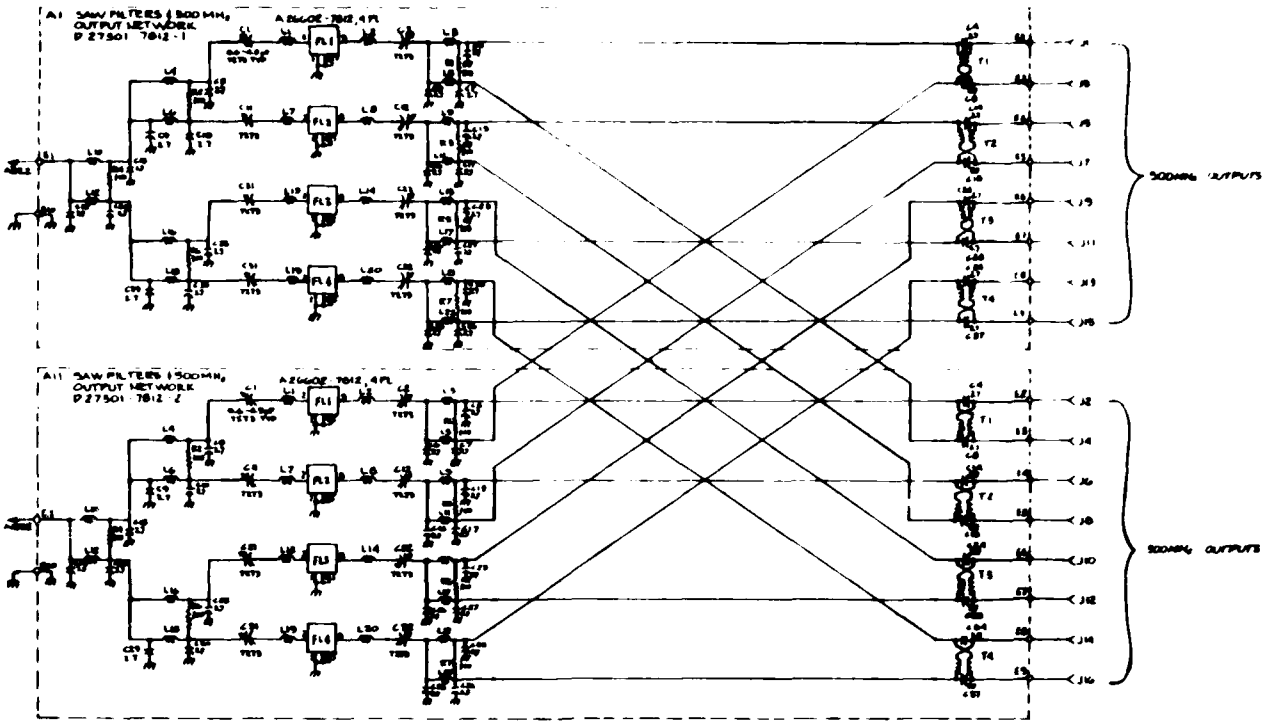


Figure 2. SAW Resonator, Schematic Diagram

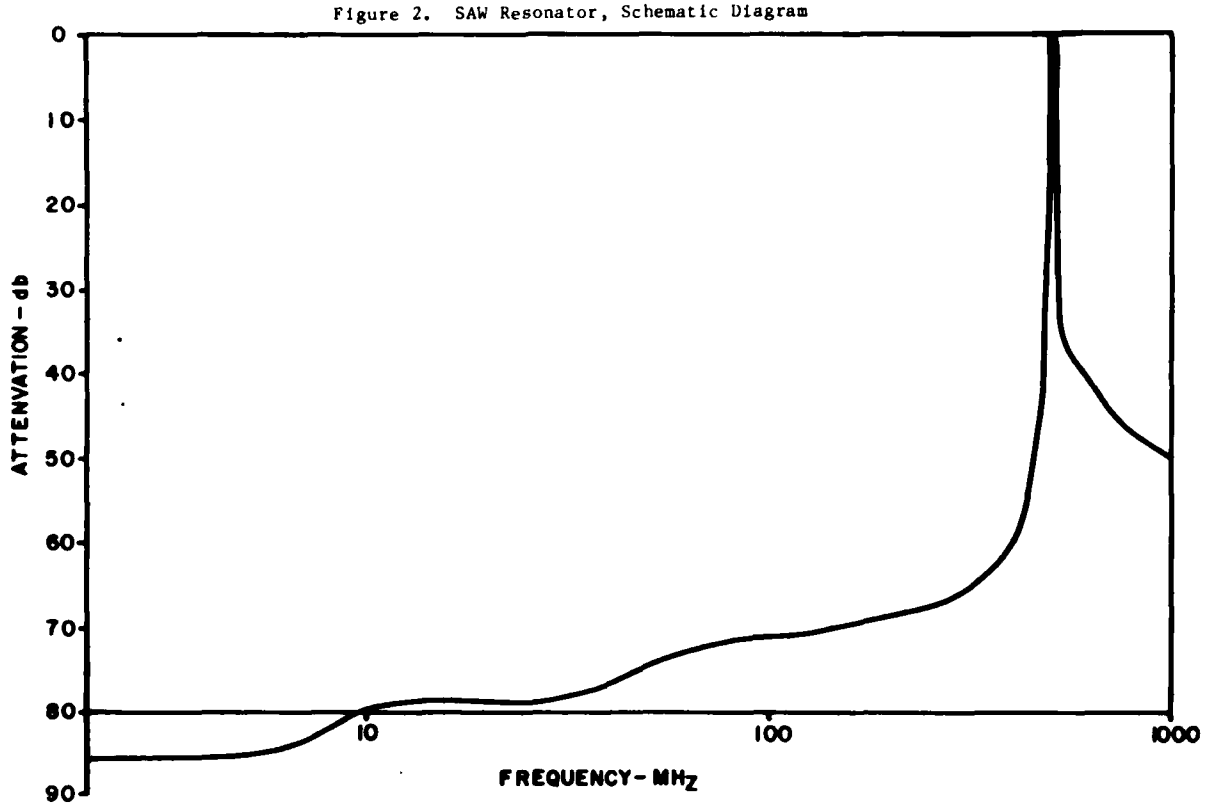


Figure 3. SAW Resonator Response, Wideband

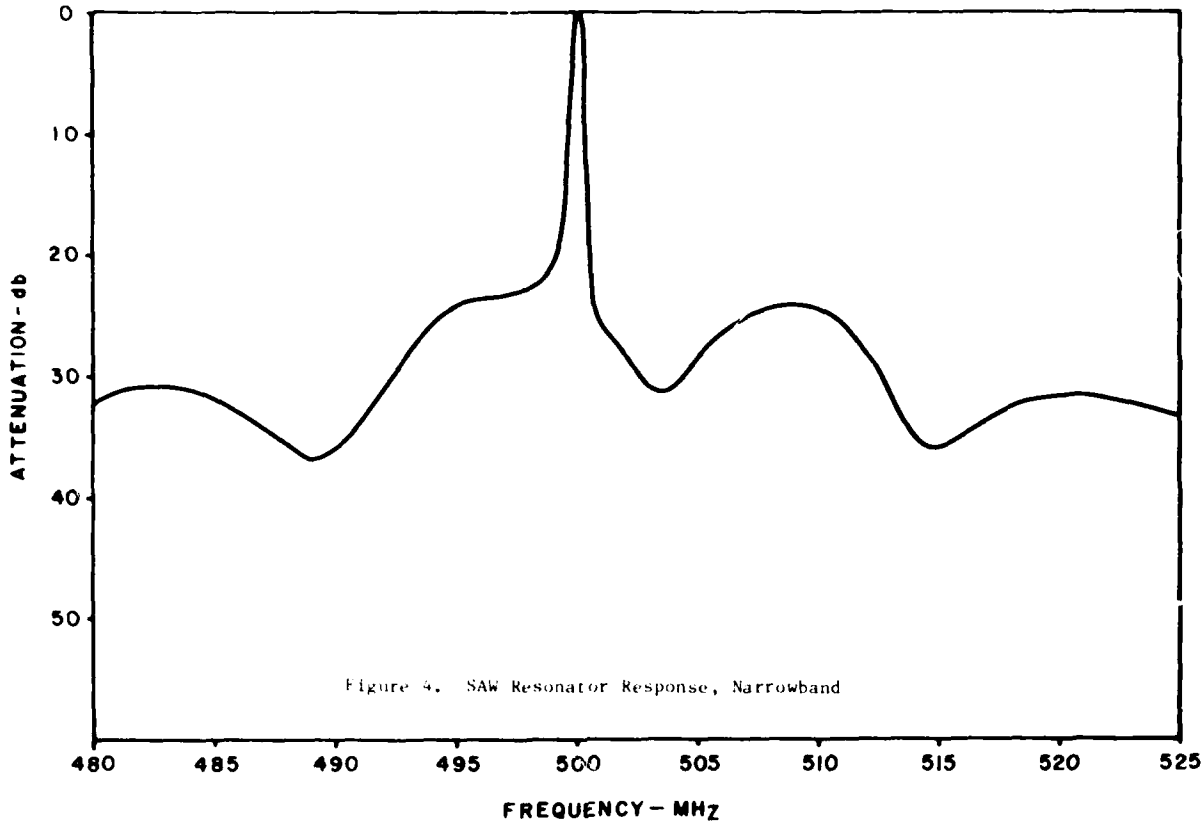


Figure 4. SAW Resonator Response, Narrowband

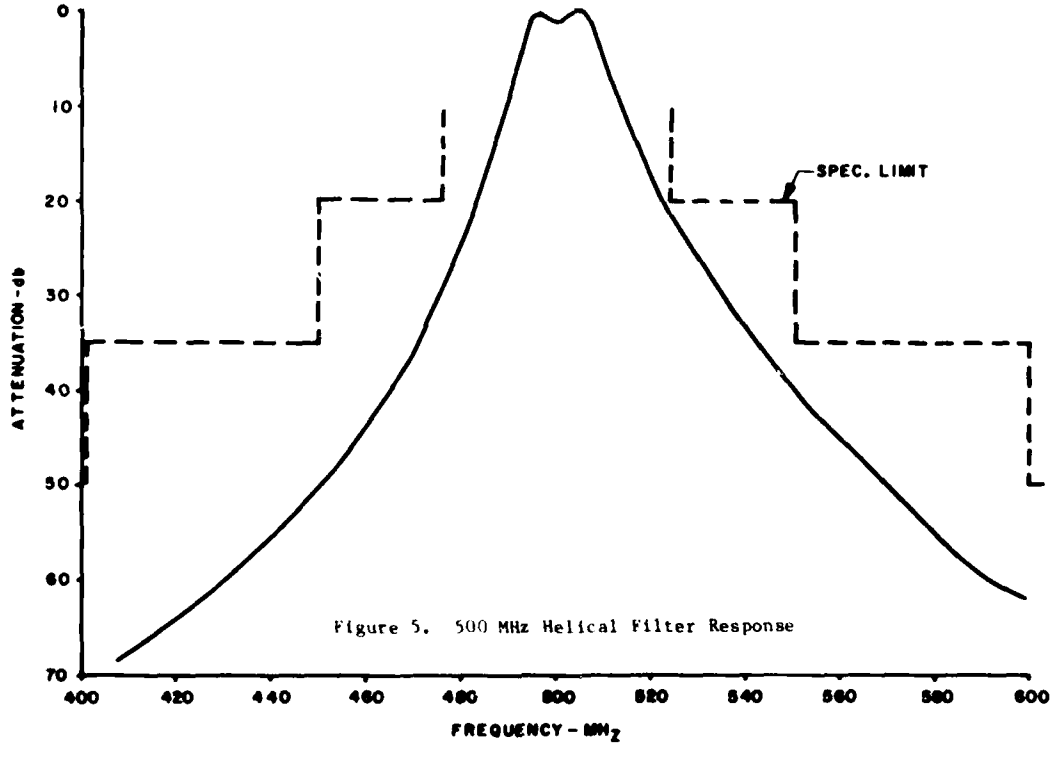


Figure 5. 500 MHz Helical Filter Response

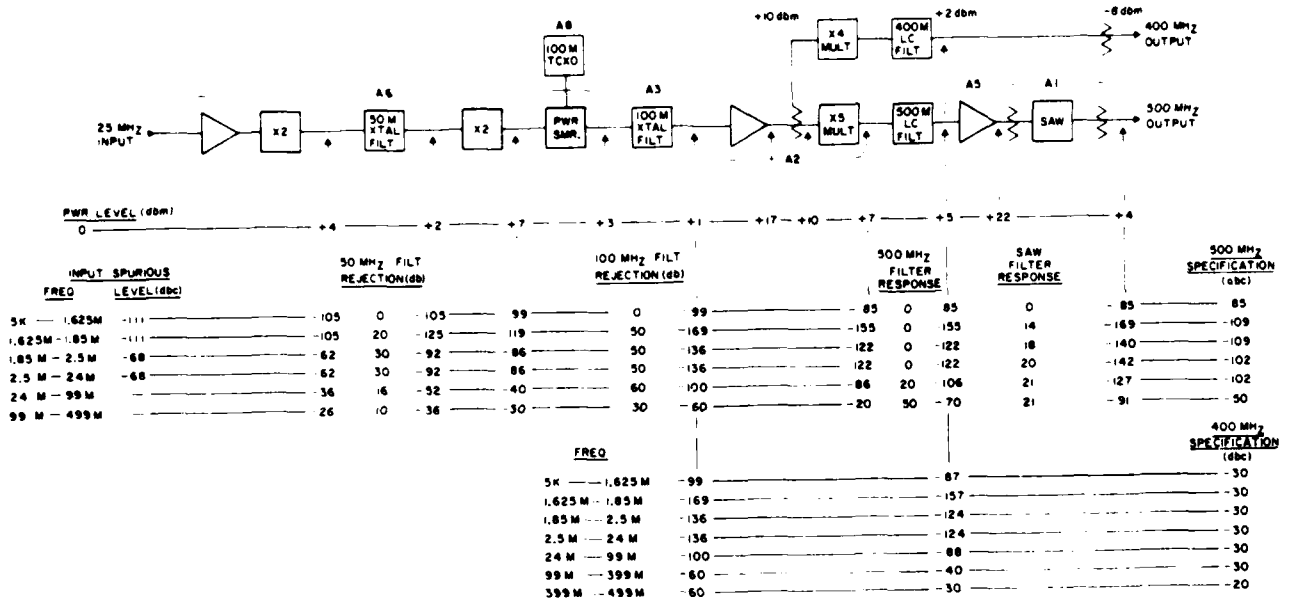


Figure 6. Spurious Analysis, Low Noise 500 MHz Frequency Source

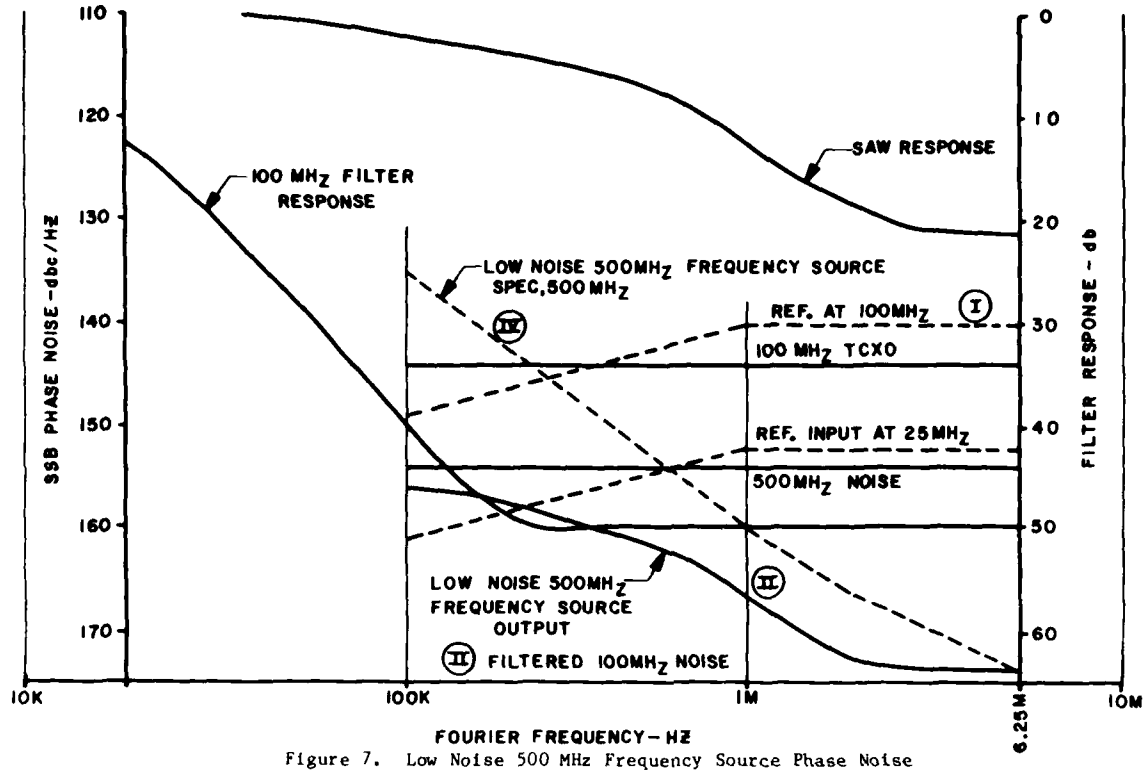


Figure 7. Low Noise 500 MHz Frequency Source Phase Noise

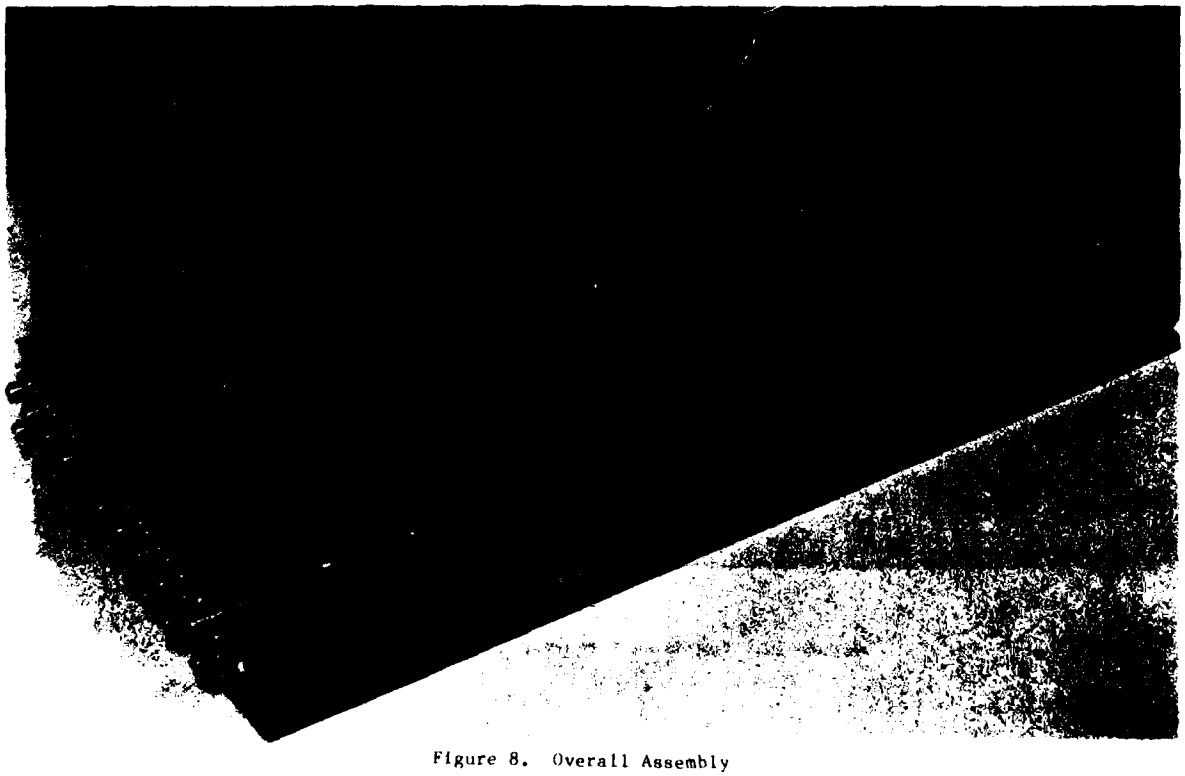


Figure 8. Overall Assembly

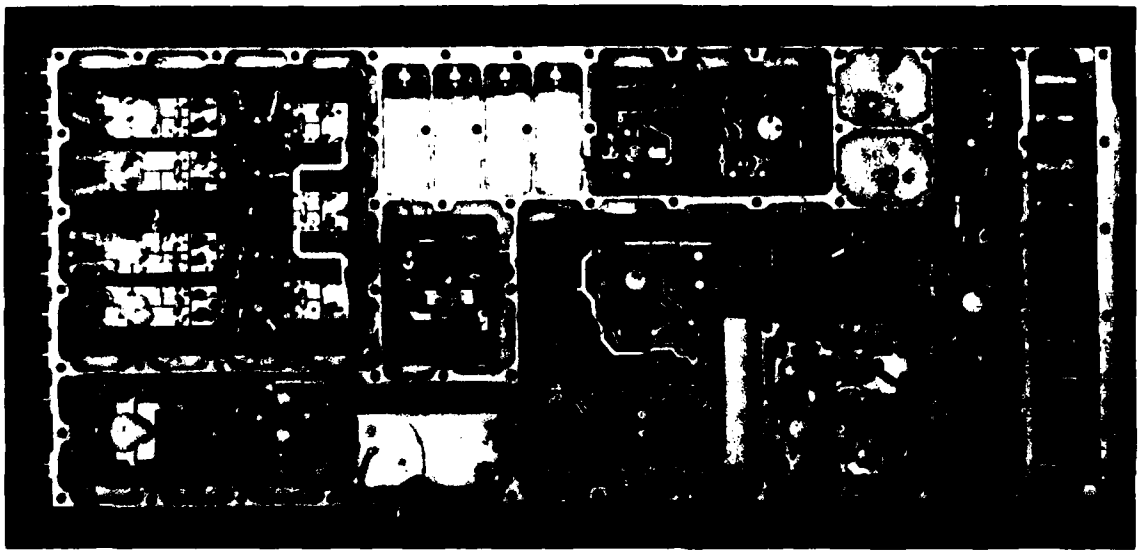


Figure 9. Low Noise Frequency Source, Top View



Figure 10. Low Noise Frequency Source, Bottom View

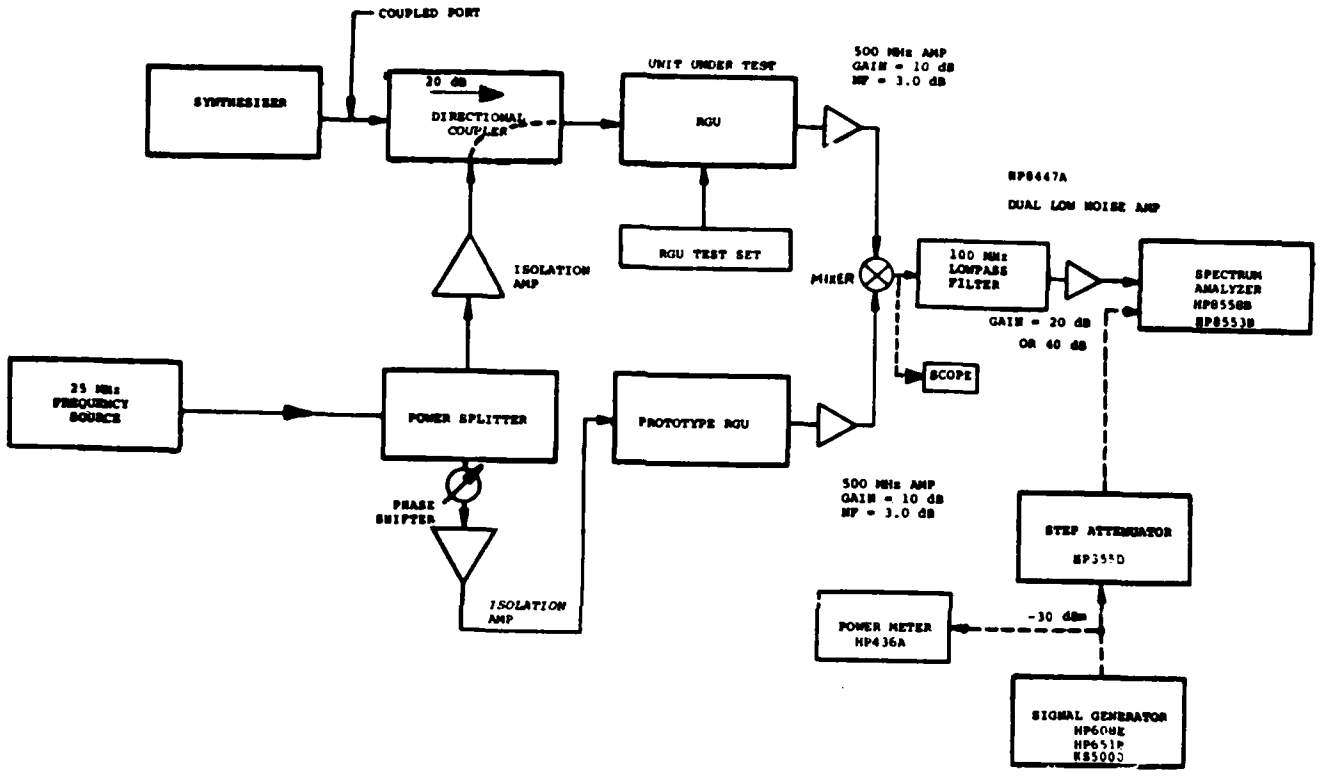


Figure 11. Test Set-Up, Phase Noise

TABLE 1
PERFORMANCE CHARACTERISTICS

PARAMETER	SPEC. LIMIT		PERFORMANCE	
SPURIOUS OUTPUTS (500MHz)	5K - 1.625M, -85 dbc 1.625M - 2.5M, -102 dbc 2.5M - 99M, -102 dbc 99M - 499M, -50 dbc		-85 dbc -108 dbc -108 dbc -85 dbc	
SPURIOUS OUTPUTS (400MHz)	5K - 399M, -30 dbc		-33 dbc	
HARMONICS	-20 dbc (500MHz) -20 dbc (400MHz)		-58 dbc -55 dbc	
		500 MHz	400 MHz	500MHz	400MHz
PHASE NOISE	Δf	(dbc)	(dbc)		
	0.1M	-135	-105	-156	-154
	1M	-160	-130	-166	-154
	2M	-166	-136	-171	-153
	6.25M	-174	-144	-174	-154
OUTPUT TO OUTPUT ISOLATION	15 db		22 db (500MHz) 19 db (400MHz)	
OUTPUT IMPEDANCE	50 Ω		50 Ω	
OUTPUT VSWR	1.5:1		1.4:1	

TABLE 2

PARAMETER	SPEC. LIMIT	PERFORMANCE
TCXO LEAKAGE TO INPUT	-100 dbm	-130 dbm
TCXO LEAKAGE TO OUTPUTS	-85 dbc	NO LEAKAGE
EXT. REF. LEAKAGE TO OUTPUT	-85 dbc	-120 dbc
OUTPUT POWER	+4 ± 2 dbm (500MHz)	+4 ± 2 dbm
	-8 ± 2 dbm (400MHz)	-8 ± 2 dbm
FREQUENCY STABILITY, ALL CONDITIONS	±5KHz, INT. (500MHz)	±3KHz
	±4KHz, INT. (400MHz)	±2.4KHz
	±0.5KHz, EXT. (500MHz)	±0.5KHz
	±0.4KHz, EXT. (400MHz)	±0.4KHz

TABLE 3

PARAMETER	SPEC. LIMIT	PERFORMANCE
D.C. POWER	18V @220Ma	18V @150Ma
POWER CONSUMPTION	4 WATTS	2.7 WATTS
RIPPLE AND NOISE VOLTAGES	NO PERFORMANCE DEGRADATION	NO DEGRADATION
TRANSIENT VOLTAGE	NO PERFORMANCE DEGRADATION	NO DEGRADATION
OVERVOLTAGE	NO PERFORMANCE DEGRADATION	NO DEGRADATION
UNDERVOLTAGE	NO PERFORMANCE DEGRADATION	NO DEGRADATION
WEIGHT	5.5LBS	5.2 LBS
RELIABILITY	0.99927	0.99939 (EXT) 0.99940 (INT)
ACCEPTANCE TEMPERATURE	20 TO 120° F	20 TO 120° F
QUAL TEMPERATURE	2 TO 130° F	2 TO 138° F
COLD TURN-ON	-20° F NO DAMAGE	-20° F NO DAMAGE

TABLE 4

ERROR BUDGET

LOW NOISE 500MHz FREQUENCY SOURCE

INTERNAL MODE

LONG TERM DRIFT (7 YEARS)	$\pm 2.5 \times 10^{-6}$
Δ VOLTAGE BUSS (2% ON 14V)	$\pm 1 \times 10^{-7}$
SETTING TOLERANCE (f_o)	$\pm 1 \times 10^{-7}$
TEMPERATURE (0 → +55°C)	$\pm 4 \times 10^{-6}$

PEAK DEVIATION $\pm 6.7 \times 10^{-6}$ INTERNAL

$\pm 1 \times 10^{-14}$ EXTERNAL

NOTES:

1. LONG TERM EXPERIMENTAL — CHROME/GOLD 137,4620
2. ΔV FROM REG. ±2% (100% FOR 5×10^5 RADS ON $\frac{\Delta V}{V}$)

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, 07703, May 1981

SPURIOUS SUPPRESSION IN DIRECT DIGITAL SYNTHESIZERS

C. E. Wheatley, III
Rockwell International
Anaheim, CA 92803

and

D. E. Phillips
Rockwell International
Cedar Rapids, Iowa 52406

Summary

This paper describes a unique method for the removal of discrete line spurs normally present in the output of direct digital synthesizers (DDS). The type of DDS considered consists primarily of an N-bit accumulator, whose contents are incremented by a value, k, at a clocking rate, F_c. Although many variations exist on the use of this technique to synthesize a periodic waveform, the simplest possible version uses only the time of occurrence of accumulator overflow to define the edges of a "square" wave. For input numbers, k, that are powers of 2, the output frequency is a direct submultiple of the clock frequency. However, for all other numbers, the circuit behaves like a fractional divider with a sequence of non-uniform output periods, which repeat with a definable pattern to give an average period equal to the desired signal's period. Unfortunately, line spurs or sidebands are created by this periodic phase modulation of the output. Historically, these lines have been reduced by using the numerical contents of the accumulator to synthesize an output waveform which has more than two levels -- e.g., a sinusoidal shape. This requires the use of digital-to-analog (D/A) converters and read-only memories (ROM) to obtain the desired wave shape, and the results are well described in the literature.

The technique described here provides for complete elimination of these discrete spurs (while retaining a binary output) by randomly varying or dithering the periods of the alternating output, while keeping the average of these periods unchanged. The method consists of adding a sequence of pseudo-random numbers to the content of the accumulator in a prescribed manner, and this process is shown to convert the discrete line spurs into a continuous noise floor, which is white out to frequencies approaching the clock frequency.

As the speed of digital accumulators continue to increase with new devices and new techniques, the noise floor realizable by this mechanization will continue to decrease at a rate of 6 dB for every doubling of the clock frequency.

Details involved in the dither process and expressions for deriving the expected value of carrier power to noise density of this synthesis technique are described. Also, expressions useful in determining the noise density associated with square waves that exhibit edge time jitters are provided. This information should prove useful to designers of logic systems where the effects of edge jitter are often not well accounted for. An important example is the generation of noise in an indirect digital synthesizer by edge instabilities of the phase detectors.

Two versions of this synthesizer using the dither technique have been breadboarded, and results confirm the predictions of the analysis. Data obtained from the breadboarded synthesizers is provided to show some of the more interesting results of this technique, as well as to define its limitations.

The Basic Concept of a Dithered Accumulator

The technique of using a digital accumulator to synthesize a selected frequency is quite well known[1], and Figure 1 shows this process in simple block diagram form. The basic idea is to increment the content (phase) of the digital accumulator at a regular clock rate, generating a frequency equal to the rate of the phase change with time. It is also well known that the resultant fundamental frequency output suffers from the presence of nonharmonic spurious spectral lines. These lines can be reduced by using D/A converters on the output of the accumulator (Figure 1) to generate sine waves. However, the spurious level will increase as the number of bits in the D/A converters decreases, reaching a maximum when only one bit is retained and the output waveform is simply a square wave.

An explanation for the spurious signals that result from the use of only one bit can be visualized by referring to Figure 2, the time process of the accumulator for a very simple example (i.e., with N = 4 and k = 3). For this example, the "ideal" output square wave would have three edges every 16 input clocks.

From Figure 2, one can observe that the overflow output is exactly the same as that which one would obtain by sampling the "ideal" continuous sine wave at the clock frequency, i.e., by passing these samples through a sample-and-hold device. Due to aliasing, this sampling process takes the spectrum of the ideal sine wave and "folds" it around the clock frequency. Since the power in the harmonics of the square wave decreases rather slowly, a great many lines to spoil the desired fundamental output. Knowing the clock frequency and the digital word, one can easily compute the amplitude and location of all spurs.

A slightly different but more common method of viewing the cause of the spurs is to observe that the output of the direct digital synthesizer is off only at the clock edge. If the output frequency is not a direct submultiple of the clock, a phase error between the ideal output and the actual output slowly increases (or decreases) until it reaches one clock period, at which time the error returns to zero and starts to increase (or decrease) again. This ramp-like phase modulation generates the same spurious spectrum as determined from the sampling viewpoint.

By periodically adding a random number to the accumulator of the direct digital synthesizer, one can destroy the coherence of the phase error from edge to edge, thereby destroying the discrete inharmonic line spectrum that results. Figure 3 illustrates the method, and the following describes why the edge-to-edge jitter becomes uncorrelated.

At each overflow of the accumulator, a random value, X, where X ranges uniformly from 0 to K-1, is added to the content of the accumulator. As seen in Figure 4, this has the effect that the accumulator will next overflow at one of two clock times, depending upon the accumulator phase error, τ_e of Figure 2) and the random number X selected.

From Figure 4, the normal phase error, τ_e, can be seen to be related to an error value, e, contained in the accumulator just prior to the normal overflow event. Clearly, e must be less than k. By adding a random number, X, the accumulator will still overflow at its normal, late edge if X is less than e, but will overflow one clock time early if X is greater than e. Now, if X is selected as specified from uniform random distribution from 0 to k-1, the probability density function of the two events is:

$$p(t_e) = \frac{e}{k} \; ; \; p(t_e - T_c) = \frac{1-e}{k} \qquad (1)$$

The expected value of the overflow point is:

$$\tau_{ave} = t_e \, p(t_e) + (t_e - T_c) \, p(t_e - T_c) \qquad (2)$$

$$= t_e \left\{\frac{e}{k}\right\} + (t_e - T_c) \left\{1 - \frac{e}{k}\right\}$$

But, since $\frac{k-e}{k} = \frac{e}{T_c}$, (from Figure 4)

$$\tau_{ave} = 0 \qquad (3)$$

Thus, the average overflow event occurs at the time where the ideal phase ramp intersects the maximum value of the accumulator. This means that the average period synthesized by the process is the correct period, and the fact that any one period is not exact introduces phase noise.

More importantly, however, the expected overflow time is not dependent upon the original phase error, τ_e. This means that, even if adjacent errors are originally correlated, adding the random value X causes adjacent accumulator resets to be independent, and all coherent spurious lines originally generated by this correlation are eliminated. The penalty will be a continuous noise spectrum due to the random choice between the two possible reset times. The following paragraphs compute the density of this noise spectrum, given the probability distribution of the overflow event.

Computation of the Spectrum for Gaussian Edge Jitter

In Appendix I, it is shown that the spectrum of a square wave that exhibits a small amount of edge time jitter may be computed if one knows the probability density function (pdf) of the time jitter, and the mathematics are relatively simple if the jitter is independent from edge to edge. Specifically, the spectrum depends upon the characteristic function of the time jitter process which is defined as:

$$F(j\omega) \triangleq E\left\{e^{j\omega\tau}\right\} = \int_{-\infty}^{\infty} p(\tau) e^{j\omega\tau} \, d\tau \qquad (4)$$

and where $p(\tau)$ is the pdf of the time of occurrence of the edge, with respect to the edge of the "true" square wave being synthesized. In the case of the jittered accumulator synthesis technique, the actual values of τ can take on only two values, one ahead of the average edge, and one behind.

The complete spectrum of the square wave as derived in Appendix I (Eq A-17 and A-18), including both discrete and continuous components, is computed from $F(j\omega)$ as:

$$S(\omega) = \frac{A^2}{2T_0} \left\{ \left|1 - F(j\omega)\right|^2 + \frac{2\pi}{T_0} \left|F(j\omega)\right|^2 \cdot \sum_{m=-\infty}^{\infty} \delta\left(\omega - \frac{m\pi}{T_0}\right) \right\} \qquad (5)$$

where A is the peak-to-peak value and $2T_o$ is the complete time period of the fundamental output square wave.

As an example, consider the case of a logic device having jitter which can be modeled as a zero mean Gaussian distribution with variance σ_t^2. In this case,

$$p(t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-t^2/2\sigma_t^2}$$

By using Equation (4) to determine $F(j\omega)$, and then substituting this into Equation (5), one can show that the output spectrum of the square wave will be:

$$S(\omega) = \frac{A^2}{T_o}\left\{\frac{1-e^{-\omega^2\sigma_t^2}}{\omega^2}\right\} + 2\pi\frac{A^2}{T_o^2} e^{-\omega^2\sigma_t^2} \sum_{\substack{m=-\infty \\ m \text{ odd}}}^{\infty} \delta(\omega - \frac{m\pi}{T_o}) \quad (6)$$

At the fundamental output frequency, ($f_o = \frac{\omega_o}{2\pi} = \frac{1}{2T_o}$) the noise power density to carrier power will be (Note 1):

$$\frac{N_o}{C} = 2\pi^2 f_o^2 \sigma_t^2 \quad \text{for} \quad \sigma_t \ll T_o \quad (7)$$

From this relationship it can be shown that for a digital device to provide an $\frac{N_o}{C}$ of -150 dBc/Hz, at a typical frequency of 1 MHz, it must exhibit an edge instability of less than 7 picoseconds.

The Output Spectrum of a Synthesizer Using Accumulator Dither

Returning to the synthesized waveform where the distribution is not continuous, $F(j\omega)$ may still be computed from Equation (4) using Equation (1) for the discrete pdf of the early and late edges. After some computational labor, the result is:

$$|F(j\omega)|^2 = 1 - \frac{4}{3}\sin^2\left(\frac{\omega t_e}{2}\right)\left\{1 - \left(\frac{t_e}{T_c}\right)^2\right\} \quad (8)$$

Note that this depends upon the error value t_e, which can vary from 0 to T_c. This means that although the average frequency is correct and independent of t_e, the continuous noise density term of Equation (5) will depend upon t_e, being maximum at $t_e = T_c/2$. If t_e were always identically 0 (or T_c), no noise would be introduced by the dither. In the general case, t_e can be expected to vary uniformly from 0 to T_c, resulting in an average value for $|F(j\omega)|^2$ of:

$$\overline{|F(j\omega)|^2} = 1 - 2/3 \sin^2 \frac{\omega T_c}{2} \quad (9)$$

Finally, under the condition that $\omega_o = \frac{\pi}{T_o} = \frac{1}{T_c}$, the spectrum of the synthesized square wave near this fundamental frequency as determined by substitution of $|F(j\omega)|^2$ into Equation (5) is:

$$S(\omega) = \frac{A^2 T_c^2}{6T_o} + \frac{2A^2}{\pi}\left\{\delta(\omega - \frac{\pi}{T_o}) + \delta(\omega + \frac{\pi}{T_o})\right\} \quad (10)$$

From the above, the noise power density to carrier power at the fundamental frequency will be:

$$\frac{N_o}{C} = \frac{\pi^2}{3}\frac{f_o}{f_c^2}, \quad (11)$$

Again, it must be emphasized that non-harmonic lines will not be present, only the noise floor defined by the first term of (10).

The power of this entirely digital technique ultimately lies in the ability of technology to provide high-speed accumulators, and provides a 6 dB N_o/C improvement for every doubling of the clock frequency. As an example, a present-day ECL synthesizer using a 100-MHz clock could deliver a 1-MHz output with an N_o/C of -95 dBc/Hz, which, although not nearly as large as can be obtained from fundamental sources, is quite sufficient to meet many practical needs. Ultimately, GaAs devices that are expected to be able to be clocked at GHz rates could provide N_o/C measures well below -100 dBc/Hz.

Experimental Results

A breadboard direct digital synthesizer (DDS) was constructed of low-power Shottky logic to test the dither effect (Figure 5). In this design a pipelined 24-bit accumulator was clocked at a 16-MHz rate and a pseudo-random number generator and control logic allowed a pseudo-random number to be periodically added to the accumulator. A dither size-select circuit was used to adjust the dither modulation to an amount just great enough to eliminate the spurious sidebands, for each value of input frequency.

The breadboard was built primarily to test circuits that will be incorporated into an LSI design using CMOS/SOS technology, rather than to optimize a TTL design. The results are not optimum for an LSI design either, but enable the principle to be checked out.

Figure 6 shows the spectrum of a typical undithered output with spurious sidebands present,

NOTE 1: This is also known as $\mathcal{L}(f)$, the frequency domain measure of phase fluctuation sidebands.

and Figure 7 shows the same output with dither. As predicted, the spurious sidebands vanish, being replaced by a uniform phase noise spectrum, flat to within less than one Hertz of the carrier.

Figure 8 plots the noise and spurious levels vs output frequency. The expected undithered spurious sideband level is commonly expressed as

$$20 \log \left(\frac{f_o}{f_c} \right) \text{ dBc} \qquad (12)$$

which is plotted as the theoretical value. Running above it in a zig-zag course is the measured spurious level of the most significant bit "square wave" output. A computer analysis also shows this same pattern. This is due to the fact that the spurious modulation pattern is the same on the leading and trailing edges, except for a varying relative phase, so there is a varying amount of reinforcement and cancellation between edges. Also, the spurious (modulation) frequency increases and decreases, going through a zero beat each time the f_c / f_o ratio is exact power of 2.

An important application of DDS is to serve as a reference for a phase-lock loop. With digital phase discriminators, only one edge of the reference wave is used, the other being supplied from the VCO, divider, etc. This results in 6 dB less noise modulating the loop (because of the correlation between edges). The spurious level is now a uniform function of a carrier frequency, since the edge interaction effects have been removed.

When dither was applied, the amount of dither modulation was determined by the most significant bit in the frequency control word, resulting in a 3-dB step per octave. There is also a 3-dB drop in noise when only one edge is used, because the two edges are uncorrelated.

The developed formula for carrier-to-noise density

$$\frac{C}{N_o} = \frac{3}{\pi^2} \frac{f_c^2}{f_o} \qquad (13)$$

is based upon a normalized 1-Hz noise bandwidth. As shown in Figure 8, this formula is rearranged and includes the system noise bandwidth. It can be seen that the noise-bandwidth-to-clock ratio is the dithered "improvement factor," as the original spurious sideband energy is uniformly spread over the band 0 to f_c, but only a portion is recognized in the system bandwidth. The graph is plotted for the commonly used voice communications bandwidth of 3 kHz.

Applications of the Dithered Accumulator

The dithered accumulator is particularly suited to LSI integrated circuits, being all digital and requiring no large ROM memory or D/A converters. Its practical value will increase as the maximum clock speeds increase with smaller-geometry CMOS/SOS and gallium arsenide circuits, reducing the noise levels. For the same output frequency, 1-MHz for example, the dithered single-edge phase noise of -49 dBc (3-kHz bandwidth) with a 16-MHz clock would drop to -62 dBc with a 64-MHz clock! Even higher clock frequencies will be possible in the future, which will allow either lower noise or higher output frequencies.

If the concept of adding noise to a system still seems strange (aren't we always trying to reduce it?), one might consider the advantage that reducing the bandwidth (for example, by following the DDS with a narrowband phase-lock loop) lowers the random noise, but close-in discrete spurious levels remain unchanged. This is an important factor in some narrowband systems such as HF tone-coded transmissions and CW telegraphy.

As direct digital synthesizers find increasing use in frequency-hopping systems, a dithered accumulator in a small, low-power, high-speed integrated circuit could be of great value in HF, VHF, and UHF communications.

References

1. Gorski-Popiek, J., Editor, "Frequency Synthesis: Techniques and Applications, IEEE Press, 1975.

Appendix 1

Duration of the spectrum of a square wave with edge time jitter

Acknowledgement

The authors wish to thank Dr. A. K. Nandi for his contributions to the analysis contained in this appendix.

Introduction

The case considered here is applicable to situations wherein the dither time process is uncorrelated (white phase noise) from edge to edge, which is appropriate for many cases of interest. In order to further simplify the mathematics, we start by analyzing an alternating sequence of time-dithered impulses. Then, be noting that a square wave can be viewed as the time integral of a train of alternating impulses, we easily obtain the desired prediction for the square-wave power spectrum. A most important consequence of the edge jitter phenomenon is that a continuous spectrum (noise) around dc is created in the case of the square wave.

Spectrum of Randomly Jittered Impulses (Uncorrelated Samples)

Consider the situation in Figure A-1 which shows a periodic alternating sequence of impulses of strength $\pm$ A. Successive impulses are separated by t_o. In Figure A-2, the same impulse sequence is shown with edge jitter added. The spacing between successive impulses is now a random variable.

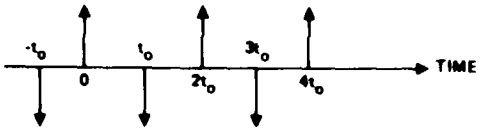


Figure A-1. Ideal Sequence of Impulses

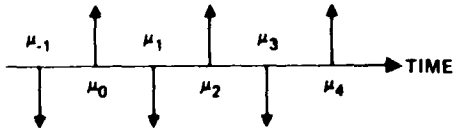


Figure A-2. Jittered Sequence of Impulses

For the jittered sequence, one can write

$$u(t) = \sum_{n=-\infty}^{\infty} (-1)^n A\,\delta(t-nt_o - \mu_n) \qquad (A\text{-}1)$$

As a specific example, we will consider the case where the μ_n's are uncorrelated, identically distributed continuous Gaussian random variables such that

$$E\mu_n = 0 \qquad (A\text{-}2)$$

$$E\mu_n^2 = \sigma^2; \quad \sigma \ll t_o \qquad (A\text{-}2)$$

$$E\mu_k\mu_\ell = 0 \quad \text{if } k \neq \ell \qquad (A\text{-}2)$$

However, other distributions are acceptable and can be handled using the procedures described below. For the ideal situation in Figure A-1, all μ_n's are zero.

An often used expression (Reference A-1) for computing power spectral density of a wide-sense stationary process is

$$S_u(\omega) = \lim_{T\to\infty} \frac{1}{2T} E\left\{U_{2T}(j\omega)\, U_{2T}(-j\omega)\right\} \quad -\infty < \omega < \infty \qquad (A\text{-}3)$$

where units are in power per Hertz and:

$U_{2T}(j\omega)$ = Fourier transform a finite section of length 2T of a sample function

$U_{2T}(-j\omega)$ = complex conjugate of $U_{2T}(j\omega)$

E = statistical expectation

Now, let

$$T = Nt_o \qquad (A\text{-}4)$$

On the average, one is considering a truncated stream of jittered impulses ($\sim$(2N+1)impulses) during the computation of the Fourier transform mentioned above. Thus,

$$U_{2T}(j\omega) = \int_{-\infty}^{\infty} u_{2T}(t)\, e^{-j\omega t}\,dt$$

$$= \int_{-\infty}^{\infty} \left(\sum_{n=-N}^{N} (-1)^n A\,\delta(t-nt_o-\mu_n)\right) e^{-j\omega t}\,dt$$

$$= \sum_{n=-N}^{N} (-1)^n A\, e^{-j\omega(nt_o-\mu_n)} \qquad (A\text{-}5)$$

$$\therefore\ U_{2T}(-j\omega) = \sum_{n=-N}^{N} (-1)^n A\, e^{+j\omega(nt_o-\mu_n)} \qquad (A\text{-}6)$$

and

$$E\left\{U_{2T}(j\omega)\, U_{2T}(-j\omega)\right\}$$

$$= E\left\{A^2 \sum_{k=-N}^{N} \sum_{\ell=-N}^{N} (-1)^{\ell+k}\, e^{-j\omega(k-\ell)t_o} e^{-j\omega(\mu_k-\mu_\ell)}\right\}$$

$$= A^2 \sum_{k=-N}^{N} \sum_{\ell=-N}^{N} (-1)^{\ell+k}\, e^{-j\omega(k-\ell)\,t_o}$$

$$\cdot\ E\left\{e^{-j\omega(\mu_k-\mu_\ell)}\right\}$$

In order to evaluate the above expectation involving summations over k and ℓ, it is convenient to examine Figure A-3.

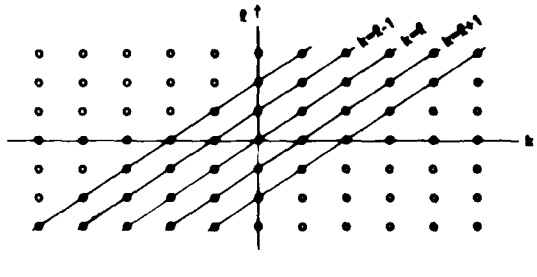


Figure A-3

Reference A-1: J. A. Aseltine, "Transform Methods in Linear Systems Analysis," McGraw Hill, New York, 1958.

One can write equation (A-7) as follows:

$$E\left\{U_{2T}(j\omega)\ U_{2T}(-j\omega)\right\} = A^2\left|(2N+1) + \sum_{k=-N}^{N}\sum_{\ell=-N}^{N}(-1)^{\ell+k}\ e^{-j\omega(k-\ell)t_o} \cdot E\left\{\prod_{i\neq k} e^{-j\omega(\mu_k-\mu_\ell)}\right\}\right| \qquad (A\text{-}8)$$

With $k\neq\ell$, one obtains

$$E\left\{e^{-j\omega(\mu_k-\mu_\ell)}\right\} = E\left\{e^{-j\omega\mu_k}\right\}\ E\left\{e^{+j\omega\mu_\ell}\right\} = F(-j\omega)\ F(+j\omega) \qquad (A\text{-}9)$$

where $F(j\omega) \triangleq E\left\{e^{j\omega\mu_\ell}\right\}$

= characteristic function of μ_ℓ

Using Equation (A-9) in Equation (A-8) and summing along lines of constant $(k-\ell)$*, one easily finds

$$E\left\{U_{2T}(j\omega)\ U_{2T}(-j\omega)\right\} = A^2\left|(2N+1) + |F(j\omega)|^2 \cdot \left\{-2N\left(e^{-j\omega t_o}+e^{+j\omega t_o}\right) + (2N-1)\left(e^{-2j\omega t_o}+2j\omega t_o\right)..\right\}\right| \qquad (A\text{-}10)$$

Adding and subtracting a term corresponding to n=0, one obtains

$$E\left\{U_{2T}(j\omega)\ U_{2T}(-j\omega)\right\} = A^2\left|(2N+1)\ (1-|F(j\omega)|^2) + |F(j\omega)|^2\left\{(2N+1) - 2N\left(e^{-j\omega t_o}+e^{+j\omega t_o}\right) + (2N-1)\left(e^{-2\omega t_o}+e^{+2\omega t_o}\right)..\right\}\right| \qquad (A\text{-}11)$$

Recalling now the basic formula for spectral density given in Equation (A-3):

$$S_u(\omega) = \lim_{T\to\infty}\frac{1}{2T}E\left\{U_{2T}(j\omega)\ U_{2T}(-j\omega)\right\} = \lim_{T\to\infty}\frac{1}{2Nt_o}E\left\{U_{2T}(j\omega)\ U_{2T}(-j\omega)\right\} = \frac{1}{t_o}A^2\left|(1-|F(j\omega)|^2) + |F(j\omega)|^2 \cdot \sum_{m=-\infty}^{\infty}(-1)^m\ e^{-j\omega mt_o}\right|$$

$$S_u(\omega) = \frac{A^2}{t_o}\left|(1-|F(j\omega)|^2) + |F(j\omega)|^2 \cdot \frac{2\pi}{t_o}\sum_{n=-\infty}^{\infty}\delta\left(\omega-\frac{n\pi}{t_o}\right)\right| \text{ n odd} \qquad (A\text{-}12)$$

This is a most general result, from which the power spectrum can be evaluated. Note that both discrete and continuous (noise) terms exist in the spectrum.

Spectrum of Square Wave with Independently Jittered Edges

Consider a square wave of magnitude $\pm$ A/2 indefinitely in time in both directions whose edges are jittering, and let the edge jitter statistics be described by Equation (A-2). Since the impulse train of Equation (A-1) can be obtained by differentiating this square wave, the power spectrum of the square wave is easily obtained by multiplying the impulse spectrum (A-12) by $1/\omega^2$. Thus,

$$S_{sq\ wave}(\omega) = \frac{1}{\omega^2}S_u(\omega) \qquad (A\text{-}13)$$

$$S(\omega) = \frac{A^2}{t_o}\left(\frac{1-e^{-\omega^2\sigma^2}}{\omega^2}\right) + \frac{2\pi A^2}{t_o^2\omega^2}\ e^{-\omega^2\sigma^2}\sum_{\substack{n=-\infty\\(n\ odd)}}^{\infty}\delta\ \omega-\frac{n\pi}{t_o} \qquad (A\text{-}14)$$

Note that the period, T, of the fundamental square wave = $2T_o$. As above, this spectrum has both a noise term and discrete terms, but the noise term is maximum at ω=0, reaching a value of

$$\frac{2A^2\sigma^2}{T}$$

*From Figure 3, it is clear that for
$(k-\ell) = \pm1, \pm3, \ldots \Rightarrow (-1)^{k+\ell} = -1$ and for
$(k-\ell) = \pm2, \pm4, \ldots \Rightarrow (-1)^{k+\ell} = +1$

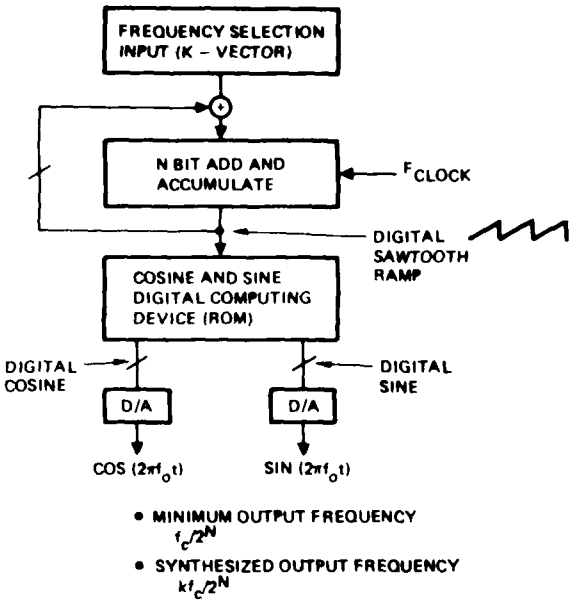


Figure 1. Direct Digital Synthesizer — Simplified Standard Block Diagram

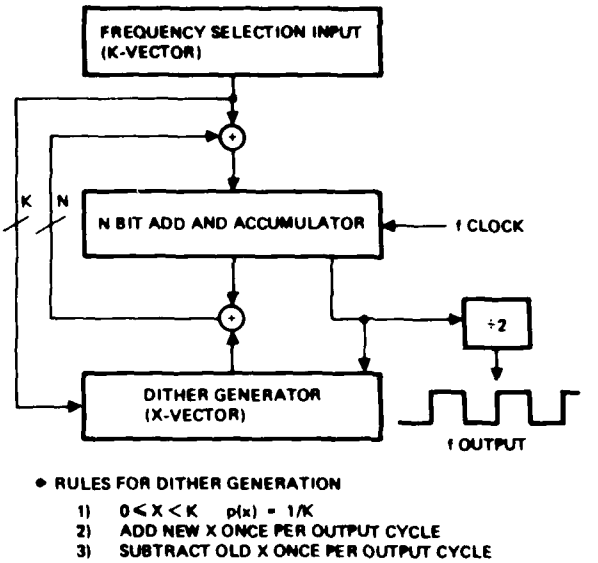


Figure 3. Direct Digital Synthesizer — Dithered Accumulator

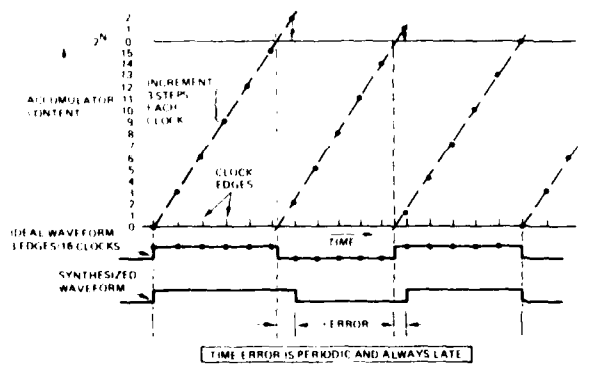


Figure 2. Accumulator Contents vs Time — No Dither (Simple Case N = 4, K = 3)

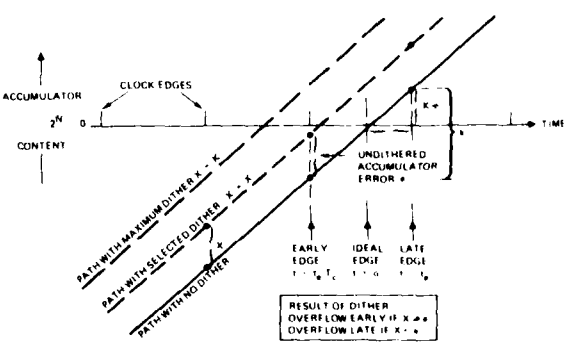


Figure 4. Accumulator Contents vs Time (Result of Random Dither)



Figure 5. Dithered Accumulator Breadboard

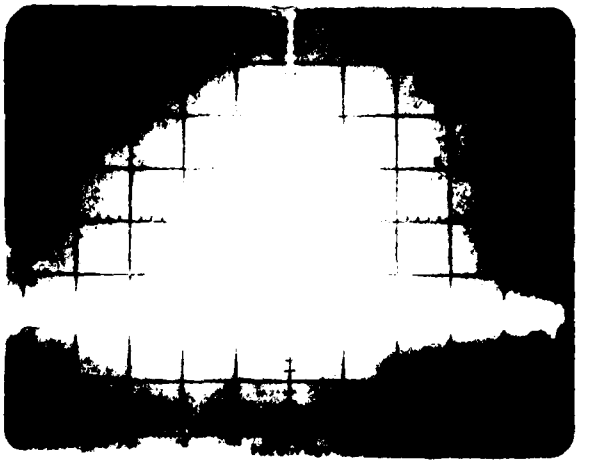


Figure 7. Digital Synthesizer Output with Random Dither

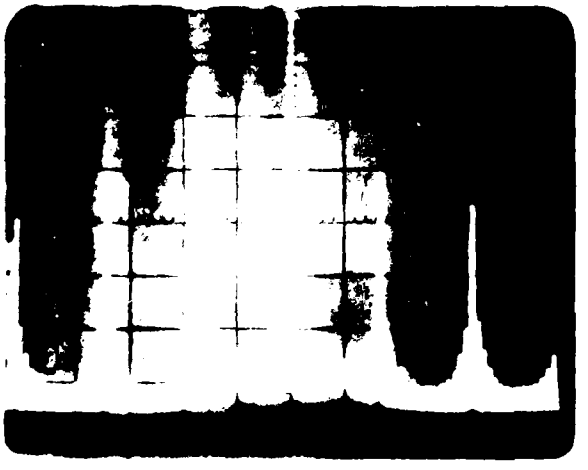


Figure 6. Digital Synthesizer Output without Dither

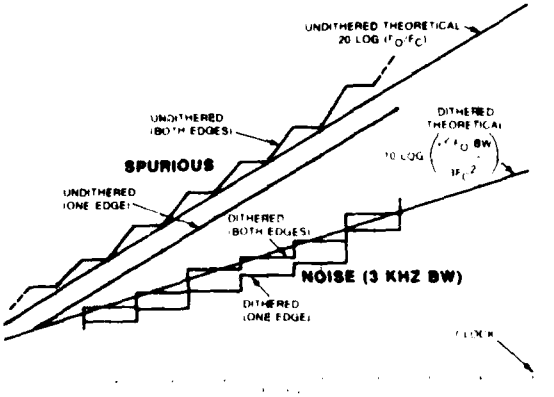


Figure 8. Carrier-to-Noise Density

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

ADVANCED SAW-LSI FREQUENCY SYNTHESIZER*

D.J. Dodson, M.Y. Huang, M.D. Brunsman

TRW
Defense and Space Systems Group
Redondo Beach, CA 90278

Summary

Fast hopping frequency synthesizers are a key component of frequency hopped spread spectrum systems. TRW has recently completed development of a prototype, direct mix-and-divide frequency synthesizer [1,2] which is ideally suited for such applications. The synthesizer generates frequencies from 1296 to 1536 MHz in 3 MHz steps and is capable of switching between frequencies in less than 50 ns. Surface acoustic wave (SAW) and RF-LSI devices are used extensively throughout the design to produce this extremely fast, small synthesizer [3-7].

Architecture Summary

A block diagram of the synthesizer is shown in Figure 1. A mix-and-divide architecture has been

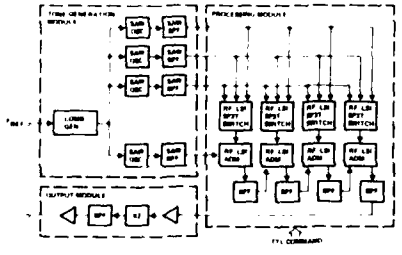


Figure 1. SYNTHESIZER BLOCK DIAGRAM

chosen for this application because of the repeated use of identical components (within the processing module) which allows the designer to take advantage of the circuit complexity available with RF-LSI circuitry. Mix-and-divide synthesizers also offer great flexibility in the choice of frequency range, number of frequencies, and step size. As shown in Figure 1, the synthesizer consists of three modules: a tone generation module, a processing module, and an output module. Four fundamental frequencies are produced in the tone generation module. Within the processing module three of these are selected in RF-LSI SP3T switches and subsequently mixed and divided in the RF-LSI analog multipliers and dividers. The resulting output of the processing module is 81 frequencies with 1.5 MHz spacings at one-half the desired output frequency. The output module doubles the frequency, amplifies, and filters the output of the processing module to produce the desired 81 frequencies from 1296 to 1536 MHz in 3 MHz steps.

*This work was supported in part by the U.S. Army Electronics Research and Development Command Contract No. DAAB07-78-C-2992.

Hardware Description

The generation of the four primary frequencies in the tone generation module is accomplished with a bank of injection locked SAW oscillators. Injection locking of SAW oscillators was investigated during the development of this synthesizer and has been previously reported [8]. It has been found that SAW oscillators obey the theory of injection locking developed by Adler [9] with some modification. Injection locking bandwidth of a SAW oscillator was found to be a strong function of phase slope or delay.

Based on this understanding of injection locking, a bank of oscillators was built with free-running frequencies at the four primary frequencies. The bank was injection locked to the output of an SRD comb generator. Injection locking bandwidth of these circuits has been measured and found to agree very closely with theory.

The output of each oscillator is filtered through a SAW filter to reduce spurious signals produced by the comb generator. The output of one of the oscillators is plotted in Figure 2. Note

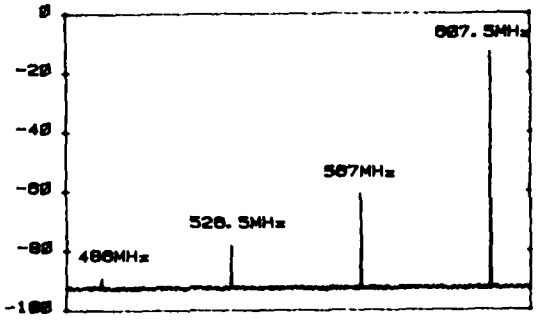


Figure 2. OSCILLATOR OUTPUT

that the spurious signals are reduced to less than -45 dBc. In this circuit the filter was actually identical to the oscillator delay line and had relatively poor rejection. If required, greater rejection of spurious signals could be achieved through the use of SAW output filters with steeper skirts.

The processing module is the heart of the synthesizer. The circuit consists of four SP3T RF-LSI switches as shown in Figure 3, four ADM (amplifier-divider-mixer) circuits as shown in Figure 4, and interstage bandpass filters. A photograph of the module is shown in Figure 5. The circuitry is mounted on a 10-layer, multilayer circuit board so the signals can be routed without coupling. The swept output of this module is shown in Figure 6. Note that the output power is not flat and actually shows discontinuous jumps in power as various switch inputs are selected. No attempt was made to produce a constant output power at this point in the circuit since the output module was being driven into saturation.

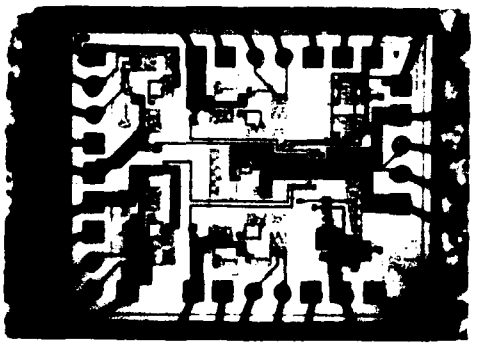


Figure 3. SP3T RF-LSI SWITCH

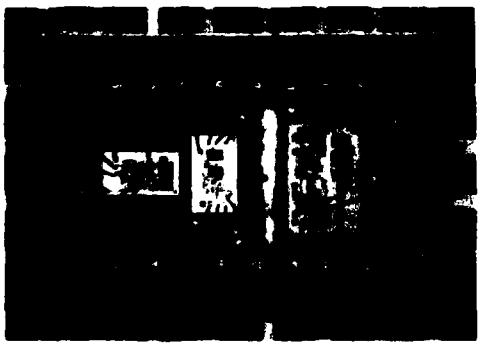


Figure 4. ADM CIRCUIT

The photograph of the output module is shown in Figure 7. The saturation characteristics and frequency responses of this circuit are shown in Figures 8 and 9. Note that the lowest power from the processing module will drive this circuit into saturation, eliminating the need for constant output power at the output of the processing module.

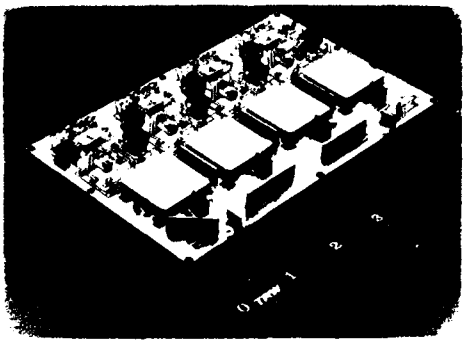


Figure 5. PROCESSING MODULE

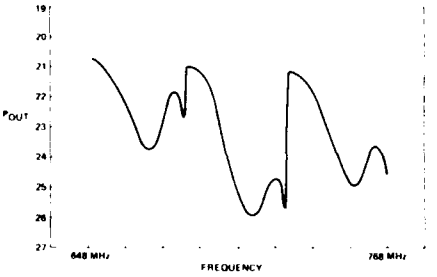


Figure 6. PROCESSING MODULE SWEPT OUTPUT

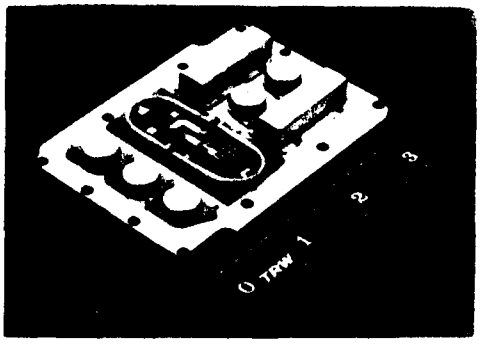


Figure 7. OUTPUT MODULE

A photograph of the integrated synthesizer is shown in Figure 10. The circuit requires approximately 70 $in.^3$ in its present configuration but could be reduced to as little as 10 $in.^3$ through the use of hybrid circuits.

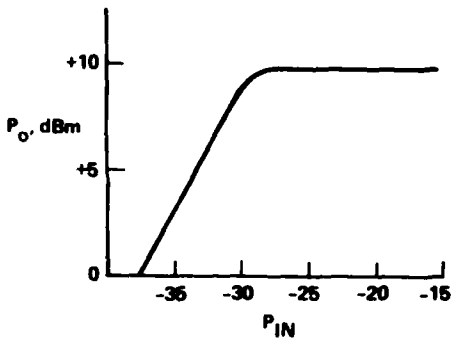


Figure 8. OUTPUT MODULE SATURATION

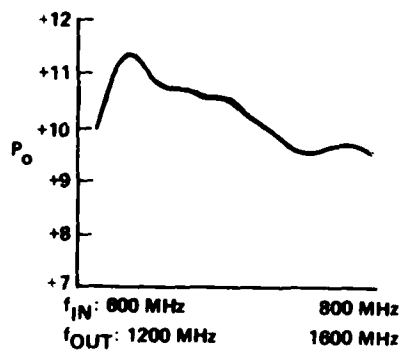


Figure 9. OUTPUT MODULE FREQUENCY RESPONSE

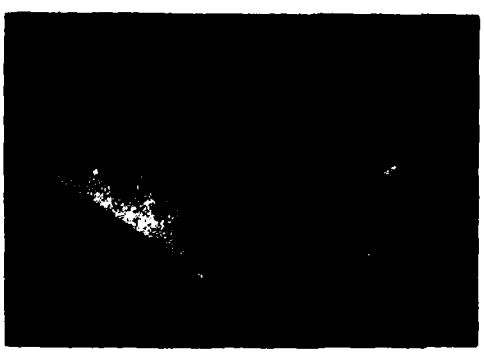


Figure 10. INTEGRATED SYNTHESIZER

Performance Summary

Performance of the synthesizer is shown in Figures 11 through 14. Figure 11 shows the swept output. Note that the output power follows the frequency response of the output module. A single frequency output is shown in Figure 12. Spurs at approximately -32 dBc can be seen. The principle contributions to spurious signals were from two unexpected sources. Coupling in the ADM package and on the multilayer board were both higher than anticipated. Care was taken with the RF switch to launch onto the chip from coplanar waveguide [10] and with this technique spurs were limited to approximately -60 dBc. A similar approach must be taken with the ADM. The multilayer board requires improved grounding and impedance matching. Improvements to the design of this board are currently being investigated.

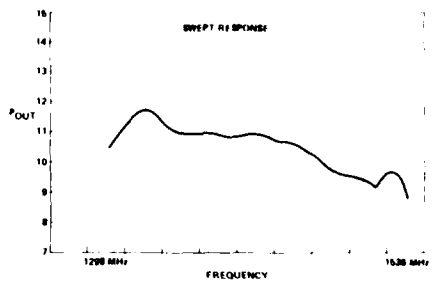


Figure 11. SWEPT OUTPUT RESPONSE

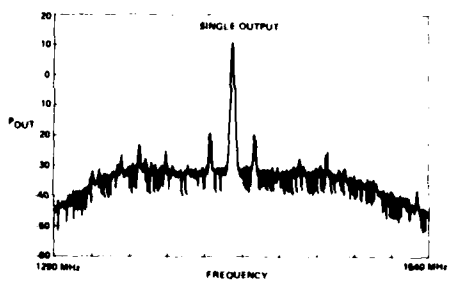


Figure 12. SINGLE FREQUENCY OUTPUT

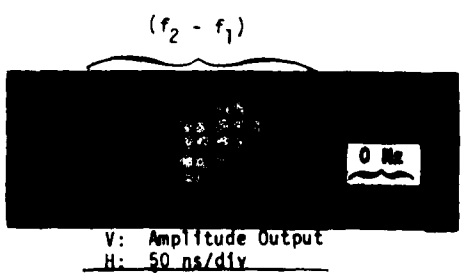


Figure 13. SWITCHING SPEED

Measured switching speed for the synthesizer is shown in Figure 13. This photograph was taken by hopping the synthesizer from 1353 to 1368 MHz and mixing its output with a coherent 1368 MHz. Here we see that the output changes from a 15 MHz beat note to 0 Hz in about 15 ns. Finally, phase noise density is shown in Figure 14.

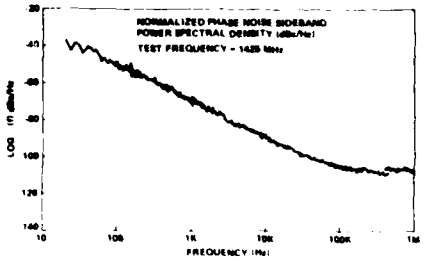


Figure 14. PHASE NOISE DENSITY

Conclusion

The work described here has demonstrated that for applications where speed is essential, a direct mix-and-divide synthesizer utilizing high frequency filters, mixers and dividers, is an excellent approach. Such synthesizers are very flexible in output range, number of frequencies, and step size. and are in no way limited to the few frequencies produced by this synthesizer. A similar TRW design produces 256 frequencies with 1 MHz spacing.

This work has also demonstrated excellent applications for SAW and RF-LSI circuitry. Both enhance the performance and reduce the size of the synthesizer. SAW filters offer excellent rejection of spurious signals in a very small volume and yield very stable oscillators. Similarly, RF-LSI circuits realize complex processing functions in very small volume.

Table 1 is a summary of the performance and potential capabilities of this synthesizer. Improvement in performance implied - particularly spurious level - is fairly straightforward. By improving packaging and the design of the multilayer board it is felt that spurious rejection in excess of 60 dB can be accomplished at the output of the processing module. Reducing size would require the circuits be built as hybrids instead of using the conventional construction techniques shown here.

Table 1. PERFORMANCE SUMMARY

PARAMETER	PERFORMANCE	CAPABILITY	COMMENTS

References

[1] R.R. Stone, Jr., and H.F. Hastings, "A Novel Approach to Frequency Synthesis", Frequency, Sept/Oct. 1963.

[2] R.C. Dixon, Spread Spectrum Systems, John Wiley and Sons, Inc., 1976.

[3] A.J. Budreau and P.H. Carr, "Direct SAW Frequency Synthesizer", 1977 Ultrasonics Symp., Paper C2.

[4] A.J. Budreau, et.al., "Frequency Synthesizer Using Acoustic Surface Wave Filters", Microwave Journal 17, pp. 65-59, March 1974.

[5] A.J. Slobodnick, et.al., "SAW Filters for Frequency Synthesis Applications", 1976 Ultrasonics Symposium.

[6] D.J. Dodson, M.Y. Huang, and R.S. Kagiwada, "Application of SAW Devices to Frequency Synthesizers", Proc. SPIE 24th Annual Tech. Symp., Paper No. 239-30, San Diego, CA, July 28 - Aug. 1, 1980.

[7] D.J. Dodson, M.Y. Huang, and R.S. Kagiwada, "Fast Hopping SAW Frequency Synthesizer", 1980 NTC Symposium.

[8] D.J. Dodson, M.D. Brunsman, and R.W. Kerr, "Injection Locked SAW Oscillators", 1980 Ultrasonics Symposium.

[9] R. Adler, "A Study of Locking Phenomena in Oscillators", Proc. IRE, Vol. 34, pp. 351-357, June 1946.

[10] C.P. Wen, "Coplanar Waveguide: A Surface Strip Transmission Line Suitable for Non-reciprocal Gyromagnetic Device Applications", IEEE Trans. MIT-17, Dec. 1969.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, 07703, May 1981

SC-CUT QUARTZ CRYSTAL UNITS IN LOW-NOISE OSCILLATOR APPLICATION AT VHF

D.J. Healey, III, S.Y. Kwan
Westinghouse Electric Corporation
D&ESC, Baltimore, Maryland

SUMMARY

Crystal-controlled VHF oscillators are widely used in signal generators that must provide good short-term frequency stability and low phase noise. These oscillators utilize AT-cut crystals operating on an overtone mode.

Further reduction of phase noise is difficult to achieve because of the amplitude-frequency effects exhibited by AT-cut crystal units.

Doubly rotated crystals, such as SC-cut plates, exhibit significantly less amplitude-frequency effects than do AT cut. It, therefore, is feasible to operate the SC-cut crystal units at higher currents than are allowable in AT-cut crystal units. Provided that internal phase fluctuations associated with the oscillator do not increase significantly as crystal current is increased, improvement in short-term frequency stability and noise of the crystal-controlled VHF signal generator is expected when using SC-cut crystal units.

An examination of the practicality of obtaining lower noise oscillators at 100 MHz when using third overtone SC-cut crystals rather than fifth overtone AT-cut crystals was made. Measurement of the amplitude frequency dependence of the SC-cut and AT-cut crystal units was made. The results indicate that crystal unit power dissipation can be increased 6- to10-dB above levels that have been used with the AT-cut units.

A review of noise sources associated with bipolar junction transistor oscillators indicates that second- and third-order distortion is responsible for the dominant flicker of phase exhibited in practical oscillators. The frequency correction term arising in solution of the vanderPol equation gives an insight into requirements for reducing flicker of frequency.

The bridged-tee oscillator is suitable as the oscillation sustaining circuit for an SC-cut VHF crystal unit. Noise sources of this circuit, including the amplitude-to-phase conversion of amplitude-modulated flicker noise, was considered and an estimate of $\mathcal{L}(f)$ expected for such an oscillator was obtained. Measurement of a pair of identical oscillators was then made and reasonable agreement is found to exist between the predicted $\mathcal{L}(f)$ and measured $\mathcal{L}(f)$.

It is found that the SC-cut VHF crystal does provide substantial improvement of VHF signal generator $\mathcal{L}(f)$ in comparison with AT-cut VHF crystal-controlled oscillators.

LINEARIZED MODEL OF OSCILLATOR PHASE NOISE

Figure 1 shows the block diagram of an oscillator modeled as a feedback amplifier configuration containing an amplifier in the forward "μ" path and a frequency selective filter in the feedback "β" path. The Barkhausen condition is described by figure 2. Figure 1 circuit can oscillate at frequencies for which the Barkhausen condition is satisfied. In a practical oscillator, μ is dependent on the amplitude, V, of the voltage applied to its input. The noise model of figure 1 does not account for nonlinearity of μ. μ is simply considered to be precisely the value necessary to satisfy the Barkhausen condition and independent of oscillation amplitude.

The model of figure 1 was first proposed by Leeson[1] and its validity confirmed by Sauvage[2]. In the model, $S\delta_{\phi i}(f)$ represents the spectral density of phase modulation associated with a signal passing through the $\mu\beta$ transmission path. This spectral density includes an $f^{-\alpha}$ dependency arising from low-frequency noise sources and nonlinearity in the $\mu\beta$ components and an f^{0} dependency arising from additive noise in the vicinity of the oscillation frequency. Shot noise and thermal noise sources contribute to the f^{0} dependency. Low frequency shot and thermal noise also can contribute to the f^{0} as well as $f^{-\alpha}$ type $S\delta_{\phi i}(f)$. This will cause an apparent noise figure, F, as used by Leeson and others to appear higher than the F otherwise used in the expression for $S_{\phi}(f)$.

In high Q oscillators, such as crystal-controlled oscillators, the $f^{-\alpha}$ dependent $S\delta_{\phi i}(f)$ when $\alpha = 1$ results in frequency flicker of the oscillator signal within the feedback bandwidth. The f^{0} dependent $S\delta_{\phi i}(f)$ results in white frequency fluctuation. Beyond this bandwidth, the output signal observed at A exhibits additive $S\delta_{\phi i}(f)$. The $S_{\phi}(f)$ of the oscillator signal at A then is as shown in figure 3. If the oscillation signal is observed at B in figure 1, then the additive phase signal at A is further attenuated by the transmission characteristic of the β circuit.

The expression for the oscillator output signal phase is

$$S_{\phi}(f) = S\delta_{\phi i}(f)\left[1 + \left(\frac{f_o}{2Q}\right)^2\left(\frac{1}{f}\right)^2\right]$$

$S_{\phi}(f)$ = one sided spectral density of phase

where f_o = oscillator frequency

Q = passive Q of the oscillator (dominated by phase shift property of β circuit)

f = offset frequency (modulating frequency)

$S\delta_{\phi i}(f)$ can be viewed as caused by a number of random amplitude, random phase voltages added to the oscillator signal voltage, since $x \int_{f_1}^{\infty} S_{\phi}(f) \propto$ Signal Power. This would be the case for additive noise. Then, since a single component of noise added to the oscillator signal can be considered as a result of equal amplitude and phase modulation, one-half the noise constitutes the phase fluctuation. This leads to the concept that $S\delta_{\phi i}(f)$ can be expressed as

$$\frac{1}{2} S\delta_{\phi i} \approx \frac{FkT}{2P_S}\left(1 + \frac{f_c}{f^{\iota}}\right) = \left(\frac{K}{f^{\iota}} + \frac{FkT}{P_S}\right)$$

where F is a noise factor

ι constant between 1 and 2

P_S is the oscillation frequency signal power at input of sustaining circuit

f is the offset frequency (modulating frequency)

f_C is a cutoff frequency

The F is not the usual noise figure of the sustaining amplifier. The reason is that nonlinearities are present in the practical oscillator and the simple oscillator noise model is a linear one. The non-linearity always present in practical electrical oscillators results in frequency up-conversion of low-frequency noise. The low-frequency noise includes flicker and white noise. Down-conversion of wideband noise also is present.

SOURCES OF NOISE IN BI-POLAR JUNCTION TRANSISTORS

Figure 4 shows an equivalent circuit representation for the transistor. Note that several parameters are nonlinear, being current and voltage dependent.

The flicker noise is modeled by a second fluctuation current generator paralleling the base current shot noise current generator. I_b and I_c are average values of base current and emitter current.

Frequency conversion of low-frequency noise caused by the oscillator signal commutation results in dominantly amplitude modulation. To influence $S\delta_{\phi i}(f)$, there must be conversion from amplitude to phase modulation. In a previous paper[1], it was theorized that second-order distortion would up-convert the low frequency noise to band surrounding the oscillator signal frequency as amplitude modulation, which then would be converted to phase modulation via the source impedance plus base resistance and base-emitter capacitance. In wideband amplifiers, this mechanism has been verified. In oscillator application, this is found not to be the dominant mechanism.

In addition to second-order distortion, there is third-order distortion.

The second harmonic current of the bipolar transistor VHF oscillation sustaining circuit, however, is nearly in quadrature with the fundamental. Therefore, the amplitude flicker modulation at the second harmonic appears as nearly all phase modulation of the oscillator signal. To minimize the "multiplicative" noise in an oscillator circuit, therefore, the nonlinearity must be made as small as possible and i_f must be as small as possible.

For a realistic quantitative measure of the flicker of phase as well as white phase noise in $S\delta_{\phi i}(f)$, a transient solution of the simple circuit shown in figure 5 was made.

The transistor was modeled to have $h_{FF} = 70$, $f_T = 1200$ MHz using a charge-controlled model.

The Z_E of 45 ohms is representative of the crystal unit impedance in a bridged-tee-oscillator circuit configuration. Self-limiting operation is investigated by examining the collector current as a function of the amplitude of the input signal to the circuit of figure 5.

Figure 6 shows a plot of the collector current for V_1 = 1 volt. Note that current limiting occurs for this condition. (The simple circuit of figure 5 is analyzed to avoid problems with initial conditions required to achieve an inexpensive solution. When a circuit involving L, C, R is subjected to transient solution, an excessive amount of computation is necessary in order to reach the steady-state solution, unless nearly exact initial conditions are provided for currents and voltages associated with the memory elements; i.e., L and C.)

Transient solutions were obtained for various values of V_1 from 0.1 to 1.0 volts and, at each value, a small change in V_4 was made and transient solution obtained. Fourier analysis of the resulting collector current then was made.

Figure 7 shows typical printout of the Fourier analysis for V_1=0.9 volts, V_4=0 volts; V_1= 0.9 volts, V_4=0.01 volts.

Note that the relative phase of the collector-current second harmonic is nearly in quadrature with the fundamental component.

Also note that the effect of an equivalent low-frequency input of 0.01 volts added to V_1 results in change of the second harmonic by 0.06 mA. The fundamental changes by 0.12 mA. Figure 8a shows a fundamental wave and a second-harmonic wave in which relative phase is zero degrees. The zero crossings of the sum of these two waves is unaffected by the amplitude changes of either the fundamental or second-harmonic component waves.

In figure 8b, a 90-degree lag of the second harmonic relative to the fundamental is shown. For this condition, the time of zero crossing of the composite waveform is affected both by the amplitude variations of the fundamental and second-harmonic component waves. Phase modulation of the oscillator-sustaining-circuit current thus occurs due to amplitude modulation of the transistor current by the low-frequency-noise current because of the relative phase that exists between fundamental and second-harmonic components of the current.

Reducing the level of oscillation voltage from 1.0 to 0.1 volts results in change of relative phase to 72.6 degrees.

It then appears that $\delta\phi_{oi}$ caused by low-frequency noise associated with the transistor will decrease as the oscillation amplitude is decreased.

The expansion of the current as a power series of V_i; i.e., $i = a_1 V + a_2 V^2 + a_3 V^3 + \ldots$ where $V = V_o \cos\omega_o t + V_n \cos\omega_n t$, shows that the amplitude modulation of the fundamental arises from the a_1, a_3, a_5,... terms and the amplitude modulation of the second harmonic arises from a_3, a_5, a_7,... terms. For the case where terms due to nonlinearity greater than fifth degree can be neglected, the result is

$$\text{FUNDAMENTAL SIDEBANDS (SECOND ORDER)} \quad \left(a_2 V_o V_n + \frac{3}{2} a_4 V_o^3 V_n + \frac{3}{2} a_4 V_o V_n^3\right) \cos(\omega_o + \omega_n)t$$

$$\text{SECOND HARMONIC SIDEBANDS (THIRD ORDER)} \quad \left(\frac{3}{4} a_3 V_o^2 V_n + \frac{5}{4} a_5 V_o^4 V_n + \frac{15}{8} a_5 V_o^2 V_n^3\right) \cos(2\omega_o \pm \omega_n)t$$

To minimize the amplitude of the sidebands, V_o and the odd-order-distortion components should be as small as possible.

Negative current feedback is useful in reducing a_3, a_5, etc., but a nonlinearity is still demanded by the vanderPol equation in order that nearly-sinusoidal, constant-amplitude sustained oscillations be obtained.

The reason for the second-harmonic phase lagging the fundamental appears to be due to the nature of the base-emitter junction impedance. If the fundamental and second harmonic frequencies were both well below the beta cutoff frequency, the phase shift could be greatly reduced. Much reduced multiplicative noise then should be experienced in a 5-MHz oscillator using the same transistor used in the 100-MHz oscillator.

VANDERPOL EQUATION

The frequency correction term arising in solution of the vanderPol equation also is an indication that the amplitude fluctuation will cause frequency fluctuation. The nonlinear differential equation of the form

$$\frac{d^2x}{dt^2} + x = \varepsilon f(x)\frac{dx}{dt}$$

obtained from the integro-differential equation

$$\frac{dx}{dt} + \int x dt = \varepsilon F(x)$$

is known as the vanderPol equation[4]. The integro-differential equation results directly from consideration of the circuits shown in figure 9. (a_1 and a_3 are assumed real in usual solution of the vanderPol equation.)

Solution of the vanderPol equation is limited to constraint that $\varepsilon \ll 1$. This provides solution for circuits that produce electrical harmonic oscillation (very nearly sinusoidal voltage and current). The circuits are quasi-linear; i.e., the nonlinear conductance (or nonlinear reactance) terms are small compared to susceptance (or reactance) terms. Although such conditions must exist, the curvature of the nonlinear current voltage characteristic can still be very large at the region of interest.

The solution is assumed to be of the form

$$x = a \sin (t+\theta)$$

where a and θ are unknown functions of t. Several quantitative methods of solution then exist to find a and θ[5,6]. All methods yield approximate solutions.

The steady-state solution yields

$$x = a_1 \sin \left[\left(1 - \frac{\epsilon}{16}\right)t + \theta_1\right] + \frac{\epsilon}{5} a_1 \cos 3\left[\left(1 - \frac{\epsilon}{16}\right)t + \theta_1\right]$$

The frequency, thus, is found to depend on the value of a. ϵ is related to the excess gain of the oscillator.

The solution shown above is for normalized angular frequency of 1 radian/second.

In denormalized form, the correction term is:

$$1 - \frac{\omega}{\omega_0} = \frac{\epsilon^2}{16} = \frac{(a_1 - G)^2}{16} \frac{L}{C}$$

$$\frac{\Delta f}{f_0} = \frac{(a_1 - G)^2}{16} \frac{L}{C} = \frac{(b_1 - R_1)^2}{16} \frac{C}{L}$$

High Q reduces Δf, but fluctuation of a_1 results in Δf fluctuation that depends on a_1. The excess gain therefore, must be as small as possible to prevent amplitude fluctuation of a from being converted to significant frequency fluctuation.

THE VHF SC-CUT CRYSTAL-CONTROLLED OSCILLATOR RESONATOR

VHF third-overtone SC-cut crystal units were obtained from two crystal unit manufacturers. Measurement of the amplitude frequency characteristics of these crystal units and of the fifth-overtone AT cut units presently used was made. Figures 10 and 11 show typical characteristics obtained for these two types of crystal units. The reduced nonlinearity exhibited by the SC cut as compared to AT cut crystal units indicates degradation of K as P_x is increased will be less for SC cut crystal units than for AT cut crystal units. Measurement of the crystal unit parameters at drive level of -20 dBm also were made using the HP4191A RF Impedance Analyzer. Figure 12 shows the measurement arrangement.

The HP4191A was controlled by a 9825A computer and measurement of R+jX, G+jB, Y $\angle\theta$ and Z $\angle\theta$ was made of the crystal unit over a range of frequencies to determine crystal unit parameter values. With fixture and static capacitance approximately antiresonated, relatively simple relationships shown in figure 13 are used to obtain R_1, C_1, and Q of the crystal unit from HP4191A measured quantities. C_0 plus C_{12} plus C_{22} is obtained by subtraction of fixture susceptance when measurement is made a few percent below crystal unit resonant frequency. The active C_0 determined from theory is found to be much smaller than the measured crystal-unit static capacitance. Inactive static capacitance thus appears to be significant. (Plates were 0.25-inch diameter with 0.1-inch diameter electrodes.) Figures 14, 15, and 16 show tabulated results of the parameter values as determined for the SC-cut crystal units using the HP4191A.

SUSTAINING CIRCUIT

A suitable sustaining circuit is provided by the bridged-tee circuit. Figure 17 shows the basic circuit. This circuit can be analyzed as a feedback oscillator in which the crystal unit provides the dominant part of the β circuit. It can also be analyzed as an LC Colpitts oscillator stabilized by the phase shifting property of the crystal unit. In this case, the $S_\phi(f)$ is improved because the effective transconductance exhibits the crystal-unit phase-shift property as well as a decrease in magnitude with frequency change in accordance with

$$Z_X = R_1 + j \frac{dx}{df} (f - f_s)$$

Figure 18 shows the measured Z_X for one of the SC-cut crystal units. In the SC-cut crystal-controlled oscillator the Barkhausen condition must not be satisfied at the B mode frequency under any condition, such as when energizing the oscillator circuit or during steady-state condition. The phase shifting characteristic and magnitude of the antiresonant impedance provided at the collector by the L, C_1, C_2, C_3, R circuit are sufficient to guarantee such conditions if $R/\Delta L > 10$ and the excess gain is less than 3.

CAD analysis[7] of the actual oscillator circuit is useful in computation of $\mathcal{L}(f)$ for real nonlineat oscillator circuits.

$\mathcal{L}(f)$ CHARACTERISTICS

Two oscillators were constructed and measurement of $S_{\phi_{osc1}} + S_{\phi_{osc2}}$ was made using arrangement of figure 19.

Figure 20 shows the results of the measurements as well as the values of $\mathcal{L}(f)$ that were calculated. Reasonable agreement exists between predicted and observed values of $\mathcal{L}(f)$. At large offset, f, additive noise introduced by circuits that follow the oscillators significantly influence the observed $S_\phi(f)$ of the signal source.

Figure 20 compares the results with the results given by Pegeot at the 34th ASFC. Note that a reduction in phase noise exhibited by the oscillator output slightly greater than the increase in crystal unit dissipation is obtained.

A comparison also was made between the SC-cut and third-overtone and fifth-overtone AT-cut crystal units operating at +6 dBm and 0 dBm crystal unit dissipation. These results are shown in figure 21.

REFERENCES

1. D.B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum", Proc. IEEE, V54, page 329, February 1966.

2. G. Sauvage, "Phase Noise in Oscillators: A Mathematical Analysis of Leeson's Model", IEEE Transactions on Instrumentation and Measurement, Vol. I.M. -26, No. 4, Dec. 1977.

3. D.J. Healey, III, "Flicker of Frequency and Phase and White Frequency and Phase Fluctuations in Frequency Sources", Proc. 26th Annual Symposium on Frequency Control, 1971, pp 29-42.

4. A. VanderPol, "The Nonlinear Theory of Electric Oscillations", Proc. IRE V22, 1934, pp 1051-1086.

5. Keller, E.G., "Analytical Methods of Solving Discrete Nonlinear Problems in Electrical Engineering", Transactions of A.I.E.E., Vol 60, 1941, pp 1194-1200.

6. Nicholas Minorsky, "Nonlinear Oscillations", D. Van Nostrand Co., Inc. 1962.

7. D.J. Healey, III, "Use of Computer Aided Design (CAD) Technique to Determine Oscillation Amplitude in Non-Linear Harmonic Oscillators" (to be published).

8. C. Pegeot, G. Sauvage, "UHF Oscillator Using SC Cut Quartz Crystal With Low Noise Performance and High Long Term Stability", Proc. 34th Annual Symposium on Frequency Control, pp. 233-236.

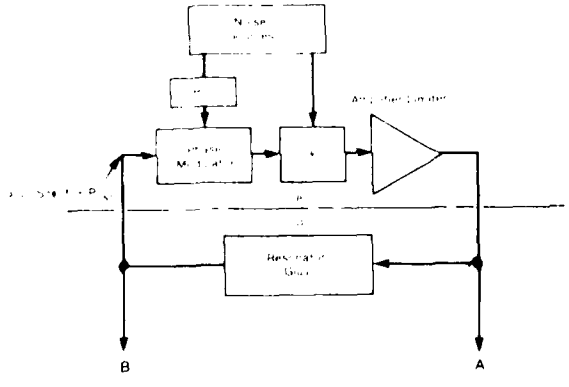
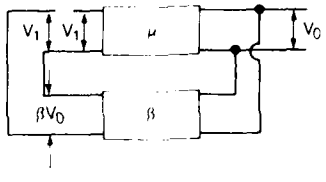
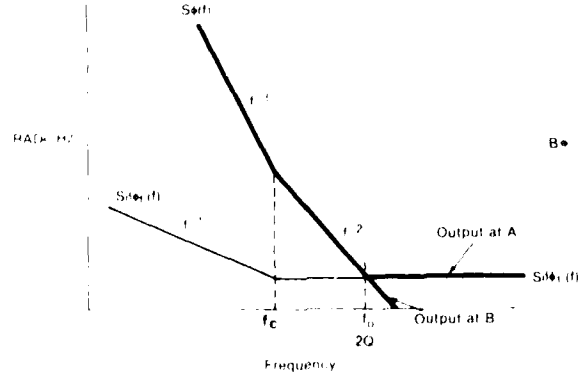


Figure 1. Model of Harmonic Oscillator



- Input Lead Is Broken (and Normal Passive Circuit Terminations Provided at Break Point)
- V_1 Applied as Input to μ
- If Return Voltage $\beta V_0 = V_1$ in Absolute Value and Phase, the Restored Circuit Will Function as an Oscillator. (Simply Nyquist Condition)

Figure 2. The Barkhausen Condition



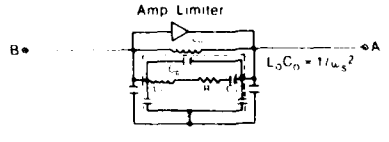


Figure 3. Oscillator Phase Noise Spectrum

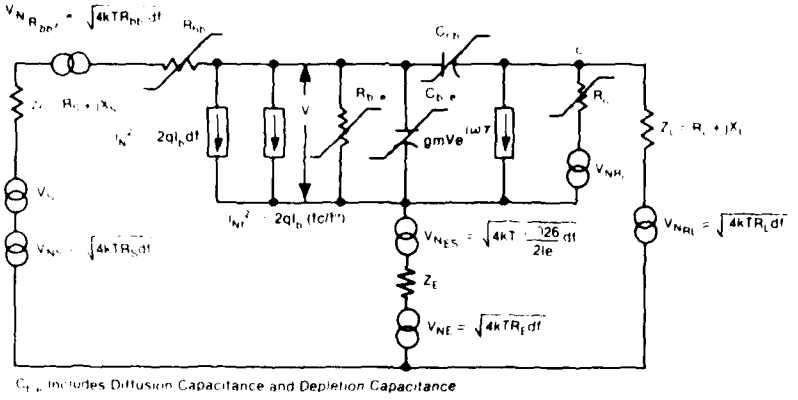


Figure 4. Transistor Circuit Large Signal Operation and Noise Sources

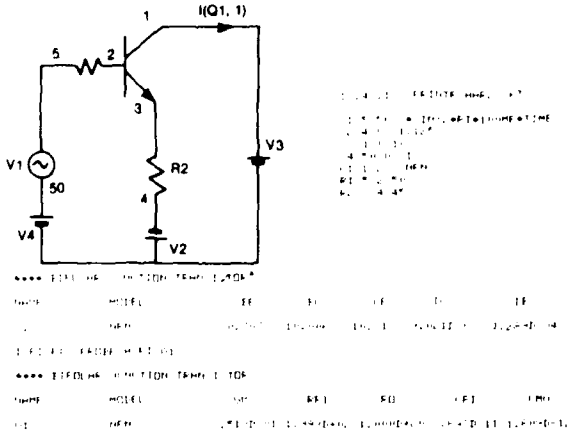


Figure 5. CAD Circuit "HAR 2"

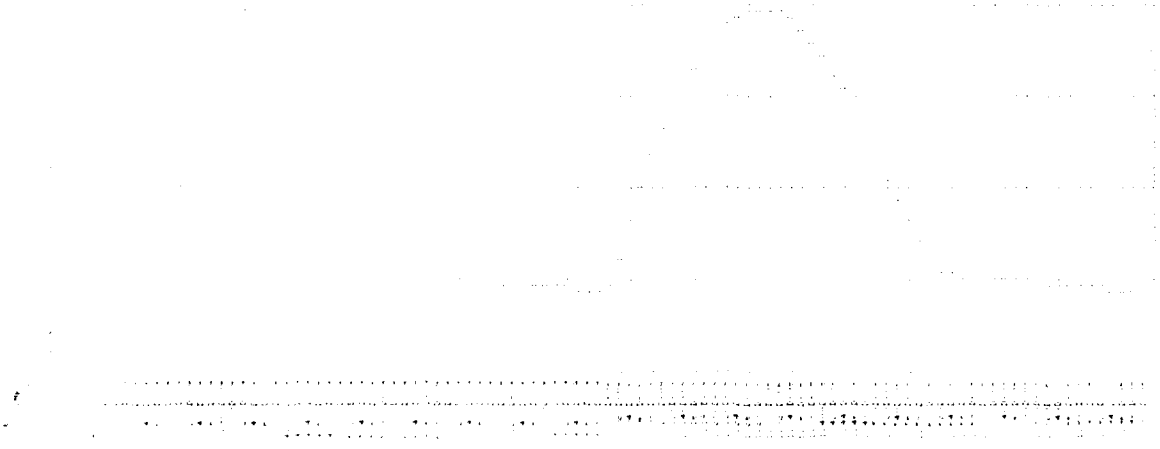


Figure 6. Collector Current of HAR 2; V4=0, V=1

FOURIER COMPONENTS OF TRANSIENT RESPONSE: I(Q1 1) V1 = 0.9 VOLTS, V4 = 0

DC COMPONENT = 1.092D-02

HARMONIC NO	FREQUENCY (HZ)	FOURIER COMPONENT	NORMALIZED COMPONENT	PHASE (DEG)	NORMALIZED PHASE (DEG)
1	1.000D+08	1.486D-02	1.000000	-16.566	0.0
2	2.000D+08	3.099D-03	0.208569	-106.854	-90.288
3	3.000D+08	1.393D-03	0.093722	-53.459	-36.892
4	4.000D+08	3.746D-04	0.025210	-65.095	-48.528
5	5.000D+08	4.664D-04	0.031382	-78.696	-62.129
6	6.000D+08	2.861D-04	0.019250	-56.925	-40.359
7	7.000D+08	1.616D-04	0.010876	-77.418	-60.852
8	8.000D+08	1.590D-04	0.010697	-74.327	-57.761
9	9.000D+08	9.653D-05	0.006496	-66.058	-49.492

FOURIER COMPONENTS OF TRANSIENT RESPONSE: I(Q1 1) V1 = 0.9 VOLTS, V4 = 0.01 VOLTS

DC COMPONENT = 1.106D-02

HARMONIC NO	FREQUENCY (HZ)	FOURIER COMPONENT	NORMALIZED COMPONENT	PHASE (DEG)	NORMALIZED PHASE (DEG)
1	1.000D+08	1.498D-02	1.000000	-16.478	0.0
2	2.000D+08	3.036D-03	0.202649	-106.476	-89.998
3	3.000D+08	1.398D-03	0.093300	-52.507	-36.029
4	4.000D+08	3.751D-04	0.025037	-58.613	-42.135
5	5.000D+08	4.505D-04	0.030070	-75.884	-59.406
6	6.000D+08	2.911D-04	0.019431	-52.839	-36.361
7	7.000D+08	1.559D-04	0.010406	-69.357	-52.879
8	8.000D+08	1.547D-04	0.010329	-69.097	-52.619
9	9.000D+08	9.814D-05	0.006552	-58.039	-41.561

Figure 7. Fourier Components of HAR 2

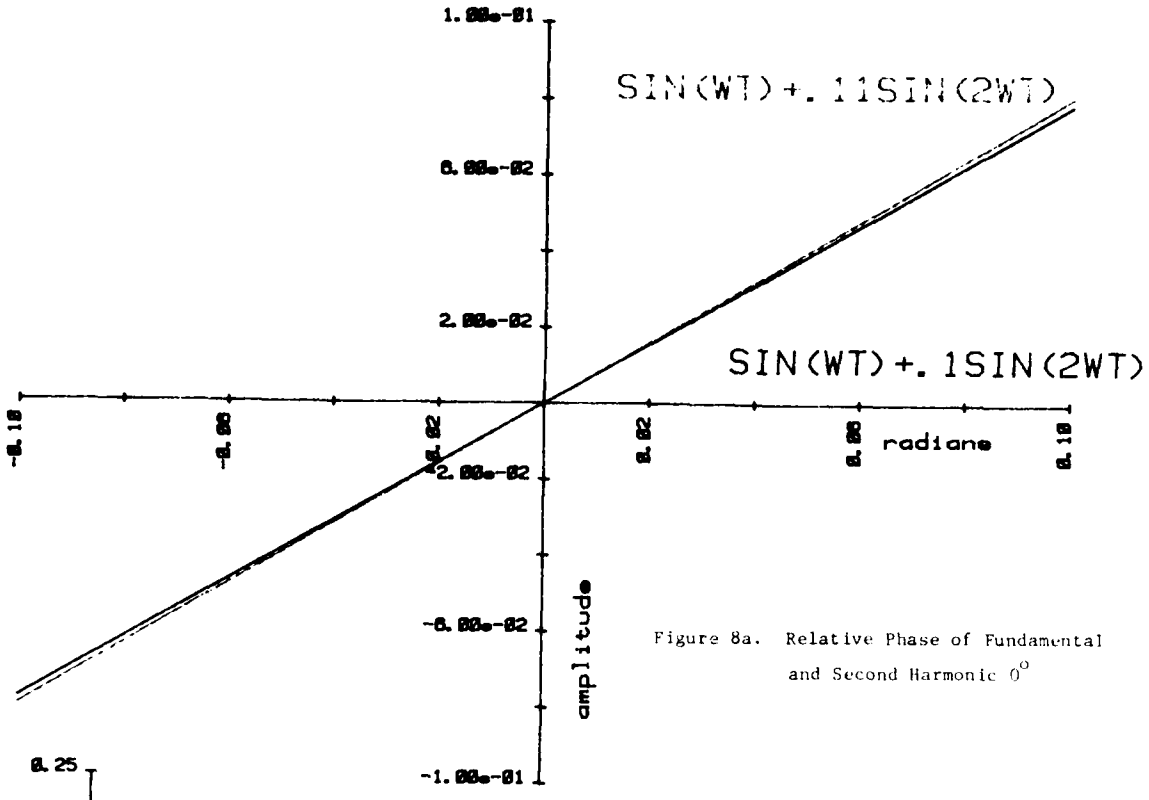


Figure 8a. Relative Phase of Fundamental and Second Harmonic 0^{o}

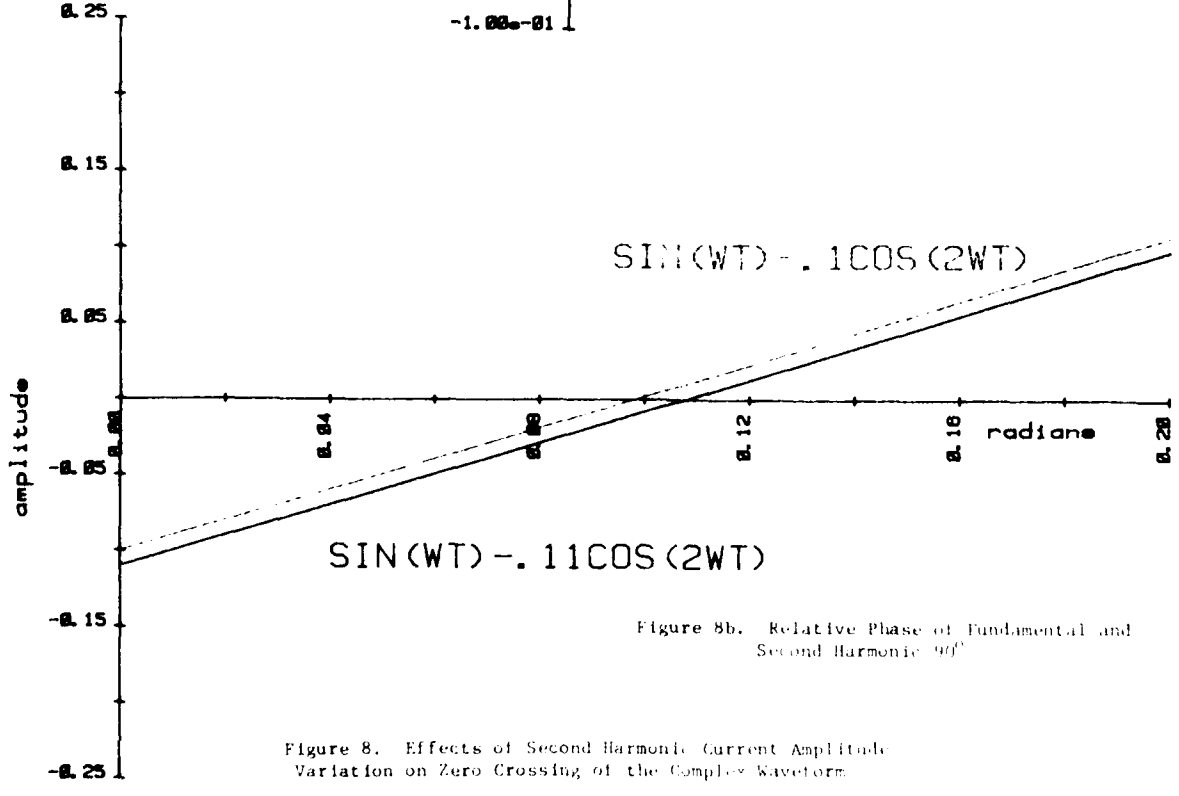


Figure 8b. Relative Phase of Fundamental and Second Harmonic 90^{o}

Figure 8. Effects of Second Harmonic Current Amplitude Variation on Zero Crossing of the Complex Waveform

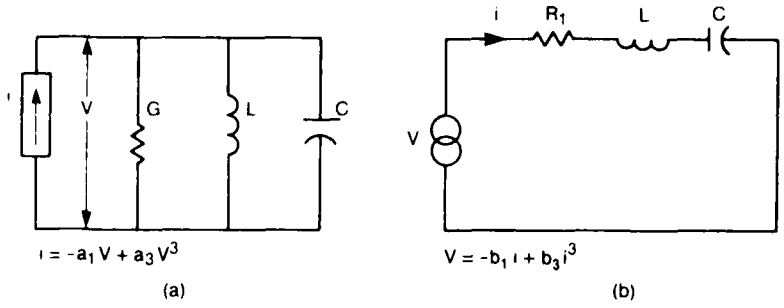


Figure 9. Simple Electrical Oscillators

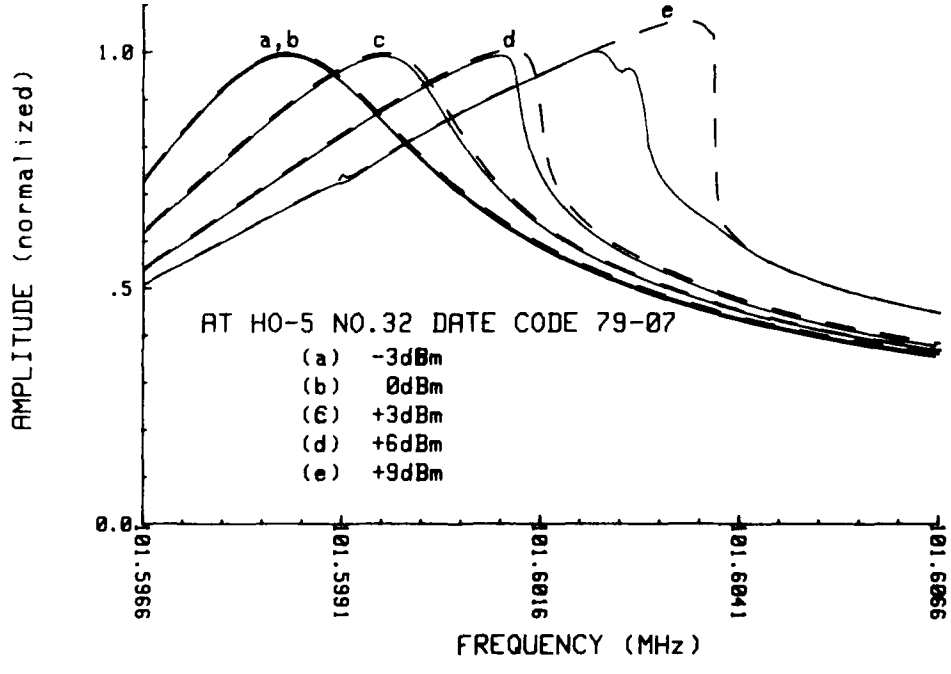


Figure 10. Large Signal Transmission Quartz Crystal Unit

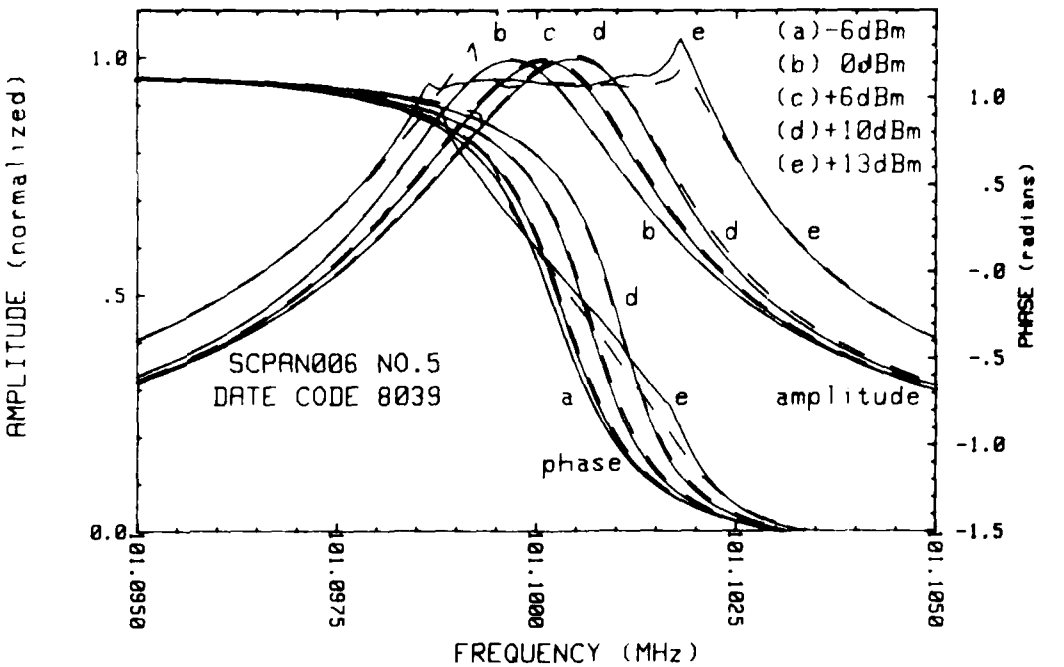


Figure 11. Large Signal Transmission Quartz Crystal Unit

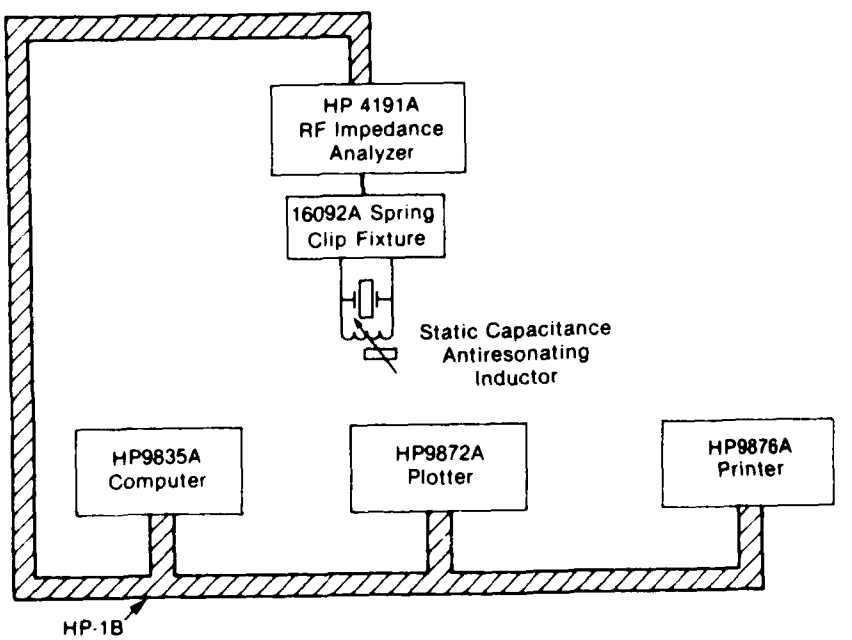
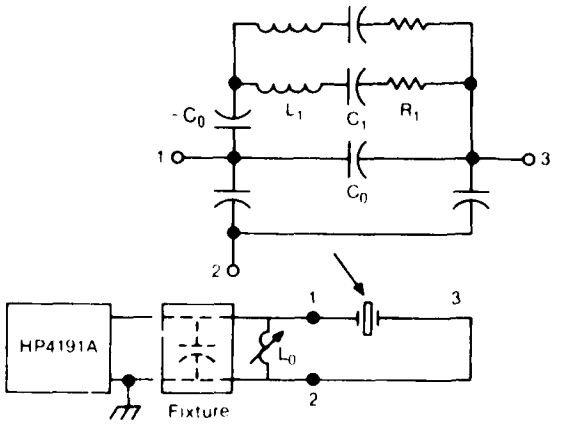


Figure 12. Arrangement Used for Measurement of Crystal Unit Parameters



L_0, Self Capacitance < 1 pF, Q>100, Antiresonates
Total Static Capacitance.
Admittance Measurement. Find f for Maximum g. Use as f_s

Impedance Measurement $Z\angle\theta$

$$\text{Determine } Q = \frac{f_s}{2}(d\theta/df)_{max}$$

Impedance Measurement R + jX

$$\text{Determine } C_1 = \frac{1}{\pi f_s^2}(dx/df \text{ in Vicinity of } f_s)$$

$$\text{Determine } R1 = \frac{1}{2\pi f_s Q C_1} = \frac{dx/df}{d\theta/df}$$

Figure 13. Some Relationships Useful in Obtaining Crystal Unit Parameters

Unit No	Nom. Freq (MHz)	Q_{XTAL}	R1 "C" Mode (OHMS)	C_1 "C" Mode (femto-farads)	dx/df OHMS/Hz	dθ/df (RAD/Hz)
5	101.1	57323	46.8	0.58296	0.05342	0.001123
6	101.1	72518	31.7	0.68174	0.04568	0.001436
7	101.1	78300	29.4	0.68731	0.04531	0.001549
8	101	95748	31.7	0.5101	0.06117	0.001896
9	101	89132	28.5	0.60635	0.05136	0.001765
10	101	96606	26.4	0.60717	0.05129	0.001913
11	101	82618	28.1	0.67524	0.04612	0.001636

Figure 14. Tabulation of Parameters for Third Overtone SC-Cut Crystal Units

S/N	B/C (dB)	Spur1/C	Spur2/C	F_{Spur}-F_{SC} (Hz)	F_{Spur}-F_{SC} (Hz)
008	1.94	8	11	18000	40000
009	4.65	10	11	14000	22000
010	0.4	6	6	16000	42000
011	0.4	15	12	14000	33000

Electrode Diameter 0.1 Inch
Plating: Chrome, Silver, Nickel

Figure 15. Resistance Ratio in dB of Unwanted Modes to "C" Modes

Unit No.	R_x (OHMS) From Analyzer	R_x (OHMS) Calculated From Measured dx/df and dθ/df
5	46.8	47.11
6	31.7	31.81
7	29.4	29.25
8	31.7	32.2
9	28.5	29.1
10	26.4	26.81
11	28.1	28.19

a. Crystal Unit Resistance

Unit No.	CO (pF) (Calculated From C1)	CX + CH + CO (pF) (From Direct Measurement)	CO-CH (pF) Piezo's Measurements
5	2.597	4.85	4.38
6	3.037	5.14	4.34
7	3.062	5.12	4.48
8	2.272	5.02	4.44
9	2.701	5.18	4.44
10	2.705	4.96	4.32
11	3.008	4.92	4.34

b. Static Capacitance

Figure 16. Comparisons of Direct Parameter Measured Value With Value Derived From Indirect Measurements

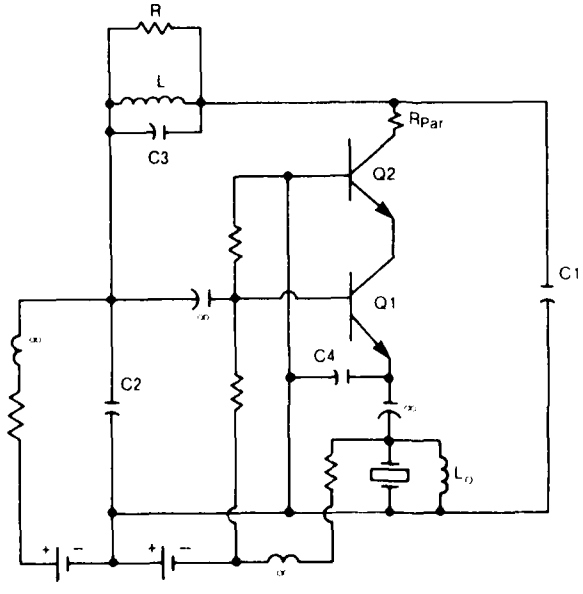


Figure 17. Typical Self-Limiting Bridged-Tee Crystal-Controlled Oscillator

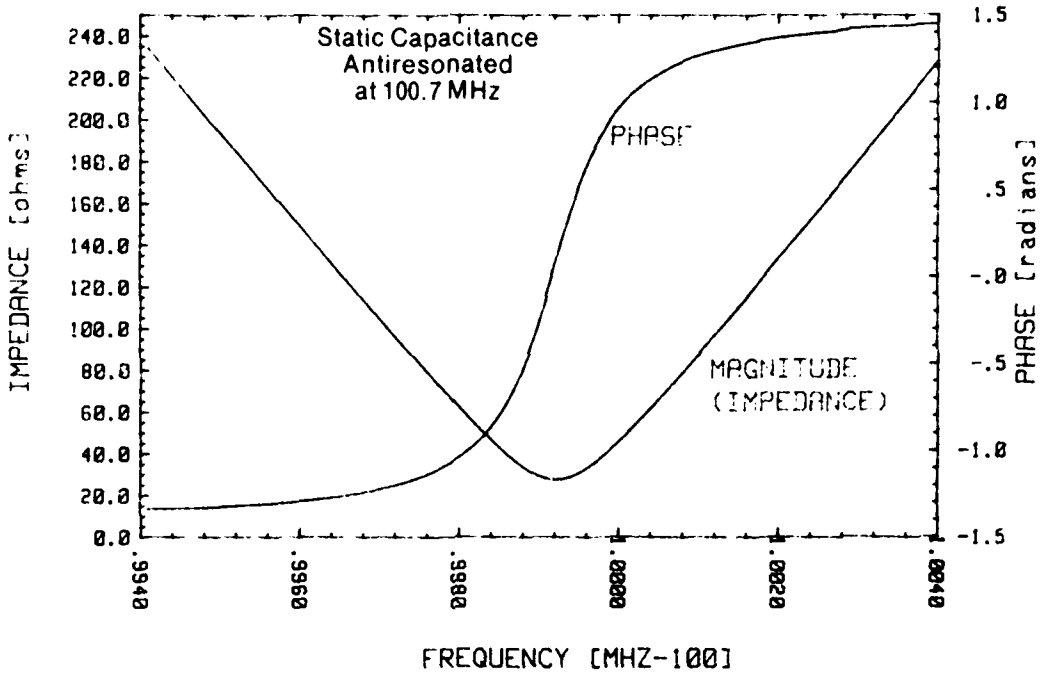


Figure 18. Impedance SCPAN5 No. 10 101.0MHz Overtone SC

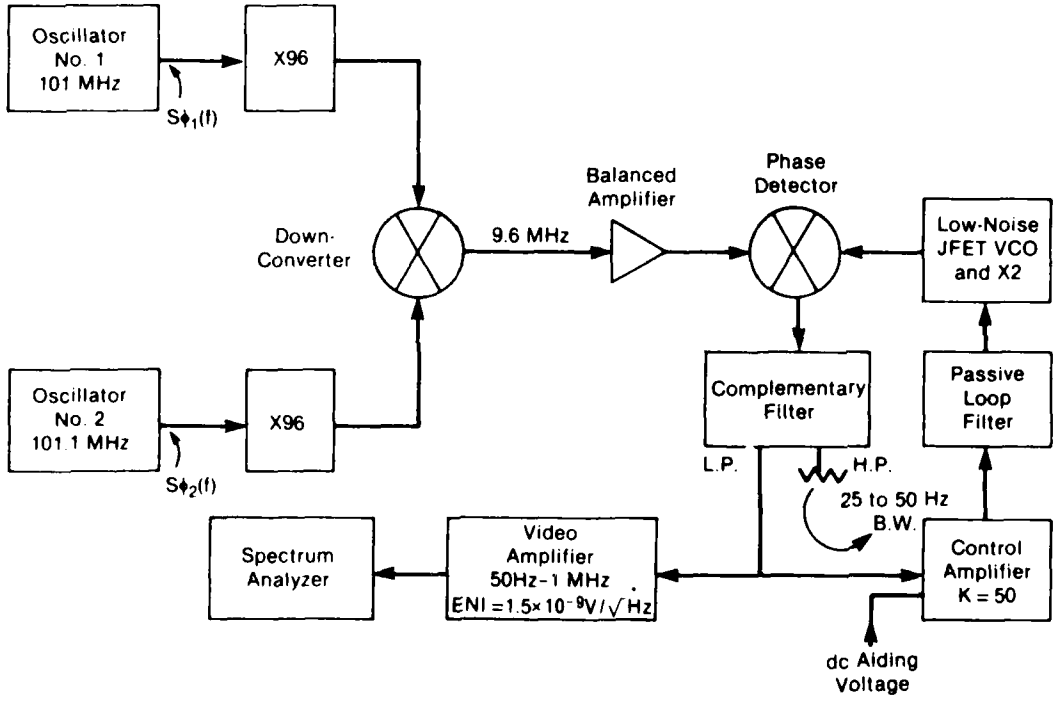


Figure 19. Arrangement Used To Obtain Measure of $S\phi_1(f) + S\phi_2(f)$

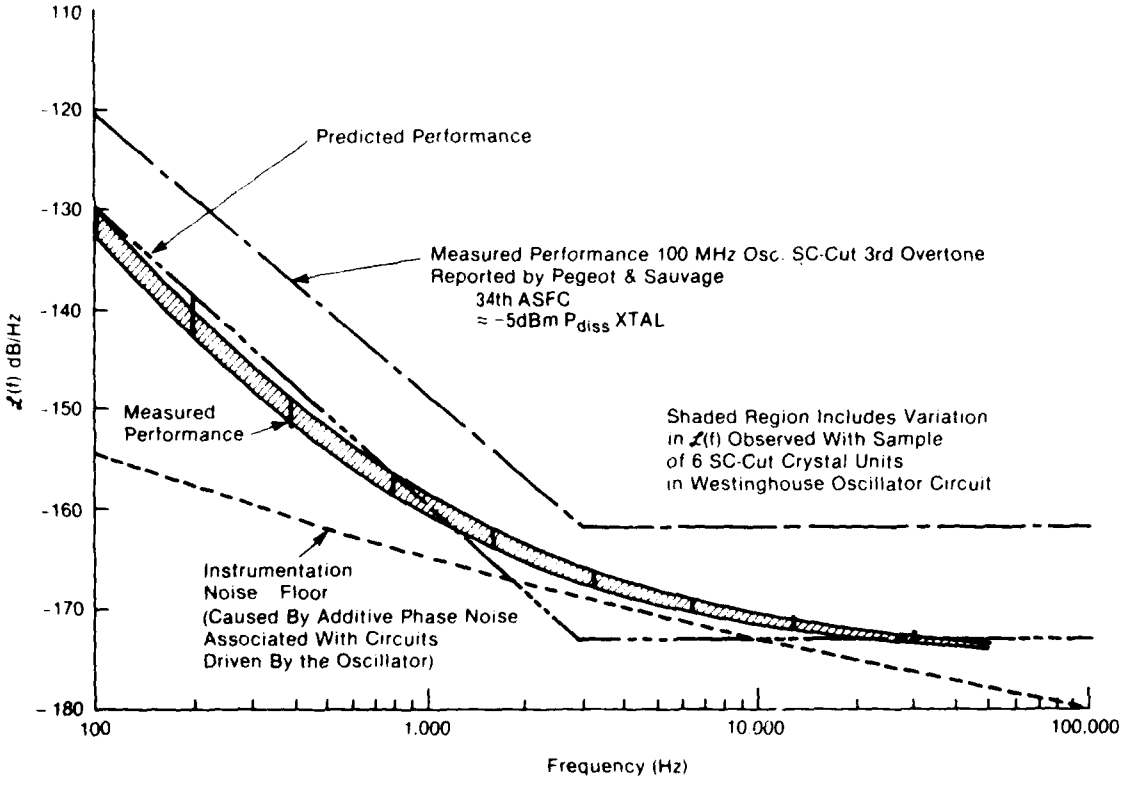


Figure 20. $\mathcal{L}(f)$ Characteristics of Third Overtone SC-Cut Crystal-Controlled Oscillators at Approximately 100 MHz

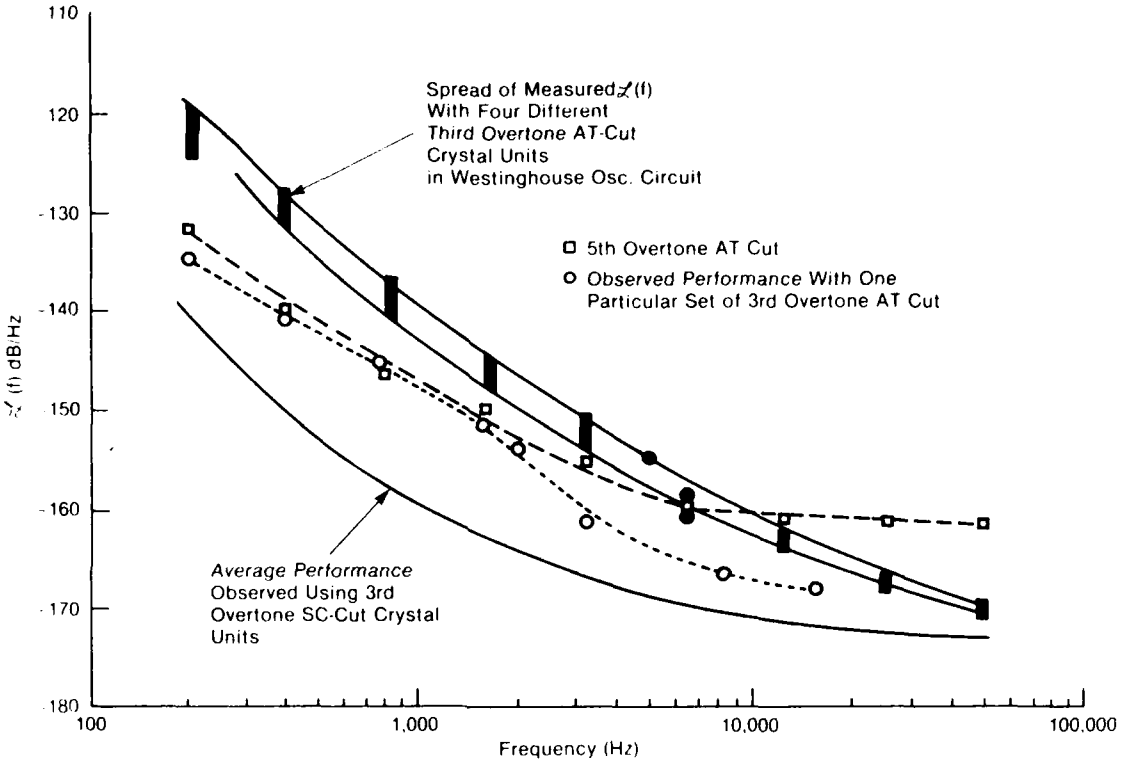


Figure 21. Comparison of $\mathcal{L}(f)$ Displayed by Oscillator Using SC-Cut and AT-Cut Units

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

DIGITAL TEMPERATURE COMPENSATION OF CRYSTAL OSCILLATORS

A. M. Renard - K. Barnhill

Raytheon Company
Equipment Division Laboratories
Sudbury, MA 01776

Summary

This paper looks at the feasibility of producing digital temperature compensated crystal oscillators. The test station is an automatic system designed for testing 20 to 35 crystals at one time. The results of a Table-Look-Up compensation method and a 7th order equation compensation method are compared. Both methods produced results within a ± 0.05ppm window of the desired frequency.

Key words (for information retrieval) Crystal Oscillators, Digital Compensation, Temperature Compensation.

Introduction

Digital temperature compensation of crystal oscillators will provide a means for producing low cost, precision oscillators. An experiment conducted to prove the feasibility of this produced and reproduced data that remained within ± 0.05ppm of the desired frequency. The data obtained suggests a resolution of ± 0.01ppm is possible in the near future. To gain this precision and reproducibility at a reasonable cost, the following 3 design goals were established:

1. The smallest increment of compensation would not cause a frequency deviation greater than the desired accuracy.
2. The system would be automatic and be able to produce crystal oscillators in production quantities.
3. The system would compensate for temperature effects over a wide operating temperature range, avoiding the need for an expensive oven.

Test System

The Automatic Crystal Test Station (ACTS) is made up of IEEE-488 bus compatible instruments. A diagram of the equipment configuration is shown in Figure 1. The ACTS System monitors frequencies, voltages and temperatures of the crystal oscillator, stores the data and calculates the required compensation to maintain the resolution desired.

In the future, the ACTS System could also automatically transfer the tabulated data to a memory chip or microprocessor contained in the oscillator package; the microprocessor could then control the temperature compensation process.

Test Technique

One complete test consisted of two temperature cycles. For both temperature cycles the oven temperature was ramped from -35°C to 85°C. During the first temperature cycle, a voltage across a varicap was adjusted until the actual crystal frequency came within ± 0.01ppm of the desired frequency. At this frequency, the voltage and temperature were recorded. This same process was repeated continuously as the temperature ramped up. On the average, 1200 pairs of points (temperature and voltage) were collected and plotted. The results of one run is shown on the voltage versus temperature curve in Figure 2.

In between temperature cycles, 35 of the 1200 pairs of points were selected to determine the coefficients for a 7th order equation that best fitted the data. For comparison purposes, several of the tests used 5th order equations to calculate the correct compensating voltages.

During the second pass, the temperature was continuously sampled and several methods were used to determine the proper varicap voltage. In one method used, known as Table-Look-Up, each temperature reading corresponded to a specific compensating voltage designated by the pairs of points collected in the first temperature cycle. The Table-Look-Up method is strictly a brute force approach. A refinement to this method includes linear interpolation between any two adjacent points. In addition, interpolation was tried with less than 100 evenly spaced points, of the original 1200 points. The purpose was to determine the trade-off between the size of the data base and the accuracy of its compensating voltage calculations.

Another method used to determine the proper voltage across the varicap was with 5th and 7th order equations. The advantage of this approach was that it required almost no memory space.

Results

A frequency deviation of ± 0.05ppm was achieved with both the 7th order equation and the 1200 point Table-Look-Up with interpolation. The 7th order equation, however, produced a smoother overall output and was therefore rated better than the Table-Look-Up. Comparisons of the other compensating methods are given below as well as discussions of significant factors and problems encountered in achieving these results. The outputs obtained during the experiment are shown in Figures 3 and 4; Figure 5 shows the frequency versus temperature characteristics of the uncompensated crystal.

The smallest incremental voltage change from the power supply was 1mv. A 1mv step change across the varicap produced a 0.01ppm jump in frequency. Initially the software incremented the power supply in 10mv steps. As shown in Figure 3, the resolution of the 10mv step was 0.1ppm. This was a significant change in accuracy showing that a 1mv step is required for the desired frequency resolution.

The frequency variations using the 1200 point Table-Look-Up produced poor results on the cold side. Variations of 0.1ppm were common with one peak showing a 1ppm error. Adding linear interpolation between adjacent points reduced all frequency errors to ± 0.05ppm. Interpolation between every 20th data point produced a frequency deviation less than ± 0.05ppm (graph is not shown). This is a significant reduction in data points. Using Table-Look-Up with linear interpolation, only 60-70 points need to be stored. This large reduction in required storage space makes storage required for Table-Look-Up comparable to the storage requirement for the 5th and 7th order equations.

A comparison between the results from the 5th order equation and 7th order equation is shown in Figure 4. The 5th order equation produced an approximate ± 0.2ppm frequency variation. The 5th order equation response is slower than the faster changing voltage versus temperature curve of the crystal oscillator. The 7th order equation had no problems following the voltage versus temperature curve and appeared to slightly smooth the output. Specifically the points around -30°C to -15°C and at 0°C on the 7th order equation output do not contain the sudden jumps that the Table-Look-Up trace exhibits.

The most disturbing problem encountered was the 8°C periodic ripple discovered in the 5th order, 7th order and Table-Look-Up outputs. (In the Table-Look-Up trace, this ripple was swamped out on the cold side by other errors but was noticeable in the 20°C to 40°C range.) This problem appeared to be caused by small temperature variations between the thermistor and crystal. Presently, the thermistor is located about ¼ inch away from the crystal. Both the crystal and thermistor are embedded in an insulating material. In order to improve this situation, a different packaging arrangement is required to allow the thermistor to track the crystal temperature more accurately.

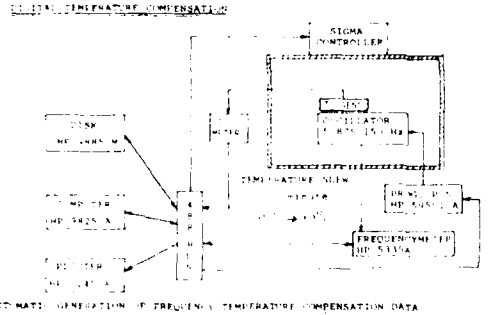
Conclusion

The ability to improve the frequency deviation from ± 0.05ppm to ± 0.01ppm depends on two factors. First, the crystal oscillator must not have any anomalies; spurious responses cannot be corrected with the techniques presented here. Secondly, the thermistor, or other temperature sensing device, must track the crystal temperature more accurately.

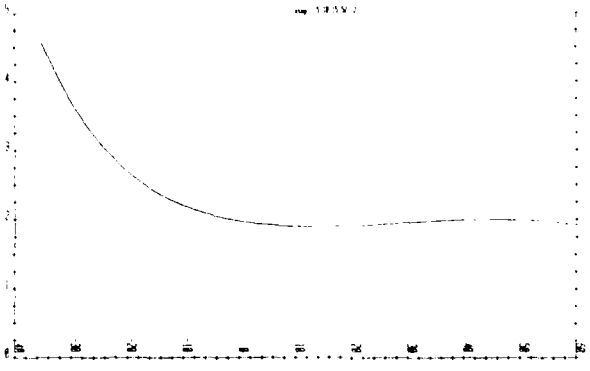
The tests have shown that production cycles of 20 to 35 crytals per 4 hour test cycle are feasible. The Table-Look-Up approach can handle 20 crystals per test cycle, and the 7th order equation compensation can handle 35 crystals per test cycle. These production numbers correspond to the maximum number of collectible data points divided by the number of required data points.

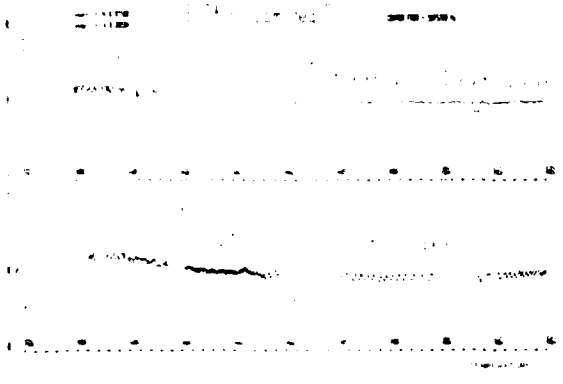
When comparing the 7th order equation and the Table-Look-Up methods, the 7th order equation was found to be smoother and produced excellent results. However, if the slope of the characteristic temperature curve of another crystal is greater than the one used for these tests, a higher order equation may be needed to maintain the desired accuracy. The Table-Look-Up does not have this problem and, therefore, has the advantage of compensating dissimilar crystals in the same pass.

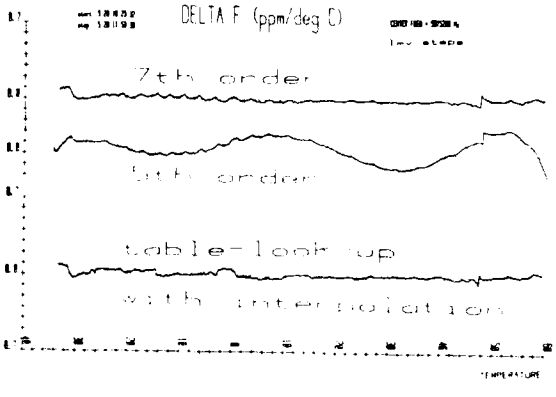
We would like to thank T. Wagner and T. Snowden of General Electric, Neutron Division, and Dr. J. Vig of the United States Army Electronics Research and Development Command (ERADCOM) for their technical support in building this test station. This investigation was partially funded by the Navy and ERADCOM.

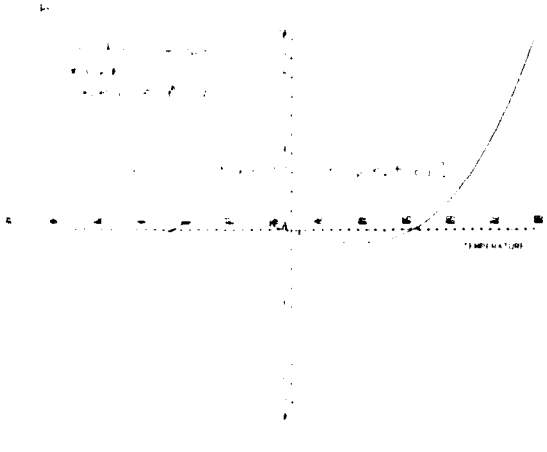


AUTOMATIC GENERATION OF FREQUENCY TEMPERATURE COMPENSATION DATA
AUTOMATIC CURVE FITTING COMPUTATION (LINEAR REGRESSION UP TO TEN ORDERS)
AUTOMATIC PLAYBACK AND X-Y PLOTTING OF BOTH DIGITAL COMPENSATION METHODS









Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

CHARACTERIZATION OF FREQUENCY FLUCTUATIONS BY CROSSCORRELATION AND BY USING THREE OR MORE OSCILLATORS

J. Rovera-Gerard, G. Petit, M. Olivier, J.J. Gagnepain

Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25030 Besançon - France

Summary

The individual stability of an oscillator can be checked by using two auxiliary oscillators and by operating the three oscillators two by two. The direct application of this principle gives erroneous results, if the three comparisons are not made simultaneously. However the results are affected by the measurement system noises. New methods are proposed for comparing three oscillators. They use the concepts of crossvariance and crosscorrelation. They enable to characterize individually the oscillators and they reduce the influence of system noises, when uncorrelated. Therefore better resolution can be achieved in the frequency or phase fluctuation measurements.

Introduction

To measure the frequency fluctuations of an oscillator, either in time domain or in frequency domain, needs a reference, which is in fact also another active oscillator. The result of the measurement is a combination of the fluctuations of the oscillator under test and of the reference. In time domain, one obtains the sum of the variances. This is a drawback, especially when we compare two oscillators, whose frequency fluctuations are of the same order of magnitude.

To obtain independantly the fluctuations of each oscillator, the "triangulation" method has been proposed ; it consists in using a third oscillator to perform three comparisons of the oscillators two by two. But this method gives erroneous results, because it supposes that the three variances, at the three comparisons, are measured exactly and not simultaneously.

The method which is presented for measurements of short term stability uses simultaneously the outputs of the three oscillators and the triangulation relations to obtain after reduction of the data, the variance of each oscillator. However as in all cases, the precision of the measurement is still affected by two phenomenon. On the one hand, the finite number of samples, which are processed, leads to a statistical dispersion of the results. It is necessary to determine the confidence interval of the experimental variances to enhance the precision. On the other hand, the calculation of the Allan variance of the beat frequency fluctuations between two oscillators is equivalent to an autocorrelation method, therefore the result involves entirely the noise of the measurement system ; this limits also the precision of the measure.

To avoid these difficulties two other methods are proposed. A new algorithm is introduced. It is called "crossvariance", and it is very similar to the well known crosscorrelation function. One of its advantages is to reject the influence of the measurement system noises. Direct application of crosscorrelation can also be made. Both methods are equivalent. The differences only appear in the measurement systems and in the signal processing.

Comparisons of three oscillators with Allan variance

Principle : three oscillators a, b, c are compared two by two as shown on fig. 1. Their frequency fluctuations are enhanced by three frequency difference multipliers. The output of each multiplier delivers a beat frequency centered at 1 MHz. The frequency fluctuations are multiplied by 10^n, where n is an integer which can be chosen from 0 to 9. The three frequencies are measured with three counters, which are simultaneously operated. After statistical average the three variances σ^2_{ab}, σ^2_{bc} and σ^2_{ac} are calculated on a computer. Each of these quantities involve the contributions of the two compared oscillators. Therefore they are directly related to the individual variances of the fractional frequency fluctuations of each oscillator σ^2_a, σ^2_b and σ^2_c.

$$\sigma^2_{ab}(\tau) = \sigma^2_a(\tau) + \sigma^2_b(\tau)$$

$$\sigma^2_{bc}(\tau) = \sigma^2_b(\tau) + \sigma^2_c(\tau)$$

$$\sigma^2_{ca}(\tau) = \sigma^2_c(\tau) + \sigma^2_a(\tau)$$

It is important to note that these experimental variances are not true variances but only estimated variances, because of the finite number of samples taken into account in the statistical average.

Let ν_a, ν_b and ν_c be the mean frequencies of the oscillators and $\Delta\nu_a(t)$, $\Delta\nu_b(t)$, $\Delta\nu_c(t)$ their fluctuations. The estimated Allan variance obtained by comparison of oscillator a and b is

$$\sigma^2_{ab}(\tau) = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} [(y^a_{i+1} - y^b_{i+1}) - (y^a_i - y^b_i)]^2 \qquad (1)$$

where τ is the time lag between two samples, m is the finite number of samples and y_i is the fractional frequency fluctuation averaged over a time τ

$$y^a_i = \frac{1}{\tau} \int_{t_i}^{t_i+\tau} \frac{\Delta\nu_a(t)}{\nu_a} dt \qquad (2)$$

The subscript a refers to oscillator a.

Eq. (1) can be written

$$\sigma^2_{ab} = \sigma^{2(ab)}_a + \sigma^{2(ab)}_b = \frac{1}{2(m-1)} \sum_i (\Delta^a_i - \Delta^b_i)^{ab} \qquad (3)$$

where the contribution $\sigma^{2(ab)}_a$ and $\sigma^{2(ab)}_b$ of each oscillator appear. The $^{(ab)}$ superscript is added to indicate that the individual variance σ^2_a was obtained in the (ab) comparison. And $\Delta^{a,ab}_i = (y^a_{i+1} - y^a_i)^{ab}$.

By similar the results of the two other comparisons one can evaluate the quantity :

$$\sigma^2_a = \sigma^2_{ab} + \sigma^2_{ac} - \sigma^2_{bc} \qquad (4)$$

which is equivalent to

$$\sigma^2_a = \sigma^{2(ab)}_a + \sigma^{2(ac)}_a + \sigma^{2(ab)}_b - \sigma^{2(bc)}_b + \sigma^{2(ac)}_c - \sigma^{2(bc)}_c - \frac{1}{m-1} \sum_i (\Delta^a_i \Delta^b_i)^{(ab)} + (\Delta^a_i \Delta^c_i)^{(ac)} - (\Delta^b_i \Delta^c_i)^{(bc)} \qquad (5)$$

If the simultaneity of measurements is not satisfied the terms $\sigma^{2(ab)}_a$ and $\sigma^{2(ac)}_a$ or $\sigma^{2(ab)}_b$ and $\sigma^{2(bc)}_b$ calculated from samples y^a_i and y^b_i taken at different times may be quite different. The estimated variance σ^2_a cannot be obtained with any confidence and after reduction of data results let sometimes appear negative values of the variances.

But if the condition of the simultaneity of the measurements is verified :

$$\sigma^{2(ab)}_a = \sigma^{2(ac)}_a = \sigma^2_a$$

$$\sigma^2_a = \frac{1}{2(m-1)} \sum_i (\Delta^a_i \Delta^b_i)^{(ab)} + (\Delta^a_i \Delta^c_i)^{(ac)} - (\Delta^b_i \Delta^c_i)^{(bc)} \qquad (6)$$

However if the noise sources of the oscillators are independant the last term go to zero (generally as $m^{-1/2}$) when the number m of samples is increased. The uncertainty due to this last term can be evaluated for commonly encountered spectra of frequency fluctuations[1].

On fig. 2 is illustrated the influence of the simultaneity of the measurements. In table I are given the results of a series of the three simultaneous comparisons between the three oscillators (the three comparisons were repeated three times). For simultaneous measurements the data are left in table II in the order they appear in table I. By using the three simple relations the individual variances of each oscillator are calculated and given in table III. In this case it appears that the results are coherent and the dispersion rather small. But if the comparisons are not simultaneous, and this was artificially obtained by using in table II combinations of the results of the measurements the calculated individual variances have very large dispersion and sometimes even negative values appear.

Influence of spurious noise sources

The beat frequency between oscillators a, b is affected by a frequency noise $n^{ab}(t)$ which can be due to spurious noise sources in the electronics of the measurement system. Taking this noise into account one obtains

$$\sigma^2_{ab} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} [(y^a_{i+1} - y^b_{i+1} + n^{ab}_{i+1}) - (y^a_i - y^b_i + n^{ab}_i)]^2 \qquad (7)$$

where

$$y^a_i - y^b_i + n^{ab}_i = \frac{1}{\tau} \int_{t_i}^{t_i+\tau} [\frac{\Delta\nu_a}{\nu_a} - \frac{\Delta\nu_b}{\nu_b} + n^{ab}(t)] dt \qquad (8)$$

The development of relation (7) lets appear an autocorrelation of the noise $n^{ab}(t)$. This term can be denoted as $\sigma^{2(ab)}_n$ with :

$$\sigma^2_n(ab) = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} [n^{ab}_{i+1} - n^{ab}_i]^2 \qquad (9)$$

Such a quantity limits the precision in the determination of the variance of each oscillator. One obtains also crosscorrelation terms due to interaction between the noise $n^{ab}(t)$ and the noise of the oscillators. They can be denoted as $\sigma^{2\,(ab)}_{an}$.

$$\sigma^{2\,(ab)}_{an} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (y^a_{i+1} - y^a_i)(n^{ab}_{i+1} - n^{ab}_i) \qquad (10)$$

with the hypothesis of independance of the noises of the oscillators and of the measurement system, these terms will disappear when a very large number of samples is used. For a finite number of samples, they contribute to the dispersion of the measured variances.

Crossvariance method

Principle : a new algorithm is proposed for calculating the variance. This so-called "crossvariance" is equivalent to crosscorrelation func-

tion and it will be shown that this method enables to reduce the influence of the different noise sources of the measurement system and to reach directly the individual variance of each oscillator.

The two-sample-crossvariance is defined by analogy with the Allan variance

Allan variance :

$$\sigma_\alpha^2(2,T,\tau) = \langle_i(\frac{y^\alpha_{i+1}-y^\alpha_i}{2})^2\rangle \qquad (11)$$

Crossvariance :

$$\sigma^2_{\alpha\beta}(2,T,\tau) = \langle_i(y^\alpha_{i+1}-y^\alpha_i)(y^\beta_{i+1}-y^\beta_i)\rangle \qquad (12)$$

This crossvariance is calculated by taking the statistical average of the product of samples obtained from two simultaneous comparisons.

An estimate of this crossvariance with regard to two signal α and β is given by averaging over a finite number m samples :

$$\sigma^2_{\alpha\beta}(2,T,\tau) = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (y^\alpha_{i+1}-y^\alpha_i)(y^\beta_{i+1}-y^\beta_i) \qquad (13)$$

Using three oscillators one has to introduce the crossvariance $\sigma^2_{ab,cd}$ relative to the beat frequency between oscillators a and b and the beat frequency between oscillators c and b. The estimation is given by the relation :

$$\sigma^2_{ab,cb} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} [(y^a_{i+1}-y^a_i)-(y^b_{i+1}-y^b_i)]^{(ab)} \qquad (14)$$
$$[(y^c_{i+1}-y^c_i)-(y^b_{i+1}-y^b_i)]^{(cb)}$$

where a, b and c refer to the three oscillators and (ab), (cb) to the two comparisons. This can also be written

$$\sigma^2_{ab,cb} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (\Delta^a_i)^{ab}(\Delta^c_i)^{cb} - (\Delta^a_i)^{ab}(\Delta^b_i)^{cb} \qquad (15)$$
$$- (\Delta^b_i)^{ab}(\Delta^c_i)^{cb} + (\Delta^b_i)^{ab}(\Delta^b_i)^{cb}$$

Using the hypothesis of simultaneity, the last term represents an estimation of the Allan variance of the oscillator b, that is σ^2_b. The other terms of the development are associated to the crossvariance between oscillators a and c, a and b, and, b and c

$$\sigma^2_{ab,cb} = \sigma^2_b + \sigma^2_{ac} - \sigma^2_{ab} - \sigma^2_{bc} \qquad (16)$$

This globaly reduces to σ^2_b if the oscillators are independant. Therefore one obtains directly the individual variance of oscillator b, without any additional reduction of data.

Influence of spurious noise sources :

The two beat frequencies between oscillators a and b and oscillators c and b are affected by noises $n^{ab}(t)$ and $n^{cb}(t)$. Cautions must be taken in the realization of the experimental device to achieve independance of these two noise sources.

One obtains the crossvariance between the two beat frequencies :

$$\sigma^2_{ab,cb} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} [(y^a_{i+1}-y^a_i)-(y^b_{i+1}-y^b_i)+n^{ab}_{i+1}-n^{ab}_i]^{(ab)} \qquad (17)$$
$$[(y^c_{i+1}-y^c_i)-(y^b_{i+1}-y^b_i)+n^{cb}_{i+1}-n^{cb}_i]^{cb}$$

The development of this relation gives three different terms :

a) the previous term of interest which is the crossvariance between the two beat frequencies in the absence of additive noise.

b) terms of crossvariance between one oscillator and the noise of the measurement system

$$\sigma^{2(cb)}_{a,n} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (y^a_{i+1}-y^a_i)(n^{cb}_{i+1}-n^{cb}_i) \qquad (18)$$

The hypothesis of independance between these two kind of noises will give a nonvanishing result only for a small number of samples.

c) a term of crossvariance between the two measurement system noises :

$$\sigma^2_{n^{(ab)},n^{(cb)}} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (n^{ab}_{i+1}-n^{ab}_i)(n^{cb}_{i+1}-n^{cb}_i) \qquad (19)$$

which is also reduced.

Therefore in the hypothesis of independance of the noise sources of the different parts of the total system, the crossvariance $\sigma^2_{ab,cb}$ gives a good estimate of the variance of the oscillator (b). In fact the different noise sources remain more or less correlated, but in any case the influence of the measurement system noise sources is strongly reduced as shown on fig. 3, especially if the number of samples is large. This is possible for short term stability measurements. In these experiments the integration times are $10^{-1}s<\tau<100s$. Thus about 200 samples can be stored without excessive duration of the total measure. For both methods the measurement system noises of fig. 3 decrease as τ^{-2}. This limitation is due to the frequency difference multiplier noises. The resolution could be improved by using in these frequency difference multipliers better voltage controlled oscillators.

This crossvariance method was used with three oscillators, the frequency fluctuations of which were artificially increased in order to make these oscillators less stable than a fourth oscil-

lator of good quality, which in this case could be used as an absolute reference. On fig. 4 are shown the results of stability measurements on the three oscillators by using the crossvariance measurement system with simultaneous comparisons, or by using Allan variance method and direct comparison with the more stable reference oscillator. Both methods are in good agreement and this demonstrates the validity of the crossvariance method.

Fig. 5 presents the results of stability measurements on three good quality quartz oscillators by means of the crossvariance method.

Crosscorrelation method

The three oscillators a, b and c are still used. The method consists in locking oscillators a and c on oscillator b with phase lock loops as shown on fig. 6. The phase lock loops act as phase demodulators for Fourier frequencies outside of the bandwidth of the loop. Consequently the error voltages $V_{ab}(t)$ and $V_{cb}(t)$ are respectively proportional to the instantaneous phase difference between oscillators a and b, and, c and b.

Then one proceeds to the calculation of the crosscorrelation function between $V_{ab}(t)$ and $V_{cb}(t)$. In fact one obtains again only an estimation of the true correlation function : because of a finite integration time T :

$$C_{ab,cb}(\tau) = \frac{1}{T}\int_{t}^{t+T} [\phi_a(t) - \phi_b(t)]^{(ab)} [\phi_c(t-\tau) - \phi_b(t-\tau)]^{(cb)}\, dt \qquad (20)$$

The development of this expression yields terms of crosscorrelation of the form

$$C_{\alpha\beta}(\tau) = \frac{1}{T}\int_{t}^{t+T} \phi_\alpha(t)\, \phi_\beta(t-\tau)\, dt \qquad (21)$$

which go to zero when T goes to infinity, with the hypothesis of independant phase fluctuations of the three oscillators, and a term which is proportional to the autocorrelation function $C_{bb}(\tau)$ of the phase fluctuations of oscillator b.

$$C_{bb}(\tau) = \frac{1}{T}\int_{t}^{t+T} \phi_b(t)\, \phi_b(t-\tau)\, dt \qquad (22)$$

The variance of the frequency fluctuations directly follows[4]

$$\sigma^2_{\phi b}(\tau) = \frac{2}{\tau^2}\,[C_{bb}(0) - C_{bb}(\tau)] \qquad (23)$$

If necessary the phase noise spectrum can also be obtained by Fourier transform of the autocorrelation function[4].

The influence of the spurious noise sources in the crosscorrelation method is very similar to what was described in the crossvariance method : one introduces in the crosscorrelation function $C_{ab,cb}(\tau)$ the noises $n_{(ab)}$ and $n_{(cb)}$ which perturb the signals at the output of the phase lock loops :

$$C_{ab,cb}(\tau) = \frac{1}{T}\int_{t}^{t+T} [\phi_a(t) - \phi_b(t) + n_{(ab)}(t)]^{(ab)} [\phi_c(t) - \phi_b(t) + n_{(ab)}(t)]^{(cb)}\, dt \qquad (24)$$

After development, three kinds of noise terms appear :

- crosscorrelation of the phase noises of the oscillators, as previously

$$C_{\alpha\beta}(\tau) = \frac{1}{T}\int_{t}^{t+T} \phi_\alpha(t)\, \phi_\beta(t-\tau)\, dt \qquad (25)$$

- crosscorrelation of the phase noises of one oscillator and of the measurement system :

$$C_{\alpha,n_{(\beta\gamma)}}(\tau) = \frac{1}{T}\int_{t}^{t+T} \phi_\alpha(t)\, n_{(\beta\gamma)}(t-\tau)\, dt \qquad (26)$$

- crosscorrelation of the noises of the two measurement systems :

$$C_{n_{(ab)}n_{(cb)}}(\tau) = \frac{1}{T}\int_{t}^{t+T} n_{(ab)}(t)\, n_{(cb)}(t-\tau)\, dt \qquad (27)$$

Here again no autocorrelation term of noises exists and therefore with the hypothesis of independant noise sources, all the crosscorrelation terms cancell with unfinite or at least large integration time. Unfortunatly the noise sources of the systems are not completly decorrelated. Nevertheless a 10 dB improvement can be easily obtained with this method with respect to the regular autocorrelation methods. This is shown on fig. 7, where are presented the phase noise spectra of the measurement system, obtained after Fourier transform.

Conclusion

This study shows that comparisons between three (or more) oscillators cannot be correctly made without simultaneity of the three measurements. And even with this requirement the measured variances still are to be considered as random functions because of the finite number of frequency samples taken into account in the statistical average.

The use of a crossvariance method or of a crosscorrelation method enables on the one hand to determine the individual stability of each oscillator, and not as with the previous methods a combination of frequency instabilities of two oscillators, and on the other hand to reduce the influence of the measurement system noise and therefore to improve the resolution.

This opens the possibility of characterizing a good frequency standard by comparing it to two other oscillators of lower qualities.

References

(1) J.E. Gray, D. Allan, "A method for estimating the frequency stability of an individual oscillator", 28th AFCS, (1974).

(2) F.L. Walls, S.R. Stein, J.E. Gray, D.J. Glaze, "Design considerations on state of the art signal processing and phase noise measurement systems", 31th AFCS, (1977).

(3) P. Lesage, C. Audoin, "Characterization of frequency stability : uncertainty due to the finite number of measurements", IEEE Trans. Instrum. and Meas., IM-22, n° 2, 157-161, (1973).

(4) L.S. Cutler, C.L. Searle, "Some aspects of the theory and measurement of frequency fluctuations in frequency standards", Proc. IEEE, vol. 54, 136-154, (1966).

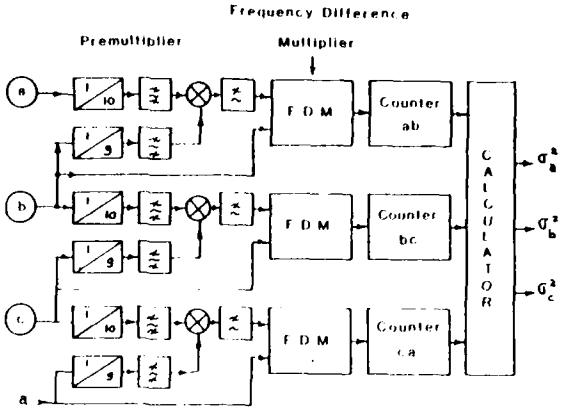


Fig. 1

Crossvariance measurement system

measurements	1	2	3	Time
σ^2_{ab}	1326	1090	1264	
σ^2_{bc}	129	143	125	
σ^2_{ca}	1213	967	1161	

Table I

combinations of non simultaneous measurements | simultaneous measurements

σ^2_{ab}	1326	1326	1090	1090	1264	1264	1326	1090	1264
σ^2_{bc}	143	125	129	125	129	143	129	143	125
σ^2_{ca}	1161	967	1161	1213	967	1213	1213	967	1161

Table II

$$\begin{cases} \sigma^2_a = \dfrac{\sigma^2_{ab} + \sigma^2_{ca} - \sigma^2_{bc}}{2} \\ \sigma^2_b = \dfrac{\sigma^2_{bc} + \sigma^2_{ab} - \sigma^2_{ca}}{2} \\ \sigma^2_c = \dfrac{\sigma^2_{ca} + \sigma^2_{bc} - \sigma^2_{ab}}{2} \end{cases}$$

calculated values of the individual variance of each oscillator

σ^2_a	172	1084	1061	1089	1051	1167	1205	957	1150
σ^2_b	154	242	29	5	215	97	121	133	114
σ^2_c	-11	-47	100	124	-84	47	8	10	11

Table III

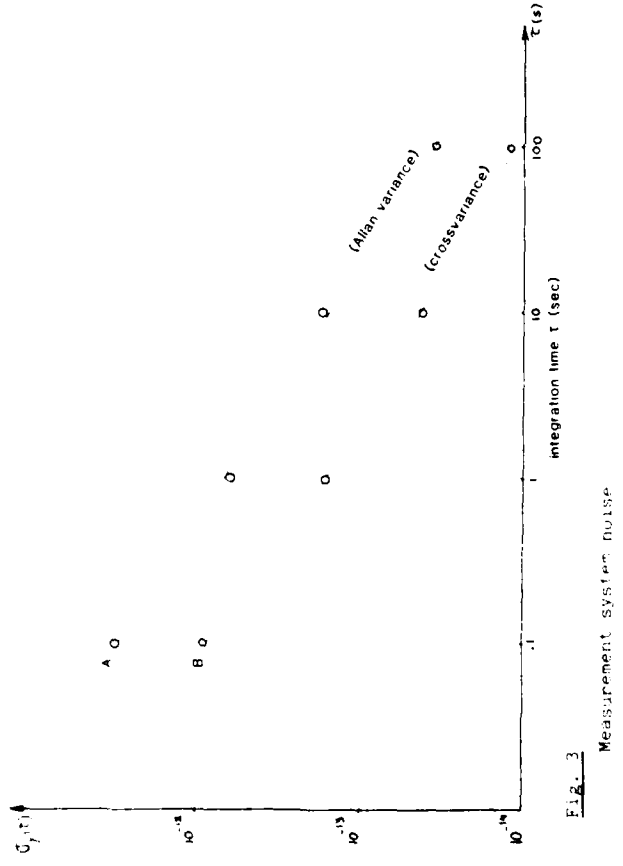


Fig. 3

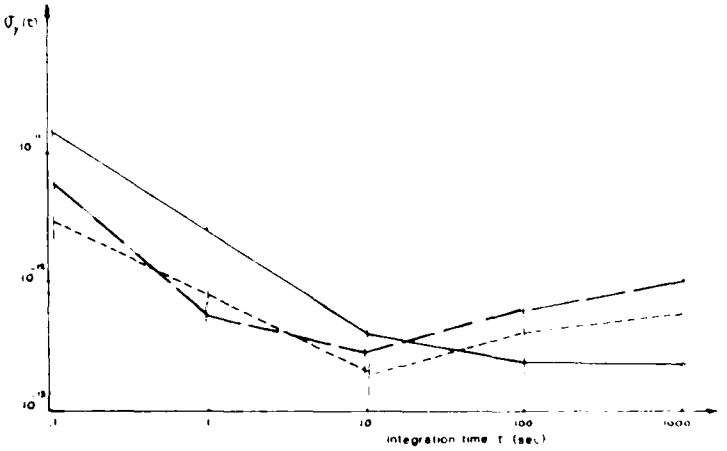


Fig. 5

Measurement of the stability of three oscillators with the crossvariance method

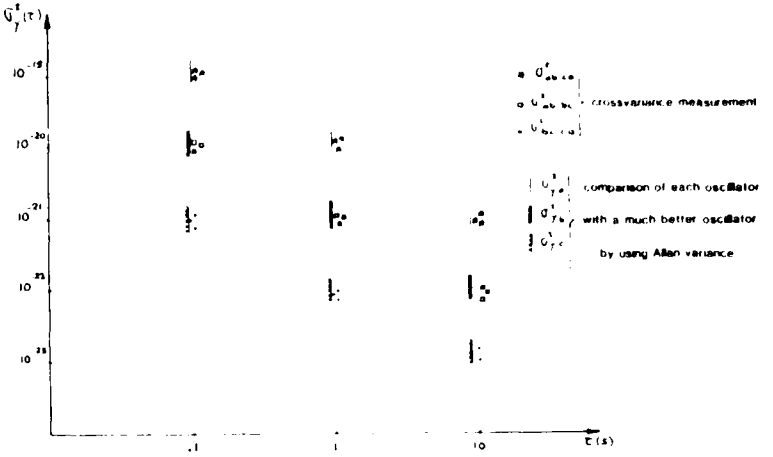


Fig. 4

Measurement of the stability of three oscillators with the crossvariance method or by comparison with a better reference oscillator

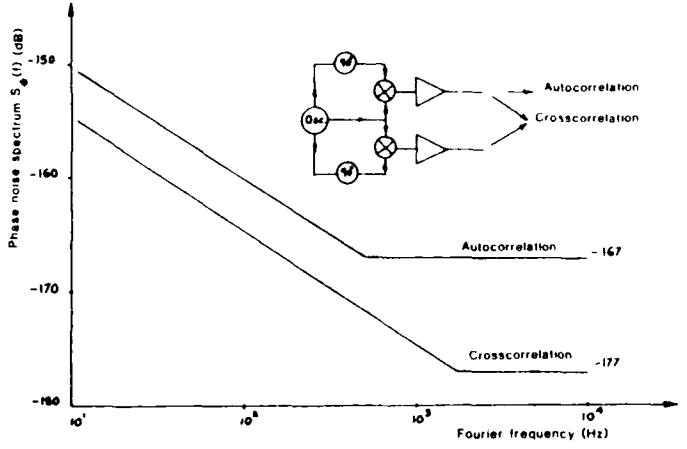


Fig. 7

Phase noise spectra of the autocorrelation and crosscorrelation measurement systems

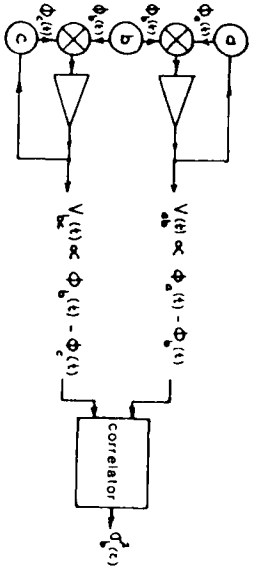


Fig. 6

Crosscorrelation measurement system

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, 07703, May 1981

AN ULTRA-HIGH RESOLUTION FREQUENCY METER

J. J. Snyder

Center for Absolute Physical Quantities
National Bureau of Standards
Washington, D. C. 20234

Summary

A novel instrument for measuring the frequency of a periodic signal contaminated by phase noise is presently being developed at the National Bureau of Standards. This frequency meter averages overlapping time intervals using a simple algorithm implemented with standard logic circuits. Because of the signal-to-noise improvement inherent in the averaging process, the standard deviation of a single measurement contaminated by white phase noise is proportional to $\tau^{-1.5}$, where τ is the time for the measurement. In contrast, the measurement uncertainty using a standard frequency counter is proportional to τ^{-1} in the white phase noise regime[1]. For many potential applications of the frequency meter, the measurement uncertainty due to contaminating noise may thereby be reduced by several orders of magnitude in comparison with a measurement over the same time interval using presently available instruments.

Introduction

For this analysis, we assume the test signal to be given by

$$V(t) = V_0 \sin[2\pi\nu t + \phi(t)] \qquad (1)$$

where ν is the nominal frequency of the test signal, and where we explicitly neglect amplitude noise and assume the phase noise $\phi(t)(|\phi| << 1)$ to be white. The reference or clock signal at ν_0 is assumed to be noise free, and is taken to be the higher of the two frequencies. The reference frequency drives a clock counter whose value at any instant defines a time scale (or equivalently, a phase scale) t. The phase of the test frequency is used to derive a set of time values $\{t_i\}$ from the clock counter, where t_i is, e.g., the time of the i^{th} positive-going zero crossing of the test signal. In the absence of noise, the frequency ratio of the two signals is the slope of t_i vs. i. The effect of white phase noise in the test signal is to introduce uncorrelated errors in $\{t_i\}$ as shown in Figure 1.

Background

A conventional high resolution counter[2] counts the clock for a time interval τ, defined as a predetermined number (n-1) of cycles of the test signal. If the clock counter was initially cleared, then at the end of the time interval τ, it contains the value N, and the frequency of the test signal is given by

$$\nu = \frac{n-1}{N}\,\nu_0 \qquad (2)$$

This method is equivalent to estimating the slope of the data in Figure 1 by drawing a straight line through the first and last points. If the clock quantization error of ±1 count is the dominant source of white phase noise, then the normalized frequency uncertainty equals the numerical resolution

$$\frac{\delta\nu_Q}{\nu} = \frac{1}{N} \simeq \frac{1}{\nu_0\tau} \qquad (3)$$

Alternatively, if the measurement is limited by white phase noise in the test signal, the normalized uncertainty is

$$\frac{\delta\nu_N}{\nu} = \frac{\delta\phi}{2\pi N} \simeq \frac{\delta\phi}{2\pi\nu_0\tau} \qquad (4)$$

where the phase fluctuation amplitude is $\delta\phi$. For both situations, the measurement uncertainty is proportional to τ^{-1}.

Frequency Meter Algorithm

The frequency meter incorporates a simple hardware digital filter based on an algorithm previously used to find zero-slope points of interferometer fringes[3], and similar to an algorithm independently proposed by Gamlen[4]. As shown schematically in Figure 2, the data $\{t_i\}$, (i=1 to 2n) are divided into n overlapping time intervals $\{\Delta t_i\}$, each of length n cycles of the test signal where n is a predetermined value. Thus each of the n intervals is approximately half the length of the total interval. For white phase noise, the n intervals may be considered independent and their average length is given as

$$\langle N \rangle = \frac{1}{n}\sum_{i=1}^{n} \Delta t_i \qquad (5)$$

The uncertainty in the computed average is reduced by the averaging process as $\sqrt{n}$ in comparison with the uncertainty in each interval. The frequency of the test signal is computed similarly to Eq. 2:

$$\nu = \frac{n}{\langle N \rangle}\,\nu_0 \qquad (6)$$

Implementation

The algorithm may be implemented with commercially available TTL logic circuits. As shown in Figure 3, a clock at frequency ν_0 ($\nu_0 \gtrsim 20$ MHz) drives a counter which defines accumulated time. Strobe pulses derived from the test signal ($\nu \gtrsim 1$ MHz) are used to sequentially latch the instantaneous values of the clock counter without interrupting its operation. During the total measurement time $\tau \approx (2n-1)/\nu$, the strobed values of the clock counter are first subtracted from an initially cleared digital accumulator for the first n cycles of the test signal and then added to the accumulator for the last n cycles. At the end of the measurement interval τ, the accumulator contains the value

$$A = -\sum_{i=1}^{n} t_i + \sum_{i=n+1}^{2n} t_i = \sum_{i=1}^{n} \Delta t_i = n\langle N\rangle \qquad (7)$$

where t_i is the instantaneous value of the clock counter at the time of the i^{th} strobe pulse, and Δt_i is the length of the i^{th} interval in units of $1/\nu_0$. The test frequency is found from Eqs. (6) and (7) to be

$$\nu = \frac{n\nu_0}{A} \qquad (8)$$

The numerical resolution of the frequency meter is $1/A = 1/n\langle N\rangle$, where $\langle N\rangle$ is the average value of the n overlapping intervals. The uncertainty due to clock quantization noise is

$$\frac{\sigma_Q}{\nu} = \frac{1}{\langle N\rangle\sqrt{n}} = \frac{2}{\nu_0\tau}\sqrt{\frac{2}{\nu\tau}} \qquad (9)$$

and the uncertainty due to white phase noise in the signal is

$$\frac{\sigma_N}{\nu} = \frac{\sigma_\phi}{2\pi\langle N\rangle\sqrt{n}} = \frac{2\sigma_\phi}{2\pi\nu_0\tau}\sqrt{\frac{2}{\nu\tau}} \qquad (10)$$

From Eqs. (3) and (4), we see that for the same measurement interval $\tau \approx (2n-1)/\nu$, the frequency meter reduces the uncertainty due to clock quantization or other white phase noise by $\sim\sqrt{n}/2$ in comparison with conventional frequency counting techniques. In addition, unlike the situation with conventional counters the numerical resolution is a factor of $\sqrt{n}$ finer than the clock quantization uncertainty. The benefit of this very high numerical resolution is that quantization noise in a series of measurements will appear to be continuous rather than discrete.

Digital Filter Analysis

A conventional frequency counter that measures only the phase difference over the measurement interval τ may be described as a digital filter consisting of a negative-going delta function at t=0 and a positive-going delta at $t=\tau$. In the frequency domain, this filter has a simple $\sin(\omega)$ response; for small ω the filter differentiates the accumulated phase to give the frequency whereas for large ω the response oscillates between $\pm$ some constant. The time-domain character of the digital filter used in the frequency meter is a single cycle of a square wave of amplitude one and period τ. As shown in Figure 4, this filter has a frequency response of $\sin(\omega)^2/\omega$, and therefore also acts as a simple differentiator for small ν, but in addition is a low-pass filter with a cut-off at $\omega \sim 1/\tau$. It is the reduction in the high-frequency phase noise that is responsible for the improved resolution of the frequency meter.

There are a number of interesting extensions of these very simple digital techniques that involve convolutions of $\sin(\omega)$ (phase difference) filters with $\sin(\omega)/\omega$ (phase accumulation) filters that give additional stages of high-frequency phase noise reduction and/or different degrees of differentiation at the origin. For example, the $\sin(\omega)^3/\omega$ filter of Figure 5 gives directly the first difference of the frequency. This filter has been shown by Allan and Barnes[5] to resolve the ambiguity in the standard definition of the Allan variance between white phase and flicker phase noise in an oscillator.

Experimental Results

A preliminary version of the frequency meter has been tested by measuring the frequency of a 21 kHz signal generated by a low-phase-noise synthesizer[6] in comparison with its 10 MHz internal reference. Both the Allan variance for a conventional frequency counter and a similar variance for the frequency meter were computed using measurement times τ ranging from $\sim 10^{-3}$ to $\sim 10^{3}$ seconds. In Figure 6 sigma, the square root of the conventional Allan variance, has a slope of τ^{-1} due to the white phase noise of the clock quantization error. Sigma as measured by the frequency meter has a slope of $\tau^{-1.5}$ out to about 1 second where it changes to τ^{-1}. As shown by Allan and Barnes[5], the slope of τ^{-1} identifies a regime dominated by flicker phase noise. The abrupt change of the slope to τ^{+1} at about 100 seconds is due to a frequency noise peak with a period of about one hour and amplitude of ~ 3 microhertz. This frequency error is shown in Figure 7. Its source is unknown but is believed due to temperature cycling within the synthesizer.

A hardware implementation of the frequency meter is under construction using standard TTL logic. This instrument will use clock frequencies above 20 MHz to measure frequencies as high as 1 MHz. An immediate application[7] of this instrument will be to the development of interferometric laser wavelength comparators with resolution exceeding a part in 10^{11}.

References:

[1] J.A. Barnes et al., "Characterization of Frequency Stability," NBS Technical Note 394, October 1970.

[2] For example, the Hewlett Packard Model 5345A. The use of company or brand names is for illustrative purposes only and does not imply any product endorsement by the National Bureau of Standards.

[3]J.J. Snyder, "Algorithm for Fast Digital Analysis of Interference Fringes," Appl. Opt. 19, 1223 (1980).

[4]M.W. Gamlen, to be published.

[5]D.W. Allan and J.A. Barnes, "A Modified 'Allan Variance' with Increased Oscillator Characterization Ability," 35th Proceedings of the Annual Symposium on Frequency Control, U.S. Army Electronics Command, Ft. Monmouth, NJ. Copies available from Electronic Industries Association, 2001 "I" Street, NW, Washington, D.C. 20006.

[6]Fluke Model 644A Frequency Synthesizer. The use of company or brand names is for illustrative purposes only and does not imply any product endorsement by the National Bureau of Standards.

[7]J.J. Snyder, T. Baer, L. Hollberg and J.L. Hall, presented at the Conference on Lasers and Electro-Optics, Washington, D.C., June 1981.

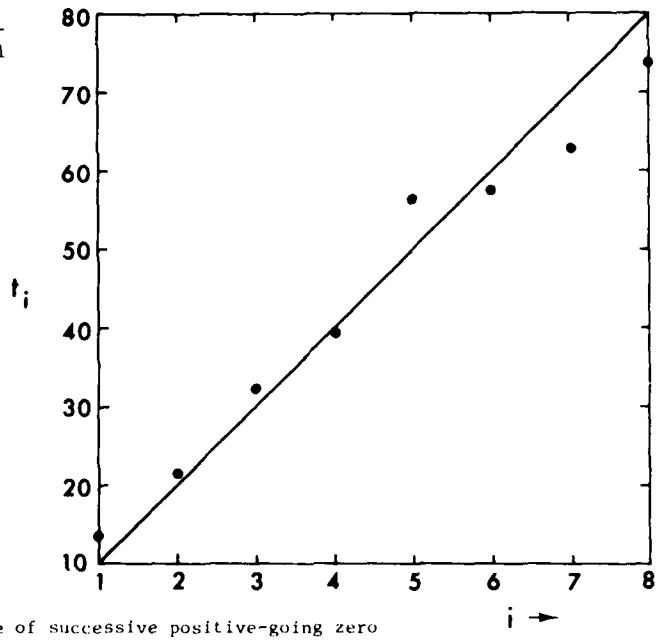


Figure 1: Plot of accumulated time of successive positive-going zero crossings of a simulated test signal. The clock frequency is 10 times the test signal frequency, and the test signal has white phase noise with standard deviation $\delta\phi = 0.6$.

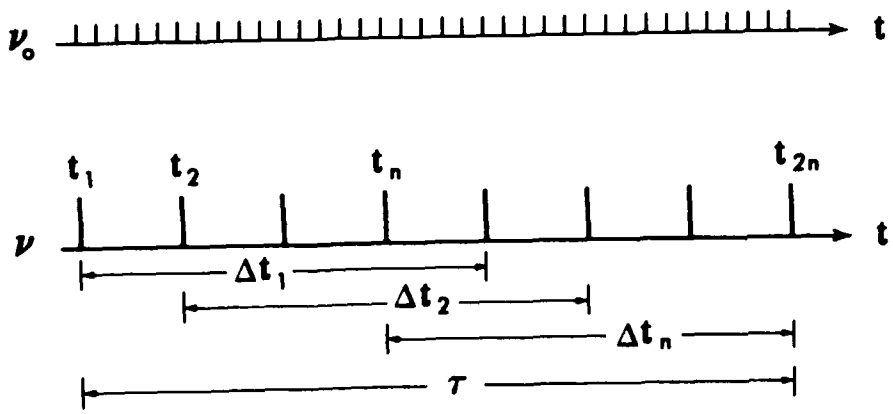


Figure 2: Characterization of frequency meter algorithm. The time scale is furnished by a counter counting the clock frequency ν_o. The total measurement interval $\tau = (2n-1)/\nu$ is subdivided into n overlapping intervals $\{\Delta t_i\}$.

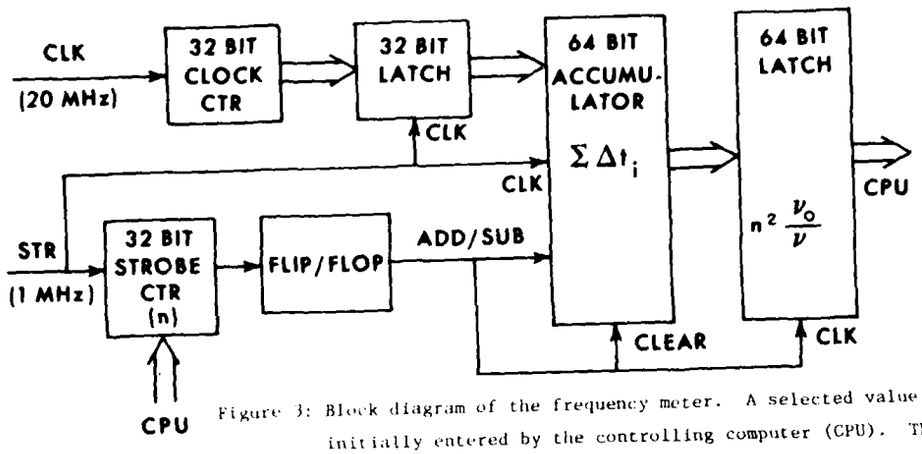


Figure 3: Block diagram of the frequency meter. A selected value for n is initially entered by the controlling computer (CPU). The value of the accumulator at the end of the measurement cycle is read by the computer and converted into a measurement of the strobe (STR) frequency using the known value of the clock (CLK) frequency.

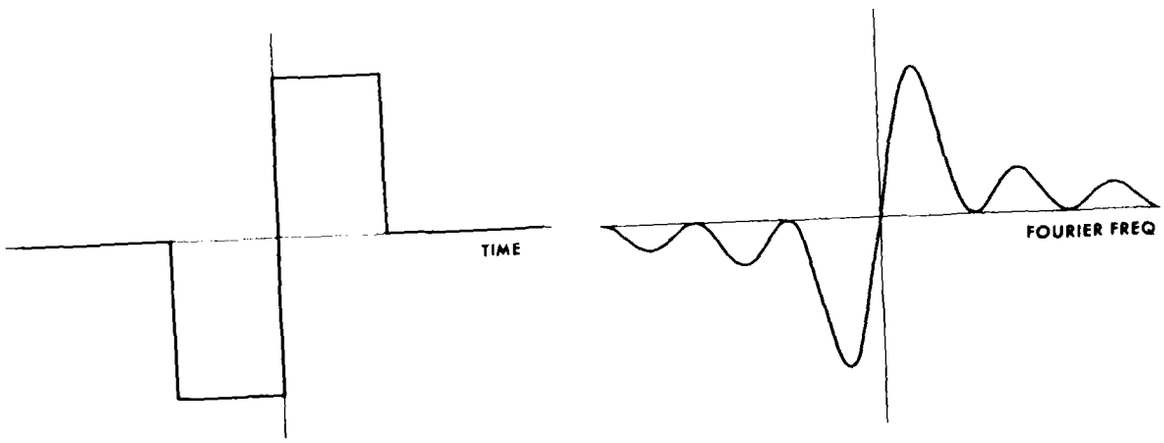


Figure 4: Transform pair for the frequency meter. The impulse response (a) is a single cycle of a square wave, implying a frequency response (b) of $\sin(\omega)^2/\omega$.

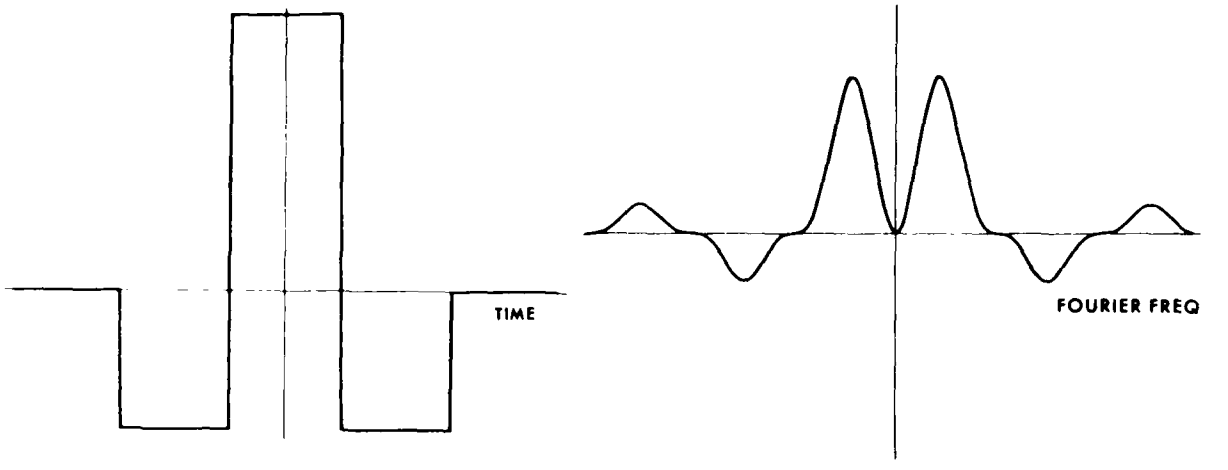


Figure 5: Transform pair for a frequency difference (variance) filter. This filter distinguishes between white phase and flicker phase noise[6]. Impulse response (a) and frequency response (b).

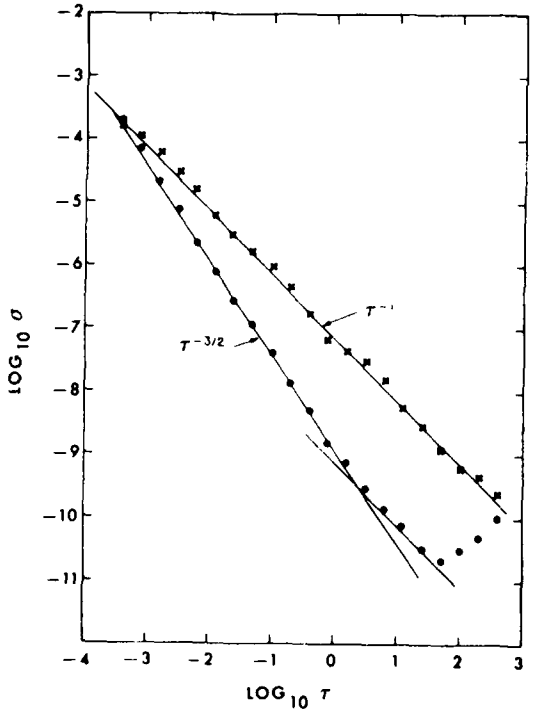


Figure 6: Frequency stability of a synthesized 21 kHz signal compared with its internal reference at 10 MHz. Sigma, the square root of the conventional Allan variance (crosses) decreases as τ^{-1} for white or flicker phase noise. Sigma for the frequency meter (dots) decreases as $\tau^{-1.5}$ for white phase noise and as τ^{-1} for flicker phase noise[7]. The increase in sigma for τ > 100 seconds is due to a frequency noise peak with a period one hour (see Figure 7). Instrumental dead time 0.1 seconds.

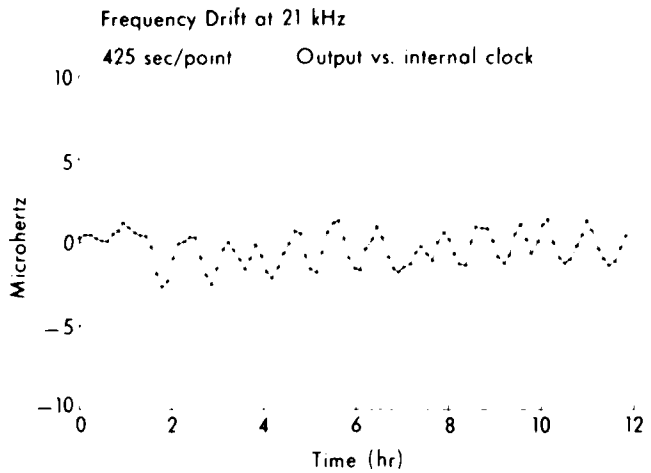


Figure 7: Frequency drift of a synthesized 21 kHz signal compared with its internal reference at 10 MHz. Source of the drift is unknown.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

A MODIFIED "ALLAN VARIANCE" WITH INCREASED OSCILLATOR CHARACTERIZATION ABILITY

David W. Allan and James A. Barnes

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

Summary

Heretofore, the "Allan Variance," $\sigma_y^2(\tau)$, has become the de facto standard for measuring oscillator instability in the time-domain. Often oscillator frequency instabilities are resonably modelable with a power law spectrum: $S_y(f) \sim f^{\alpha}$, where y is the normalized frequency, f is the Fourier frequency, and α is a constant over some range of Fourier frequencies. It has been shown that for power law spectrum $\sigma_y^2(\tau) \sim \tau^{\mu}$, and that $\mu = -\alpha-1$ for $-3<\alpha< +1$, where τ is the nomimal sample time over which each value of y is measured. The modified "Allan Variance" developed in this paper yields $\mu = -\alpha-1$ for all α in the range $-3<\alpha$, which removes the previous ambiguity: $\mu = -2$ for $+1<\alpha$. In other words, with the modified "Allan Variance" one can easily distinguish between white phase noise ($\alpha = +2$) and flicker phase noise ($\alpha = +1$) -- commonly occurring for the short term instabilities of quartz crystal oscillators and active hydrogen masers.

Key Words. Flicker Noise; Frequency Stability; Oscillator Noise Modeling; Power Law Spectrum; Time-Domain Stability; White Noise.

Introduction

The random fluctuations in precision oscillators may often be characterized by a power law spectrum:

$$S_y(f) = h_\alpha f^\alpha \quad , \tag{1}$$

where y is the normalized frequency deviation, f is the Fourier frequency, h_α is the intensity of the particular noise process, and α is constant over some range of f. The typical values of α are +2 (white noise phase modulation, PM); +1 (flicker noise PM); 0 (white noise frequency modulation, FM); -1 (flicker noise FM); and -2 (random walk FM). The Allan variance, as it has come to be known,[1] has been demonstrated as a very useful statistical tool for characterizing these various random processes with the exception that if α = +1 or +2, the dependence on τ is nominally the same, i.e., $\sim \tau^{-2}$. It is not at all uncommon for white PM and flicker PM to occur in precision oscillators for τ of the order of one second and shorter. The modified Allan variance, as developed in this paper, depends as τ^{-2} for α = +1 and as τ^{-3} for α = +2. This yields a clear distinction in the time domain between these heretofore somewhat ambiguous processes.

Definition of "Allan Variance" and Related Concepts

Define y, the normalized frequency deviation, as

$$y(t) = \frac{\nu(t) - \nu_o}{\nu_o} \quad , \tag{2}$$

where $\nu(t)$ is the output frequency of the oscillator being studied, and ν_o is nominally the same frequency, but of a reference oscillator assumed for the moment without loss of generality to be better than the test oscillator. The time deviation from some arbitrary origin (t = 0) is the integral of the frequency deviations (from that origin):

$$x(t) = \int_0^t y(t') \cdot dt' \tag{3}$$

The $i^{\underline{th}}$ average frequency deviation over an interval, τ, is

$$\bar{y}_i = \frac{x_{i+1} - x_i}{\tau} \tag{4}$$

where the assumption is made that the time deviation measurements are nominally spaced τ apart.

The "Allan Variance" is defined as:

$$\sigma_y^2(\tau) = \tfrac{1}{2} \langle (\bar{y}_{i+1} - \bar{y}_i)^2 \rangle \ , \qquad (5)$$

where the brackets "< >" denote infinite time average. Using equation (4), one may write:

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2} \langle (x_{i+2} - 2x_{i+1} + x_i)^2 \rangle \ . \qquad (6)$$

It has been shown that typically $\sigma_y^2(\tau)$ varies as τ^{μ}, and that $\mu = -\alpha - 1$ for $-3 \leq \alpha \leq +1$.[1,2] Hence, we see one of the dimensions of usefulness of $\sigma_y^2(\tau)$, i.e., ascertaining the dependence on τ allows an estimate of α (the power law spectral type of noise). However, if $\alpha \geq +1$, then $\mu \cong -2$, and the τ dependence becomes somewhat ambiguous as to the type of noise in this region. It is interesting to note that in the region $\alpha \geq +1$, $\sigma_y^2(\tau)$ is bandwidth (f_h) dependent; i.e., the bandwidth of the measurement system will affect the value of $\sigma_y(\tau)$, and furthermore, one may use the bandwidth dependence[3] to determine the value of α (see also Appendix Ref. 2).

Development of the Modified Allan Variance

One may also write $\sigma_y^2(\tau)$ in terms of a generalized autocovariance function:

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2} [4U_x(\tau) - U_x(2\tau)] \ , \qquad (7)$$

where

$$U_x(\tau) = 2[R_x(0) - R_x(\tau)], \qquad (8)$$

and where

$$R_x(\tau) = \langle x(t+\tau) \cdot x(t) \rangle \ , \qquad (9)$$

the classical autocovariance function of x(t). Using the Fourier transforms of generalized functions, one may determine the coefficients relating the power spectral density to $\sigma_y^2(\tau)$. Ref. 1 gives these relationships. It is of interest to note that $U_x(\tau)$ has the following approximate form in the region $\alpha \geq +1$ (see Appendix Ref. 2):

$$U_x(\tau) \sim a(\alpha) \left[\left(\frac{1}{2\pi f_h} \right)^{-\alpha+1} - \tau^{-\alpha+1} \right] \qquad (10)$$

Hence, one notes that by changing the reciprocal bandwidth as well as τ, one affects $\sigma_y^2(\tau)$ in similar ways, depending on the value of α. From this, one should be able to deduce the value of α, since the bandwith dependence becomes stronger for α moving positive from +1, and the τ dependence becomes stronger as α moves negative from +1. One can change the bandwidth in the hardware or in the software. In the past, it has typically been done in the hardware.[3] James Snyder[4] has shown that it is relatively easy to change the bandwidth in the data processing by a clever technique and we have followed his lead. In particular, we have chosen a new variance analysis scheme which coincides with the Allan variance at the minimum sample time, τ_o, (i.e., minimum data spacing), but which changes the bandwidth in the software as the sample time, τ, is changed.

Each reading of the time deviation, x_i, has associated with it an intrinsic nominal (hardware) measurement system bandwidth, f_h. Define $\tau_h = \frac{1}{2\pi f_h}$; and similarly we may define a software bandwidth, $f_s = f_h/n$, which is 1/n times narrower than the hardware bandwidth. This software bandwidth can be realized by averaging n adjacent x_i's; $\tau_s = n\tau_h$, where $\tau_s = 1/f_s$. We have defined a modified Allan variance which allows the reciprocal software bandwidth to change linearly with the sample time, τ:

$$\text{Mod } \sigma_y^2(\tau) = \frac{1}{2\tau^2} \left\langle \left[\frac{1}{n} \sum_{i=1}^{n} (x_{i+2n} - 2x_{i+n} + x_i) \right]^2 \right\rangle \qquad (11)$$

where $\tau = n\tau_o$. Eq. 11 clearly coincides with Eq 6 for n = 1. One can see that, in general, we have formed a second difference of three time readings with each of the three being an average of n of the x_i's (with non-overlapping averages). As n increases, the (software) bandwidth decreases and this bandwidth varies just as $f_s = f_h/n$.

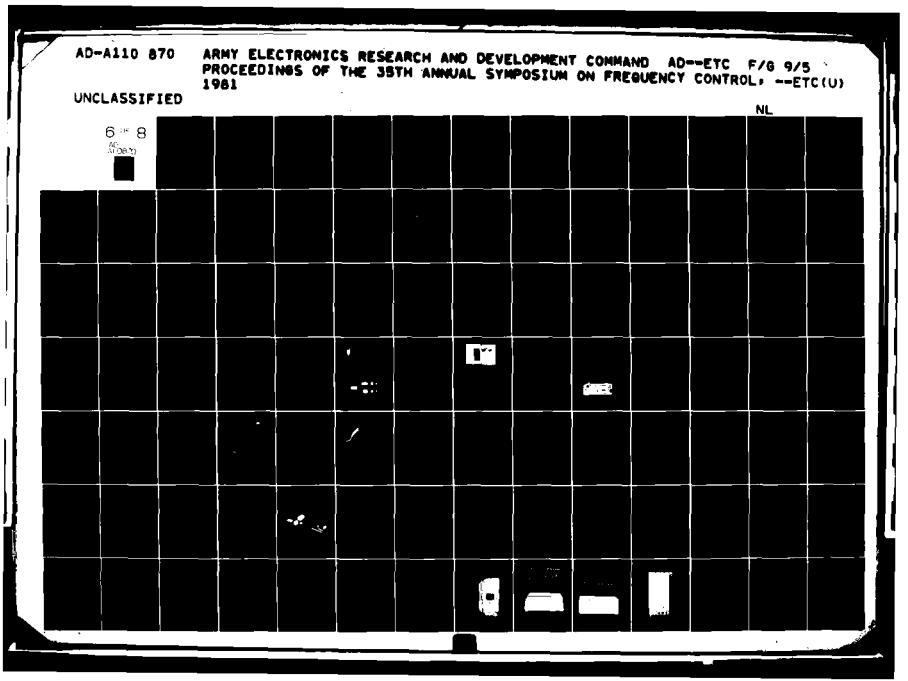
For a finite data set of N readings of x_i (i = 1 to N), we may write an estimate

AD-A110 870 ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND AD--ETC F/G 9/5
PROCEEDINGS OF THE 35TH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL, --ETC(U)
1981

UNCLASSIFIED

NL

6 OF 8
AD
A110870



$$\text{Mod } \sigma_y^2(\tau) = \frac{1}{2\tau^2 n^2 (N-3n+1)} \cdot \sum_{j=1}^{N-3n+1} \left[\sum_{i=j}^{n+j-1} (x_{i+2n} - 2x_{i+n} + x_i) \right]^2 \quad (12)$$

Eq. 12 is easy to program, but takes more time to compute than for $\sigma_y(\tau)$. This is only of significance for the smaller computer or handheld calculator.

Comparisons, Tests, and Examples of Usage of the Modified Allan Variance

We simulated various power law noise processes, and applied Eq. 12. Shown in Fig. 1 are the resulting τ-dependences of the modified Allan variances for α = -2, -1, 0, +1, and +2. The solid lines drawn are the anticipated or theoretical slopes for the particular noise process. One sees excellent agreement for white noise PM and for flicker noise PM, and nominal agreement for the others.

One can express Eq. 11 in terms of the generalized autocovariance function:

$$\text{Mod } \sigma_y^2(\tau) = \frac{1}{2\tau^2 n^2} \left\{ [4U_x(\tau) - U_x(2\tau)] \cdot n + \sum_{i-1}^{n-1} (n-i)[-6U_x(i\tau_o) + 4U_x((n+i)\tau_o) + 4U_x((n-i)\tau_o) - U_x((2n+i)\tau_o) - U_x((2n-i)\tau_o)] \right\} . \quad (13)$$

In the range $-3 \leq \alpha \leq +1$, one may write:

$$U_x(\tau) = a(\alpha) \cdot \tau^{-\alpha+1} , \quad (14)$$

which when substituted in Eq. 13, and using Eq. 7, yields

$$\text{Mod } \sigma_y^2(\tau) = \sigma_y^2(\tau) \left\{ \frac{1}{n} + \frac{1}{n^2 \, 4n^{-\alpha+1} - (2n)^{-\alpha+1}} \cdot \sum_{i=1}^{n-1} (n-i) \cdot \left[-6i^{-\alpha+1} + 4(n+i)^{-\alpha+1} - (2n+i)^{-\alpha+1} + 4(n-i)^{-\alpha+1} - (2n-i)^{-\alpha+1} \right] \right\} \quad (15)$$

Since we know that $\sigma_y^2(\tau)$ is well behaved in this range and $\mu = -\alpha - 1$, it is of interest to look at the ratio:

$$R(n) = \text{Mod } \sigma_y^2(\tau) / \sigma_y^2(\tau) .$$

As stated before, at $n = 1$ ($\tau = \tau_o$) the ratio is unity. One can evaluate Eq. 15 with a computer. A reasonable empirical fit may be formed, which is good to 0.5% or better of Eq. 15:

$$R(n) = \frac{q + pn^E - p}{qn^E} \quad (16)$$

which approaches p/q asymptotically as n approaches infinity, and is within 1% of p/q for $n \geq 8$. Listed in Table 1 are the empirical values of p, q, and E and the quality of fit for the appropriate power law noise processes.

TABLE 1

Noise Type Name	α	p	q	E	fit	Mod $\sigma_y(\tau)/\sigma_y(\tau)$ $\tau \gg \tau_o$
White FM	0	1	2	2	perfect	.707
Flicker FM	-1	99.9	148	2.35	½%	.821
Random Walk FM	-2	33	40	2.35	<½%	.908
Flicker Walk FM	-3	1	1	--	perfect	1

The results of Table 1 are in reasonable agreement with simulated results of Fig. 1(a) through 1(e). The last row in Table 1, "flicker walk" frequency modulation, is out of the range of applicability of α, but the ratio, R(n), is still convergent.

The $U_x(\tau)$ function for flicker noise PM is extremely complicated and has not been developed, but one can arrive at an empirical value for it. The $U_x(\tau)$ function is derivable for the other power law spectral processes. Table 2 gives the relationships between the time domain measure Mod $\sigma_y^2(\tau)$ and its power law spectral counterpart, given in Eq. 1. Also listed in the right hand column of Table 2 are the asymptotic values of $R(n)$:

TABLE 2

Noise Type Name	α	Mod $\sigma_y^2(\tau)$	Comment	$R(n)$ $n \to \infty$
White PM	+2	$h_2 \cdot \frac{3 f_h}{(2\pi)^2 n \tau^2}$	Exact	1
Flicker PM	+1	$h_1 \cdot \frac{1.038 + 3\ln(\omega_h \tau)}{(2\pi)^2 \tau^2}$	Empirical	1
White FM	0	$h_0 \cdot \frac{R(n)}{2\tau}$	Exact	0.5
Flicker FM	-1	$h_{-1}\ 2\ln(2) \cdot R(n)$	Empirical; Exact Available	0.674
Random Walk FM	-2	$h_{-2} \cdot \frac{(2\pi)^2 \tau}{6} \cdot R(n)$	Empirical; Exact Available	0.824

It is clear from Table 2 that Mod $\sigma_y^2(\tau)$ is very useful for white PM and flicker noise PM, but for $\alpha < +1$ the conventional Allan variance, $\sigma_y^2(\tau)$, gives both an easier-to-interpret and an easier-to-calculate measure of stability.

It is interesting to make a graph of α versus μ for both the ordinary Allan variance and the modified Allan variance. Shown in Fig. 2 is such a graph. This graph allows one to determine power law spectra for non-interger as well as interger values of α. The dashed line for the modified Allan variance has been intentionally moved to the left in Fig. 2 because for small values of n the value of μ will appear to be slightly more negative that for $\sigma_y^2(\tau)$, even though for large n, they both approach the same slope (i.e., the same values of μ). In fact, in the asymtotic limit, the equation relating μ and α for the modified Allan variance is

$$\alpha = -\mu - 1, \text{ for } -3 < \alpha < +3 . \qquad (17)$$

The value of $\mu = -4$ for $\alpha = +3$ was verified empirically with simulated data, and it appears that for $\alpha > +3$, μ remains at -4.

A direct application for using the modified Allan variance recently arose in the analysis of atomic clock data as received from a Global Positioning System (GPS) satellite. We were interested in knowing the short-term characteristics of the newly developed, high-accuracy NBS/GPS receiver, as well as the propagation fluctuations. Fig. 3 shows both $\sigma_y^2(\tau)$ and Mod $\sigma_y^2(\tau)$ for comparison. Using Mod $\sigma_y^2(\tau)$, we can tell that the fundamental limiting noise process involved in the system is white noise PM with the exciting result that averaging for four minutes can allow one to ascertain time difference to better than one nanosecond excluding other systematic effects.

Conclusion

We have developed a supplemental measure, the "Modified Allan Variance" (Mod $\sigma_y^2(\tau)$), which has very useful properties when analyzing oscillator or signal stability in the presence of white noise phase modulation or flicker noise phase modulation. *It also works reasonably well as a stability* measure for other commonly occuring noise processes in precision oscillators.

We would recommend that for most time domain analysis, $\sigma_y^2(\tau)$ should be the first choice. If $\sigma_y^2(\tau)$ depends on τ as τ^{-1}, then the modified Allan variance can be used as a substitute to help remove the ambiguity as to the noise processes.

Acknowledgments

The authors are deeply grateful for stimulating discussions with Dr. James J. Snyder, and for the useful comments of Dr. Robert Kamper and Mr. David A. Howe.

References

1. James A. Barnes, Andrew R. Chi, Leonard S. Cutler, Daniel J. Healey, David B. Leeson, Tomas E. McGunical, James A. Mullen, Jr., Warren L. Smith, Richard L. Sydnor, Robert F. C. Vessot, and Gernot M. R. Winkler, Proc. IEEE Trans. Instrum. Meas. IM-20, 105 (1971); also published as NBS Tech. Note 394 (1971).
2. David W. Allan, Proc. IEEE 54, 221 (1966).
3. R. F. C. Vessot, Proc. IEEE 54, 199 (1966).
4. James J. Snyder, Proc. 35th Annual Symp. on Freq. Control (1981), to be published.

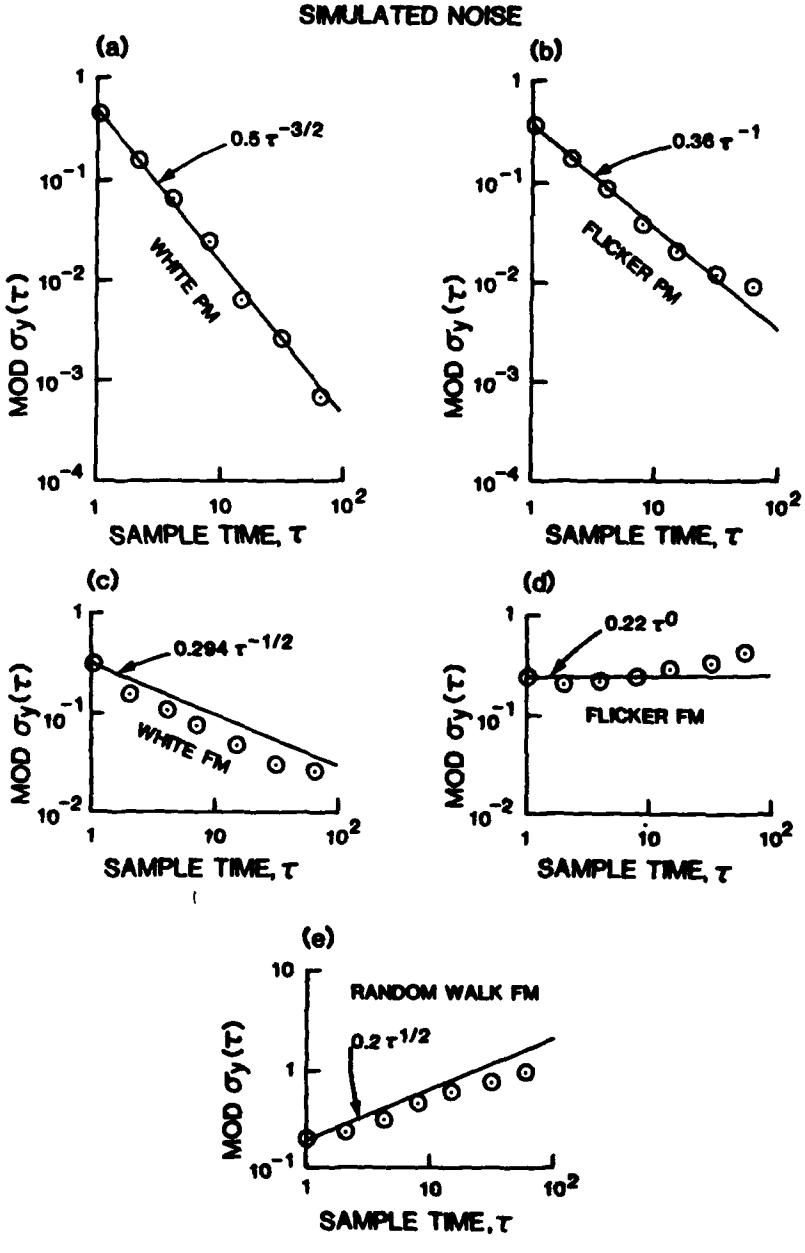


Fig 1a-e. Mod $\sigma_y(\tau)$ using Eq. 12 was calculated for different sample times for independently generated and simulated noise processes, which were white phase noise, flicker phase noise, white frequency noise, flicker frequency noise, and random walk frequency noise, respectively. Mod $\sigma_y(\tau)$ was computed for 399 data points in each case. One sees the excellent fit to the theory for white phase noise and flicker phase noise, an important new contribution in the ability to characterize oscillators having these noise processes.

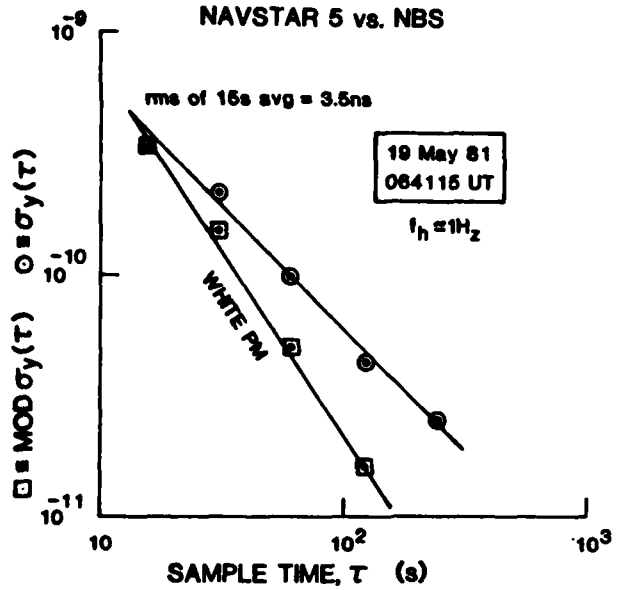


Fig. 2. A plot showing the mapping of the α vs. μ, both for the standard Allan variance and the modified Allan variance, where α is the exponent for the power law spectral density on the frequency fluctuations and μ is the exponent on the sample time, τ. Mod $\sigma_y(\tau)$ nominally follows the equation $\mu = -\alpha-1$, even for $\alpha < +1$, thus removing the previous ambiguity for power law processes where in this region.

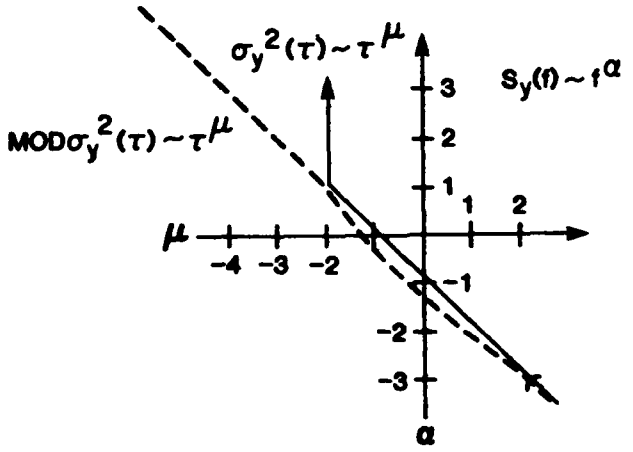


Fig. 3. An application of Mod $\sigma_y(\tau)$ analyzing an NBS/GPS receiver output showing that it is white noise FM, whereas $\sigma_y(\tau)$ would not allow one to ascertain directly the value of α, i.e., the power law spectral noise type.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

RELATION BETWEEN 1/f NOISE AND Q-FACTOR IN QUARTZ RESONATORS AT ROOM AND LOW TEMPERATURES, FIRST THEORETICAL INTERPRETATION

J.J. Gagnepain, J. Uebersfeld, G. Goujon
Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté-Besançon
32, av. de l'Observatoire - 25000 Besançon - France

and

P. Handel
Physics Department, University of Missouri, St Louis, MO 63121

Summary

The ultimate stability of quartz crystal oscillators is given by the frequency noise of the quartz resonator, which corresponds in the time domain to a floor. The measurement of the resonance frequency fluctuations of the resonator itself can be made only if the influence of the electronics noise is reduced. By using a π transmission network and a balanced phase bridge the driving source phase fluctuations can be rejected by 50 dB to 60 dB. This measurement system was used at room temperature with a large number of pairs of resonators, and a correlation between 1/f frequency noise levels and Q-factors was found, following a $1/Q^4$ law.

In order to extend the measurements to very low temperatures, a new measurement system was studied, which allows to operate with pairs of non identical resonators. Thus 1/f frequency noise was measured at 4K and at 1K. A large decrease of the noise is observed corresponding to the increase of the Q-factors.

A theoretical explanation of the $1/Q^4$ law is attempted. It is based on three-phonon interaction processes, which are at the origin of the acoustic attenuation and of variations of velocity. Both phenomena are related by means of the thermal phonon relaxation time.

Introduction

1/f noise is a fascinating universal phenomenon, which as a rule appears in almost all physical systems, as well in the natural ones as in electronic devices. It is generally agreed upon, that its origin and its mechanisms still remain not well understood.

Quartz crystal resonators do not escape from 1/f noise, and the ultimate stability of oscillators will be given by the level of the 1/f fluctuations of the resonator's resonance frequency.

With an oscillating loop it is difficult to distinguish in the output signal the respective contributions of the electronics (amplifiers, buffers, ...) and of the quartz crystal resonator to the phase fluctuations. A new technique was presented by F. Walls[1], which allowed to measure the inherent frequency fluctuations of a quartz crystal in a passive π transmission network with a large rejection of the noise usually associated with electronics. But as it will be shown this interesting technique requires pair of resonators at the same frequencies. Therefore it is not adapted for measurements at very low temperatures. Nevertheless this technique is appropriate to measurements at room temperature, and was used to study a large number of crystals.

Frequency noise measurements at room temperature

The measurement system which has been used by F. Walls is represented on fig. 1. A frequency source with a good spectral purity drives in transmission two crystals, as identical as possible. A phase bridge composed of a double balanced mixer and a 90° phase shifter delivers a voltage proportional to the relative phase fluctuations of the two transmitted signals. By careful adjustments of the resonator frequencies and Q-factors the phase bridge is balanced in order to reject the phase fluctuations of the source by 50 dB to 60 dB.

Measurements were performed on pairs of identical resonators in a frequency range from 1 MHz to 25 MHz, including different technologies of fabrication, various sizes and shapes of electrodes and mountings. Typical frequency noise spectra are shown on fig. 2. These results already were presented at the FCS[2]. At lower frequencies a $1/f^2$ frequency random walk spectrum appears. It can be attributed to temperature fluctuations ΔT following the relation $\Delta f/f_o = a\,\Delta T/\Delta t$ where a is the dynamic temperature coefficient[3-4]. This noise can be minimized by using very accurate thermostats[5].

The levels of 1/f noise measured at 1 Hz from the carrier, are plotted as a function of the Q-factors on fig. 3.

By linear regression with logarithmic scale curves were fitted to those experimental points, following the relation between the fractional frequency power density spectrum, the Q-factor and the Fourier frequency

$$S_y(f) = \frac{\lambda}{Q^n} \cdot \frac{1}{f} \qquad (1)$$

The values obtained for the level coefficient λ and for the Q-exponent n depend on the selected experimental points among the available values. After mathematical treatment it was obtained that $3.8 < n < 4.3$. Therefore it is proposed to consider that the most probable value will be n=4. However a small variation of n value induces a drastic change of the level coefficient λ. This is not surprising on account of the number of decades involved in the Q-factor and frequency scales. For the previous range of n values the corresponding values of λ are $0.1 < \lambda < 63$. A reasonable fitting would correspond to λ close to unity when n is an integer equal to 4. Therefore the averaged phenomenological law which is proposed is

$$S_y(f) = \frac{1}{Q^4} \cdot \frac{1}{f} \qquad (2)$$

This is the first and only one law which was found between 1/f noise levels and resonator parameters. Attempts were made to correlate 1/f noise and the resonance frequency, but without decisive success.

As pointed out these data were collected from different quartz crystal units, and therefore it is difficult to assume that even at constant frequency, only one parameter -the Q-factor- was modified. The acoustic attenuation of sound waves in a crystal being temperature dependant, the second step of the study consisted in cooling down to liquid helium temperature a pair of resonators and comparing the noise levels at the different temperatures as a function of the Q-factors.

Frequency noise measurements at low temperatures

A typical curve of the quartz crystal internal losses (inverse Q-factor) as a function of temperature is represented on fig. 4. The peak of maximum loss which appears at 20K is due to phonon-phonon interactions. Below this temperature the losses decrease and therefore large increase of Q-factors can be obtained.

The resonance frequency is also changed, as shown on table I.

crystal units	f(300°K - f(4.2°K)
1	5 502 Hz
2	6 025 Hz
3	5 589 Hz
4	5 730 Hz
5	5 644 Hz
6	5 773 Hz

Table I : Frequency differences between room and liquid helium temperatures of different 5 MHz 5 th overtone resonators

Between room temperature and liquid helium temperature the total variation is of the order of 5-6 KHz for 5 MHz AT-cut resonators, but not sufficiently reproductible from one resonator to another one. This means that two resonators with identical frequencies at room temperature, will present frequency differences, when at 4K, which can be as large as a few 100 Hz. This is due to small differences of the crystal blank orientations. But such a frequency gap is too large to be cancelled out by means of tuning capacitors.

Thus, a new measurement system adapted to pairs of resonators with different frequencies was studied. Its schematic diagram is represented on fig. 5. Each of the two resonators is driven by one of the side bands of a low noise signal,which is amplitude-modulated. The modulation frequency f is obtained from a frequency synthesizer, whose frequency is divided by a factor 4000 in order to purify its spectrum. At the output of the first mixer the intermediate frequency is affected by the total frequency noise of the two resonators. The influence of the driving source phase noise is cancelled out by balancing the Q-factors of the resonators. A second mixing with a 2f reference frequency delivered by a digital phase shifter coupled to the divider gives at its output a noise voltage proportional to the resonator's frequency fluctuations. A filter cancells out the 4f frequency component. A FFT spectrum analyzer gives directly the $S_y(f)$ spectrum of the two resonators. One another advantage of such a system is to enable easier and faster calibrations, just changing by a small amount the modulation frequency and measuring the corresponding DC voltage shift at the mixer output.

The 1/f noise of a pair of 5 MHz - 5th overtone - AT-cut quartz resonators was measured at 300K, 4.2K and 1K. The difference between their resonance frequencies at 4.2K was equal to 1 KHz. The noise levels reduced to one resonator are given on table II.

Temperature	300°K	4.2°K	1°K
Q-factor	2.4×10^6	5×10^6	9×10^6
S_y(1Hz)	1.6×10^{-24}	1.4×10^{-25}	3.5×10^{-26}

Table II
1/f frequency noise levels
as a function of temperature

These results confirm the dependance of 1/f noise versus Qfactor. The data do not strictly follow the $1/Q^4$ law. This will be commented in the next section.

Theoretical approach

The lattice anharmonicities, due to the non-linear nature of the interatomic bonds and the corresponding three-phonon interaction processes give to the crystal the properties of thermal expansion and finite thermal conductivity[6].

Calculation of the attenuation can be made by using two different theoretical methods. In the Landau-Rumer theory[7] the sound wave is considered as a beam of phonons which are scattered by collisions with thermal phonons. This theory does not take into account interactions between thermal phonons and is valid only in the low temperature range, when the phonon life time is large enough and therefore the energy and momentum well defined. At higher temperatures the life time becomes shorter. The uncertainties of energy and momentum increase and the selection rules arising from energy and momentum conservation laws break down[8]. In this case, following the theory of Akhieser[8] the sound wave is represented macroscopically in the classical form. The thermal phonon system is disturbed from equilibrium by the sound wave strain, but tends to return to equilibrium because of the collisions. The phonons are localized and their mean free path has to be small when compared to the sound wave length.

$$\hbar\Omega \ll k_B T \tag{3}$$

where Ω is the angular frequency of the sound wave and T the absolute temperature. $\hbar$ and k_B respectively are the Planck and Boltzmann constants. A detailed description of the method can be found in ref. 9. Only the main usefull results will be summerized here.

The thermal energy of the crystal of volume V is the sum of the discrete energies stored in each normal mode over all the normal modes k and branches j of the lattice

$$E_{th} = \sum_{kj} [N(kj) + \tfrac{1}{2}] \hbar\,\omega(kj) \tag{4}$$

N(kj) is the number of phonons with frequency $\omega(kj)$ in the mode kj.

The mechanical potential energy due to the strains induced by the sound wave is

$$E_{sound} = C_{\alpha\beta}\,\eta_{\alpha\beta} + \tfrac{1}{2}\,C_{\alpha\beta\gamma\delta}\,\eta_{\alpha\beta}\,\eta_{\gamma\delta} + \cdots \tag{5}$$

$C_{\alpha\beta}$ and $C_{\alpha\beta\gamma\delta}$ have the dimension of elastic constants.

The strain derivative of the total energy gives the stress

$$T_{\alpha\beta} = (\partial/\partial\eta_{\alpha\beta})\,(E_{th} + E_{sound}) \tag{6}$$

When in equilibrium the number of phonons in mode kj is given by the BoseEinstein distribution

$$n(kj) = (\exp[\hbar\omega(kj) / k_B T] - 1)^{-1} \tag{7}$$

and this equilibrium corresponds to $T_{\alpha\beta} = \eta_{\alpha\beta} = 0$. Therefore

$$T_{\alpha\beta} = C_{\alpha\beta\gamma\delta}\,\eta_{\gamma\delta} + \frac{1}{V}\sum_{kj}[N(kj)-n(kj)]\hbar\,\frac{\partial\omega(kj)}{\partial\eta_{\alpha\beta}} + \frac{1}{V}\sum_{kj}[n(kj)+\tfrac{1}{2}]\hbar\,\frac{\partial^2\omega(kj)}{\partial\eta_{\alpha\beta}\partial\eta_{\gamma\delta}} \tag{8}$$

The stress equation of motion in a material reference system

$$\rho\,\frac{\partial^2 u_\beta}{\partial t^2} = \frac{\partial T_{\alpha\beta}}{\partial a_\alpha} \tag{9}$$

and the linearized strain-displacement relation

$$\eta_{\alpha\beta} = \tfrac{1}{2}\left[\frac{\partial u_\alpha}{\partial a_\beta} + \frac{\partial u_\beta}{\partial a_\alpha}\right] \tag{10}$$

have plane wave solutions with wave vector k and frequency

$$u_\beta = u_o\,e_\beta(KJ)\,e^{i(2\pi\vec{K}.\vec{a}-\Omega t)} \tag{11}$$

$$N(kj) = n(kj) + \Delta N(kj)\,e^{i(2\pi\vec{K}.\vec{a}-\Omega t)} \tag{12}$$

In the absence of phonons the solution has the simpler form

$$u^o_\beta = u^o_o\,e^o_\alpha(KJ)\,e^{i(2\pi\vec{K}.\vec{a}-\Omega_o t)} \tag{13}$$

The correction due to the phonons can be calculated by using a perturbation method. This leads to

$$\Omega = \Omega_o + \frac{2\pi^2\hbar}{\rho V\Omega_o}\,e_\alpha(KJ)\,e_j(KJ)\,K_\beta K_\delta \sum_{kj}\frac{\partial^2\omega(kj)}{\partial\eta_{\alpha\beta}\partial\eta_{\gamma\delta}} [n(kj)+\tfrac{1}{2}] - \frac{\pi i\hbar}{\rho V\Omega_o}\,e_\alpha(KJ)\,K_\beta \sum_{kj}\frac{\partial\omega(kj)}{\partial\eta_{\alpha\beta}}\,\frac{\Delta N(kj)}{u_o} \tag{14}$$

The attenuation α is given by

$$\alpha = -\mathrm{Im}\,\frac{\Omega}{s_o(KJ)} \tag{15}$$

and the new phase velocity of the sound wave

$$s = \mathrm{Re}\,\frac{\Omega}{2\pi K} \tag{16}$$

s_o is the umperturbed phase velocity.

These quantities are completely determined when $\Delta N(kj)$ is known. This quantity corresponds to the response of the phonon reservoir to the sound wave. This response is evaluated by using the Boltzmann transport equation

$$\frac{\partial N(kj)}{\partial t} = \frac{1}{2\pi} \frac{\partial \omega(kj)}{\partial \eta_{\alpha\beta}} \frac{\partial \eta_{\alpha\beta}}{\partial a_\gamma} \frac{\partial N(kj)}{\partial k_\gamma} - v_\alpha(kj) \frac{\partial N(kj)}{\partial a_\alpha} + \left[\frac{\partial N(kj)}{\partial t}\right]_{coll} \quad (17)$$

The term $\left[\frac{\partial N(kj)}{\partial t}\right]_{coll}$ is most important because it introduces the collisions between phonons. Without collisions and for $\Omega\tau >> 1$ (where τ is the mean thermal phonon relaxation time) and $\hbar\Omega << k_B T$ the Akhieser theory goes back to the Landau-Rumer theory.

The mean phonon relaxation time τ is defined by

$$\left[\frac{\partial N(kj)}{\partial t}\right]_{coll} = - \frac{N(kj)-n(kj)}{\tau} \quad (18)$$

Three different types of collisions must be distinguished :

- normal processes (N), in which the total energy and the total quasi-momentum of the phonons are conserved.

- umklapp processes (U), in which only the total energy is conserved, but not the quasi-momentum.

- elastic scattering (E), corresponding to collisions of phonons with impurities, defects, etc... The scattering being elastic, the frequency does not change but the wave vector and the polarization can be altered. The total energy is conserved. However, the number of phonons of any particular frequency cannot change.

N processes alone cannot restore equilibrium, and some E and U processes are necessary to give a redistribution of the quasi-momentum. E processes with some N and U processes can also bring the phonon distribution back to equilibrium. Thus the attenuation α and the velocity change $s-s_o$ take the general form

$$\alpha = \frac{CT\Omega_o \langle \gamma^2_{eff} \rangle}{2\rho s_o^3} \frac{\Omega_o \tau}{1+\Omega_o^2\tau^2} \quad (19)$$

$$s-s_o = \frac{CT\langle \gamma^2_{eff} \rangle}{2\rho s_o} \frac{\Omega_o^2 \tau^2}{1+\Omega_o^2\tau^2} \quad (20)$$

C is the specific heat. γ is an effective Grüneisen constant which is a function of the Grüneisen parameter

$$\gamma(kj) = - \frac{1}{\omega(kj)} \frac{\partial \omega(kj)}{\partial \eta_{\alpha\beta}} e_\alpha(KJ) K_\beta \quad (21)$$

This function depends on the involved processes. The average $\langle \gamma^2 \rangle$ is given by

$$\langle \gamma^2 \rangle = \frac{\sum_{kj} \gamma^2 n(kj) [n(kj+1)] \omega^2(kj)}{\sum_{kj} n(kj) [n(kj+1)] \omega^2(kj)} \quad (22)$$

Using the low frequency Young modulus $E_o = \rho s_o^2$, the quantity $\Delta E = CT\langle \gamma^2 \rangle$, and the relation between attenuation and Q-factor

$$\alpha = \frac{\Omega_o}{2Qs_o} \quad (23)$$

and between velocity and frequency

$$\frac{s-s_o}{s_o} = \frac{\Omega-\Omega_o}{\Omega_o} \quad (24)$$

one obtains

$$\frac{1}{Q} = \frac{\Delta E}{E_o} \frac{\Omega_o \tau}{1+\Omega_o^2\tau^2} \quad (25)$$

$$\frac{\Omega-\Omega_o}{\Omega_o} = \frac{\Delta E}{2E_o} \frac{\Omega_o^2\tau^2}{1+\Omega_o^2\tau^2} \quad (26)$$

A fluctuation $\delta\tau$ of the relaxation time will produce a frequence fluctuation $\delta\Omega$ of the resonator resonance frequency Ω_o

$$\frac{\delta\Omega}{\Omega_o} = \frac{\Delta E}{E_o} \frac{\Omega_o^2\tau^2}{(1+\Omega_o^2\tau^2)^2} \frac{\delta\tau}{\tau} \quad (27)$$

Introducing the power spectral density $S_y(f)$ of the fractional frequency fluctuation $y = \delta\Omega/\Omega_o$, and using eq. (25) in eq. (27) yields

$$S_y(f) = \frac{1}{Q^4} \left[\frac{E_o}{\Delta E}\right]^2 S_{\delta\tau/\tau}(f) \quad (28)$$

which exhibits the $1/Q^4$ law.

It is now to be assumed that τ is fluctuating with a 1/f power spectral density. This is not demonstrated yet, but is in accordance with the fluctuations of a cross section which follow a 1/f spectrum as demonstrated in the quantum 1/f noise theory by P. Handel[10-11].

A rough evaluation of the Grüneisen constant γ can be made either by using the relation of the thermal expansion constant[6] or by means of the third order elastic constants of the macroscopic theory[12]. In both cases it is found that for quartz γ is of the order of 10. Therefore $(\Delta E/E_o) \approx 10^{-5}$. This shows from the phenomenological eq.

(2), that for high quality resonators ($Q\ 2\times10^6$) and for Fourier frequencies equal to 1 Hz, $\mathcal{L}_{\delta\tau/\tau}$ (1Hz) is approximately equal to 10^{-5}. This is reasonable if one considers that $\delta\tau/\tau = -\delta Q/Q$.

The previous relations were established with the high temperature condition $\Omega\tau \ll 1$. It can be remarked that the $1/Q^4$ law remains true also for $\Omega\tau>1$. Near room temperature, attenuation and Q-factor are independant of temperature. This appears in eq. (24) where $\Delta E \ll T$ and $\tau \ll T^{-1}$ on account of the temperature dependance of thermal conductivity.

At low temperature the Landau-Rumer method must be used. Generally the theory indicates for the attenuation a temperature dependance $\propto T^4$. This is roughly what is observed on fig. 4 below the phonon-phonon relaxation peak temperature. If one considers the experimental values of table II, the measured Q-factors do not follow this law. There is a threshold, which is attributed to the losses coming from the resonator plating and mounting and which are temperature independant. Thus two different loss phenomena are present at the same time at low enough temperatures. However temperature is implicitely still included in eq. (27)- For instance specific heat is temperature dependant below the Debye temperature. These two facts can explain why the experimental values of 1/f noise levels given in table II do not exactly follow the $1/Q^4$ law.

Conclusion

In the limit of its accuracy, the phenomenological relation between 1/f frequency noise and Q-factor can be used to predict the ultimate stability of quartz resonators and oscillators. The theoretical model gives confirmation of the $1/Q^4$ dependance. It is based on the assumption of 1/f fluctuations of the relaxation time. This is to be verified, but is not unrealistic in the perspective of the quantum 1/f noise theory, since relaxation time is determined by the cross sections of the elementary processes involved in the interactions.

This work was supported in part by the U.S. Air Force under Grant AFOSR-80-105, the US Office of Naval Research and the Direction des Recherches, Etudes et Techniques of France.

References

(1) F.L. Walls, A.E. Wainwright, "Measurement of the short term stability of quartz crystal resonators and the implications for crystal oscillator design and applications", IEEE Trans. Inst. Meas., IM-24, n° 1, 15, (1975).

(2) J.J. Gagnepain, "Fundamental noise studies of quartz crystal resonators", 30th Ann. Freq. Cont. Symp., 84, (1976).

(3) A. Ballato, J.R. Vig, "Static and dynamic frequency-temperature behavior of singly and doubly rotated oven-controlled quartz resonators", 32nd Ann. Freq. Cont. Symp., 180, (1978).

(4) G. Théobald et al, "Dynamic thermal behavior of quartz resonators", 33rd Ann. Freq. Cont. Symp., 239, (1979).

(5) G. Marianneau, J.J. Gagnepain, "Digital temperature control for ultrastable quartz oscillators", 34th Ann. Freq. Cont. Symp., 52, (1980).

(6) J.J. Gagnepain, "Nonlinear properties of quartz crystal and quartz resonators : a review", 35th Ann. Freq. Cont. Symp., 1981.

(7) L. Landau, G. Rumer, Phys. Z. Sowjetunion, 11, 18, (1937)

(8) A. Akhieser, "On the absorption of sound in solids", J. Phys. (USSR), 1, 227, (1939).

(9) H.J. Maris, "Interaction of sound waves with thermal phonons in dielectric crystals", Physical Acoustics, vol. VIII, Academic Press, (1971).

(10) P.H. Handel, "Quantum theory of 1/f noise", Physics Letters, 53A, n° 6, 438, (1975).

(11) P.H. Handel, "1/f noise - an infrared phenomenon", Phys. Rev. Letters, 34, n° 24, 1492, (1975).

(12) R. Brugger, "Generalized Grüneisen parameters in the Debye model", Phys. Rev. 137, n° 6A, (1963).

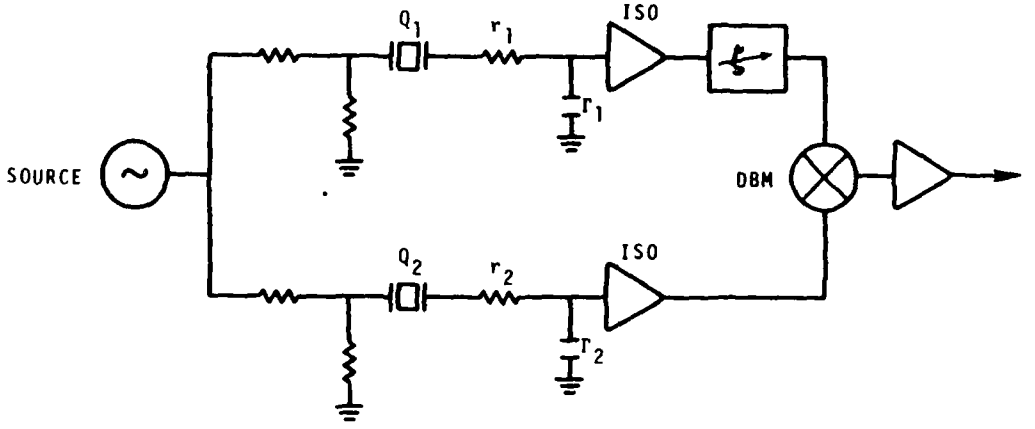


Fig. 1

1/f noise measurement system for pairs of identical resonators at room temperature

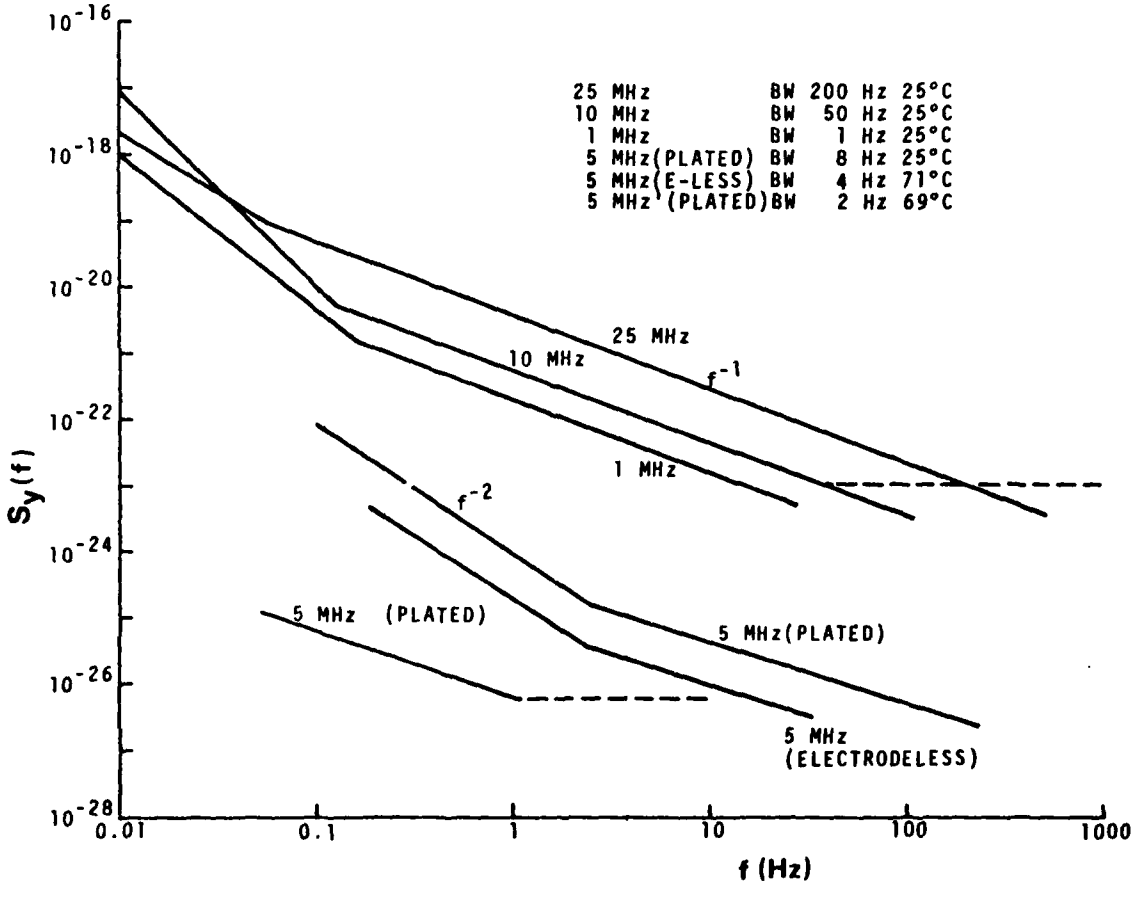


Fig. 2

Fractional frequency fluctuation power density spectrum of quartz resonators

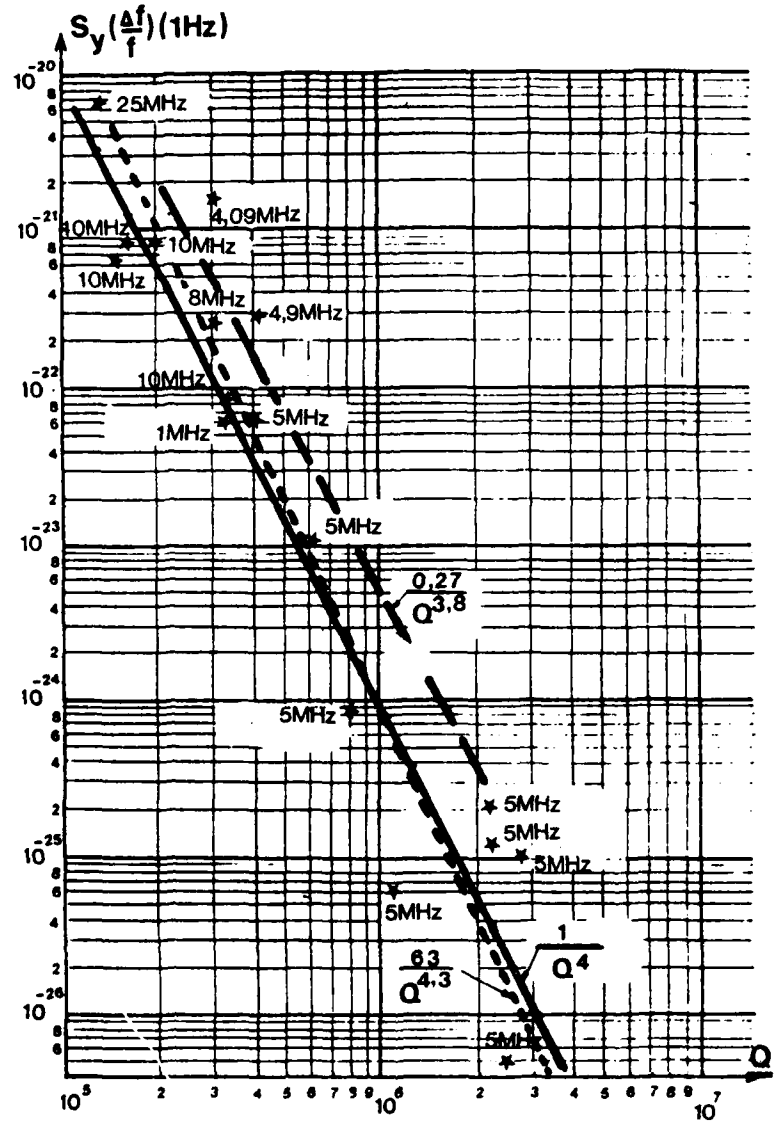


Fig. 3

1/f frequency noise level at 1 Hz from the carrier versus Q factor

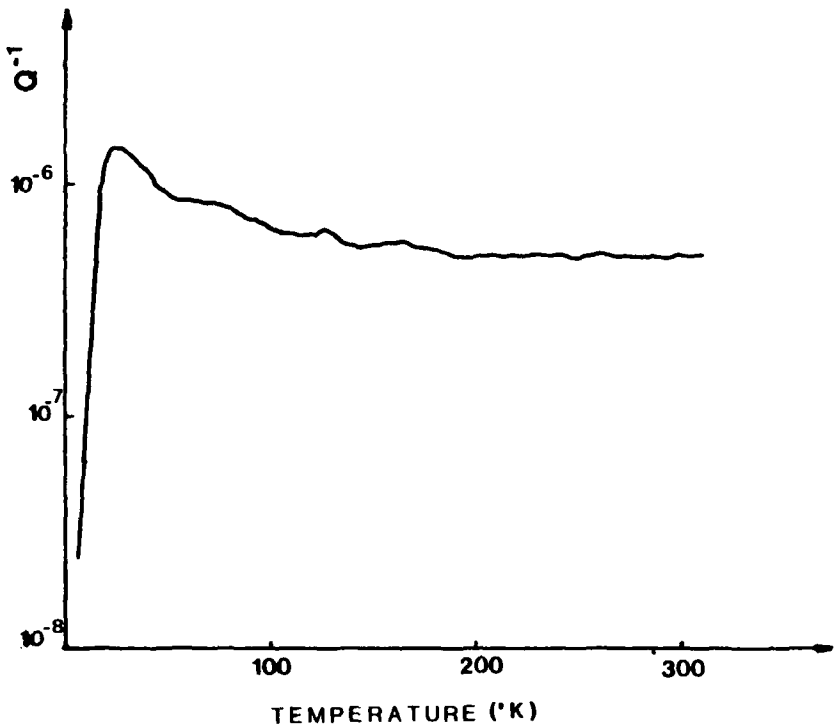


Fig. 4

Internal losses of quartz crystal versus temperature

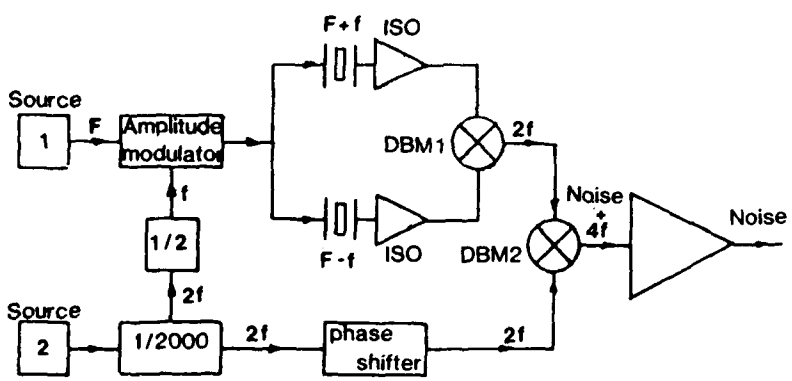


Fig. 5

1/f frequency noise measurement system for resonators at low temperatures

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

1/f FREQUENCY FLUCTUATION OF A QUARTZ CRYSTAL OSCILLATOR AND TEMPERATURE FLUCTUATION

Y. Noguchi, Y. Teramachi, and T. Musha

Department of Applied Electronics,
Graduate School at Nagatsuta,
Tokyo Institute of Technology
Nagatsuta, Midori-ku, Yokohama 227, Japan

1. SUMMARY

Frequency fluctuations of 5-MHz 5th-overtone plano-convex, AT-cut and IT-cut, quartz crystal oscillators, held in a temperature-stabilized oven, were measured with reference to a Cs oscillator (HP 5061A); temperature fluctuations were measured simultaneously using a Pt resistor.

The power spectral density of the temperature fluctuation was proportional to $1/f^2$, where f is Fourier frequency.

The power spectral density of the frequency fluctuation was separated into three parts; one being proportional to $1/f^2$, one to 1/f and the third being frequency independent. The 1/f part and the frequency independent part were in quantitative agreement with the power spectral density estimated from the temperature fluctuation.

A coherence function between the frequency and temperature fluctuations was calculated. The frequency fluctuation is correlated with the temperature fluctuation in the $1/f^2$ and frequency independent regions.

However, the coherence is lower in the range where the 1/f spectrum is observed. The observed 1/f fluctuation is larger than the instrumentation noise, and hence is not electric noise but is attributed to the frequency fluctuation itself.

2. INTRODUCTION

Quartz crystal oscillators are used in a wide range of applications because of their small size, power consumption, etc. One of their big merits is the superiority of their short term stability, but for long term stability they are inferior to other high stability oscillators. It is well known that the long term stability of a quartz oscillator is largely degraded by temperature fluctuations, but the exact mechanism is not yet known.

In the present paper, the transfer of temperature fluctuations to frequency fluctuations and the correlation between the frequency and temperature fluctuations are studied. The quartz oscillator was placed in a precisely temperature-stabilized oven in order to hold its ambient temperature at the turn-over point. When the temperature is varied with time, the oscillation frequency varies in a spiky manner, called "thermal shock" (1-3). In recent years, various studies on this phenomenum have been reported (4-6), and it has been proposed that a transfer function is effective for representing the effect on the oscillation frequency to a small temperature variation. (7)

Little research has been reported on frequency fluctuations in quartz oscillators. In 1963, Attkinson et. al. reported that the power spectral density was proportional to $1/f^{1.4}$ (8), where f is Fourier frequency. From this, it seemed that the 1/f fluctuation (9), which is widely observed in semiconductors, resistors, etc., would also occur in quartz. In 1977, Gegnepain et. al. reported that the 1/f power spectral density in the range higher than 1 Hz was inversely proportional to Q^4, where Q is the Q-value of the quartz resonator. (10)

In the present paper, the temperature and frequency fluctuations in quartz oscillators were simultaneously measured from 10^{-2} Hz to 10^{-6} Hz, and it was shown that the power spectral density of the frequency fluctuation can able to separated into three parts; one being proportional to $1/f^2$, one to 1/f and the other being frequency independent. The $1/f^2$ portion is related to the temperature fluctuations by the static frequency-temperature coefficient α, and the frequency independent part branch is related to the temperature fluctuations by the dynamic frequency-temperature coefficient β, while the residual 1/f part does not correlate with the temperature fluctuations at all.

3. MEASUREMENT

3-1 Measurement System

The measurement system shown in Fig.1 was developed. The beat period method was used to measure the frequency fluctuation, the frequency from the oscillator under test being multiplied with a much higher stable frequency and the frequency variation calculated from the counted period variation. Because the quartz oscillator under test was relatively stable, the reference oscillator had to be very highly stable, and so a Cs atomic oscillator (HP 5061A) was used for the reference.

The temperature fluctuation was measured with a D.C. resistance bridge, using a Pt resistor (sealed in glass, 4.5mmϕx 26mm length, R(0°C) = 100Ω). The output signal after the D.C. resistance bridge and the first stage amplifier was small, so they were placed in the oven with the oscillator in order to reduce thermo-electromotive voltages ,and reduce changes in circuit constants and amplifier offset drift due to temperature.

The noise from the temperature measurement system was dominated by the 1/f noise from the

first stage amplifier connected to the Pt bridge. This noise level made it impossible to measure temperature fluctuations of frequencies higher than 10^{-3} Hz (see Fig.7). THe amplified temperature signal from the oven was amplified, the D.C. offset was removed, and it was passed through low pass filter, which eliminated aliasing that might otherwise occur in the subsequent spectral analysis. The signal was then converted from analogue to digital form and sampled synchronously with the beat period counter. The digital data was punched out on paper tape by a punch controlled by an 8080A-CPU and was later analyzed on a mini-computer.

3-2 Method

Three quartz crystal units were prepared:

1. AT-cut HC-6/U T_0=75.5°C $Q=2.4\times10^6$ diameter of blank: 14.2φ
2. AT-cut TO-3 T_0=62°C $Q=2.3\times10^6$ diameter of blank: 16φ
3. IT-cut TO-3 T_0=58.8°C $Q=2.3\times10^6$ diameter of blank: 16φ

All units were 5-MHz 5th-overtone plano-convex type. The analysis was carried out in two steps, as below:

1) Estimate of Transfer Function

The temperature of the oven was varied high and low cyclically with a sufficient long period and an amplitude range of between 0.03 deg. p-p and 0.1 deg. p-p. The frequency and temperature fluctuation were measured simultaneously and the transfer function from temperature to frequency was obtained. The difference between high and low temperature was made large enough that the signal to noise ratio would be good, and reliable estimates were obtained.

2) Estimate of Correlation

Single and double ovens were prepared, the power spectral densities of the temperature fluctuations of which are shown in Fig.7. A quartz unit was then placed in the oven, the temperature of which was kept at various fixed values. The frequency and temperature fluctuations were measured at the same time, the cross spectral density of the two obtained, and divided by the roots of the power spectral densities of the frequency and temperature fluctuation.

4. THERMAL SHOCK

When the ambient temperature variation of a quartz oscillator is slow, the frequency variation obeys a cubic equation (11). However, as the temperature variation gets faster the frequency variation becomes spiky. This phenomenon is called themal shock (1-3).

This behaviour has previously been described by Ballato et. al. (12), but a somewhat different formulation is employed in the present paper:

$$\Delta f = \alpha\Delta T + \frac{d\Delta T}{dt} \qquad (4\text{-}1)$$

$$\Delta T = T(t) - T_0$$

where α is the static frequency-temperature coefficient and β is the dynamic frequency-temperature coefficient.

5. RESULTS

5-1 Estimate of Transfer Function

The estimate of a transfer function was made in order to see how the frequency variation would be influenced by the temperature variation, as stated in section 3-2. The transfer function H(f) is represented by the form below:

$$H(f) = |H(f)|\exp(-j\phi) \qquad (5\text{-}1)$$

where the absolute value H(f) is called the gain factor and the argument of the transfer function is called the phase factor.

Transfer functions are shown in Figs. 2 and 3 for the HC-6/U quartz crystal. In the lower regions of the figures, as the static frequency-temperature coefficient α approaches zero, the gain factor in Fig. 2 approaches a straight line inclined at 45° and the value of phase factor in Fig. 3 approaches π/2. This characteristic curve is given from differentiators. In the frequency range higher than 10. Hz, the gain factor is lowered and the phase factor is rotated, so there is a delay factor. The sign of the static frequency-temperature coefficient in the lower range makes the phase converge to π or 0, from (-) or (+), respectively.

These observations are in eq. (5-2):

$$H(f) = (\alpha + j2\pi f\beta)\frac{(1 + j2\pi f\tau_3)}{(1+j2\pi f\tau_1)(1 + j2\pi f\tau_2)} \qquad (5\text{-}2)$$

It should be noted that the term $(\alpha + j2\pi f\beta)$ is equal to eq. (4-2). τ_1, τ_2, τ_3, are the delay times: τ_1 is the delay time from the ambient temperature of the quartz unit to the temperature of the quartz itself; τ_2 is the delay time from the ambient temperature of the Pt resistor to that of the quartz unit; and τ_3 is the delay time from the ambient temperature of the Pt resistor to the measured temperature. The heat flow during the measurement is shown in Fig. 4, where

$$H_1(f) = \frac{1}{1 + j2\pi f\tau_1}$$

$$H_2(f) = \frac{1}{1 + j2\pi f\tau_2} \qquad (5\text{-}3)$$

$$H_3(f) = \frac{1}{1 + j2\pi f\tau_3}$$

The transfer function was estimated both using and not using a copper block, in which the quartz unit and a Pt resistor are placed. The difference between the trensfer fundtions of the IT-cut and the AT-cut were measured using the two TO-3 quartz crystals. In Figs. 5 and 6, data points marked A and B are obtained from the AT-cut, and those marked C and D are from the IT-cut. A and C were measured using the copper block, while B and D did not use it. Since the copper block increases the thermal coupling between the quartz unit and the Pt resistor, it reduces the value of τ_2, as is clear from in Figs. 5 and 6.

Using Figs. 3, 5, and 6, the constants in eq. (5-2) were estimated, and are shown in Table 1. It is thought that the differences in τ_1 are due to the differences in the unit type, HC-6/U or TO-3. Differences in τ_1 influence the value of the dynamic frequency-temperature coefficient β.

5-2 Power Spectral Density of Temperature Fluctuation

The power spectral densities of the temperature fluctuation are shown in Fig. 7, and are proportional to $1/f^2$ for both ovens. In the power spectral density of the temperature fluctuation in the single oven there is a time constant below 10^{-3} Hz, but in the double oven this cannot be seen above the noise. Thus in both ovens, it is in the frequency range below 10^{-3} Hz where it is possible to measure the correlation between temperature and frequency fluctuations.

The temperature-controlling feed-back gain of an oven is always in some measure nonlinear, but the nonlinearity is here not important in the temperature range of the present measurements (75.5°C - 58°C = 17.5°C).

5-3 Power Spectral Density of Frequency Fluctuation

The power spectral densities of the frequency fluctuation were measured using the copper block, above.

5-3-1 1/f Frequency Fluctuation

The power spectral densities of the frequency fluctuation of two quartz oscillators, the AT-cut and IT-cut TO-3 units, were measured in the single oven, and are shown in Fig. 8. The coherence between the frequency and temperature fluctuation are shown in Fig. 9. The dotted lines in Fig. 8, are calculated from eq. (5-4):

$$P_{\Delta F}/F = |H(f)|^2 P_{\Delta T} / (5\ \mathrm{MHz})^2 \quad (5\text{-}4)$$

where $P_{\Delta F}/F$ is the power spectral density of the frequency fluctuation, $P_{\Delta T}$ is that of the temperature fluctuation, and H(f) is the estimated transfer function in eq.(5-2). The calculation uses the values of τ_s and β in Table 1 and values of $\alpha_{AT} = 0.025$, $\alpha_{IT} = 0.0015$, and $\alpha_{IT} = 0.024$.

In Fig. 8, it is seen that the measured power spectral density of the frequency fluctuation of the AT-cut quartz unit fits the calculated curve well and is proportional to $1/f^2$ at lower frequencies. Also in Fig. 8, the measured power spectral densities of the frequency fluctuation of IT-cut units fit calculated curves well, above 10^{-4} Hz, but when the α is small, the measured curve separates from the calculated curve at a slope proportional to 1/f at frequencies lower than 10^{-4} Hz. Thus, for the power spectral densities of the frequency fluctuation of the AT-cut and IT-cut quartz units, the measured curve fits the calculated one well at above 10^{-4} Hz, where the spectra are flat.

In Fig. 9, it is seen that the coherences go sharply down, due to the time constant of the Pt resistor itself and noise, above 10^{-3} Hz. The coherence obtained with the AT-cut unit is high, while the coherence obtained with the IT-cut unit in the small α extends to lower below 10^{-4} Hz, where the measured power spectral density of the frequency fluctuation is proportional to 1/f. It can therefore be assumed that the 1/f frequency fluctuation is not correlated with the temperature fluctuation.

It would be wrong to assume that the coherence measured with an AT-cut unit would always be high or that measured with on IT-cut unit would fall, there being no basic difference in the coherence measured using amy cut. In the present results, the static frequency-temperature coefficient was accidentally large in the measurements on the AT-cut unit, and so the coherence was high.

5-3-2 1/f and $1/f^2$ Frequency Fluctuation with temperature Fluctuation

The frequency fluctuation of the IT-cut unit in a double oven was measured near the turnover point over a long period, about two months. Fig. 10 shows the power spectral densities of the frequency fluctuation. The dotted line in Fig. 10 is calculated from eq. (5-4), with α_{IT}=0.05. Fig. 11 shows the coherence. The measurement was carried out with temperature kept at two points, one at 58° C, which was near but a little different from the turn-over point, and the other at 62°C, which was considerably different from the turn-over point.

At 58°C, 1/f frequency fluctuation is observed over a wide range, from 10^{-3} Hz to 5×10^{-5}Hz.

At 62° C, the static frequency-temperature coefficient has a larger value, α_{IT} =0.16. Therefore, as is shown by the upper line in Fig. 10, the level of the $1/f^2$ portion goes up, and its range gets wider, and the range the 1/f portion gets narrower.

The coherence inthe range the $1/f^2$ region is high, and that in the 1/f region is low. The frequency independent region, which was seen in Fig. 8, is not observed in Fig.10. This is explained by the frequency independent region being generated by the dynamic frequency-temperature coefficient β. The level to be measured is the shoulder-like break in the dotted line, which is here below the 1/f region level.

5-3-3 Summary of Section 5-3

The power spectral density of the frequency fluctuation is divided into three points:

1. A $1/f^2$ dependence is introduced by the static frequency temperature coefficient α. The coherence of the frequency fluctuation and the temperature fluctuation in this range is high;

2. A frequency independent region is introduced by the dynamic frequency-temperature coefficient β. The coherence in this range is high;

3. A 1/f branch is not introduced in the calculations, as stated above. The coherence of the frequency fluctuation and the temperature fluctuation in this range goes down.

The 1/f frequency fluctuation appears when the static and dynamic frequency -temperature coefficients, α and β, respectively, are lowered. It is of course, necessary to reduce the level of the temperature fluctuations in the oven, and it was impossible to observe 1/f frequency fluctuations with the AT-cut unit in the single oven.

Fig. 12 shows the present results compared with other power spectral densities of frequency fluctuations in references (8),(10) and (13). The 1/f frequency fluctuation due to Gagnepain et. al.(10) is to be noted, as the level of 1/f frequency fluctuation in Fig. 12 of the present measurements fits an extrapolation this curve well. It can thus be

assumed that the mechanism of 1/f frequency fluctuation generation in quartz oscillators is common to both.

It is difficult to exclude the possibility that 1/f frequency fluctuation is correlated with the temperature fluctuation, since only the spatially averaged temperature fluctuation was measured, and it is possible that 1/f frequency fluctuation may be correlated with the local thermal gradient fluctuation. However, it is not easy to measure local thermal gradient fluctuations in the quartz itself without disturbing it with the measurement.

6. CONCLUSIONS

The frequency fluctuation of a quartz oscillator and the temperature fluctuation of a Pt resistor, which was placed near the quartz, were measured simultaneously, the measurement system being designed paying attention to makig the noise from it as small as possible. The following conclusions are obtained:

1) The power spectral density of the temperature fluctuationis proportional to $1/f^2$, where f means Fourier frequency;

2) The power spectral density of the frequency fluctuation can be divided into three parts; one being proportional to $1/f^2$, one to 1/f, and the third being frequency independent;

3) The $1/f^2$ part is related to the static frequency-temperature coefficient α, and the frequency independent to the dynamic frequency-temperature coefficient β;

The coherence between the frequency and temperature fluctuations is high in the Fourier frequency ranges where these two are observed, while the coherence is low in the ranges where the 1/f dependence is observed;

4) The level of the 1/f dependence depends on neither the working temperature nor the temperature fluctuation, but its observability depends on the magnitudes of the 1/f and frequency independent levels.

ACKOWLEDGEMENTS

The authors gratefully ackowledge the advice of Prof. Emeritus H. Fukuyo, Assoc. Prof. N. Ohura, and Dr. N. Kuramochi of Tokyo Institute of Technology. They also wish to thank members of the Time Division of Tokyo Astronomical Observatory for offering a Cs oscillator as a reference oscillator.

REFERENCES

1. A. W. Warner,"Ultra-Precise Quartz Crystal Frequency Standards", IRE Trans. Instrumentation, Vol. I-7, Nos. 3,4, December 1958, pp. 185-188.

2. A. W. Warner and D. L. White, "An Ultra-Precise Standard of Frequency", Eleventh Interium Report on Contract DA 36-039 SC-73078 to US Army Signal R&D Laboratory, Fort Monmouth, NJ, July 1959,42p.

3. A. W. Warner, "Use of Parallel-field Excitation in the Design of Quartz Crystal Units", Proc. 17th AFCS, May 1963, pp. 248-266.

4. J. A. Kusters, "Transient Thermal Compensation for Quartz Resonators", IEEE Trans. Sonics Ultrason., Vol. SU-23, No. 4, July 1976, pp. 273-276.

5. R. J. Munn, "Warm-up Characteristics of Oscillators Employing 3. M.C. Fundamental Crystals in HC-27/U Enclosures", Proc. 19th AFCS, April 1965 pp. 658-668.

6. R. Holland, "Nonuniformly Heated Anisotropic Plates: I. Mechanical Distortion and Relaxation" IEEE Trans. Sonics Ultrason., Vol. SU-21, No.3, July 1974, pp. 171-178.

7. Y. Teramachi, M. Horie, H. Kataoka, and T. Musha, "Frequency Response of a Quartz Oscillator to Temperature Fluctuation", Proc. 33rd AFCS, May 1979, pp. 235-238.

8. W. R. Attkinson, L. Fey, and J. Newman, "Spectrum Analysis of Extremely Low Frequency Variations of Quartz Oscillators", Proc. IEEE 1963.

9. T. Musha, "On 1/f Fluctuation", J. Appl. Pys. (Japanese), Vol. 46, No. 12, p. 1149 (1977).

10. J. J. Gagnepain, J. Uebersfeld, "1/f Noise in Quartz Crystal Resonators", Proc. of Symposium on 1/f Fluctuations, July 1977.

11. R. Bechman, A. D. Ballato, and T. J. Lukaszek, "Higher-order Temperature Coefficients of the Elastic Stiffnesses and Compliances of Alpha-Quartz", Proc. IRE, 50, 1812 (1962).

12. A. Ballato and J. R. Vig, "Static and Dynamic Frequency-Temperature Behavior of Singly and Doubly Rotated, Oven-controlled Quartz Resonators", Proc. 32nd AFCS.

13. D. W. Allan, "A Review of Flicker Noise Frequen- Instabilities in Precision Frequency Standards", Proc. of Symposium on 1/f Fluctuations, July 1977.

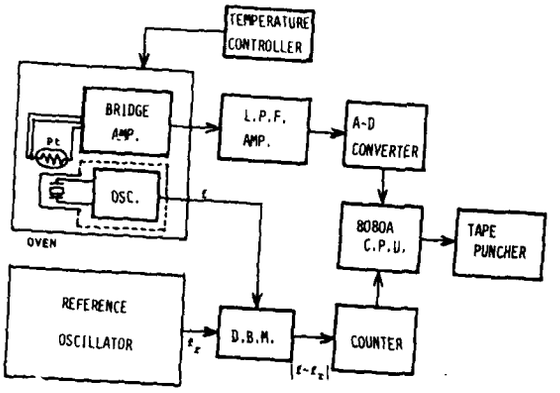


Fig. 1 Measurement System

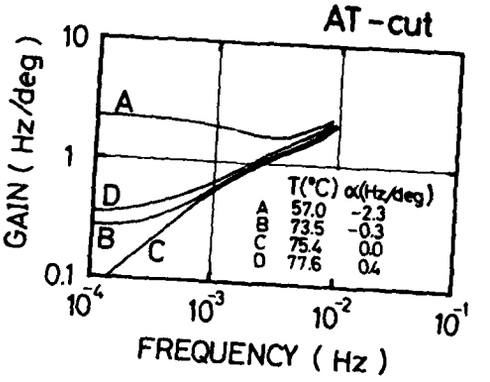


Fig. 2 Gain factor of transfer function as a function of Fourier frequency.

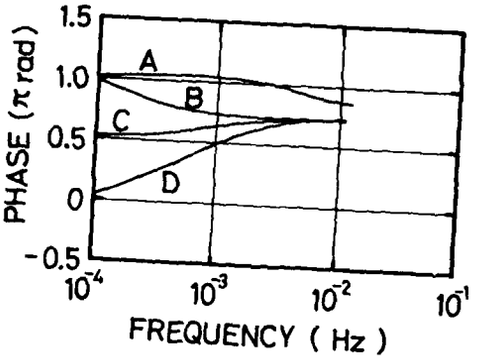


Fig. 3 Phase factor of transfer function as a function of Fourier frequency.

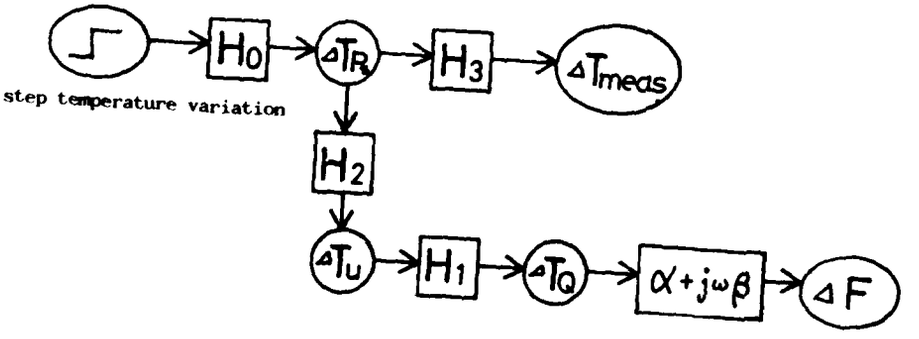


Fig. 4 Heat flow in the measurements.

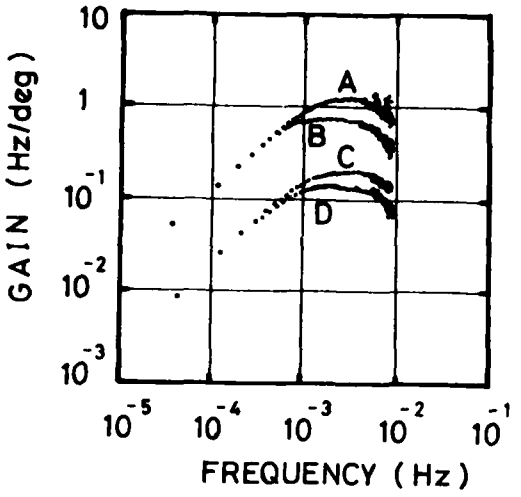


Fig. 5 Gain factor of transfer function as a function of Fourier frequency

Quartz \ Const.		β Hz/deg/sec	β sec/deg	τ_1 sec	τ_2 sec	τ_3 sec
AT HC6U	(A)	-120	-2.4×10^{-5}	40	100	50
AT TO-3	(A)	-175	-3.5×10^{-5}	150	100	50
	(B)	-175	-3.5×10^{-5}	150	50	50
IT T03	(A)	-30	-6×10^{-6}	150	100	50
	(B)	-30	-6×10^{-6}	150	50	50

Table 1 Constants of transfer function:
(A) - using a copper block;
(B) - not using a copper block.

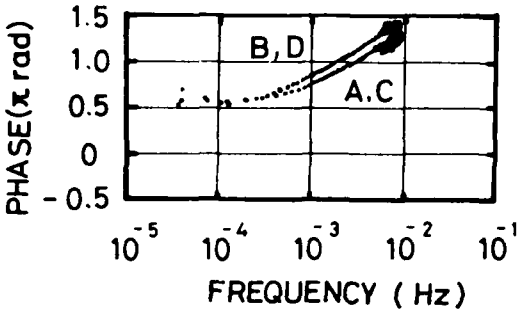


Fig. 6 Phase factor of transfer function as a function of Fourier frequency

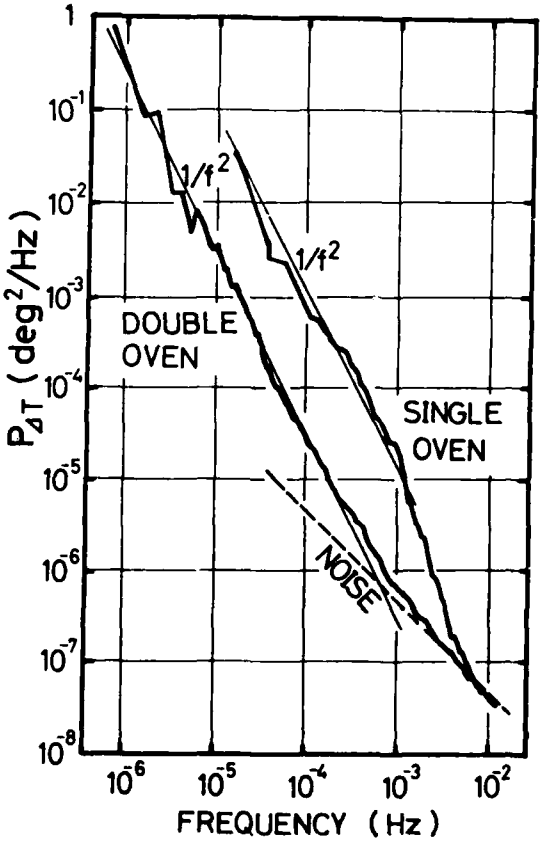


Fig. 7 Power spectral densities of temperature fluctuation as a function of Fourier frequency

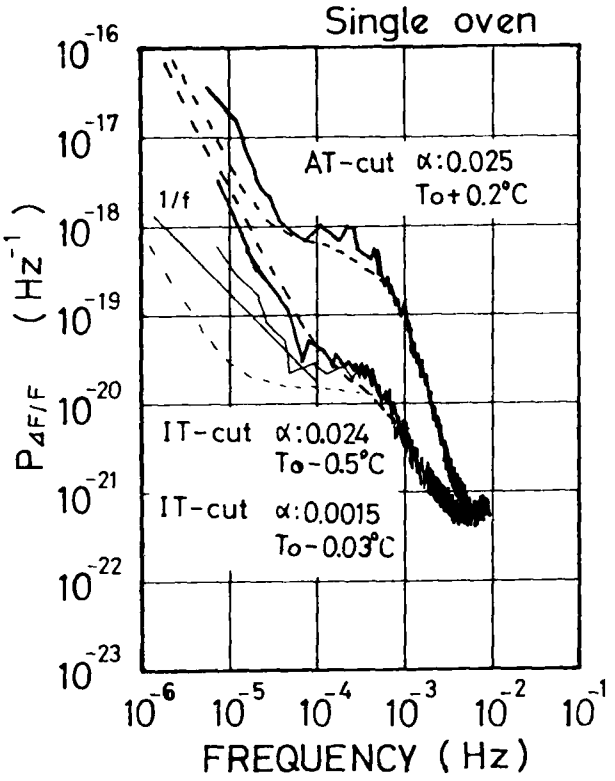


Fig. 8 Power spectral densities of frequency fluctuation as a function of Fourier frequency, in the single oven.

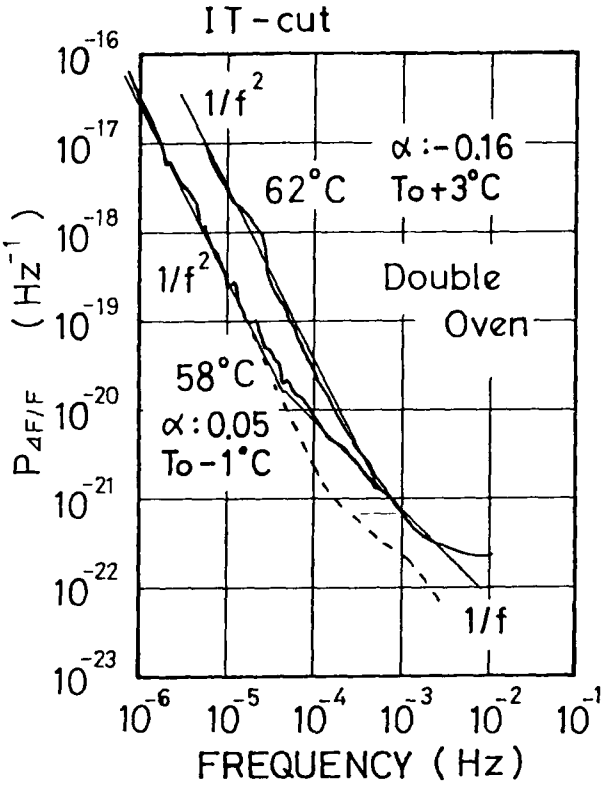


Fig. 10 Power spectral densities of frequency fluctuation as a function of Fourier frequency, in the double oven.

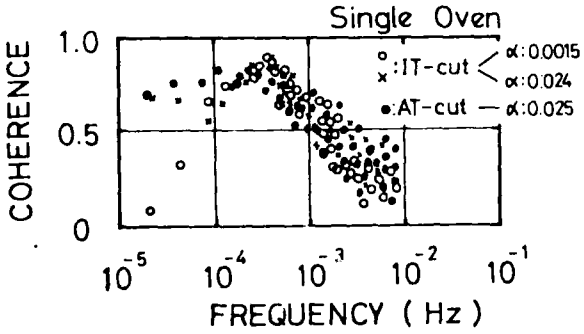


Fig. 9 Coherence as a function of Fourier frequency, in the single oven.

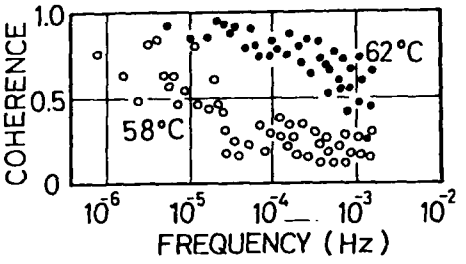


Fig. 11 Coherence as a function of Fourier frequency, in the double oven.

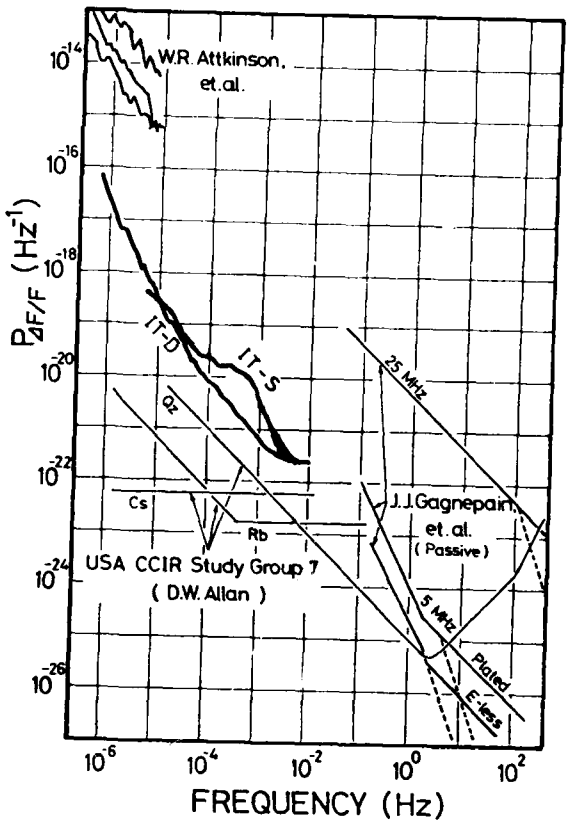


Fig. 12 Power spectral densities of frequency fluctuation of various oscillators as a function of Fourier frequency.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

FREQUENCY RETRACE OF QUARTZ OSCILLATORS

Ferdinand Euler and Nicholas F. Yannoni

Rome Air Development Center
Deputy for Electronic Technology
Hanscom AFB, MA 01731

Summary

Frequency retrace measurements are reported on oven controlled quartz oscillators utilizing AT and SC cut plated and BVA resonators. Prior to full aging, the retrace error is added to the aging effect. With well-aged resonators, after one or several on-off cycles, the frequency settles at a new level characteristic for intermittent operation. Severe frequency shifts have sometimes been found after the the first restart following prolonged continuous operation. SC cut resonators appear to show distinctly smaller retrace errors than AT cut.

Key Words

Crystal oscillators, quartz, resonators, frequency retrace, AT cut, SC cut, BVA resonators

INTRODUCTION

Quartz oscillators deployed in certain tactical situations undergo repeated shutdown-restart cycling between missions which may last from several hours to a few days. For this type of application the frequency retrace characteristics of the oscillator are of high interest. With the appearance of the SC cut resonators frequency retrace may show marked improvement. For instance, H. Jackson[1] was able to achieve reduced retrace error ranging from 2×10^{-9} to 5×10^{-10} with SC cut resonators of ceramic flatpack design. Besson and Peier[2] reported that their new mounting structure of BVA crystrals led to retrace improvement exceeding an order of magnitude, showing retrace error values of 2 or 3×10^{-10}. Similar results were presented by Besson, Graf and Emmons.[3]

These findings have prompted us to subject oscillators utilizing BVA and plated crystals of both SC and AT cut to a sequence of controlled power interruptions and measure the resulting frequency response. The objective of this exercise is to examine the frequency retrace, collect data on currently available oscillator and resonator types and find the relationship between retrace and other parameters such as length of interruption, resonator aging, time after restart and the warm-up characteristics of the crystal oven. The results presented here are preliminary, limited by the number of sample crystals and oscillators studied so far. The data will be evaluated to identify problems and opportunities with respect to intermittent oscillator operation.

EXPERIMENTAL PROCEDURE

Oven controlled 5 MHz quartz oscillators were subjected to sequences of 4 to 15 repeated power interruptions lasting typically 1, 2, or 3 days and occasionally 4 or 6 hours. After each restart the oscillator frequency was measured continuously by mixing the test signal with a constant reference frequency and feeding the beat frequency, approximately 1 kHz, to a Hewlett-Packard (HP) model 5345 counter. Both reference and counter were phase-locked to a cesium standard. The automatic measurements were controlled by an HP 9825 desktop computer and recorded on magnetic tape. Sampling times of the frequency measurements ranged from 7-14 sec for 3 hour, to 150-300 sec for 3 day runs.

The data reported here were collected on 3 SC and 1 AT cut BVA resonators, and on plated resonators from Frequency Electronics, Inc. (FEI), as well as on 3 Austron oscillators. These samples are designated here as follows: BVA1 through BVA3 are the three SC cut and BVA7 the one AT cut BVA type resonators placed in oscillators from Frequency & Time Systems, Inc. (FTS). The FEI resonators, located in FEI oscillators are labeled

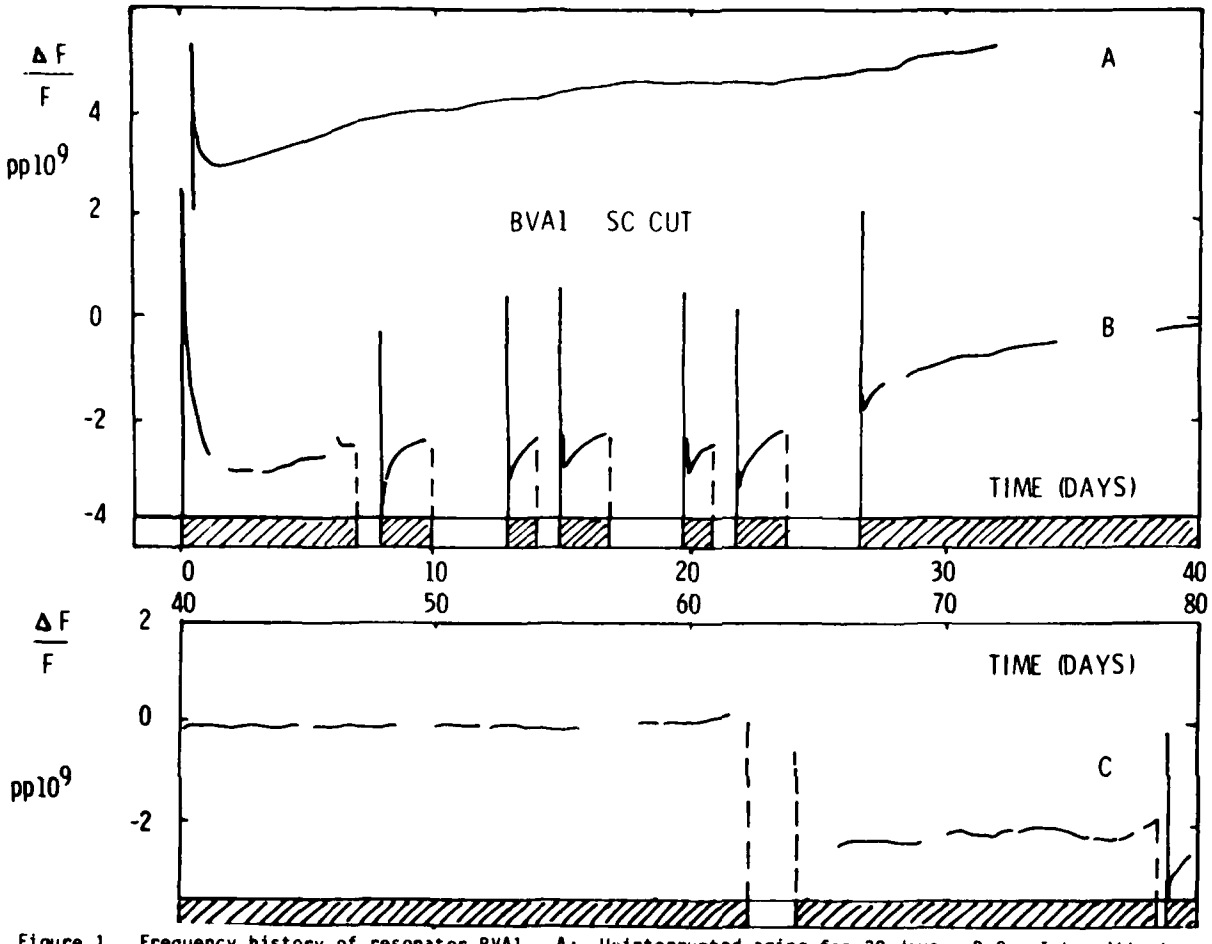


Figure 1. Frequency history of resonator BVA1. A: Uninterrupted aging for 32 days. B,C: Intermittent operation (on: shaded, off: open) for 80 days with 8 sequential restarts R1 through R8. Fractional frequency change $\Delta F/F = 0$ is based on the frequency measured 3 h 15 m after turn-on (B and C only, A is displaced for clarity). Line breaks during on-time indicate operation without data collection.

F1 through F5 for AT, and FS1 through FS4 for SC cut. A1, A2 and A3 are the symbols used for the Austron oscillators. The AT cut crystals were 5th and the SC 3rd overtone resonators.

RESULTS AND DISCUSSION

The fractional frequency offset for sample resonator BVA1, an SC cut, is shown as a function of time in Fig. 1 for continuous as well as intermittent operation. The continuous aging history of curve A was followed by 6 months of undocumented intermittent activity before the retrace experiments shown in curves B and C. Curve B shows the behavior after restarts R1 through R6. The on-off cycling was started after 7 days of continuous operation. The retrace curves appear to follow an overall trend similar to that of the continuous aging curve A. When the unit was fully aged, as after 63 days and then cycled off and on, as in curve C, the retrace is offset by -2×10^{-9} but thereafter behaves again like a fully aged unit.

A superposition of the retrace curves for the first three hours of operation after restarts R1 through R6 is shown in Fig. 2. Characteristic behavior involves a moderately small end-of-warmup positive frequency overshoot characteristic for SC cut resonators. The unit then settles down to a fairly constant frequency. We note that tests R1 to R6 all lie within a fractional frequency deviation bandwidth of 2×10^{-9} after 1 hour of operation. Figure 3 gives extended time performance for the same experiments. Retrace runs R1 through R5 lie within an bandwidth of less than 1×10^{-9} for the 2 day period shown. The larger offset and upward trend of R6 can be attributed to the ongoing aging as shown also in Fig. 1 B. Retrace

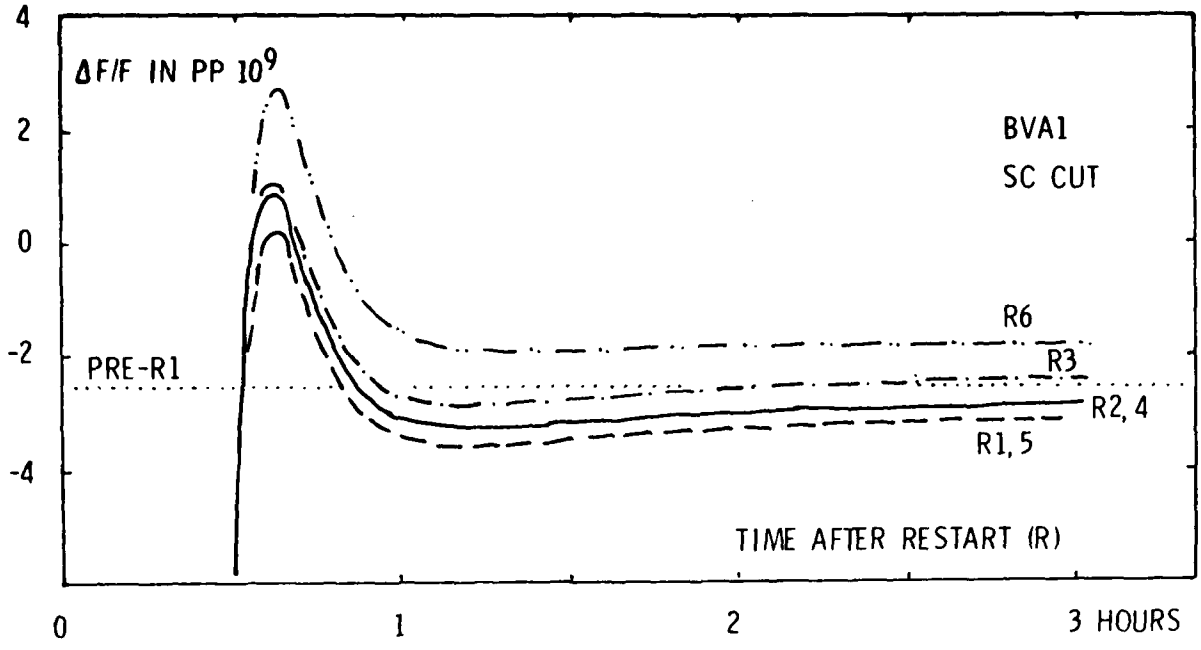


Figure 2. Superimposed short-term retraces after restarts R1 through R6 of resonator BVA1. $\Delta F/F = 0$ as in Fig. 1. Dotted line PRE-R1 is the $\Delta F/F$ level before the first interruption.

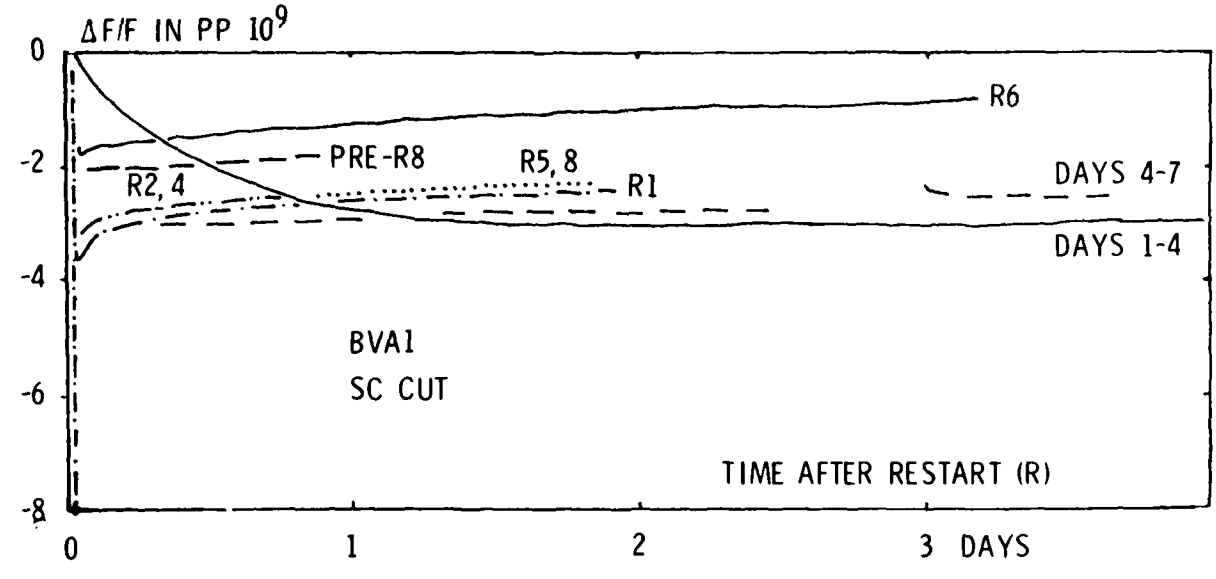


Figure 3. Superimposed 1 to 3 day retraces after restarts R1 through R8 of resonator BVA1. $\Delta F/F = 0$ as in Fig. 1. Aging and stabilization curves DAY 1-4 and DAY 4-7 show the history prior to R1. R2,4 and R5,8 curves are too close to show separately. R3 lies between the R1 and R2,4 curves. R7 omitted because of incomplete data. Dashed line PRE-R8 is the $\Delta F/F$ level prior to interruption 8 (see Fig. 1 C).

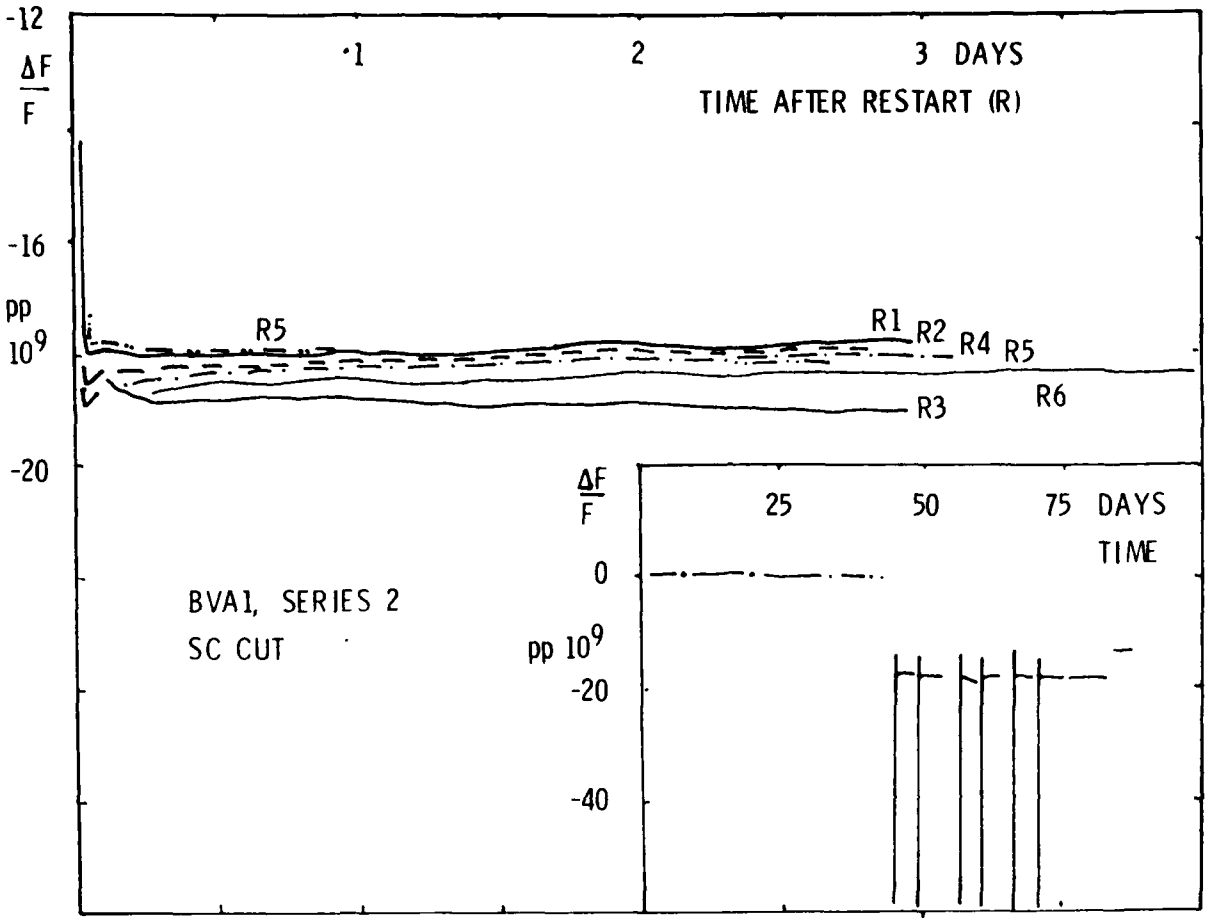


Figure 4. Superimposed retraces from a second series of restarts R1 to R6 of resonator BVA1. Day 0 of the inset is 70 days after completion of the measurements shown in Fig. 1. $\Delta F/F = 0$ was reset manually for this series.

between successive restarts is, in this series, 2×10^{-10} or less with few exceptions, in agreement with refs. 2 and 3.

Data for a second series of restarts with BVA1 is shown in Fig. 4. There were about 100 days of continuous operation between R8 of the first and R1 of the second series. The crystal can be considered well aged at this time. The 6 superimposed retrace curves are shown in the main figure. For moderate time intervals, up to 3 days and beyond, the curves are virtually flat and lie within a bandwith of 1×10^{-9}. Retrace during the first few hours, not shown here, is similar to the non-aged case (Fig. 2). Comparison of Figures 3 and 4 does indicate that the aging trends visible in series 1 retraces are practically removed by the extended aging prior to series 2.

Both series exhibit a drop in frequency when the restart was the first one after protracted operating periods (Fig. 1 C and the inset of Fig. 4). The drop was 17×10^{-9} in the latter and 2×10^{-9} in the former case. Additional measurements are required to establish whether or not such relatively large changes are usually induced by the first restart after extended continuous oscillation.

Data for a second SC cut BVA resonator is shown in Fig. 5. The restart curves of this non-aged unit differ from those of BVA1, given in Fig. 3. The recovery characteristic is seen to extend over a range exceeding 1×10^{-8} with equilibration times in the order of a few days. The differences between curves R1 through R4 reflect the aging shown in the inset.

Figures 6,7 and 8 show data obtained on BVA7, an AT cut crystal. At the end of the oven warmup, after less than 1 hour of operation, the unit exhibits as shown in Fig.6, the deep frequency minimum characteristic for AT cut resonators.

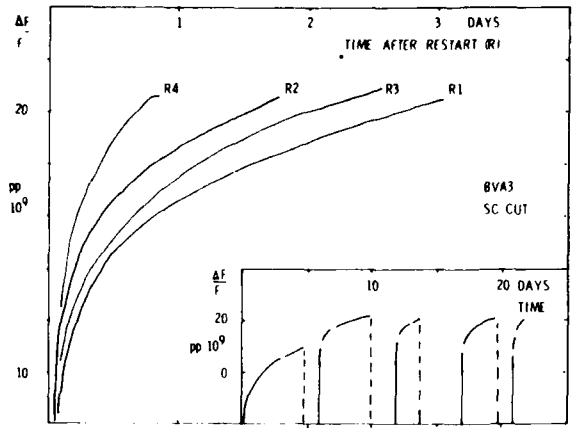


Figure 5. Superimposed retraces after restarts R1 to R4 of resonator BVA3. $\Delta F/F=0$ is based on the frequency measured 1.5 days after turn-on.

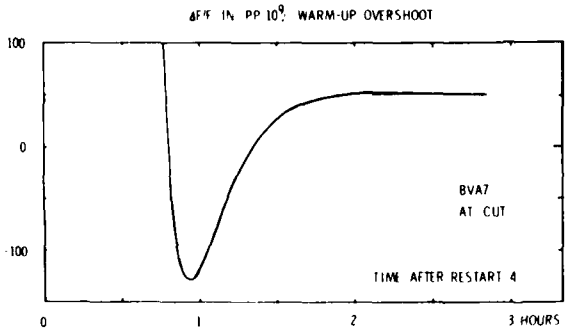


Figure 6. Frequency minimum at end of crystal oven warmup, resonator BVA7, after restart R4. $\Delta F/F=0$ is based on the frequency measured 1 day after the first turn-on.

This minimum is attributed to the large difference between dynamic and static frequency-temperature curves. A superposition of 4 retrace curves is shown in Fig. 7. With this resonator, the large minimum is followed by a relatively shallow maximum about 2 hours after restart. From one restart to the next, these maxima differ by several parts in 10^9 and follow an aging curve. Moderate and long-term results are given in Fig. 8. In the major part of the figure, the retrace is seen to steady down in about 12 hours with all five restarts falling into a bandwidth of approximately 5×10^{-9}. The inset shows a gradual frequency rise which is characteristic of an aging curve. An interesting facet of these data is that the rising aging trend is in all cases begun by a descent, indicating that the retrace data of this unit sit on top of the aging curve.

The positive overshoot before settling

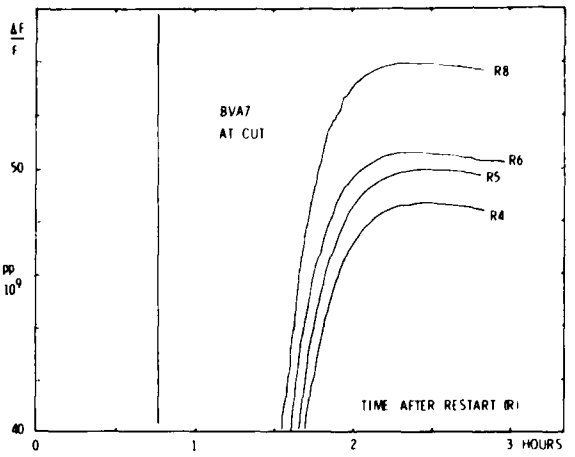


Figure 7. Superimposed short-term retraces after restarts R4 to R8 of resonator BVA7. The vertical line 47 min after restart shows the frequency drop into the end-of-warmup minimum. $\Delta F/F=0$ as in Fig. 6. No data were taken until 3 hours after R7.

down can be attributed to several possibilities; among them are the temperature lag between crystal and oven, the precision with which the oven is set to the turnover temperature of the crystal and the interplay of these factors with the dynamic and static frequency-temperature characteristics.

In the remaining portion of this section, a comparison of retrace characteristics between BVA and plated AT and SC cut resonators will be presented. Figures 9 and 10 show the variety of retrace curve shapes and Fig. 11 compares retrace error data between successive restarts.

Figure 9 shows curves selected from AT and SC cut resonator short-term retrace measurements. They are displaced vertically to show the variety of slopes found. The fast warmup combined with flat retrace shown by oscillator A1 (AT cut) and resonator BVA1 (SC cut) is very desirable. With 1 to 3 day intermittent operation, shown in Fig. 10, the short-term slopes pose no severe difficulties and, e.g., A2 might be preferred over A1 which shows diurnal fluctuations. In this time regime, well aged crystals tending to show flat retrace curves such as BVA1, FS1 and FS3 are clearly preferred.

The actual retrace error between restarts is not always easy to determine. For our purposes we decided to base it on measurements taken 1 to 3 days after restart. If the two successive retrace curves were not parallel we averaged the fractional frequency differences. First retraces after prolonged continuous operation were excluded from the comparison, because they showed larger frequency changes which should be analyzed separately. Finally, the long-term aging component was subtracted from the measured differences. After this conditioning the scattered points shown in

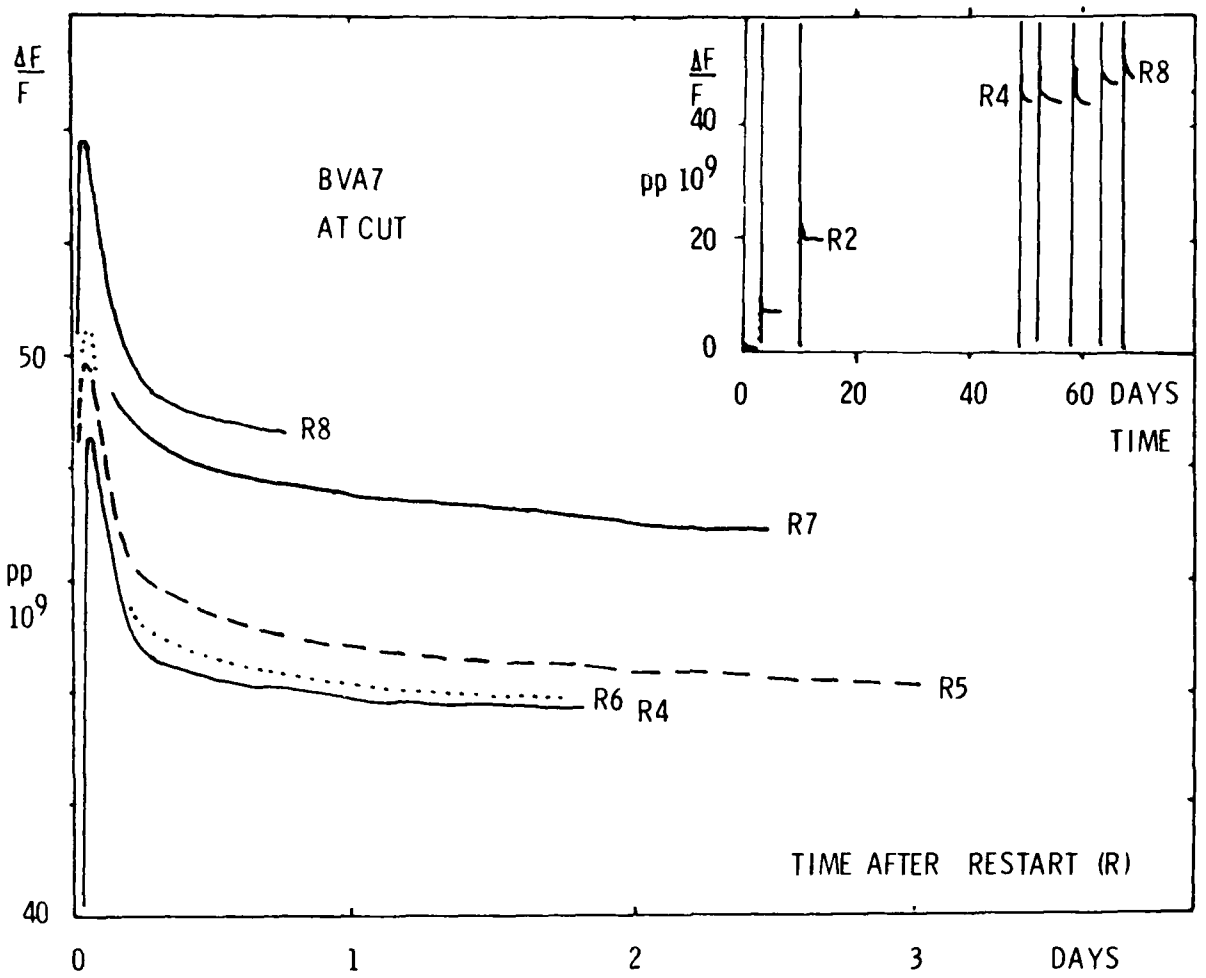


Figure 8. Superimposed 1 to 3 day retraces after restarts R4 to R8 of resonator BVA7. $\Delta F/F=0$ as in Fig. 6. Note opposing trends in long-term aging and retraces shown in the inset.

this conditioning the scattered points shown in Fig 11 were obtained. We note that the AT cut data show substantially larger dispersion than those for SC cut, presumably related to the inherent reduced stress sensitivity of the latter. Also the AT cut does not show any noticeable difference in distribution as a function of interruption time. The SC cut resonators do exhibit a slight downward trend which appears to vary directly with length of turn-off. Within each cut category, no significant retrace error differences were found between BVA and plated resonators and between oscillators of different manufacturers.

CONCLUSIONS

Frequency retrace measurements on a limited number of quartz oscillators utilizing BVA and plated resonators have been reported. If the resonator is in its aging phase the retrace error is added to the aging drift, whereas with well aged resonators the frequency seeks a new level characteristic for intermittent operation. The new level may occur after the first interruption following prolonged continued operation with some crystals but may establish itself only after several on-off cycles with others. Large, interruption-induced frequency shifts have been observed occasionally. In general, retrace errors show distinctly less spread with SC cut than with AT cut resonators. Comparison of the retrace curve slopes shows the need for careful selection of crystals, oscillators and ovens. Retrace as close as a few parts in 10^{10}, although definitely feasible, cannot yet be reliably obtained on a production basis.

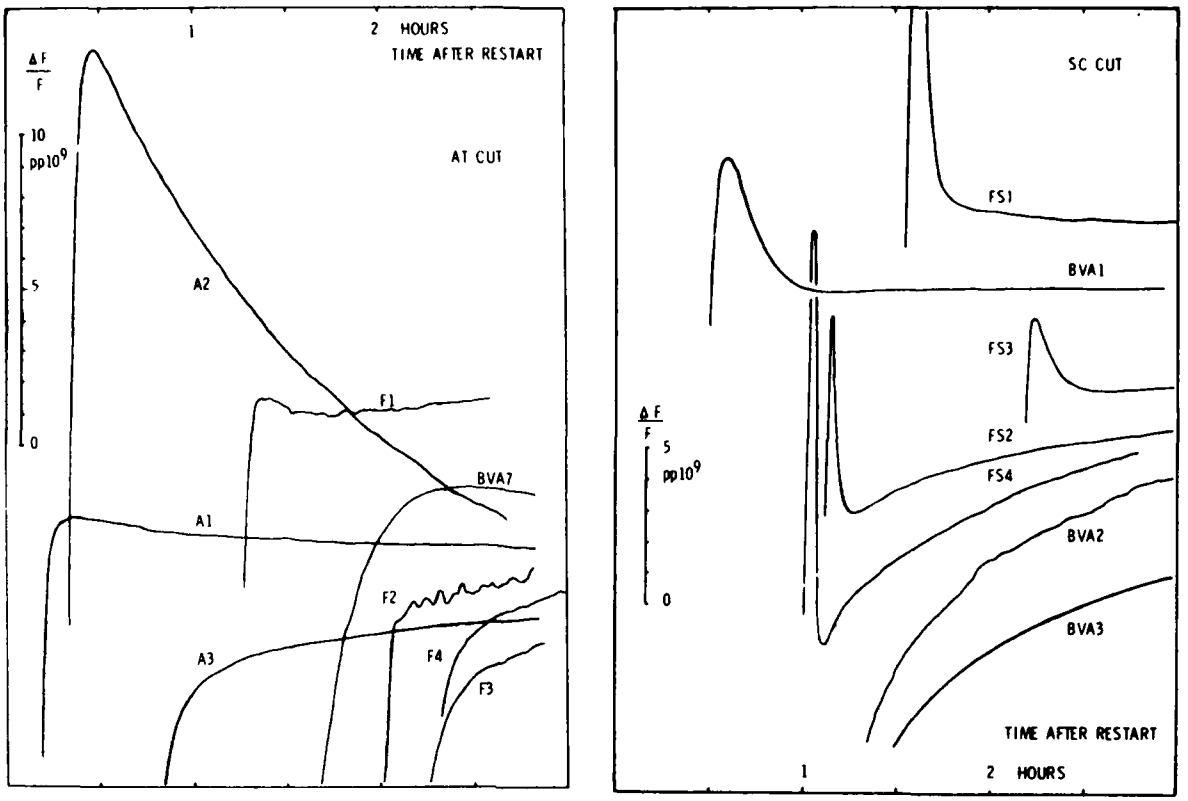


Figure 9. Short-term retrace curves found with AT and SC cut resonators. All curves are displaced vertically to show the variety of slopes encountered. With AT cut resonators, the endof warmup minima have been omitted for clarity.

ACKNOWLEDGMENTS

The authors wish to recognize the conscientious technical assistance of Mr. Paul A. Ligor. They also wish to thank Dr. Don Emmons for stimulating and helpful discussions and the loan of an oscillator.

REFERENCES

1. H. W. Jackson, Proc. 34th AFCS, 449 (1980).
2. R. J. Besson and U. R. Peier, Proc. 34th AFCS, 175 (1980).
3. R. J. Besson, P. G. Girardet, E. P. Graf, and D. A. Emmons, Proc. 34th AFCS, 457 (1980).

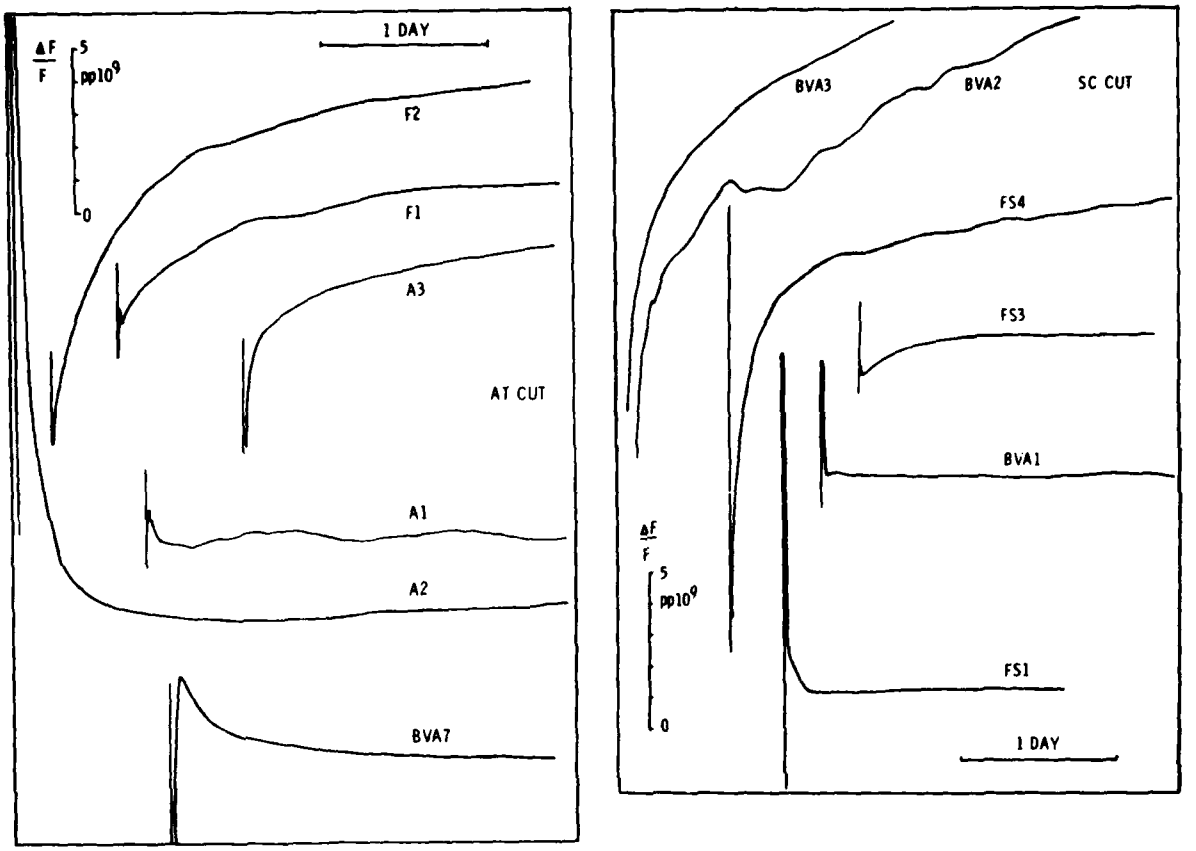


Figure 10. One to three day retrace curves found with AT and SC cut resonators. All curves are on the same relative scales as indicated. They are displaced vertically to show the variety of slopes and horizontally for clarity.

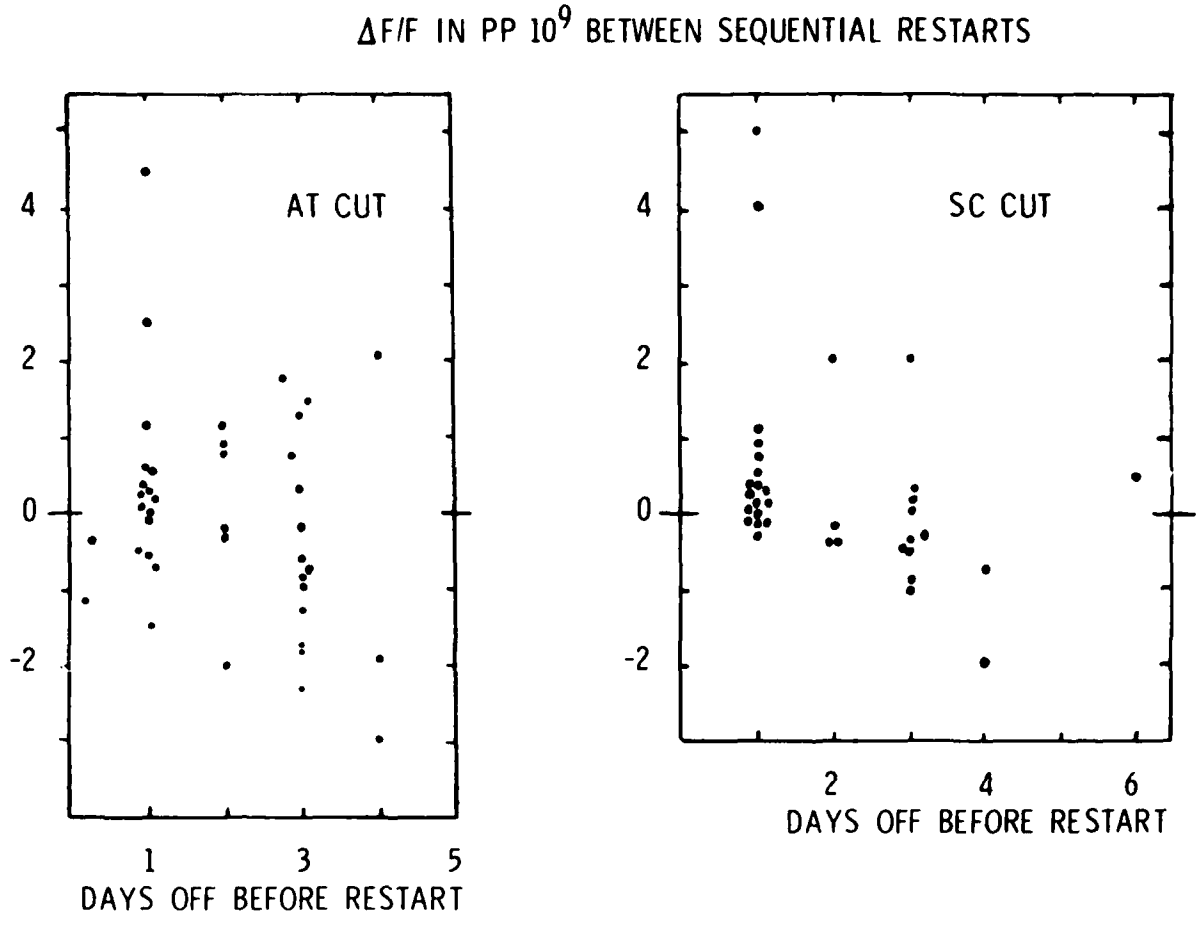


Figure 11. Fractional frequency difference $\Delta F/F$ between retraces after successive restarts, displayed as a function of the interruption length before each restart, for AT and SC cut resonators.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

FREQUENCY CONTROL REQUIREMENTS FOR 800 MHZ LAND MOBILE COMMUNICATION

Robert Kinsman and David Gunn

Motorola Inc. Communications System Division
Schaumburg, IL 60196

ABSTRACT

In 1968, proceedings were initiated by the Federal Communications Commission (FCC) with the goal of making more frequency spectrum available for Mobile Communications use. As a result of this study on FCC Docket 18262, UHF-TV channels 70-83 covering the frequency range of 806 to 890 MHz were converted for use as Land Mobile Communications Channels. Currently 30 MHz of spectrum is available for conventional and trunked Mobile Radio use and 40 MHz has been allotted for Mobile Radio Telephone Systems. The Radio Telephone band is restricted for use in a "cellular" system which allows for reuse of frequencies, thus maximizing the number of users in a given area. Two experimental systems are now in operation in Chicago and the Washington-Baltimore area. Other segments of the band are being held in reserve for future use. Currently the FCC is Considering the allocation of a frequency band at 929-932 MHz for wide area paging operations.

The authorization for use of these frequency bands has created new business opportunities for the radio manufacturers and new challenges for the design engineer. This equipment requires a 2 to 1 improvement in frequency-temperature stability and tighter tolerance selectivity as compared to current 450 MHz radios.

This paper describes the requirements of this new frequency band, the type of equipment which is being supplied, and how it affects Frequency Control Technology.

INTRODUCTION

Since its inception in the early 1930's, the use of mobile two-way radio equipment has steadily increased. Today there are over 6.1 million transmitters licensed for land mobile use in the United States alone. As available spectrum is used up, particularly in the large cities, increasing pressure has been exerted on the FCC to authorize use of new frequency bands. In response to these needs, a study was initiated by the FCC in 1962 to review the allocation of of frequency spectrum in the 806 to 947 MHz band. This study on FCC Docket 18262 resulted in the conversion of UHF-TV channels 70-83, covering the frequency range of 806-890 MHz, for use as land mobile communication channels.[1] In this new band, 30 MHz of spectrum was made available for conventional and trunked land mobile radio use and 40 MHz for land mobile telephone systems. In addition, the FCC is considering the allocation of channels at 929-932 MHz for use in wide-area paging operations.

Recent developments in high frequency synthesizers and RF power devices as well as tighter tolerance frequency sources have made it possible to build reliable, efficient radios in this frequency band at a reasonable cost.

To achieve maximum utilization of this new frequency band, two new system concepts are authorized. The "trunked" system will allow many users to share up to 20 channels and will require a radio which will automatically switch to an open channel. The radio telephones will use a "cellular" system which allows for reuse of frequencies in a given area.

For the conventional and trunked systems, channel spacing is the same as in the 450 MHz band. Thus, for comparable operation, frequency tolerances are essentially cut in half. In the transmitter, the frequency tolerance is established by FCC rules. Other problems due to oscillator sideband noise are created because of the higher operating frequency. In the receiver, stability of the IF filters, in addition to the local oscillator, must be considered to insure proper operation under all conditions. The selectivity of the receiver is also affected by the sideband noise of the local oscillator. A description of these factors, how they affect overall radio performance, and methods of measurement will be included in this paper.

HISTORY[2]

The early development of mobile radio was spearheaded by various police departments who obviously would benefit by instant communications with moving

vehicles. The first successful use of mobile receivers ' attributed to the Detroit Police Department in 1928. Operating on 94 Meters, receivers were built which could withstand the rigors of the automobile environment. In 1933, two-way radios were successfully placed in operation by the Bayonne, NJ Police Department. This equipment operated at 33.1 MHz without crystal control.

On October 13, 1937, FCC order #19 was issued allocating 29 channels for police department use in the frequency range of 30.58 to 39.9 MHz. By this time, the use of crystal control was generally accepted and in 1938, the FCC established a frequency tolerance of $\pm$ .05% for all transmitters operating above 30 MHz. In 1940, the state of Connecticut established a state wide system for the State Police Department. This system was the first to be converted from AM to FM to take advantage of FM's superior rejection of AM type interference from ignition noise and flutter in a moving vehicle. This system was the forerunner of today's land mobile radio industry.

In 1940 with only a few thousand transmitters in use, interference was minimal. However, since that time efforts have been concentrated on more efficient spectrum utilization as user demand increased. This has resulted in the opening of higher frequency bands and the splitting of channel spaces as the development of new technology made the needed equipment practical. For example, the 150 MHz band was originally established for 120 KHz channel spacing. This was then split to 60 KHz and again to 30 KHz. The need for more channels was met in some of our major cities by converting UHF-TV channels 14 thru 20 in the 470 to 512 MHz band for limited use. The dramatic increase in the use of mobile radio is illustrated by the number of licensed transmitters authorized over the past 30 years.

Year	Transmitters
1950	180,000
1960	1,300,000
1970	3,500,000
1980	6,100,000

As a result of the latest FCC proceedings, additional spectrum is now available in the 806-947 MHz band. By providing for use of the latest computer technology, even more efficient spectrum utilization is achieved through trunking and the cellular radio telephone system.

NEW 800 MHZ SYSTEMS

The recent actions of the FCC as a result of the study on FCC Docket 18262 now provide an additional 70 MHz of frequency spectrum for land mobile radio use. The allocation of these frequencies is shown in Figure I. Interspersed in the 806 to 947 MHz band are several reserve bands which will be considered for allocation in the future. The systems which are now authorized are described here.

Conventional Dispatch System

The operation of the new 800 MHz radios in the conventional dispatch system is identical to that used in the lower frequency bands. Radios are supplied with from 1 to 4 channels depending on the needs of the user. The radio operator manually selects the channel to be used and must monitor the selected channel before transmitting. The mobile unit typically has plug-in crystal oscillators for each operating channel and has a nominal power output of 35 watts with a working range of about 15 miles radius from the base station. For wider area coverage, repeaters are used. Hand held radios are also in use with typical power outputs of 1.5 watts.

Trunked Dispatch Systems[3]

The technique of trunking has been used for many years in the telephone system. Now with the advances made in microprocessors, large scale integration (LSI), and digital technology, it has become practical to apply the technique to land mobile communications systems. Simply stated, trunking is the automatic sharing of a small number of channels among a relatively large number of users. Channel assignments are automatically assigned by a control computer which can equalize the radio traffic over all the channels and therefore accommodate more users per channel. This also simplifies operation for the user because of shorter access time and no need to monitor before transmitting. The mobile unit is a multichannel radio (5 to 20) that utilizes a digital frequency synthesizer and has a nominal power output of 35 watts. The repeaters are also multichannel with a nominal 125 watt power output. Digital signalling is used for control commands in the system.

Cellular Mobile Telephone Systems[4]

In a cellular system, the coverage area is divided into cells. Each cell is assigned a number of channels to be used by mobiles within the cell. All cells are under the control of a master computer that assigns channels, power outputs, and automatic transfers as a mobile passes from one cell to another. Because of the controlled operations, channel frequencies can be re-used in non-adjacent cells. System expansion can be accomplished by splitting a cell into several smaller cells to allow increased user density. A cellular system mobile unit uses a digital frequency synthesizer to provide operation on 666 channels. Power output is 3 - 5 watts although some systems are being designed to accommodate 1 watt portable units. The signalling system is common for all systems, so that a user can roam to any area in the United States and still obtain service. Each cell site has multiple receivers and transmitters and is connected by land lines to the computer master switch that controls all

cells. The nominal power of a cell transmitter is 25 watts. Average cell sizes vary from 3 to 8 miles in radius.

FREQUENCY CONTROL ASPECTS

Transmitter frequency tolerances as established by the FCC are shown in Table 1 for all currently authorized Land Mobile Radio use. These tolerances must be held under all operating conditions. The rated system deviation and channel spacing are also specified by the FCC. These are shown in Table 2 for the narrow band FM systems now used throughout the country. For conventional systems, FM deviation is now standardized at +5 KHz on all bands but the channel spacings are a result of channel splitting from when the bands were originally opened. The historical impact on technology can be observed from Table 1. As each new band was opened at a higher frequency, tighter frequency tolerances were needed. Without continued improvements in crystals, ovenized oscillators and TCXOs, it would not be possible to build the higher frequency radios in large volume at a reasonable cost.

For the new cellular mobile telephone system rated deviation is +12 KHz to permit the use of high speed signalling. Channel spacing is only 30 KHz, but adjacent channels will not be used in the same cell.

Although the basic transmitter frequency stability and channel spacing are specified by the FCC, the Land Mobile Radio Industry has worked to standardize transmitter and receiver performance. In the United States, the Electronic Industries Association (EIA) has established standard operating conditions and minimum specifications for portable and vehicular radios as well as standard test methods.[5] On an international basis, this work is being carried out by the International Electrotechnical Commission (IEC).[6] Efforts are now in progress to standardize test methods worldwide. As is often the case, it is much easier to write a specification than it is to specify a reasonable method to determine conformance.

TRANSMITTERS

In its simplest form a transmitter can be built as shown in Figure 2. In this arrangement, a frequency modulated signal is generated by use of a voltage controlled crystal oscillator. For the tighter tolerance units, temperature compensation is also included. The output of the crystal oscillator is then multiplied to the operating frequency and amplified to the desired power level. Many of the newest model radios use a frequency synthesizer particularly for trunked and radio telephone systems where automatic channel switching is required. For these, a crystal oscillator is used to determine overall frequency stability, but modulation is usually incorporated in the phase lock loop. The transmitter performance characteristics directly affected by crystal oscillator performance are described in the following section:

1. Carrier Frequency Stability

   Transmitters are required to hold the frequency tolerances shown in Table 1 over a temperature range of -30°C to +60°C to meet EIA standards. In a vehicular radio, temperature rise of the equipment is a significant factor and normal temperature specifications are extended to -30°C to +85°C. For systems requiring ±.0005% stability, various temperature compensation schemes have been used incorporating both passive and active circuitry. However, for 800 MHz equipment, ±.00025% stability must be held. To allow for other system variations, the TCXO is normally held to ±.0002%. This is the first time that TCXO's must be built to this very tight tolerance in large quantities. Whereas looser tolerance oscillators could be tested at discrete temperatures, these units must be continuously tested across the temperature range. Oscillators have been developed which incorporate custom IC's with laser trimmable resistors for temperature compensation and a system has been established which tests the units under computer control and then automatically trims the temperature control circuitry.[7]

   In addition to temperature stability, aging must also be considered. The FCC requires that all transmitters be checked annually for frequency compliance. In practice however, radios are normally checked on a 6 month cycle. To insure meeting FCC requirements, good quality crystal oscillators must age less than 1 ppm per year.

2. FM Hum and Noise Level

   FM Hum and Noise level is defined as the ratio of residual frequency modulation to standard test modulation of the transmitter output with a standard FM test receiver. The transmitter is first modulated with a standard test signal which is defined as a 1000 Hz tone at the level required to produce 60% of full rated system deviation (±3 KHz). Then the modulation is removed and the residual noise output of the test receiver is noted. The ratio of the modulated signal to the unmodulated noise is then computed. For a good quality transmitter this ratio should be at least 60 db. The residual frequency modulation is a direct result of the close-in sideband noise of the unmodulated carrier as illustrated in Figure 3. The standard test receiver has a flat passband to ±5KHz and an audio passband from 0 - 3000 hz with a 6 db per octave de-emphasis. Therefore, the

noise in the band $f_o \pm$ 5KHz will be detected. This places a severe requirement on our VCXO in Figure 2 due to the increase of sideband noise level as a function of multiplication. An 800 MHz transmitter would require an oscillator with a 6 db improvement in sideband noise to give performance equal to a 450 MHz transmitter.

3. Transmitter Sideband Noise

Transmitter Sideband Noise is defined as the undesired emission of noise by the transmitter and is measured in the adjacent channels and as far away from the carrier frequency as is practical. The noise spectrum is generated in the VCXO and is degraded as a direct function of the multiplication process. In addition the noise floor may be limited by the noise figure of the RF power amplifier. The sideband noise is measured by zero-beating the transmitter output against a high quality signal source and measuring the output with a low frequency spectrum analyzer. The EIA currently specifies a minimum of 55 db attenuation at the adjacent channels for transmitters at 450 MHz. However, more recent studies have indicated that a minimum standard of 70 db is needed.[8] To maintain this quality in an 800 MHz transmitter requires either an improved crystal oscillator or a higher frequency oscillator with identical sideband noise levels. If a synthesizer is used the same restrictions apply to the output VCO.

RECEIVERS

A typical receiver arrangement is shown in Figure 4. Commonly used IF frequencies are 10.7 MHz or 21.4 MHz in receivers which incorporate monolithic crystal filters. Eight to ten poles of selectivity are generally used with a 6 db bandwidth of 12 - 13 KHz. The local oscillator may be a crystal oscillator with multipliers or a synthesizer. The basic function of a receiver is to provide maximum sensitivity and adequate gain at the assigned channel and maximum rejection or selectivity at adjacent channels as well as other interfering signals. Accurate receiver testing is difficult because of the excellent sensitivity and narrow band selectivity of current high quality receivers. Several new test methods and definitions are under evaluation by EIA and IEC committees. Some of the major characteristics and methods of test are described below.

1. Reference Sensitivity (Usable Sensitivity

This determines the basic sensitivity of the receiver and is used as a reference for many of the other tests. The test is performed in the set up shown in Figure 5. The standard test signal at the input is frequency modulated with a 1 KHz tone. Deviation is 60% of the rated system deviation ($\pm$3 KHz). The audio output of the receiver is terminated in it appropriate load and the signal is measured with an audio distortion analyzer. The "SINAD" method of measurement is used. SINAD is an acronym for "signal plus noise plus distortion to noise plus distortion ratio" expressed in decibels where the "signal plus noise plus distortion" is the audio power recovered from a modulated radio frequency carrier, and the "noise plus distortion" is the residual.[9] This ratio is a measure of audio output signal quality. To perform the measurement, the standard test signal is set to the specified channel frequency and the signal is adjusted to a level such that the SINAD equals 12 db. This signal level is defined as the Reference Sensitivity. The EIA specifies a minimum standard of 1 microvolt for receivers in the 450 and 800 MHz band. Todays highest quality receivers achieve a sensitivity of 0.35 microvolts at 450 and 800 MHz.

2. Variation of Sensitivity with Signal Frequency (Usable Bandwidth)

In receiver design it is desirable to use as narrow an IF bandwidth as possible to limit the noise response and therefore, the ultimate sensitivity. The minimum bandwidth is determined by the deviation of the carrier (modulation) and the frequency stability of the oscillator injection. The set up for this measurement is identical to that used for the reference sensitivity. As the carrier frequency is shifted away from the nominal channel frequency, the signal level is increased to maintain 12 db SINAD. Figure 7 shows a plot of signal level versus frequency for a 450 MHz radio-telephone receiver. This particular receiver has 12 poles of selectivity with an overall 6 db bandwidth of 14 KHz. The Usable Bandwiuth is determined by the frequencies which are 6 db above the Reference Sensitivity; from Figure 7, -4.0 KHz and +4.3 KHz. The Minimum Usable Bandwidth is defined as the smaller of the frequency differences ie, +4.0 KHz for our example. Thus the Minimum Usable Bandwidth is affected by filter symmetry and injection oscillator frequency accuracy in addition to overall receiver bandwidth. To insure system operation within the Minimum Usable Bandwidth, temperature stability of the transmitter, receiver, local oscillator and IF filters must all be accounted for. For our example, this translates into a total accumulative drift of +4.7 ppm at 850 MHz. EIA standards are still under discussion for this specification.

3. Adjacent Channel Selectivity

To measure Adjacent Channel Selectivity,

a second signal source is connected to the combining network in Figure 5. Signal generator number 1 is set on the channel frequency and is modulated with the standard test signal (1 KHz at ± 3 KHz deviation). The signal level is set 3 db above the Reference Sensitivity. Signal generator number 2 is set at the adjacent channel frequency and is modulated at 400 Hz with ± 3 KHz deviation. The signal level is increased until the receiver SINAD is reduced to 12 db. This measurement is performed on both the upper and lower adjacent channels. The Adjacent Channel Selectivity is defined as the ratio of the interfering signal to the Reference Sensitivity.

In high frequency receivers, the Adjacent Channel Selectivity is limited by a phenomenon called "desensitization" as well as by the selectivity of the IF filter. In this condition, the sideband noise of the local oscillator mixes with an interfering signal. Essentially the roles of the local oscillator and the interfering signal are reversed. Thus a receiver may have 100 db of IF selectivity at the adjacent channel and yet only achieve 90 db under test due to "desense". It has been found experimentally that the sideband noise of the local oscillator must be 6 to 10 db better than the selectivity desired. Thus, for a 90 db receiver, the sideband noise of the injection frequency should be at least 100 db in the adjacent channel (10 KHz bandwidth) or 140 db/Hz. In a typical 450 MHz receiver using a fundamental mode crystal oscillator and a X24 multiplier, the oscillator sideband noise must be 168 db/Hz at the adjacent channel.

The EIA requires a minimum Adjacent Channel Selectivity of 70 db. High quality receivers typically achieve 80 or 90 db.

4. Spurious Response Immunity (Spurious Response Attenuation)

Spurious attenuation is measured in exactly the same way as adjacent channel selectivity. Spurious responses are principally caused by mixing of harmonics of the input signal and the local oscillator. These can be predicted mathmatically and are minimized by an appropriate combination of IF frequency and front end selectivity in the receiver design. The EIA minimum standard is 85 db. Typical high quality receivers achieve 100 db.

5. Intermodulation Immunity (Intermodulation Spurious Response Attenuation)

Intermodulation spurious responses are generated in a receiver when two or more undesired signals are mixed in a non-linear portion of the receiver. The test is performed in the test circuit of Figure 5 with two additional signal generators connected to the combining network. Generator 1 is set at the nominal channel frequency with the standard test modulation. Generator 2 is unmodulated and is set at the adjacent channel frequency ($F_a = f_a + \Delta f$). Generator 3 is unmodulated and is set at a frequency twice the adjacent channel away from the nominal frequency ($f_b = f_o + 2\Delta f$). Interference occurs due to mixing of harmonics of f_a and f_b in such a way that in-band frequencies are generated. For example:

$$2f_a - f_b = f_o$$

The test is performed by setting Generator 1 to the Reference Sensitivity and then increasing its level by 3 db. Generators 2 and 3 are set at equal signal levels. This level is then increased until the receiver SINAD is reduced to 12 db. The Intermodulation Immunity is defined as the ratio of the level of Generator 1 or 2 to the Reference Sensitivity.

Intermodulation Immunity may be limited by non- linearity in the monolithic crystal filters.[10,11] For high quality receivers, the first crystal filter in the IF (adjacent to the first mixer) must meet very stringent requirements. These parts are normally processed specially or selected from standard parts. The EIA minimum standard for Intermodulation Immunity is 60 db. High quality receivers typically achieve 80-85 db.

SUMMARY

Transmitters

The FCC requirements for the new 800 MHz mobile radios specify a frequency tolerance of ±.00025% over the temperature range. This is one-half the tolerance required for current VHF and UHF radios and will require substantial improvements in design and testing of large volume production TCXO's. Oscillator sideband noise is also a critical factor in maintaining clean transmitter emissions on channel and at adjacent channels. With reference to current 450 MHz radios, the VCXO or synthesizer VCO must have a 6 db improvement in sideband noise or maintain the same sideband noise level at twice the operating frequency for proper transmitter performance.

Receivers

For conventional land mobile radios at 800 MHz the Rated System Deviation and Channel Spacing are identical to the 450 MHz band. Thus the

required selectivity will not change for the new band. However, frequency stability and accuracy of the crystal filters and local oscillator are much more critical. To maintain operation within the Usable Bandwidth over temperature and time, the IF crystal filter must be symmetrical about f_o and the local oscillator must maintain frequency within 2 to 3 ppm under all conditions. For the Cellular Radio Telephone a new rated deviation is specified creating a need for a new family of monolithic crystal filters. Adjacent Channel Selectivity is limited by local oscillator sideband noise due to desensitization. To maintain selectivity comparable to 450 MHz receivers the local oscillator sideband noise must be improved to compensate for multiplication to the higher operating frequency.

CONCLUSIONS

The recent action by the FCC to allocate portions of the 806 - 947 MHz frequency band for Land Mobile Radio use has created 600 new channels for conventional radio users and 666 new channels for radio telephone service. Two new systems are authorized which incorporate automatic computer controlled channel selection; the trunked system and the cellular radio telephone system. Frequency control needs include a 2 to 1 improvement in oscillator stability and sideband noise over present 450 MHz radio equipment. Tighter tolerance monolithic crystal filters are required to provide more exact passband characteristics. Finally, a new family of monolithic filters is needed to meet the bandwidth requirements of the cellular radio telephone.

REFERENCES

1. Land Mobile Radio Service, Second Report and Order, Docket 18262, 46 FCC 2d 752 (1974), Federal Communications Commission, Washington D.C., 20554

2. D.E. Noble, "The History of Land Mobile Radio Communications", Proceedings of the IRE, pp. 1405 - 1414, Vol. 50, No. 5, May, 1962.

3. S. Thro, "Trunking - A New System Configuration for Fleet Dispatch Communications", Proceedings 29th IEEE Vehicular Technology Conference, pp. 302 - 306, 1979

4. M. Stern and V. Graziano, "A New Portable and Mobile Radiotelephone System Goes to Washington", Telephony, pp. 26 - 30, Vol. 193, No. 3, July 18, 1977

5. Minimum Standards for Land Mobile Communications FM or PM Receivers, 25 - 470 MHz, RS - 204-B, and Minimum Standards for Land Mobile Communications FM or PM Transmitters, 25 - 470 MHz RS-152-B, Electronic Industries Association, Washington D.C. 20006

6. Methods of Measurement for Radio Equipment Used in the Mobile Services, IEC Standard 489-3, Bureau Central de la Commission Electrotechnique Internationale, Geneva, Switzerland

7. T. Keller, D. Marvin and R. Steele, "Integrated Circuit Compensation of A.T. Cut Crystal Oscillators", Proceedings 34th Annual Frequency Control Symposium, pp. 498-503, 1980

8. J.H. McMahon, "Interference and Propagation Formulas and Tables Used in the Federal Communications Commission Spectrum Management Task Force Land Mobile Frequency Assignment Model", IEEE Transactions on Vehicluar Technology, pp. 129-134, Vol. VT-23, No. 4, November, 1974

9. IEEE Test Procedure for Frequency-Modulated Mobile Communications Receivers, IEEE Transactions on Vehicular Technology, pp. 85-99 Vol. VT-18, No. 2, August, 1969

10. S. Malinowski and C. Smith, "Intermodulation in Crystal Filters", Proceedings 26th Annual Frequency Control Symposium, pp. 180 - 186, 1972

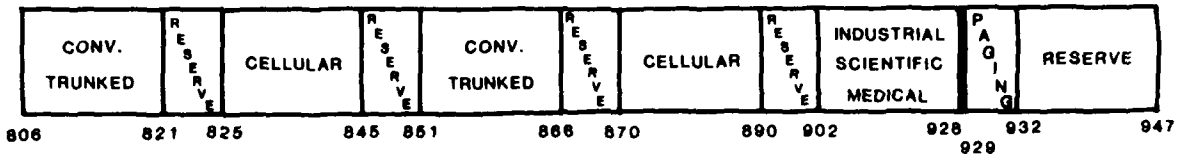
11. W. H. Horton and R.C. Smythe, "Experimental Investigations of Intermodulation in Monolithic Crystal Filters", Proceedings 27th Annual Frequency Control Symposium, pp. 243 - 245 1973

FREQUENCY RANGE	FIXED AND BASE STATION	MOBILE STATION	
		OVER 3 WATTS	3 WATTS OR LESS
25 - 50 MHZ	.002 %	.002 %	.005 %
132 - 174 MHZ	.0005 %	.0005 %	.005 %
450 - 512 MHZ	.00025 %	.0005 %	.0005 %
806 - 890 MHZ	.00015 %	.00025 %	.00025 %

TABLE 1
AUTHORIZED TRANSMITTER FREQUENCY TOLERANCES

FREQUENCY RANGE	CHANNEL SPACING	RATED SYSTEM DEVIATION
25 - 50 MHZ	20 KHZ	± 5 KHZ
132 - 174 MHZ	30 KHZ	± 5 KHZ
450 - 512 MHZ	25 KHZ	± 5 KHZ
806 - 890 MHZ (Conventional)	25 KHZ	± 5 KHZ
806 - 890 MHZ (Cellular)	30 KHZ	± 12 KHZ

TABLE 2
RATED SYSTEM DEVIATION AND CHANNEL SPACING

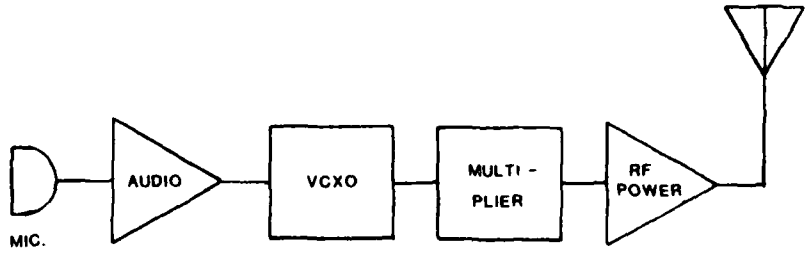


CONVENTIONAL
AND TRUNKED - 600 CH 25 KHZ/CH

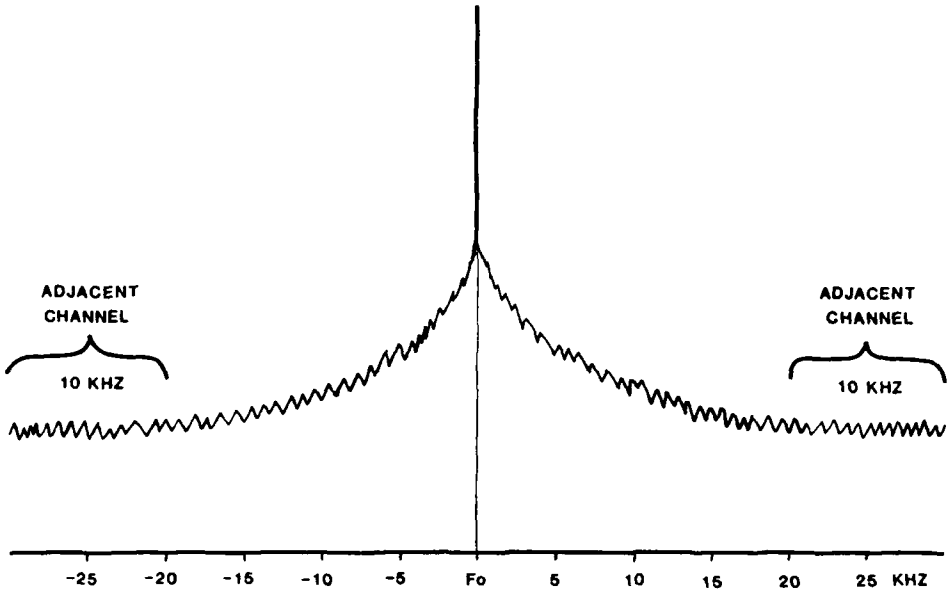
CELLULAR - 666 CH 30 KHZ/CH

806 - 947 MHZ FREQUENCY ALLOTMENT

FIGURE 1

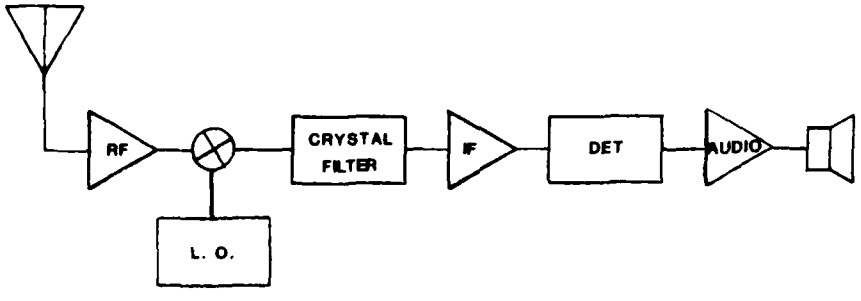


TRANSMITTER
FIGURE 2



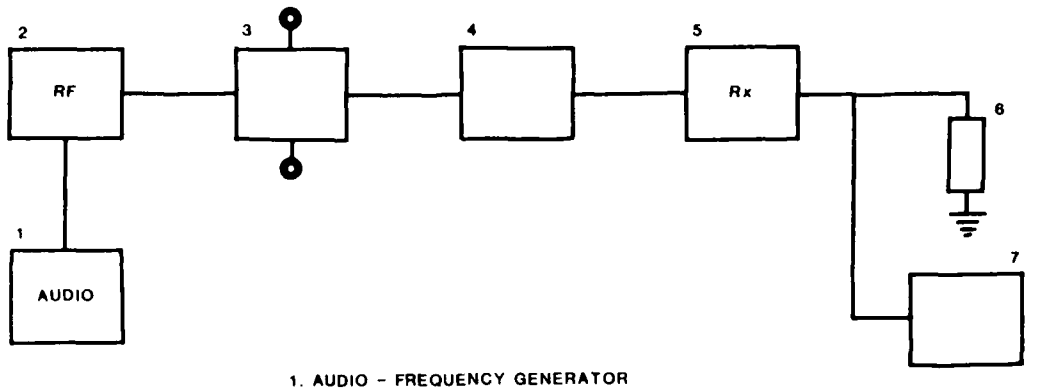
OSCILLATOR SIDEBAND NOISE

FIGURE 3



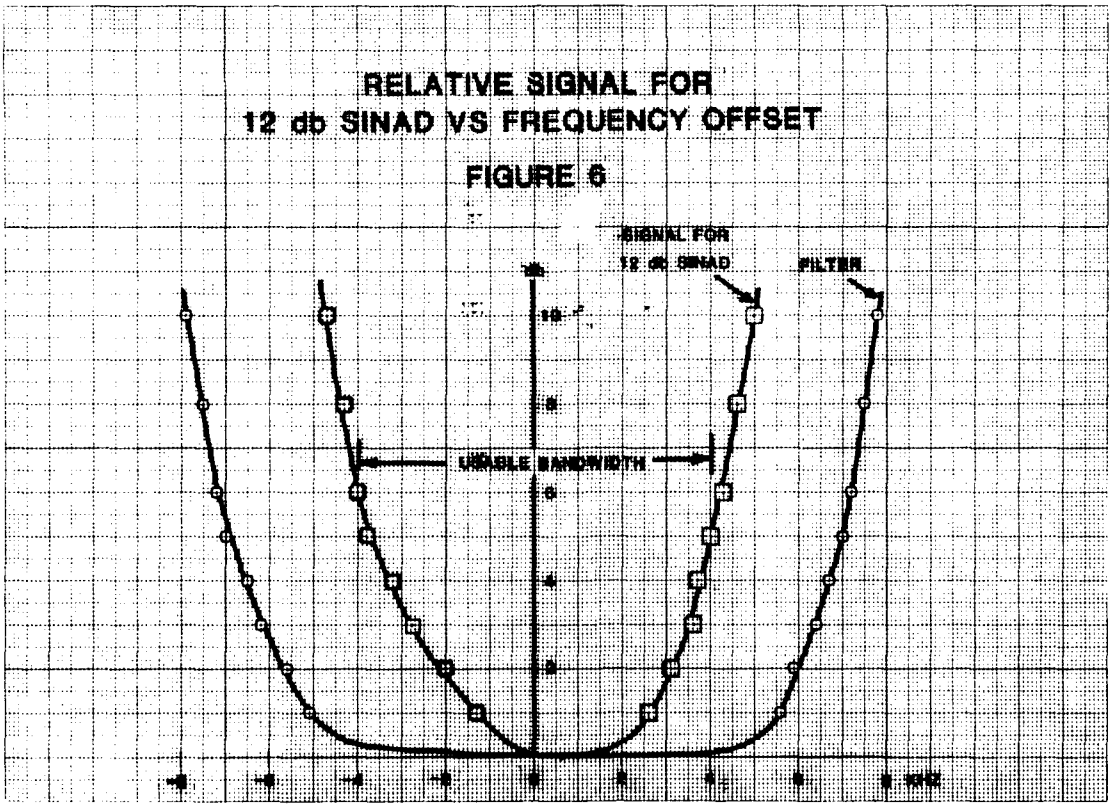
RECEIVER

FIGURE 4



1. AUDIO - FREQUENCY GENERATOR
2. RADIO FREQUENCY SIGNAL GENERATOR
3. MATCHING OR COMBINING NETWORK
4. ARTIFICIAL ANTENNA, IF REQUIRED
5. RECEIVER UNDER TEST
6. AUDIO FREQUENCY LOAD
7. NOISE AND DISTORTION METER

RECEIVER TEST SET UP.



Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

FREQUENCY STABILIZATION REQUIREMENTS FOR MODERN MILLIMETER WAVE SYSTEMS

A. E. Firkel

TRW Defense and Space Systems Group
One Space Park
Redondo Beach, California 90278

Summary

Since the advent of millimeter wave solid state devices, there has been increasing interest in incorporating them in small, lightweight sophisticated systems. The particular applications have been radar transceivers, high data rate transmission systems, secure communication networks, and atmospheric and deep space probes. The advantages of millimeter waves, i.e., higher bandwidth, higher antenna gain, smaller component size, and lower probability of intercept and jamming, are coupled with some inherent disadvantages. After consideration of the key parameters, it becomes apparent that, in the area of radar systems, the problem of oscillator stability has not received as much attention as other issues. Relatively modest phase noise reduction that is achievable by standard processing techniques has the potential of improving radar performance more substantially than corresponding advances in transmitter power, antenna gain, or receiver noise figure.

Introduction

This paper focuses on the frequency stability of millimeter wave oscillators and its impact on system performance, particularly in reference to radars. A typical radar transceiver might take the form shown in Figure 1. The received signal-to-noise ratio is given by the radar equation.

Radar Range Equation

$$P_r = \frac{P_t G^2 \lambda^2 \sigma \exp(-2\alpha R)}{(4\pi)^3 R^4} \qquad (1)$$

P_r = received power

P_t = transmitted power

G = antenna gain

λ = free space wavelength

σ = target cross section

α = atmospheric attenuation coefficient

R = range

Typical Receiver Equation

$$\left(\frac{S}{N}\right) = \frac{P_r A_p}{KTBN} \qquad (2)$$

P_r = received power

A_p = processing gain (if pulse compression or spread spectrum techniques are used)

K = Boltzmann constant

T = effective antenna temperature

B = post detection bandwidth

N = receiver noise figure

In order to obtain maximum information about the return signal, the received signal-to-noise ratio should be maximized. At present, the following performance is available at 94 GHz.

1. Low atmospheric attenuation (α = 0.4 DB/Km @ sea level.
2. Output power of several watts CW from IMPATT amplifiers involving multi-diode combiners. Pulsed power is up to 10 dB higher.
3. Antenna gains of 50-60 dB are attainable with practical apertures.
4. Single sideband receiver noise figure less than 9 dB has been attained.

It is apparent that the above system parameters further improvements to the above parameters have reached a point of diminishing return. The area which has not been addressed with as much effort and attention as the issue of processing gain. [10 log (TW/1) for spread spectrum or 10 log TW for pulse compression.]

Processing Gain Tradeoffs

Theoretically, the processing gain A_p could be capable of being increased indefinitely at the expense of integration time. In practice, as A_p increases, so does the B.E.R. For a spread spectrum radar, the limits on frequency are set by:

1. Highest bit rate achievable with conventional ECL circuitry ~ 100 MHz = f_H
2. Noise performance of the receiver. This determines the lowest useful modulation frequency, f_L, (or word rate).

Receiver FM Noise Performance

In order to investigate the effects of oscillator FM noise, the following discussion will be restricted to a biphase encoded CW radar with a pseudorandom code.

To establish the noise floor of the receiver, consider an FM modulated carrier being employed. The receiver L.O. is derived from the transmitted carrier (except for the frequency offset, i.e., the I.F.). At close range, the sideband power is suppressed within the receiver, while at longer ranges residual sidebands appear according to Figure 2. This is the same as what happens if the original carrier is contaminated by FM noise sidebands. The received noise floor density due to this effect is

$$P_N = NKT + P_s \int_{f_L}^{f_H} L(f)\, df \tag{3}$$

where P_N = noise power density

N = mixer noise figure

$L(f)$ = phase noise density

P_S = signal power in one of M spectral lines assumed to be equally spaced and of equal amplitude

$M = \frac{f_H}{f_L}$ = number of spectral lines

f_H = clock rate (highest frequency spectral line)

f_L = word rate (lowest frequency spectral line)

f_H is assumed to be much greater than f_L

Since noise adds incoherently in the correlator, the output noise power is

$$P_N^{OUT} = NKTf_H + \left(\frac{f_H}{f_L}\right) P_s \int_{f_L}^{f_H} L(f)\, df \tag{4}$$

The signal is recovered coherently by the correlator, and hence signal voltages add

$$P_s^{OUT} = \left[\sum_{i=1}^{M} \sqrt{P_s Z}\right]^2 Z^{-1} = \left(\frac{f_H}{f_L}\right)^2 P_s \tag{5}$$

Hence,

$$\left(\frac{S}{N}\right)_{OUT} = \frac{P_s \left(\frac{f_H}{f_L}\right)^2}{NKTf_H + \left(\frac{f_H}{f_L}\right) P_s \int_{f_L}^{f_H} L(f)\, df} \tag{6}$$

The $\left(\frac{S}{N}\right)$ ratio at the input, before any processing, is given by

$$\left(\frac{S}{N}\right)_{IN} = \frac{P_s \left(\frac{f_H}{f_L}\right)}{NKTf_H} \tag{7}$$

The processing gain is, thus

$$A_p = \left(\frac{S}{N}\right)_{OUT} \Big/ \left(\frac{S}{N}\right)_{IN} = \frac{P_s \left(\frac{f_H}{f_L}\right)^2}{NKTf_H + \frac{f_M}{f_L} P_s \int_{f_L}^{f_H} L(f)\, df} \times \frac{NKTf_H}{P_s \left(\frac{f_H}{f_L}\right)} \tag{8}$$

$$A_p = \frac{f_H}{f_L}\left[1 + \left(\frac{S}{N}\right)_{IN} \int_{f_L}^{f_H} L(f)\, df\right]^{-1} \tag{9}$$

Assuming an inverse power law for L(f)

$$L(f) = \frac{\alpha}{f^n} \tag{10}$$

$$A_p = \frac{f_H}{f_L}\left[1 + \left(\frac{S}{N}\right)_{IN} \frac{\alpha}{(n-1)\, f_L^{n-1}}\right]^{-1} \tag{11}$$

allowing f_L to vary as the independent variable and setting $\frac{\partial A_p}{\partial f_L} = 0$.

$$A_p^{max} = \frac{f_H (n-1)^{\frac{1}{n-1}}}{2\left[\frac{S}{N}_{IN} \alpha\right]^{\frac{1}{n-1}}} \tag{12}$$

The maximum processing gain that can be practically utilized must be consistent with an output $\left(\frac{S}{N}\right)$ ratio which yields an acceptably low B.E.R. and probability of false alarm. Assuming this to be satisfied by $\left(\frac{S}{N}\right)_{OUT} \gtrsim 20$ dB.

$$A_p \geq \frac{10^2}{\left(\frac{S}{N}\right)_{IN}} \qquad (13)$$

For f_H = 100 MHz - a practical clock frequency

$$\left(\frac{S}{N}\right)_{IN} \leq \left[\frac{(2 \times 10^{-6})^{n-1}}{n-1}\,\alpha\right]^{\frac{1}{n-2}} \qquad (14)$$

$$\therefore \quad A_p^{max} \geq 10^2 \left[\frac{(n-1)}{(2 \times 10^{-6})^{n-1}\,\alpha}\right]^{\frac{1}{n-2}} \qquad (15)$$

Consider the typical phase noise spectrum of a millimeter wave Gunn oscillator, shown in Figure 3, curve (a).

$n = 3, \quad \alpha = 10^8$

A_p^{max} = 57 dB, f_L = 100 Hz.

Consider the case of the phaselocked oscillator, with its characteristic phase noise spectrum, as shown in Figure 3, curve (b). Using the same analysis, it can be shown that

$$A_p = \frac{10^8}{f_L}\left\{1 + \left(\frac{S}{N}\right)_{IN}\left[\frac{10^8}{2f_\star^2} + 10^{-6}(f_\star - f_L)\right]\right\}^{-1}$$

$$= \frac{10^8}{f_L}\left[1 + \left(\frac{S}{N}\right)_{IN}(.005 + .1)\right]^{-1} \qquad (16)$$

$$= \frac{10^8}{f_L}\left[1 + 0.1\left(\frac{S}{N}\right)_{IN}\right]^{-1}$$

$f_\star$ = loop bandwidth

Hence A_p increases indefinitely as f_L is reduced. For the purposes of comparison, for f_L = 1 Hz, A_p = 80 dB

$$\left(\frac{S}{N}\right)_{min} = -60 \text{ dB} \qquad (17)$$

Design Criteria for Gunn Diode Phase Locks

From the previous equations, it is evident that noise around 1 kHz away from carrier should be suppressed as much as possible. This can be achieved by locking the oscillator to a harmonic of an ultrastable low frequency source. A S.A.W. oscillator operating around 1 GHz is a promising candidate. In Figure 4, curve (a) shows the noise of the fundamental S.A.W. oscillator, curve (b) shows the equivalent noise around the multiplied signal (translated by 20 log 94), and curve (c) shows the spectrum of the free running Gunn oscillator. The ideal loop should have a bandwidth of 2.2 MHz and adequate gain and phase margin at that frequency. Actual noise spectrum of the locked oscillator is significantly higher than the ideal curve because:

1. Millimeter wave oscillator noise is inherently higher than that at lower frequencies because of lower circuit Q and higher device noise. This necessitates a higher loop gain than is required at lower frequencies. The combined gain bandwidth product is difficult to realize without reduction in the phase margin.

2. Device sensitivity to power supply current and voltage fluctuations is high (e.g., at 94 GHz, a Gunn oscillator has tuning sensitivity $\simeq$ 1 GHz/volt).

3. Tuning elements are not readily available. Varactor tuning is lossy, bias tuning is non-linear, temperature and time dependent, and requires complicated electronics, optical pushing has not yet been proven suitable for this purpose.

4. Down conversion process required for the phase comparison adds more noise.

As a result, millimeter wave oscillators still require improvements in device and circuit performance. Further noise reduction at lower offsets from the carrier has been achieved by locking the reference S.A.W. oscillator to a low frequency ultrastable crystal oscillator (66 MHz), but the performance improvement is not commensurate with the resultant increase in complexity.

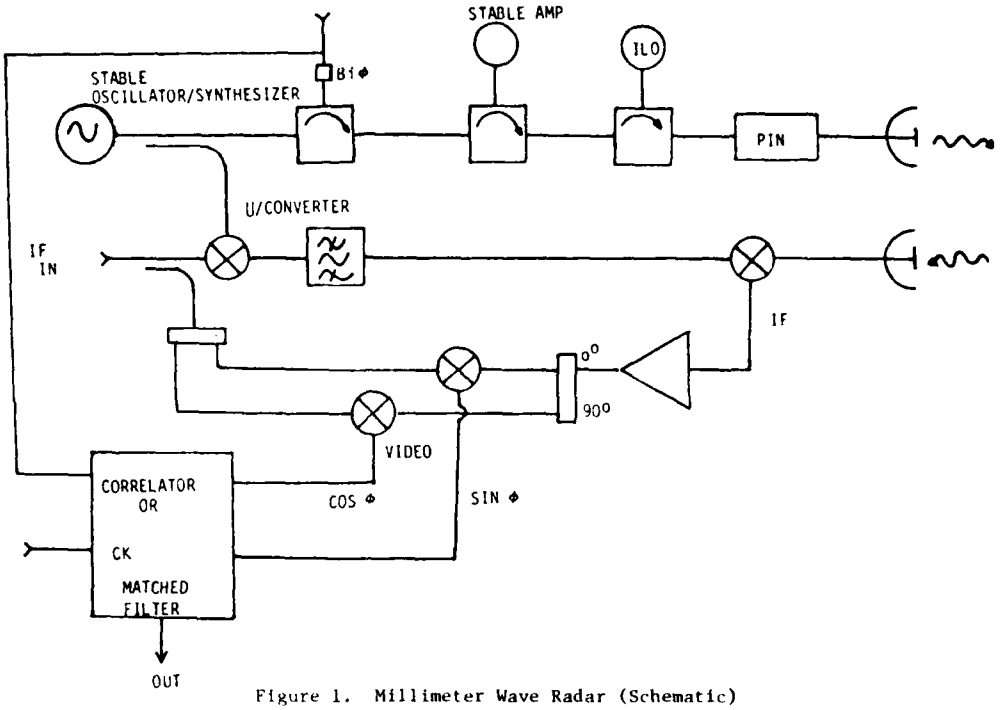


Figure 1. Millimeter Wave Radar (Schematic)

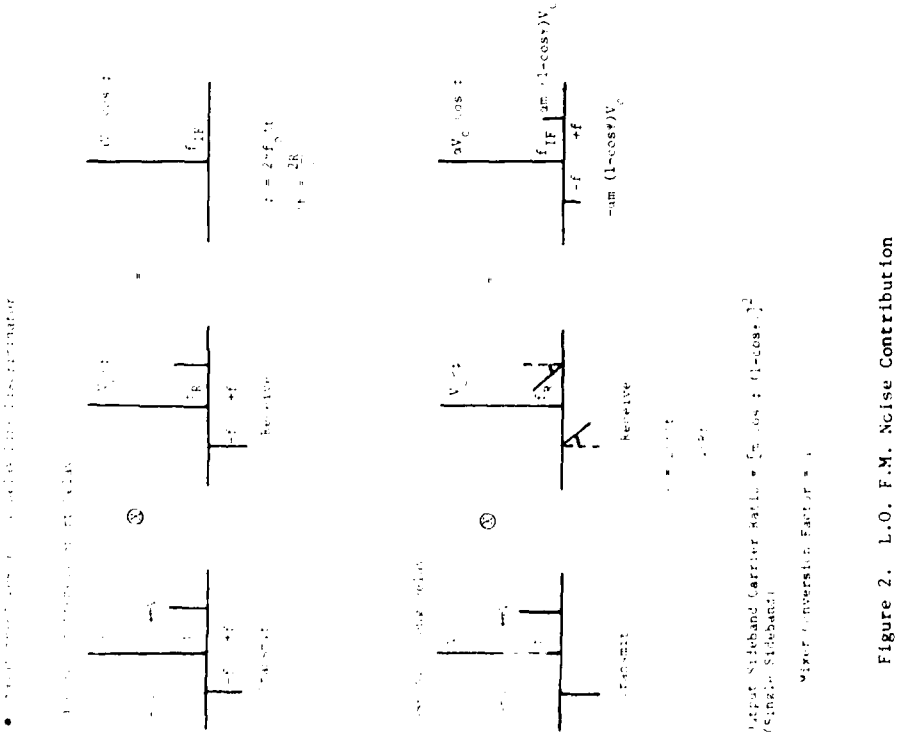


Figure 2. L.O. F.M. Noise Contribution

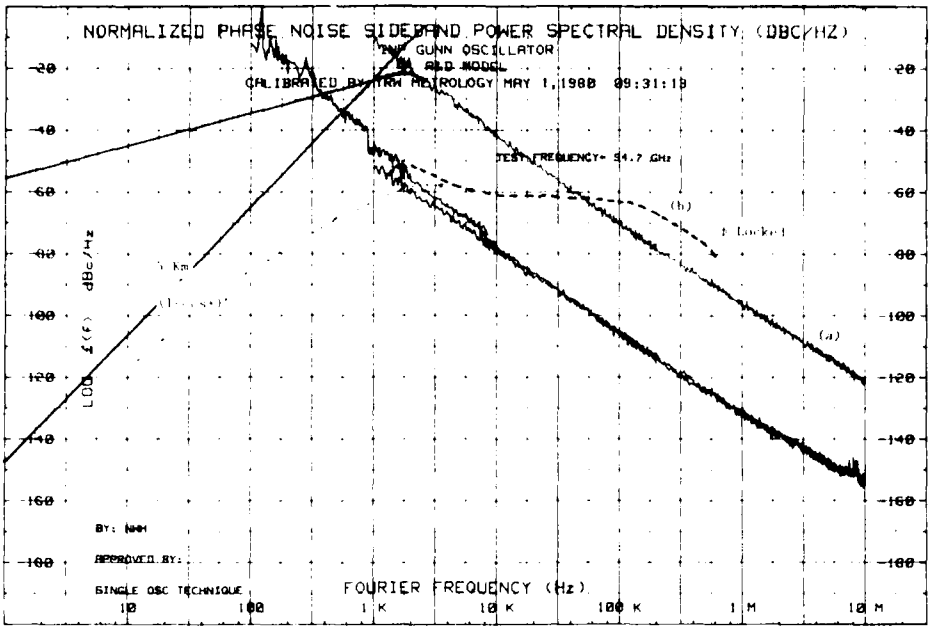


Figure 3. Gunn Oscillator Phase Noise Spectrum

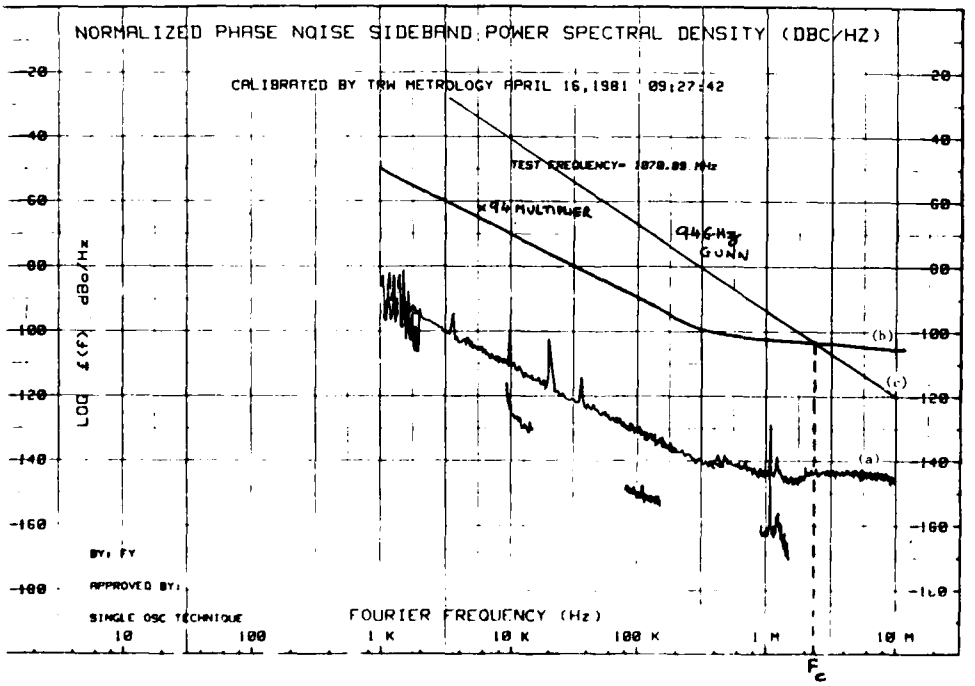


Figure 4. Phase Noise Spectra of a Reference Oscillator

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

FREQUENCY STABILITY REQUIREMENTS FOR A 95 GHz INSTRUMENTATION RADAR SYSTEM

David N. McQuiddy, Jr.
Texas Instruments Incorporated
Dallas, Texas

SUMMARY

The spectral purity characteristics that must be achieved by the frequency control mechanism of a 95 GHz coherent radar system are described. The radar system has been designed to serve as a general purpose instrument capable of supporting the collection of target and clutter backscatter data under a varying set of pulsewidth, polarization and pulse repetition interval conditions. The data is stored on magnetic tape in digital format and used to develop efficient algorithms for target detection, discrimination and classification. The system has been designed to provide a minimum of 40dB sub-clutter visibility to accomplish the algorithm development goals. Since data integrity (coherence) is almost totally dependent on system spectral purity, major design effort was focused on signal generation and control.

The system signal generation technique employs a single 100 MHz oscillator source to establish system timing, the transmit signal, and the local oscillator signal. A phase locked Gunn oscillator and injection locked CW and pulsed IMPATT oscillators are used to preserve signal coherence during transmit/receive operation.

Measured phase noise spectral density data are presented and compared with the desired performance specifications at key points in the signal path. The measurement techniques are briefly described. System processed data are shown that demonstrates greater than 40 dB sub-clutter visibility at target ranges greater than 2 km.

System Overview

Background

The application of millimeter-wave technology to electronic systems may provide effective solutions to a variety of military problems involving short-range targeting operations in clear to moderately adverse weather conditions. Millimeter-wave solutions can be implemented in the form of physically small equipment having high-resolution performance in range and angle with adequate doppler sensitivity.

Examples of military applications in which millimeter-wave technology provides attractive system alternatives include low-level interdiction strike missions in fixed or rotary-wing aircraft against ground armored vehicles. In the east European scenario, such missions are typically characterized by short time lines that support a requirement for automatic target acquisition and handoff for weapon delivery. Missile terminal-homing applications are another example. In this case, millimeter-wave technology may be used effectively in resolving high-value fixed targets or in identifying the unique signatures of selected tactical targets to enhance antiradiation homing by defeating the shutdown tactic. Finally, but of no less significance, is the area of ground-to-ground antiarmor operations that may include both fire control and battlefield IFF. In this area, millimeter-wave technology has demonstrated the potential of being an effective compliment to electro-optical (EO) systems in the adverse conditions of weather and battlefield-induced contaminants.

Within the battlefield scenario that is germane to millimeter-wave applications, the high resolution and doppler sensitivity achievable at millimeter wavelengths may permit the integration of target acquisition and first-order classification techniques (e.g., target type through doppler signatures resulting from engine vibration and tread motion) in systems that are less complex, more readily integrable into a platform and, hence, potentially less costly than longer wavelength system alternatives.

System Description

The 95 GHz instrumentation system consists of a solid-state transmitter and receiver front-end, a digital data acquisition system, and a mini-computer data evaluator. The program objective was to develop a millimeter-wave measurement system that could be used in a variety of field conditions; hence, flexibility and portability were key system implementation considerations. Figure 1 is an artist's conception of the system as it would be deployed in field operations. The RF assembly was constructed as a remote sensing unit coupled to the system controller/data acquisition system via an IF link. Two men can transport the remote assembly to a location providing the optimum field of view. The unit can be trained in both elevation and azimuthal planes. A boresight telescope allows correct target alignment. An elevation or depression angle reading is provided. The combination is a rugged, self-contained, field-portable unit that can be readily deployed to acquire key millimeter-wave data under a variety of conditions that

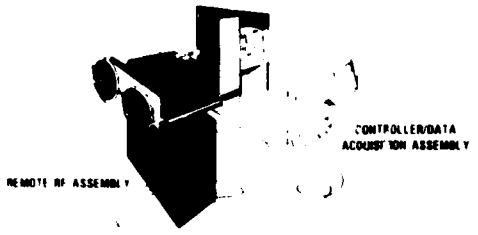


Figure 1. Transportable 95 GHz Instrumentation Radar System

are not well suited to laboratory measurements. In field operations, the instrumentation system provides radar coherent processing interval (CPI) data records for subsequent off-line processing. System operation is computer monitored, thus permitting operator selection of a variety of radar parameters and system operating conditions that meet the needs of the planned experiment.

Figure 2 shows the instrumentation system in simplified block diagram form. The system exciter, transmitter, receiver, antennas and polarization control units are located in the remote RF assembly. The remaining units are located in the van. In general, the system gathers radar data reflected from the environment within its fixed beamwidth, provides synchronous detection (I & Q), formats the incoming time series data according to pulse repetition interval (PRI), and records the data on a CPI basis. Limited data analysis can be performed near real time by the mini-computer prior to passing the data on to the digital tape recorder.

A list of key system parameters is provided in Table I. The system operates at 94.8 GHz and provides an instantaneous bandwidth of 300 MHz. The transmitter provides a minimum power of 0.5 W to the antenna input. Two antenna sizes can be used and are readily changed in the field. The 18 inch diameter antenna provides a minimum of 48dB gain with a 0.49 degree beamwidth while the 9 inch diameter antenna provides 42dB gain with a 0.98 degree beamwidth. The cassegrain antennas (two of each size) are provided with boresight aligned telescopes. Four pulsewidth values can be selected. The shortest, 5 ns, is generated by compressing a 100 ns pulse that is nonlinear frequency chirped over a 200 MHz bandwidth. The 5 ns compressed pulse provides the system with a minimum resolvable range less than 1.0 meter. A constant transmit pulse repetition frequency (PRF) is used, and selectable receive PRF is achieved via programmed receiver blanking. Thirty-two receive PRF values are provided between 1.25 and 40 KHz.

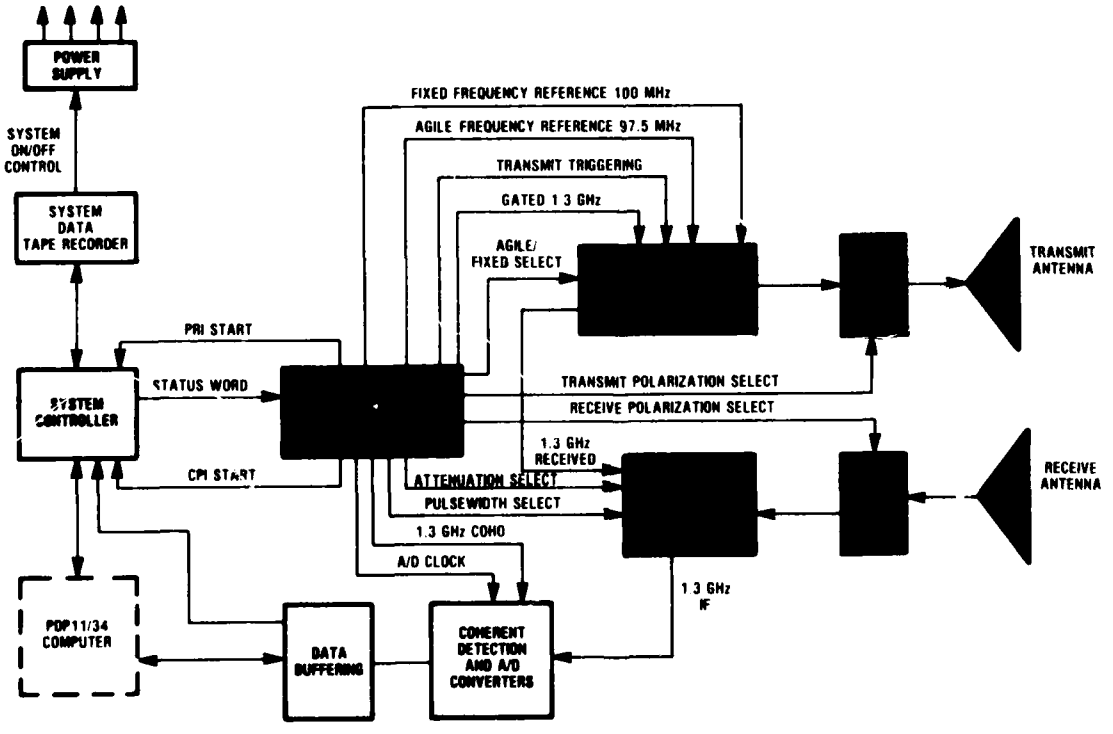


Figure 2. Simplified System Block Diagram

TABLE I. KEY SYSTEM PARAMETERS

PARAMETER	PERFORMANCE
FREQUENCY	94.8 ± 0.15 GHz
PEAK TRANSMIT POWER	0.5 WATTS MINIMUM
ANTENNA GAIN	9 INCH: 42 dB MINIMUM 18 INCH: 48 dB MINIMUM
RECEIVER NOISE FIGURE	11.5 dB MAXIMUM
PULSEWIDTHS	25, 50, 100 ns 5 ns (COMPRESSED FROM 100 ns)
SYSTEM LOSS	7.0 dB
PULSE REPETITION FREQUENCY (PRF)	
TRANSMIT	40 kHz
RECEIVE	40 kHz/(N + 1) N = 0 TO 31

The system operates bistatically and provides both coherent and noncoherent modes of operation. The system modes of operation are provided in Table II. The polarization diversity feature is implemented by electrically switching a waveguide ferrite polarizer on a pulse-to-pulse basis. The polarizer operates in a cylindrical waveguide mode and is followed by a fixed quarter-wave plate so that linear (H and V) and circular (RHC and LHC) polarizations can be attained. A switchable polarizer is contained in both the transmitter and receiver channels, allowing the transmission of one polarization state and the reception of the orthogonal state (linear or circular). Either fixed-frequency or pulse-to-pulse frequency-agile operation over a 144 MHz bandwidth may be selected with coherent operation provided in all fixed-frequency modes.

The function of the data acquisition system (DAS) is to preprocess the data collected by the remote RF assembly and record it in a format suitable for off-line processing. The initial DAS operation is synchronous detection of the receiver IF signal using the 1.3 GHz COHO reference to produce in-phase and quadrature video channels. Four high-speed 100 MHz A/D converters quantize the video data to a 4-bit level. Two A/D converters per channel are time phased to provide an I and Q channel instantaneous throughput rate at 200 MHz each. Data buffering and storage are provided by high-speed ECL memory that collects the incoming data on a PRI basis. Data reformation is accomplished by MOS bulk storage that accumulates range data from 1024 PRI's to form a coherent processing interval.

The 4-bit quantization level currently used in the DAS was a component availability limitation at the time of implementation and not a design constraint. All digital hardware in the DAS, exclusive of the A/D converters, has been designed to accept 8-bit data when 8-bit A/D converters with the high sampling rates used here become available.

A photograph of the instrumentation system hardware is shown in Figure 3. From left to right, the system components are the digital tape recorder (HP7970E), the data acquisition system and control panel, and the remote RF assembly equipped with 18 inch diameter antennas.

TABLE II. MODES OF OPERATION

- **PULSE-TO-PULSE POLARIZATION DIVERSITY**
  - **RHC, LHC, TRANSMIT RHC-RECEIVE LHC AND VICE VERSA**
  - **H, V, TRANSMIT H-RECEIVE V AND VICE VERSA**
- **PULSE-TO-PULSE FREQUENCY AGILITY**
  - **100 ns PULSEWIDTH ONLY**
  - **16 DISCRETE FREQUENCIES IN 144 MHz BANDWIDTH**
- **COHERENT I & Q CHANNELS**
  - **4-BIT A/D CONVERTERS**
  - **200 MHz CLOCK RATE**
- **COHERENT PROCESSING INTERVAL/MEMORY STORE**
  - **64 CONTIGUOUS RANGE CELLS**
  - **1024 PULSE REPETITION INTERVALS (RECEIVE PRF)**

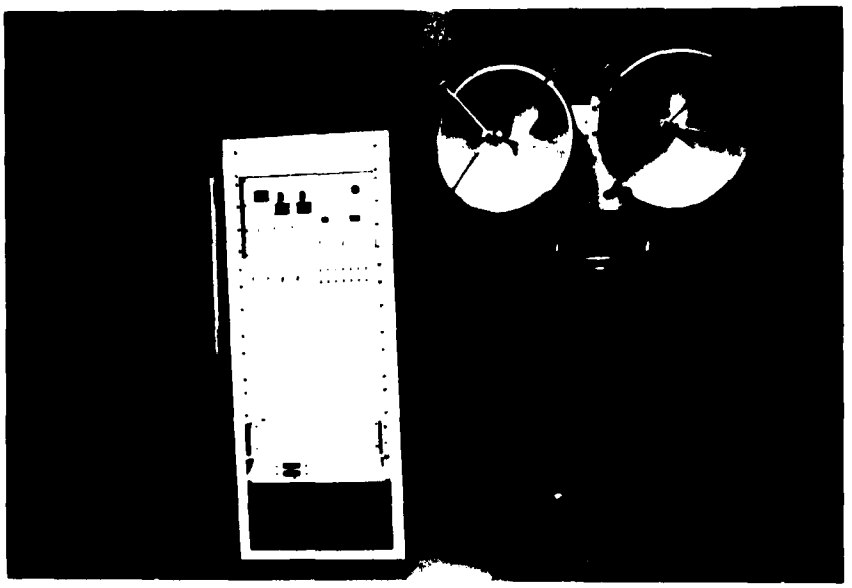


Figure 3. Photograph of System Hardware

Recent tests performed with the system using a 1 m^2 trihedral reflector have verified that a single pulse signal-to-noise ratio of 9dB can be achieved at 1 km with the 18 inch diameter antennas. The measurements were made in clear weather. The transmitter power was measured and found to be 1 W. Assuming an atmospheric path loss of 0.72 dB/km, the actual system loss calculates to be 3dB with the receiver noise figure assumed to be 11.5 dB. Applying the full processing gain of a 1024-point CPI in a clutter free environment, the 1 m^2 reflector should result in a 11 dB signal-to-noise ratio when located at 5 km range.

Spectral Purity Requirements

The system operating parameters have been selected to meet a wide variety of data collection needs. From an operational point of view, the PRF is selected to contain the doppler spectrum of the target of interest for the particular experiment being conducted. The data will be collected and the frequency spectrum from -PRF/2 to PRF/2 (or zero to PRF) will be filtered with digital transforms to characterize the target. Since the system must provide the capability to detect small moving targets in the presence (same range cell) of large stationary reflectors (discrete or point clutter), the combined transmitter and receiver local oscillator signal phase noise spectral density is constrained. The assumption is made that phase noise and not amplitude noise will be the primary mechanism degrading moving target detection. System tests have proven this assumption valid. The term sub-clutter visibility can be defined as the ratio of the large stationary reflector's backscatter cross-section to the backscatter cross-section of the smallest moving target that can detected. The system employs a range-gated sampling process to provide range resolution. The sampling process folds the phase noise of the transmitted signal into the same spectral interval occupied by the target return. The folding process degrades sub-clutter visibility and provides an additional constraint on the system signal phase noise.

Since millimeter-wave radars provide adequate detection performance at short ranges only due to atmospheric loss, the short radar transit time can be used to advantage. The advantage results if a single master oscillator is used to generate both the transmit and the receiver local oscillator signals. Correlation of the transmitted signal, delayed by the propagation time to the target and back, with the local oscillator reference signal will suppress the phase noise near the carrier. Phase noise far removed from the carrier is degraded by 3 dB. The corner frequency occurs when:

$$f_{corner} = 1/(4\ T_d)$$

where T_d is the propagation time to the target and back. Below the corner frequency phase noise is suppressed by 20 dB for each decade decrease in frequency. At the corner frequency the phase noise is 3 dB greater than the phase noise of the stable source. At two times the corner frequency, the phase noise has increased an additional 3 dB; however, the average degradation is only 3 dB when considering the spectrum out to the receiver IF bandwidth.

The system carrier-to-single sideband phase noise ratio (C/SSB) can be defined in terms of the factors discussed above. For frequencies out to the PRF/2, the ratio is set by the value of the sub-clutter visibility that is desired. Since it

is most common to express the phase noise in terms of a per unit bandwidth, the doppler filter bandwidth (doppler resolution cell) must be multiplied times the sub-clutter visibility value to btain the normalized C/SSB. If it is desired that the target signal appearing in a particular doppler resolution cell have a specified ratio to the noise in the cell, then this ratio (SNR) must be multiplied times the normalized C/SSB to obtain the improved C/SSB. Two other factors must also be considered. First, since range-gated sampling will raise the noise level in the frequency interval under consideration, it must be taken into account. This can be accomplished by requiring that the sum of the folded noise (normalized) contribute at the same level as the noise in the frequency interval under consideration. Multiplying the improved C/SSB by two then takes the folding into account. The second factor is the correlation effect associated with using a single stable source. The correlation factor can be expressed as:

$$CF=2(1-\cos(2\pi f_{offset} T_d)) \text{ for } f_{offset} < f_{corner}$$

where f_{offset} is the frequency offset from the carrier, and

$$CF = 2 \quad \text{for } f_{offset} > f_{corner}.$$

Summarizing in the form of an equation:

$$C/SSB(f_{offset}) = SCV \times DRC \times SNR \times 2 \times CF \text{ (Hz)}$$

for $2XDRC < f_{offset} < PRF/2$ where SCV is the desired sub-clutter visibility and DRC is the doppler filter bandwidth in Hertz.

For frequencies from PRF/2 to half the receiver IF bandwidth (BW_{IF}) the phase noise spectral density function must be such that range-gated sampling contribute no more than half the total noise power that appears in the frequency interval to the PRF/2. Considerable flexibility can be exercised in describing this function. One simple form is:

$$C/SSB\ (f_{offset}) = SCV \times DRC \times SNR \times 2 \times CF \times BW_{IF}/(2XPRF)$$

$$\text{for } PRF/2 < f_{offset} < BW_{IF}/2.$$

For the purpose of illustration, a particular set of system parameters have been selected. The system phase noise specification will be derived for this example and system data collected under the selected parameters will be presented later in the paper. A large stationary target is present at a range of 2.46 km. Assume that a small moving target 40 dB smaller in cross-section is present in the same range cell. Target motion is such that a 5 Hz doppler resolution cell is appropriate. The system will be operated at a PRF of 2.5 KHz and a 512-point discrete transform will be performed to produce the filtering function. The 100 ns pulsewidth sets the receiver IF bandwidth at 12 MHz. Figure 4 shows one phase noise spectral density function that will meet the requirements for moving target detection at a SNR of 1. Note that two discontinuities are present. The function requires a higher performance level from 100 KHz out to 6 MHz so that the requirements from 1.25 KHz to 100 KHz can be relaxed.

Figure 4 represents the system phase noise specification. A family of these curves can be generated for various PRF's and a worst case specification derived from them; however, the end result will not differ significantly from Figure 4.

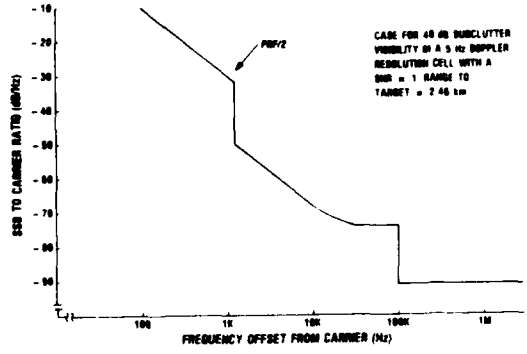


Figure 4. System Phase Noise Specification

System Signal Generation

The signal generation approach that was selected to meet the system spectral purity requirements is shown by the block diagram in Figure 5. The primary signal path is highlighted with bold interconnecting lines. All signals including the system timing are derived from a single 100 MHz stable source. The 100 MHz signal is used as a reference to phaselock a transistor oscillator operating at 7.8 GHz. This oscillator was designed to achieve low noise performance for communications applications. The phaselock bandwidth is approximately 100 KHz. At frequencies greater than the loop bandwidth, the system noise is determined by the free running characteristics of the transistor oscillator. The multiplied phase noise of the 100 MHz source sets the system performance for frequencies within the loop bandwidth. The 7.8 GHz signal drives a harmonic mixer to mix its twelfth harmonic with the output of a Gunn oscillator operating at 93.5 GHz. The 100 MHz difference signal from the mixer is compared in a phase detector with the signal from the 100 MHz stable source. The output of the phase detector is used to phaselock the Gunn oscillator. The loop bandwidth of this phaselock oscillator is slightly greater than 10 MHz. The output power of the phase locked Gunn oscillator is divided equally between two paths. One path provides the local oscillator signal for the receiver balanced mixer,

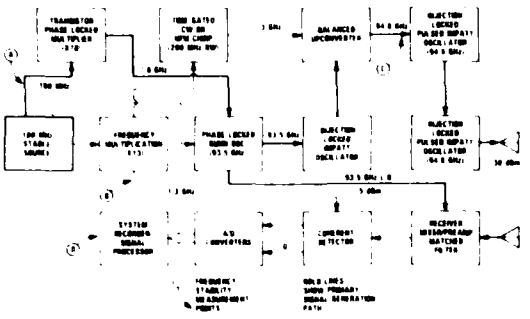


Figure 5. System Signal Generation Technique

the second path drives an injection locked IMPATT oscillator. This CW oscillator is used to amplify the small (+5dBm) signal from the Gunn oscillator to a power level of +17dBm. The injection locked bandwidth of the CW IMPATT oscillator is greater than 700 MHz. The output of this oscillator drives a single-sideband upconverter. The second input to the upconverter is a time gated 1.3 GHz signal that is obtained by multiplying a sample of the 100 MHz stable source by thirteen and then gating it with a diode RF switch. The gate time (pulsewidth) is determined by the selected system pulsewidth. The pulsed 94.8 GHz upconverter output signal is applied to a two-stage pulsed IMPATT injection locked oscillator assembly. The two oscillators are gated on for 140 ns. The pulsed injection signal is present during the latter 100, 50, or 25 ns of the 140 ns pulse depending on the system pulsewidth selection. The free running pulsed oscillator spectrum falls outside the matched receiver bandwidth and doesn't result in detectable target time sidelobes. The two-stage pulsed IMPATT oscillators amplify the upconverter output signal (+5dBm) up to the +30 dBm level at the transmitter antenna input port.

The 5 ns pulsewidth is achieved via nonlinear FM pulse compression techniques using surface acoustic wave (SAW) devices. This approach was selected over conventional linear FM pulse compression because it permits better control of time sidelobe structure and mismatch loss at the small time-bandwidth products required here (TB=20). The expanded transmit pulse is obtained by impulsing the SAW device with a 5 ns sample of the 1.3 GHz signal. The impulsed SAW device generates a 100 ns nonlinearly down-chirped signal that is phase referenced to the system stable source and covers a 200 MHz bandwidth. The down-chirp approach was selected to provide a more natural match to the thermally induced free running chirp to be expected from the pulsed IMPATT oscillators. A matching SAW compressor is located in the receiver and compresses the 100 ns transmitted signal to a value near 5 ns having time sidelobes down more than 23 dB relative to the peak of the main lobe.

The signal flow for the frequency agile mode is not shown. Briefly, a stepping synthesizer, which covers the range of 97.34 to 97.66 MHz serves as a reference (in place of the 100 MHz signal) to the Gunn phaselock assembly. A times eighty multiplied version of the synthesizer output replaces the 7.8 GHz signal to the phase locked Gunn oscillator. This produces an output signal that steps in 9 MHz steps over a 144 MHz band at 93.5 GHz.

A photograph of the transmitter assembly is shown in Figure 6. The transistor oscillator phaselock subassembly, phase locked Gunn oscillator, and receiver balanced mixer are shown on the right. The CW injection locked oscillator, balanced upconverter, and two-stage pulsed IMPATT oscillators are on the left.

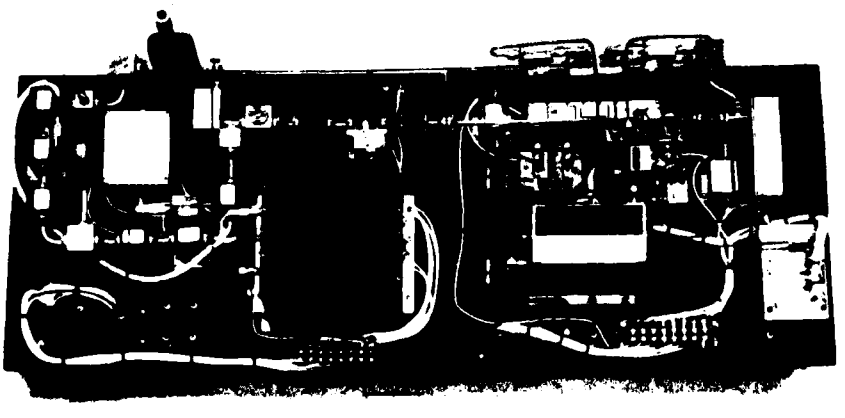


Figure 6. Photograph of Transmitter Hardware

System Stability Measurements

Stable Source Performance

The phase noise performance of the 100 MHz stable source was measured by constructing two "nearly" identical sources. One source was designed with an electronic tuning capability. The tuning mechanism was lightly coupled to prevent the introduction of extraneous noise. The two oscillators were locked together with a narrow bandwidth phaselock loop. The loop bandwidth was approximately 10 Hz. The phase detector output was amplified by a low noise FET video amplifier and applied to a low frequency spectrum analyzer (HP3585A). The spectrum of the phase demodulated signal was recorded from 20 Hz (twice the loop bandwidth) out to 100 kHz. The phase noise spectral density was then calculated assuming equal noise contributions from both sources. The results of this measurement are shown in Figure 7. Also shown in the figure is the system specification translated to 100 MHz. The translation amounts to 62.5dB. It is seen that adequate performance has been achieved.

The same measurement technique was used to evaluate the noise characteristics at 1300 MHz (Figure 5 point B). A times thirteen frequency multiplier is used on the second source to create a drive signal to a phase detector at 1300 MHz. The oscillators are phase locked at the 100 MHz frequency. Figure 8 displays the results of this measurement. Again, the measured data supports the system requirement.

Phase Locked Gunn Oscillator Performance

The phase noise was measured at 93.5 GHz by comparison with a second source that was derived completed by multiplication from a "clean" 97 MHz oscillator. The multiplied source was mixed with a sample of the system phase locked Gunn oscillator injection locked IMPATT oscillator combination to produce a difference frequency near 300 MHz. the difference signal was applied to a spectrum analyzer with a 1 kHz bandwidth. The results of this measurement adjusted to a 1 Hz bandwidth is shown in Figure 9. No secondary tests were performed to verify which of the two sources dominate; however, the two noise plateaus in the vicinity of 100 kHz and 10 MHz seem to indicate the phase locked oscillator dominates. The system specification is repeated in this figure for reference.

The measurement described above at 93.5 GHz was limited to an evaluation of the spectrum from 2 kHz out to greater than 10 MHz. To examine the "close-in" spectrum, a separate experiment was performed. Figure 10 shows a block diagram of the measurement technique that was used. The test provides a measure of the additive phase noise produced by the CW injection locked IMPATT oscillator. The system transmitter supplies a large fraction of the test signals. The only modification required to the transmitter unit was to operate the 1.3 Ghz signal into the upconverter CW rather than the normal time gated manner. For "close-in" data an HP5390 Phase Noise Analyzer was used. An HP8660C

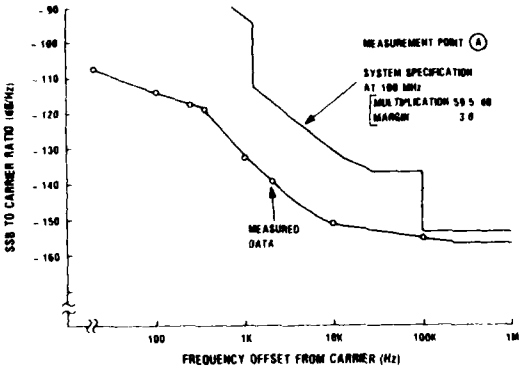


Figure 7. Measured 100 MHz Stable Source Phase Noise

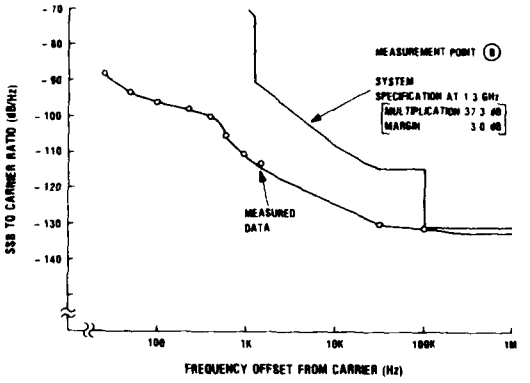


Figure 8. Measured 1.3 GHz Phase Noise

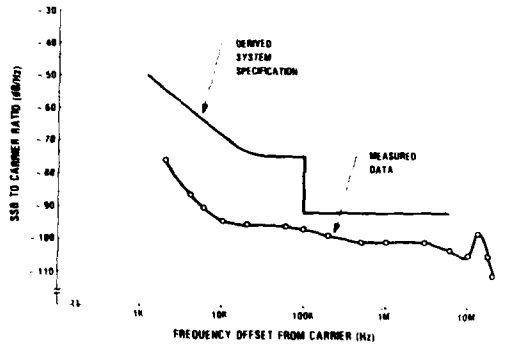


Figure 9. Gunn PLO Phase Noise at 93.5 GHz

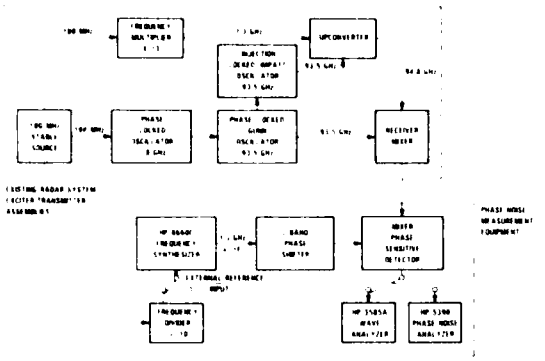


Figure 10. Measurement Technique for Injection Locked IMPATT Oscillator Additive Phase Noise at 94.8 GHz.

Frequency Synthesizer was used to produce an IF signal near 60 kHz for application to the analyzer. This allows measurements from very near the carrier out to 10 kHz. Over this same spectrum, the HP8660C was measured at 1.3 GHz and found to be nearly 20 dB lower in noise. A typical phase locked system was used to measure the additive noise further away from the carrier. The phaselock is accomplished by referencing the synthesizer to the system stable source. The effective loop bandwidth is less than 100 Hz. The result of the additive phase noise measurement is shown in Figure 11. It can be seen by comparison with Figure 9, that the additive noise is greater than the total noise in the range from 2 to 100 kHz. Obviously, this should not be; however, considering the difficulties involved in making phase noise measurements at 95 GHz, the 10 to 15 dB discrepancy is not a major surprise.

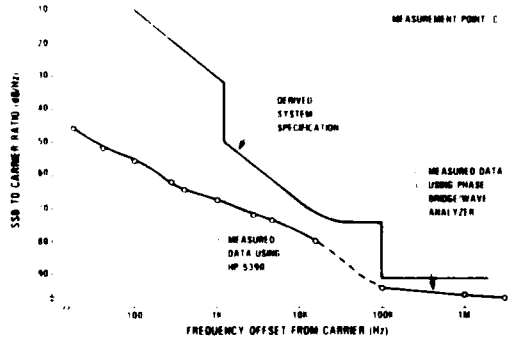


Figure 11. Measured Additive Phase Noise at 94.8GHz

System Sub-Clutter Visibility

During initial system operating tests, the instrumentation hardware was set up in an elevated test station overlooking the TI antenna range. The system was used to record radar returns from both fixed and moving ground targets. A number of runs were conducted using a variety of radar parameters. One experiment in particular was devised to evaluate sub-clutter visibility performance. A large stationary target was located at 2.46 kM and illuminated with a 100 ns pulse train at a 2.5 kHz repetition rate. A full CPI was recorded and processed off-line. The off-line processing consisted of performing a complex discrete Fourier-transform using 512 data points from the CPI. The result of the off-line processing is shown graphically in Figure 12. An expanded frequency scale was selected that shows the doppler performance near the carrier. The stationary target is seen in the zero doppler cell at a relative amplitude of +10dB. The power in the second doppler cell at 9.76 Hz has a relative amplitude of -31dB, and demonstrates a sub-clutter visibility greater than 40dB. Although not shown, the doppler spectrum was examined out to 2.5 kHz and most cells displayed greater than 40dB visibility. It should be commented here that the visibility limit could be a result of system thermal noise and not phase noise. The thermal noise floor is set by the 4-bit A/D drive requirements and the processing gain. In other words, the system phase noise performance supports the 40dB sub-clutter requirement at the minimum and could be significantly better.

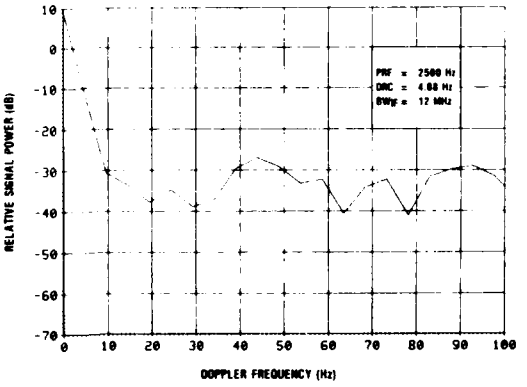


Figure 12. System Performance Data for a Large Stationary Target at 2.46 km

Conclusion

A millimeter-wave instrumentation radar system has been described. The principal components of the system are a 95 GHz transceiver, a data acquisition system, and a digital recorder. The combined elements of the system provide a high degree of flexibility in data collection and signal processing. Radar returns may be processed coherently and mapped into a variety of range, amplitude, and doppler combinations.

Specifically, phase locked Gunn oscillators have demonstrated the degree of spectral purity required to support the coherent signal processing of radar returns at millimeter-wave frequencies. A cascade chain of pulsed, injection locked, IMPATT oscillators have provided transmitter signal amplification at 95 GHz while preserving signal coherence. A phase controlled waveform using a SAW pulse compression expander has been amplified by pulsed, injection locked, IMPATT oscillators. The target return has been compressed in a matched filter providing a range resolution better than 1m. The compressed waveform has been coherently processed.

In general, it can be concluded that a solid-state millimeter-wave radar can be realized in a form supportive of high-resolution range and doppler performance.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

OPTICAL FREQUENCY CONTROL FOR WAVELENGTH MULTIPLEXED SYSTEMS

F. S. Welsh and T. S. Stakelon

Bell Laboratories
555 Union Boulevard
Allentown, PA 18103

Abstract

The capacity of a fiber-optic transmission system can be increased dramatically by optically combining light at different wavelengths from several sources onto a single fiber. The technique is analogous to the current practice of frequency division multiplexing in electrical transmission systems. Unfortunately, while the linewidth of semi-conductor laser sources is very small, uncertainties resulting from thermal drift, manufacturing variation, and other effects require a large window for each wavelength. Thus, existing technology allows only relatively poor use of the available fiber transmission bandwidth.

This paper will describe limitations on wavelength-multiplexed system channel capacity which result from instabilities of currently available sources and characteristics of optical filters and gratings. Finally, the requirements for improved source stability and multiplexer performance will be outlined and the resulting opportunities for large-capacity systems will be described.

Introduction

While the wide transmission band capabilities of fiber-optic systems are well-known, practical considerations often limit the bandwidth to lower values. In particular, the difficulties associated with generating and electronically multiplexing very high bit-rate digital signals and limitations of span length resulting from pulse distortion in the fiber place economic limits on the use of fiber bandwidth. Many of these problems can be overcome by optically combining (multiplexing) light of different wavelengths from several sources onto a single fiber. An analogous situation has existed for many years in frequency-division-multiplexed transmission systems. Figure 1 illustrates the dramatically increased capacity of the L-multiplexed coaxial systems over the past forty years.[1] This growth has resulted partly from increasing the number of coaxial tubes per sheath from 4 to 22, but even more significantly by an increase from 600 to 10,800 voice channels per tube through improvements in frequency multiplexing. For comparison, the capacity of a full fiber cable (144 fibers) which carries digital optical pulses at 90 Mb/s has a two-way channel capacity of just less than 100,000 voice circuits, somewhat less than the L5 system.

One of the practical problems limiting further expansion through time-division multiplexing (TDM) is the reduction of maximum span length as bit rate increases. This effect is illustrated in Figure 2, where fiber losses of 3 dB or 1.5 dB per kilometer and a 1.3μm laser source are assumed. These curves are for multi-mode fiber, so that the limitation in path length is largely the result of modal dispersion which results in intersymbol interference. Much longer spans at higher bit rates can be achieved using single-mode fiber, but there are additional practical problems for this case as well.

Wavelength-division multiplexing on the other hand, can complement time-division multiplexing at any bit rate and offers the potential of future growth comparable to that we have seen for the FDM systems. The optical frequencies are high enough, about 10^{14} Hz, that negligible band broadening occurs for any practical modulation frequency. This situation is in contrast to that of FDM where the bandwidth is determined by the modulation frequency. To a large extent, the improvements in capacity of FDM systems were made possible by improvements in crystal oscillators and filters. In optical systems, the laser assumes the place of the electronic oscillators and, as we shall show, even modest improvements in laser stability can present a significant opportunity for improved system design.

Figure 3 illustrates possible uses for WDM: in addition to simultaneous transmission of several channels in one direction, it is also possible to design bidirectional systems. Identical optical components can be used in either case, although the increased cross-talk requirements for bidirectional transmission will generally require better performance from the filtering technology.

It is the intent of this paper to estimate the limits to the capacity of multimode wavelength-multiplexed systems using lasers. There are three principal contributors to these limits which will be discussed: (1) the available wavelength regions, as determined by the fibers and sources, (2) the filtering technology, which determines minimum guard spacing between channels, and (3) channel width and source stability, that is, the space required for a single TDM channel. We will not consider here single-mode systems or LED-based systems.

Wavelength Regions

The wavelength regions for useful fiber optic systems are determined by available source wavelengths and low-loss regions in the fiber. Practical systems require the use of semiconductor lasers which are available in wavelength ranges of 0.8 to 0.9μm (GaAlAs) and also for 1.1 to 1.6μm (InGaAsP). Figure 4 shows a typical loss characteristic for a kilometer of fiber, together with cross-hatched areas indicating those low-loss regions where sources are available. Obviously, the edges of the three bands shown are not well-defined, but for purposes of estimating channel capacity we may roughly consider three bands of 0.80 - 0.90μm, 1.10 - 1.35μm, and 1.45 - 1.60μm. Hill has examined the capacity of wavelength multiplexed systems in the 0.80 - 0.90μm region.[2] Losses in these bands will be in the range of 1 to 5 dB/dm for cabled optical fibers.

The chromatic dispersion in the fiber will affect span length, and it is important to observe that it is not possible to optimize the dispersion over the entire band. So for long spans or very high bit rates, the need to limit dispersion to some small value may narrow the available region for multiplexing to something less than shown here. This is an important system consideration which depends on the specific application. It will not be further discussed here, however, as we are interested in determining the maximum potential multiplexed capacity.

Multiplexing Devices

There are two general classes of wavelength multiplexing or demultiplexing devices which have received extensive attention in the literature:[3] (1) those relying on optical interference filters to reflect some wavelengths while passing others, and (2) dispersive devices relying on a diffraction grating to separate the wavelengths spacially.

Figure 5 illustrates the principal of a two channel interference demultiplexer. The input light from the common (transmission) fiber is collimated by the first lens onto the dichroic filter. In the case of a long wave pass filter, this light will be reflected or transmitted accordingly as its wavelength is below or above the transition region. Reflected light is refocussed by the first lens onto one output fiber while transmitted light is refocussed by a second lens onto a second output fiber. The identical device also serves as a multiplexer. A convenient realization of this device using gradient index (GRIN) lenses is also shown in Figure 5.[4] These lenses are simply right cylinders of glass in which the index of refraction varies radically. Cut to the proper length they act as collimating lenses and allow a solid structure free of air-glass reflections. In such a configuration, the interference multiplexer should have losses less than 1 dB.

More than two channels may be accommodated by cascading two or more of these devices, or by using tilted filters and multiple reflections, usually at the price of increased loss or polarization sensitivity. The grating demultiplexer offers a more attractive alternative for larger numbers of channels. This device, illustrated in Figure 6, uses a common lens for all channels. The dispersion of the grating results in the output channels appearing in a linear array at the focal plane of the lens. The same device can also serve as a multiplexer. A gradient index lens design[5] results in a very compact structure less than 1 cm long, although in practical use the need to array the fibers and limit bend radii may result in a larger package as shown in Figure 7. One disadvantage of the grating device is that any variation in source wavelength will appear as a lateral displacement at the output which will result in additional loss at the desired fiber and possibly crosstalk into the adjacent fibers. This feature makes it desirable to use a larger output fiber than the input fiber in order to widen the bandwidth. Such a design change may preclude use of this device as a multiplexer. Hence, while well controlled source wavelengths may dictate both a grating multiplexer and demultiplexer, a system may consist of a filter multiplexer and a grating demultiplexer for optimum performance with poor or moderate control.

We have studied experimentally the cross-talk and insertion loss of these devices, and inferred the minimum guard band required between channels. A typical response for a filter device is illustrated in Figure 8. Here it may be seen that for 15 dB of crosstalk isolation the transition region is about 150Å for a high-quality interference filter. In addition, a similar margin must be allowed for filter drift and manufacturing variation. A reasonable guard band is about 300 to 400Å. For a typical grating device (Figure 9), the transition region depends on the band shape (grating, lens, fiber sizes and spacings) and requires less tolerance for variation. A minimum guard band of 2 to 3 times the channel width is a reasonable estimate for grating multiplexers. In addition, for grating demultiplexers, with a minimum guard band of about 0.4 times the channel width, crosstalk isolation improves dramatically away from the band edges which may result in substantially better system performance than for the filter devices.

Source Stability

Very large channel capacities are in principle possible with demonstrated multiplexing and demultiplexing devices. One must now consider, however, the sources themselves to allow sufficient wavelength space for each of the individual channels. Figure 10 shows schematically the various contributions to channel width arising from GaAlAs laser source variations. These variations tend to increase with increasing laser wavelength. A typical multimode laser source consists of several very narrow (~2 nm) laser lines. There is negligible broadening of this basic 10-50 nm linewidth from the information signal since the carrier frequency in the case of lightwave systems is about 10^{14} Hz, far above any present or envisioned modulation frequency. Rather, the channel width will be determined by variations about this

nominal wavelength arising from manufacturing variation, aging, and temperature fluctuations. This situation is in considerable contrast to the analog frequency division multiplexed systems where crystal oscillators have resulted in an extremely stable nominal frequency, and the modulation sidebands determine channel width.

As shown in the figure, manufacturing variation is about 20 nm. (It and others will depend on the wavelength and type of laser.) To this must be added about 5 nm of allowance for aging and about 20 nm or more to allow for a temperature coefficient of wavelength of about 3Å/°C. The total variation is thus about 45 nm; as a result only about 3 to 5 channels can be accommodated with current source technology despite the larger potential capacity of these devices.

Fortunately, all of these variations are controllable to some extent, thus offering the potential for substantial future improvements in system design. These methods of control fall into several categories:

1. Device selection. By sorting the manufactured device by wavelength, one can effectively limit the uncertainty from this source, independently from any improvements in wavelength control in manufacture, but at the expense of manufacturing yield.

2. Temperature control. Especially where wide variations in ambient temperature are expected, the control of temperature to a narrow range will effectively remove this source of wavelength variation. Generally, however, it will be necessary to use cooling devices to control the temperature to avoid excessive degradation of laser lifetime.

3. Feedback stabilization. Several methods for external control of semiconductor laser frequencies have been described.[6-10] These include the use of an internal Bragg reflector to select the feedback wavelength or the use of an external cavity either with or without wavelength-selective reflection. It is likely that for successful operation in a system a combination of feedback control and temperature control would be required, and the degree of stabilization will depend on the details of implementation. Considerable work remains in this area to produce a practical system using feedback control of laser wavelength, but the largest improvements in capacity are likely to result from this development, since aging and manufacturing variation will now be directly compensated.

Conclusions

Using the assumptions on demultiplexer device performance and laser stabilization described above, one can now compute the required wavelength allowance per channel (channel width plus guard band). This is not a firm number because of variations in system environment, required span lengths, loss margins, crosstalk requirements, and so forth. Nonetheless, it will provide a suggestion of the potential capacity improvements and those developments of highest priority.

Figure 11 shows the results of these calculations assuming the worst case of long wavelength lasers. The required allowances for current technology are indicated on the left. Filter multiplexers and demultiplexers are initially assumed. Some improvement results from laser selection, but significant improvement begins with the implementation of temperature control. The improvements resulting from ± 15Å, ± 10Å, and ± 5Å feedback control are next, resulting in an overall reduction in bandwidth per channel of a factor of 10. Where feedback stabilization begins, the resulting increase in channel capacity will be large enough to indicate the use of grating devices, so these are assumed for the second half of the progression.

Now from this information on channel widths and the assumptions on available source wavelengths, one can estimate the corresponding expected number of channels as a function of increasing source stability. This estimate is shown in Figure 12, where the ticks on the abscissa correspond approximately to the technology improvements indicated in Figure 11. The shaded regions indicate the results for grating demultiplexers assuming either 15Å or 30Å minimum linewidth allowance. The dashed lines indicate limits using only interference filter devices. As suggested by this figure, rather large increases in overall capacity may be expected in the future, with the most significant contributions coming from direct feedback stabilization of the laser sources themselves.

References

1. F. C. Kelcourse and F. J. Herr, B.S.T.J., 53, 1901 (1974).

2. A. M. Hill, Proceedings of the European Optical Communications Conference, Amsterdam, 1979.

3. W. J. Tomlinson, Proceedings of the 3rd Fiber Optical Communications Exposition, San Francisco, 1980.

4. S. Sugimoto, K. Minemura, K. Kobayashi, M. Shikada, H. Nomura, K. Kaede, A. Veki and S. Matsushita, Electron. Lett., 14, 15, (1978).

5. W. J. Tomlinson and G. D. Aumiller, Appl. Phys. Lett., 31, 169, (1977).

6. T. L. Paoli and J. E. Ripper, Appl. Phys. Letter., 25, 744 (1974).

7. F. K. Reinhart, R. A. Logan and C. V. Shank, Appl. Phys. Lett., 27, 45 (1975).

8. T. Okoshi and K. Kikuchi, Electron. Lett., 16, 179 (1980).

9. K. Utaka, K. Kobayaski, K. Kishino and Y. Suematsu, Electron. Lett., 16, 455 (1980).

10. D. J. Malyon, D. W. Smith and R. W. Berry, Electron. Lett. 16, 744 (1980).

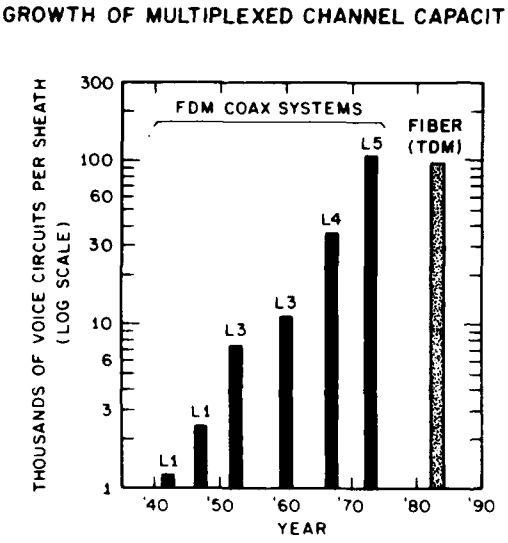


Fig. 1 Growth of transmission capacity of L-multiplexed coaxial systems and comparison to a 90 Mb/s fiber-optic system.

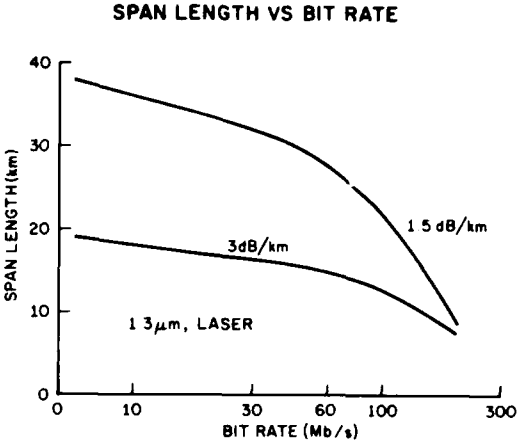


Fig. 2 Dependence of maximum span-length on bit rate for a 1.3μm laser system.

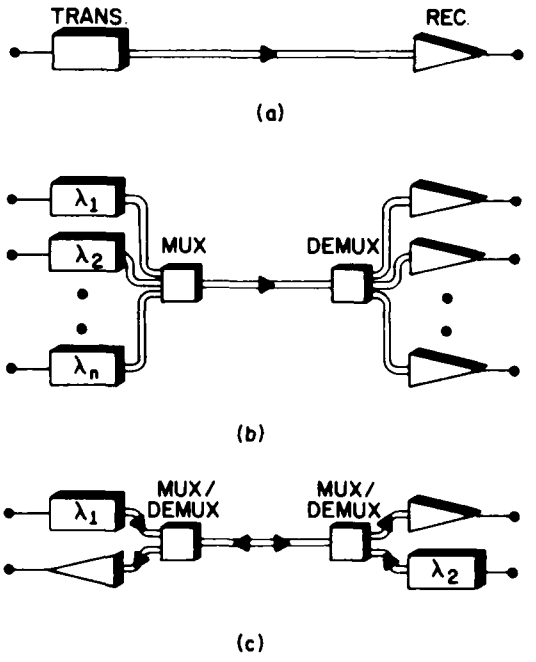


Fig. 3 System configurations for (a) single wavelength, (b) unidirectional wavelength multiplexing, (c) bidirectional wavelength multiplexing.

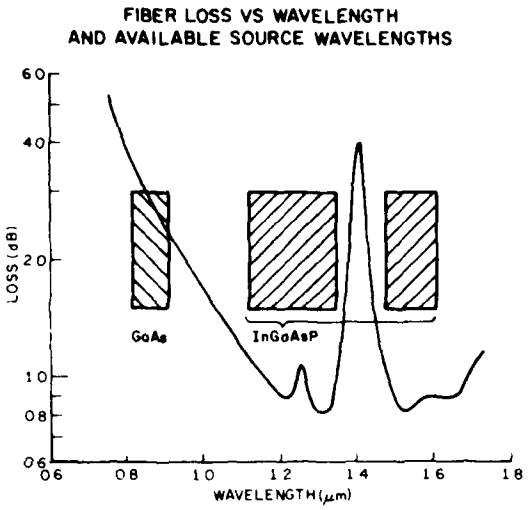


Fig. 4 Typical uncabled fiber loss versus wavelength.

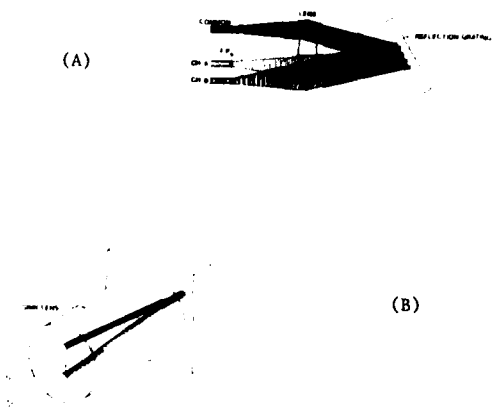


Fig. 6 Multi-channel grating demultiplexer: (a) Principle of operation, (b) GRIN-lens realization with grating affixed to prism.

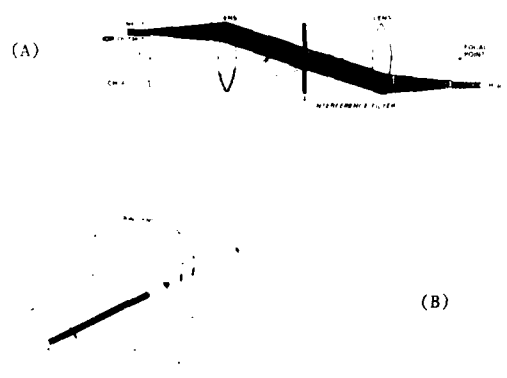


Fig. 5 Two-channel interference filter multiplexer/demultiplexer: (a) Principle of operation, (b) GRIN-lens realization.

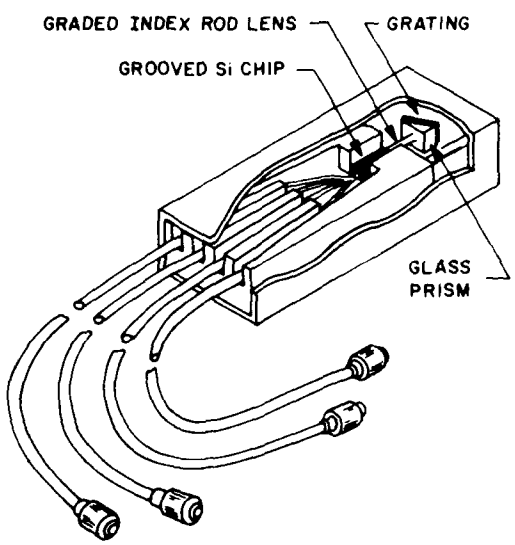


Fig. 7 Packaged version of gradient index lens grating demultiplexer.

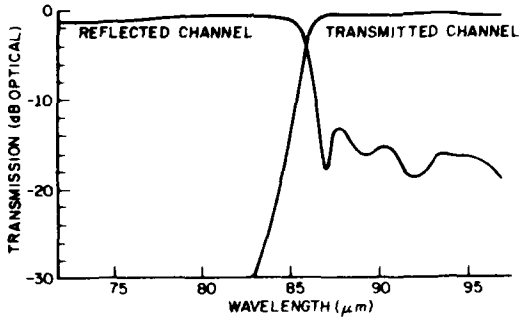


Fig. 8 Filter multiplexer transmission (reflection) versus wavelength.

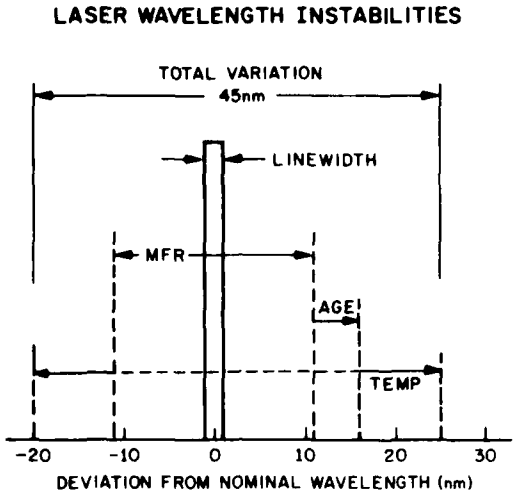


Fig. 10 Contributions to channel width arising from variation of the laser source wavelength in the short wavelength band.

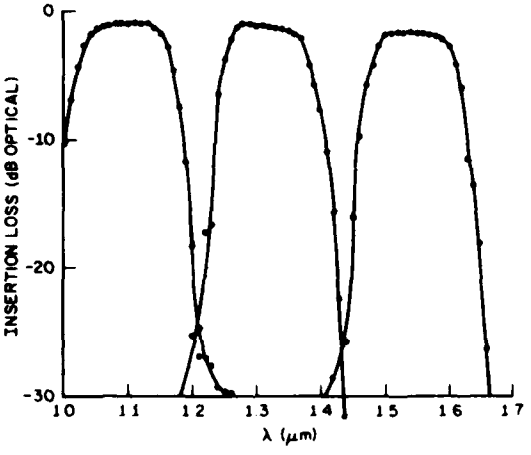


Fig. 9 Three-channel grating demultiplexer response.

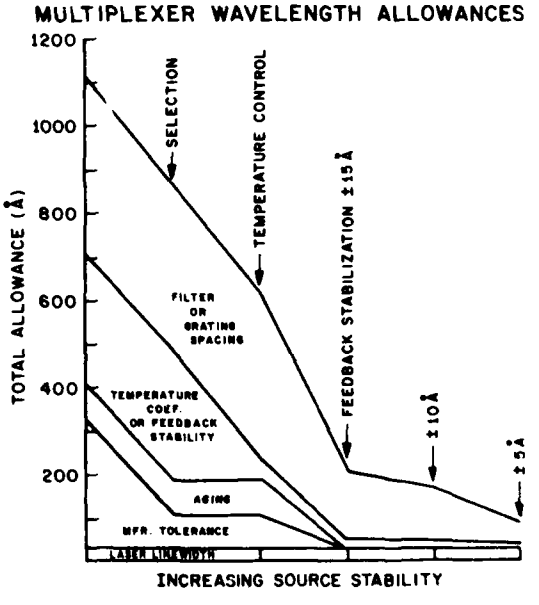


Fig. 11 Required channel allowances as a function of increasing laser wavelength stability in the long wavelength bands.

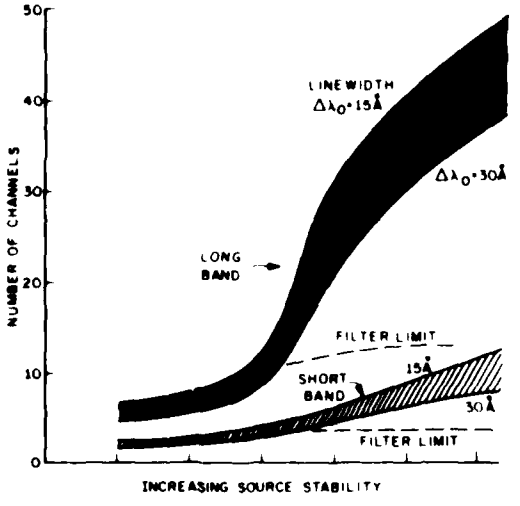


Fig. 12 An estimate of the maximum number of multiplexed channels as wavelength stability increases. The abscissa corresponds to that shown in Figure 11.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

NAVSTAR GLOBAL POSITIONING SYSTEM (GPS) CLOCK PROGRAM: PRESENT AND FUTURE

Capt Douglas M. Tennant

Air Force Systems Command
Space Division/YEZ
Los Angeles, California

The success of the Global Positioning System (GPS) satellite program rests primarily on the reliability and performance of Cesium and Rubidium Atomic Frequency Standards and Crystal Oscillators. GPS can point with pride to two years of successful orbital experience with Rubidium and one year of Cesium. Program office activites and on-orbit performance over the last six months are summarized in this paper. The status of the GPS clock program is presented with special emphasis on the status of, and unfolding plan for, the GPS User Segment.

GPS Clock Program: Present and Future

Introduction

The past six months, since the last published status of the GPS clock program, have been exceedingly busy for the Air Force, Aerospace, and vendor members of the GPS clock team.[1] There have been design reviews, clock data reviews and "buy-offs," and a clock program meeting.

Clock performance on-orbit has been outstanding, with no new clock failures. The Rubidium (Rb) standard on Navstar 3 is still going strong after more than two years of continuous operation. Early orbit concerns about the Cesium (Cs) clock in Navstar 6 have been allayed by our improved understanding of zero-g clock operations. We have recently gone to the Cs standard in Navstar 5 and thereby doubled the rate at which we obtain orbital experience in that kind of clock.

Reliability is the major GPS clock issue and is unlikely to pass away completely in the near term. In the early days of GPS, time to run clocks on the ground was difficult to obtain. Today we are dollar-limited and somewhat hobbled financially, in part, by our orbital success. Were we having more reliability problems, more money would perhaps be available to work on solutions. A reasonable case can be made, however, that all significant mechanisms of infant mortality and mission-limiting catastrophic failure have been observed and eliminated. We are made further optimistic by the fact that clock orbital data gathered and processed shows an excellent correlation with data taken during ground testing. This fact makes us feel that we are conducting legitimate long-term tests of our clock on-orbit and can therefore feel reasonably confident that they will support a nominal mission. Our fingers (and toes) are crossed.

The question of whether to proceed with Cesium, Rubidium, or both has been largely answered by a consensus of the clock experts who spent a few days with us in March 1981. The words "largely answered" are used because the feeling remains that, given the GPS application of atomic clocks, if one could be intelligent enough about modeling techniques and characterizing clock noise, then perhaps clock differences become insignificant. More on this later.

There has been crystal performance on-orbit of itself worth mentioning. Navstar 1 has relied solely upon crystal clocks for the last 16 months. All indications are that the navigational data generated by this satellite is at least marginally useful in the short term to obtain user position solutions, if not for velocity calculations.

This talk will conclude with a programmatic discussion of the GPS user segment, where it stands, and where it plans to go in the next year or two.

GPS Navigation System Tutorial

As shown in figure 1, the GPS navigation system consists of three segments: Space, Control, and User. The Space Segment is the set of orbitting satellites, each one containing redundant atomic clocks which, running one-at-a-time, provide precise timing to that satellite's Navigation Subsystem (Nav). Within the Nav, several functions are performed as follows: two L-band carriers, L1 and L2, at 1.575 GHz and 1.227 GHz, respectively, are synthesized from the 10.23 MHz clock output. Two pseudo-random noise codes (PRNC), a precision (P) code, and a clear access (C/A) code are generated based upon the clock's timing, and ephemerides and clock modeling data for the whole GPS constellation are impressed upon the PRNC in a modulo-2 scheme. The P and C/A codes are individually applied to the L1 carrier through bi-phase shift key modulation and then the two modulated signals are combined in phase-quadrature to form the output signal which is then transmitted to the User and Control Segments.

Periodic updates of each satellite's ephemeris and clock models are performed by the Control Segment, which also continuously predicts the major parameters of total GPS system performance using a Kalman Filter. The Control Segment consists of monitor stations, located in Guam, Hawaii, Alaska, and Vandenberg AFB, CA., with primary and secondary upload stations located at Vandenberg AFB, CA., and Sunnyvale AFS, CA., respectively, and a Master Control Station (MCS) at Vandenberg. Put in basic terms, the monitor stations receive the satellites' L-band Nav signals and pass them to the MCS for processing. The MCS, using a Kalman Filter, takes the raw satellite data, makes the known, systematic corrections, weighs it statistically in the context of previous data, compares it with the best available reference, and then outputs its "best guess" as to the behavior of the several major Space and Control Segment operating parameters for the next 24 hours. The model thus generated is uploaded to each satellite daily or as necessary to keep the satellites, individually and collectively, operating in a usable fashion. The reference is periodically refreshed by passing data from Kalman to the Naval Surface Weapons Center for generation of a new reference.

The several space vehicles transmit the clock and ephemeris information, pertinent to all the satellites, by impressing it upon the two L-band carriers, as described above. Special equipment of the User Segment, the User Sets, receive and process this data and develop a navigation solution in three-dimensional position and velocity, and time, for the specific user. Users may choose from an array of sets to be discussed in more detail later. To date, users have successfully employed the GPS Navigation System to determine their position accurately on land, at sea, and in the air. The Air Force has used it convincingly in exercise bombing runs, as has the Navy for instrumentation of missiles test-fired at sea.

Clock Program Status

The previous report on this subject was positive and optimistic.[1] Nothing has happened to amend that view. Our primary nemesis of times past, the Rb lamp, has caused us no new difficulties. The Aerospace Corporation has been studying the lamp failure mechanism over the past year as well as looking at ways to ameliorate the situation. Our "fix" which involved using a "known, adequate Rb fill" still appears to be a good one and the added changes of going to other lamp materials having lower affinity for Rb are positive perhaps to the point of overkill. Aerospace, in cooperation with EFRATOM, Inc., is continuing to run lamps in vacuum by way of reducing our concern in this area; among those being run is a Sapphire lamp which may well be the right final answer. Diode lasers are also being studied in this connection.

Five of six GPS satellites on-orbit are considered operational at this writing. Navstar 1 is running on its last functioning crystal clock and providing some usable navigational data. Figure 2 is recent Kalman filter data. As you can see, its drift has been negative and fairly constant at about -3 X 10^{-11}/day. Navstar 2 is no longer considered useful for navigation, but is still functioning and is being utilized for diagnostic purposes. Navstar 3's Rb clock with representative Kalman data in figure 3, went over the two year mark on 20 January 1981. Its long-term drift is low and constant, on the order of -6 X 10^{-14}/day. Navstar 4's Rb standard shows similar performance, figure 4, and likewise low, constant drift of approximately -1.2 X 10^{-13}/day. Navstar 5 has, for nearly two months, been operating on its Cs clock; operation is nominal as in figure 5 and its drift is on the order of +2 X 10^{-14}/day. The Navstar 6 Cs, operating now for over a year, is well characterized in data and is exhibiting almost no long-term drift; it is fairly noisy in the short-term, but has essentially no long-term drift at all. Figure 6 illustrates. Aerospace Corporation, Dr. Fletcher Wicker in particular, has taken orbital data from the Control Segment and generated stability calculations on the GPS clocks. Plots for Navstars 3, 4, 5, and 6 are shown in figure 7. You may recall from my PTTI talk that the combination of rapidly rising ion-pump current with falling beam current and intermittent output frequency shifts on the part of the Navstar 6 Cs gave us considerable concern last summer.[1]

The program office views with mixed emotion the fact that the frequency shifts have not been observed since August of 1980. We are, on the one hand, delighted that the frequency shifts are no longer recurring, but remain concerned over the fact that we do not know what caused them and may not have the chance to gain further perception into that problem. Beam current and ion-pump current are changing only very slowly now and a long beam tube life is predicted. Figure 8 graphically brings this matter up to date. Diagnostic testing on the ground, using the Cesium qualification unit, was unsuccessful in duplicating the problem seen on-orbit. Navstar's 7 and 8 are awaiting launch decisions and are presently configured to have three Rb clocks and one Cs standard.

GPS User Segment

An important aspect of any briefing on GPS must be to discuss the present utility and the anticipated future value of the User side of the system. The finest satellite-borne atomic clocks in the world will do us no good if the User equipment cannot make efficacious use of the data generated.

The User Segment consists of the several developmental GPS User Sets deployed in the field by the four Department of Defense

branches and the Coast Guard (to name just the primary users) which are being tested under all tactical conditions. A GPS User Set is a time and frequency measuring receiver combined with a navigation data computer. The operating principle of the set is based upon the time-of-transmission codes modulated onto the L1 and L2 carriers. When this code is compared with the time of signal reception, an apparent signal propagation time, or "pseudo-range", is determined. Using this scheme, user position could be accurately determined with three satellites if their clocks were perfectly synchronized. GPS finds it more practical to use one-way transmission from four different satellites to solve simultaneously for user position and user receiver clock phase and frequency error. The user's ability to navigate depends primarily upon the accuracy with which satellite slant-range can be determined, and the geometric configuration of the four satellites and the user set. Blending and combining the range and clock measurements with an extrapolated navigation solution in a Kalman Filter makes optimum use of all the information available. The frequency measurements are actively used in the form of a delta-range (change in pseudo-range over a short time interval) measurement, computed by integrating the apparent Doppler shift of the tracked carrier frequency.

In this developmental phase, fighter aircraft used the X-Set, a high-dynamics configuration, which accepts inputs on 4 receiver channels at the same time. That number of channels allows it to handle four satellites' range measurements simultaneously. Cargo aircraft and the like, use the Y-Set, (for medium dynamics) with one aided input channel which sequentially tracks the satellites to develop its solutions for position, velocity, and time. Lastly, for ground vehicles, the Manpack or the Z-Set (low-dynamics sets) are used by those primarily concerned with portability or low cost. Their receivers also have a single channel which is stepped through the several satellites in view to develop their navigation solutions. These sets all use Kalman Filter techniques as an integral part of their computational software. The precise internal timing needed to put the set in time-step with each satellite, is provided by quartz crystal clocks. Submarines, which may operate for long periods of time between satellite updates, need and will use Cs clocks.

Two vendors, Rockwell-Collins in Cedar Rapids, Iowa and Magnavox in Torrance, California are in head-to-head competition for the contract to build the preproduction prototype User Sets. They are both working towards Critical Design Reviews to take place this year and field testing to begin late next year. The program office is targetting for a full-scale production contract award to one of these vendors in 1984.

Results to date from the developmental User Set testing show that the system works extremely well and has tremendous potential for providing navigation support to a very wide spectrum of military and civilian users in the years to come. Recently the low-dynamic manpack set successfully navigated, in field test, to an accuracy of better than 4 meters, a remarkable accomplishment! Some typical User Set test data is shown in figure 9. Results of bombing runs done at the US Army Yuma Proving Ground with an F-4 aircraft guided by an X-Set, cannot be shared in this forum but are, I assure you, equally impressive.

An important future use of GPS will involve rather than the task of navigation, the lesser included process of time-transfer. It can be expected, according to David Allan of NBS, that time can be transferred between points, by a GPS satellite in common view, to a synchronization error of less than 10 nsec. Even the preproduction User Sets being developed by Rockwell-Collins and Magnavox, although not optimized for time transfer, are expected to distribute UTC to users with accuracies of nearly 100 nsec. Imagination alone is the limit, it would appear.

Navstar GPS Clock Plan

On 10-11 March 1981, the GPS program office convened, at Los Angeles, the third in the recent series of clock meetings the purpose of which has been to instill a "team spirit" in those agencies and vendors involved in GPS and to receive the benefit of the objective scrutiny of a panel of technical experts, a listing of whose names is a legitimate WHO'S WHO of Time and Frequency in the United States. The consensus of the group was that, for the GPS application, the Cs standard should be the clock of choice, with Rb, because of its much larger tendency to drift, relegated to a "back-up" role which would make use of its superior short-term performance.

As an aside, the difference between their tendencies to drift in frequency offset makes the difference in the way that Cs and Rb clocks are modeled in the Kalman. It has been found that the Cs drift term is, for our purposes, very nearly zero. Rubidium's drift is much larger, may be positive or negative, and lends much more uncertainty to the time prediction problem than does Cs. The appearance of the matter is that if we are to continue with Rb, then we must either find a way to limit its long-term drift, or learn how to model it better. Probably both.

As in times past, the program office was criticized for not having articles in long-term ground test and the importance of this effort was strongly reiterated.

The following planned efforts then are reflective of present program office thinking as tempered by the advice we have received:

a. Rubidium - We will continue to buy the Rockwell built Rb clock, with small design changes, and add automatic baseplate control

(good to about 0.2^{0} C diurnally) as quickly as feasible. Additionally, offline, we will continue, through prototype procurement, to work with EG&G Inc., in Salem, Massachusetts. Their efforts have already produced some extremely useful output which time does not permit my mentioning here.

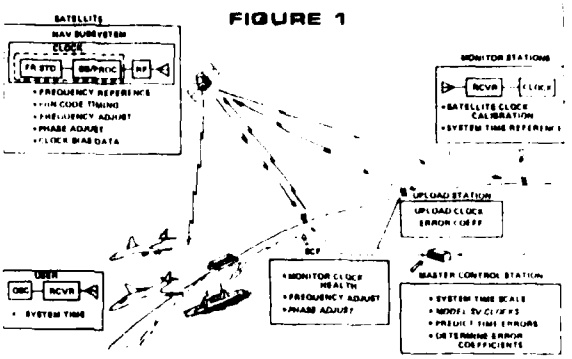
b. Cesium - We will continue to procure Cesium Clocks from Frequency and Time Systems, Inc., (FTS), through Rockwell International, to support our satellite needs. Our partners in precise timing at the Naval Research Laboratory (NRL) will continue to work, offline, in several different ways to improve the reliability and performance of the Cesium Clock. Negotiations are underway with FTS for contracts to eliminate the qualification deficiencies of our present Cs clock, to properly configure two Cs standards for long-term ground test, and to buy clock subassemblies for reliability testing. Further, NRL is actively working on Cs clock development contracts with Frequency Electronics, Inc., (FEI) in Long Island, New York and KERNCO, Inc., in Danvers, Massachusetts which will support the long-term performance, radiation-hardness, and reduced weight requirements of the GPS operational satellite.

c. Hydrogen Maser - The GPS program office views with great excitement and academic interest the NRL funded maser efforts underway with Hughes Research Laboratory in Malibu, California, the National Bureau of Standards in Boulder, Colorado, and the Smithsonian Astrophysical Observatory in Cambridge, Massachusetts. Potentials for improved navigational accuracy and performance, represented by the hydrogen maser, cause us to applaud the NRL effort and wonder again about the possibility of installing one of these marvelous machines in the Control Segment sometime soon.

In conclusion, the GPS Program Office is committed to putting the best performing and most reliable clocks available, from the state-of-the-art, on every satellite. Your comments and interest are welcome and actively solicited.

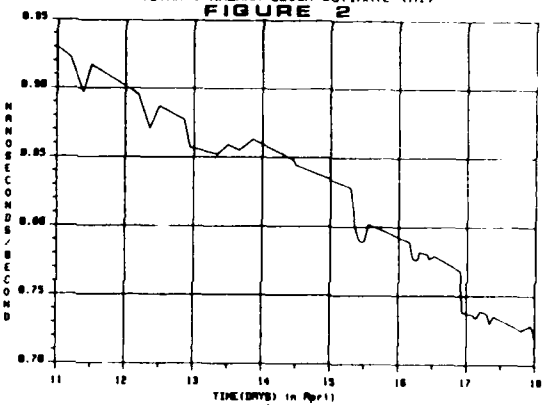
Reference:
1 D. Tennant; "Navstar Global Positioning System (GPS) Clock Program; Present and Future." 12th Annual PTTI Applications and Planning meeting. Greenbelt, MD, December 1980.

FIGURE 1

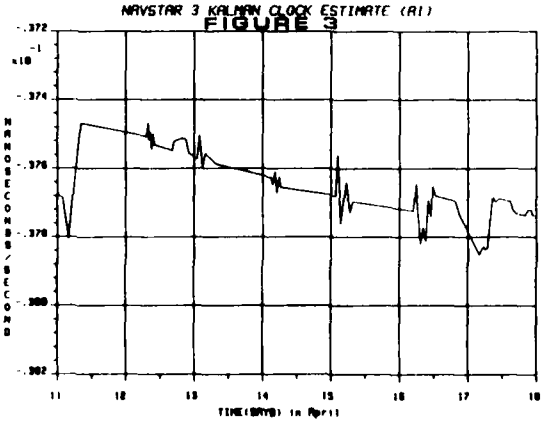


GPS CLOCK FUNCTIONS

NAVSTAR 1 KALMAN CLOCK ESTIMATE (A1)
FIGURE 2



NAVSTAR 3 KALMAN CLOCK ESTIMATE (A1)
FIGURE 3



NAVSTAR 4 KALMAN CLOCK ESTIMATE (A1)

FIGURE 4

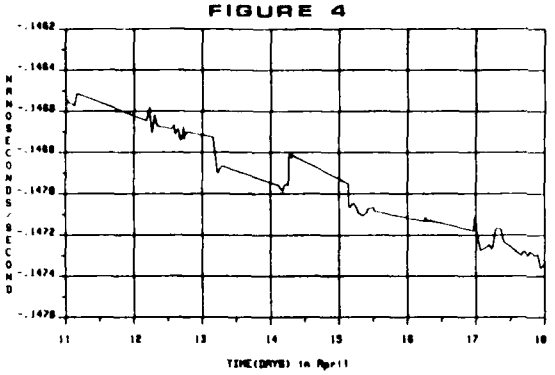


FIGURE 5

NAVSTAR 5 KALMAN CLOCK ESTIMATE (A1)

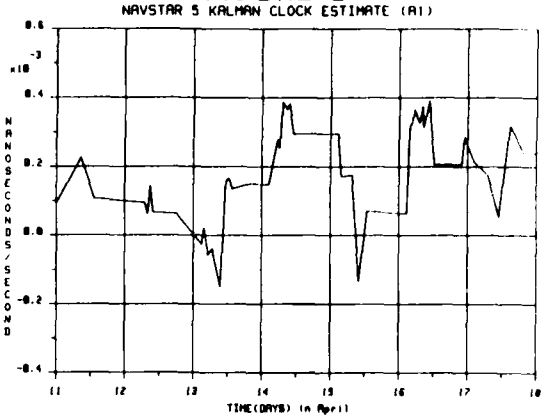
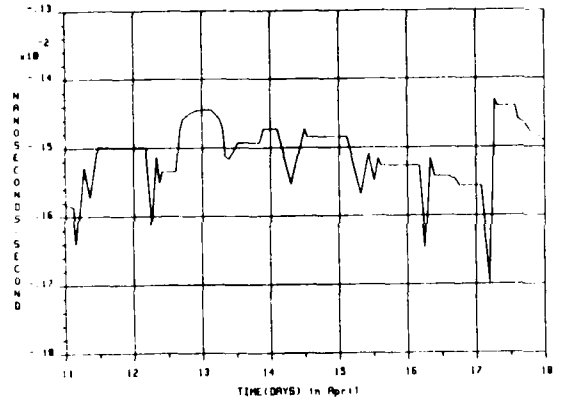


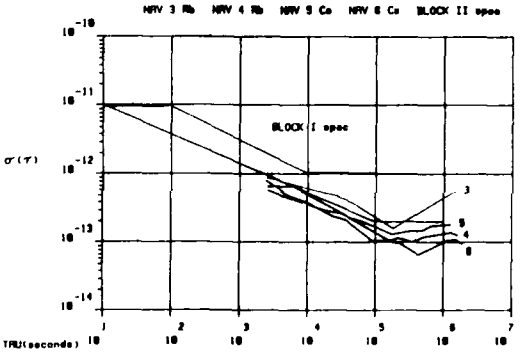
FIGURE 6

NAVSTAR 6 KALMAN CLOCK ESTIMATE (A1)



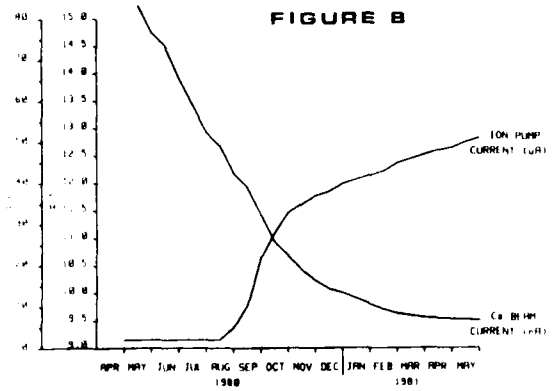
ON-ORBIT ALLEN VARIANCE CURVES

FIGURE 7



BEAM AND ION PUMP CURRENTS FOR NAV 6 SINCE LAUNCH

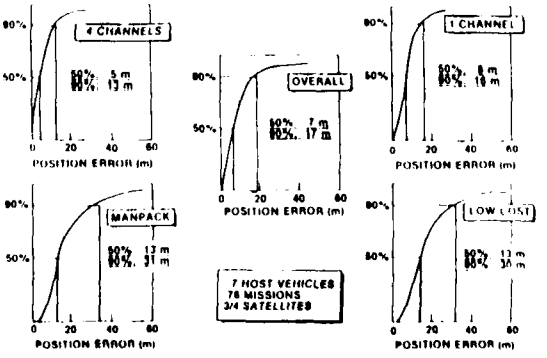
FIGURE 8



NAVSTAR Global Positioning System

PHASE I TEST RESULTS

FIGURE 9



Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

TIME DISSEMINATION USING NAVSTAR GLOBAL POSITIONING SYSTEM (GPS) PHASE IIB USER EQUIPMENT

M.D. Yakos and E.H. Hirt
Rockwell International/Collins Divisions
400 Collins Rd. N.E.
M.S. 124-310
Cedar Rapids, IA 52406

Abstract

The NAVSTAR Global Positioning System (GPS) is a satellite based navigation system that uses spread spectrum signal techniques to permit a User to obtain highly accurate, worldwide position, velocity and time information. This is achieved by measuring the difference in the time-of-arrival of a set of four satellite signals. GPS time is deterministically related to Universal Coordinated Time (UTC) such that a User can estimate UTC to submicrosecond accuracy. A key to system performance is the ability of the GPS User Equipment (UE) to estimate the time bias between UE reference clock and GPS System Time. This is accomplished by precisely controlling the hardware and the computational errors. It is therefore required that the UE employ a high quality crystal oscillator with low phase noise and predictable drift characteristics.

The application of accurate time keeping in a mobile environment is a basic principle of the GPS User Equipment Set. This Set with its receiver utilizes the information provided by the spaceborne segment of the NAVSTAR GPS system to provide the User with navigation information, GPS system time and corresponding UTC data. The complexity in mechanizing the GPS User Equipment Set time keeping function is discussed. This discussion includes both hardware and software compensation techniques. Highlighted is the utilization and effects of the crystal oscillator based frequency standard within the Set.

Introduction

The Navstar Global Positioning System (GPS) is a worldwide navigation system. A benefit of GPS often overlooked is that of precision time transfer.

The GPS system provides position, velocity and time information via satellites in 10,890 nautical mile circular orbits. The information transfer is accomplished by L-band radio frequency transmissions. A properly equipped User is able to detect and decode the RF signals and estimate his position, velocity and time. GPS system time will be known with respect to UTC to within 100 nanoseconds.

The purpose of this paper is to describe how the Phase IIB User Equipment (UE), currently under development for the NAVSTAR Joint Program Office, accomplishes the time dissemination and time transfer functions. For the purposes of this paper time dissemination and time transfer are defined as:

Time Dissemination - the reporting of time to remote Users.

Time transfer - synchronization of remote clocks to GPS system time (or UTC).

Notation

Vector quantities are denoted by underlined variables.

GPS NAVIGATION CONCEPT

The GPS system is a signal time-of-arrival navigation system. The satellites function as the transmitters operating at known locations with respect to an (ECEF) earth-centered-earth-fixed coordinate system. The User measures a range between his location and that of the satellites. Figure 1 depicts this relationship. His position is the mutual intersection of the spheres described by the range measurement and centered at the satellites. If the User knew time perfectly with respect to the satellite transmitters, his range to a satellite could be computed using (1).

$$R = c(\Delta T) \qquad (1)$$

Where:

- R = Range to the satellite, in meters
- c = Speed of light, in meters/second
- ΔT = Path delay transit time between satellite and User, in seconds.

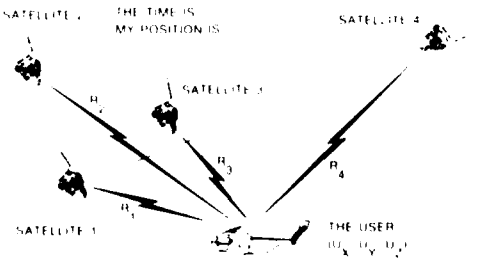


Figure 1. 3-D Navigating with the GPS.

In order to mechanize this ideal system it would be necessary for the User to have a high accuracy (atomic) clock, or to maintain a two-way phase lock with the satellites. Constraints imposed on the operational GPS preclude use of expensive atomic clocks by the User and User-to-satellite radio transmissions. The User clock operates asynchronously with the satellite clock.

It is not the intent of this paper to describe the GPS signal structure in detail as this has been covered adequately in the literature.[2] The segments of GPS are the Control Segment, the Space Segment and the User Segment. The Control Segment observes and corrects the satellite transmissions. The Space Segment is composed of an 18 satellite constellation arranged into six 55° inclination orbital planes of three satellites each. The User Segment is the beneficiary.

GPS is a one-way ranging system. The User to satellite "range", measured by the User, is actually a pseudorange defined by (2).

$$R_p = c\,\Delta T + c\,T_{ATM} + c\,CE_u \qquad (2)$$

Where:

c	=	Speed of light, in meters/second
ΔT	=	Path delay, in seconds
R_p	=	Pseudorange, in meters
T_{ATM}	=	Atmospheric delay, in seconds
CE_u	=	User clock error with respect to GPS time, in seconds

True range (1) can also be expressed as (3).

$$R = \sqrt{(X_s - X_u)^2 + (Y_s - Y_u)^2 + (Z_s - Z_u)^2} \qquad (3)$$

where:

X_s, Y_s, Z_s = Satellite position coordinates
X_u, Y_u, Z_u = User position coordinates

which results in (4).

$$R_p = \sqrt{(X_s - X_u)^2 + (Y_s - Y_u)^2 + (Z_s - Z_u)^2} - T_{ATM} + CE_u \qquad (4)$$

The GPS satellites transmit 50 bit per second navigation data describing the satellite position vector and satellite clock bias. The dual frequency L-band transmission allows the User to calculate T_{ATM}.

What remains is a one-way ranging equation containing four unknowns:

a. X_u, Y_u, Z_u the User position coordinates

b. CE_u the User clock error

Classically, this is a problem of four unknowns which require four simultaneous equations for a solution. By ranging four different satellites, the User obtains the necessary independent equations to solve for the unknowns.

The navagation and time solution are actually calculated by a Kalman filter within the GPS User Equipment (UE). The generic output of the GPS UE is position, velocity and time (PVT). So fundamental is this output that the receiver is commonly referred to as the "PVT".

Determination of Time GPS Phase IIB User Equipment

The first problem encountered when discussing time relationships within GPS is to define "time". For the purposes of GPS the primary time reference is Universal Coordinated Time. Within GPS several time references can be found:

Local Time (T_{UE}) - Uncorrected UE clock time (i.e., frequency standard)

GPS Time (T_{GPS}) - Time referenced to the Master Control Station of the Control Segment

Universal Coordinated Time (UTC) - GPS time deterministically adjusted for bias errors (i.e., leap seconds and fractional seconds)

Satellite Time (T_{SV}) - Local Time maintained at the satellite.

The GPS Phase IIB User Equipment (UE) is designed to provide precision time in several forms:

a. Time Mark
b. Data Capture
c. Control Display Unit Clock
d. Cesium Reset

A feature of the time mark output is that the "tone", fast risetime analog pulse, preceeds the "time", associated descriptive digital data. The digital data word contains both UTC and GPS time. In a sense it is similar to a WWV time of the tone announcement.

Time mark pulses occur at a nominal 1 pulse per second rate. Error of the time mark pulse is less than 150 nanoseconds relative to UTC.

The Data Capture pulse time output consists of a digital data word. This word represents the time at which an external strobe was applied to the GPS UE. Accuracy of this word is approximately 1 millisecond with respect to UTC. In case of conflict, the data capture pulse time has a higher priority than the time mark output.

The Control Display Unit (CDU) has the capability of displaying time (UTC) with a resolution of 1 second (see Figure 2).

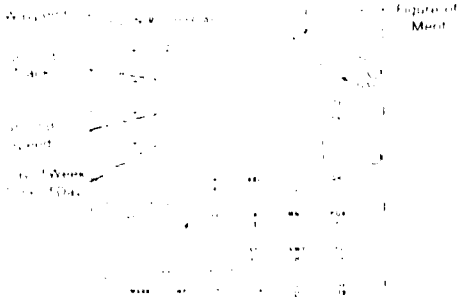


Figure 2. UTC CDU Time Readout.

The previously described time outputs are intended for time dissemination purposes. Additionally there is a class of Users who are interested in synchronizing their own precision laboratory grade atomic clocks to GPS derived UTC. For this reason a separate capability is provided by the GPS UE for time transfer Users. This capability is generically referred to as cesium reset.

Cesium Reset provides a fast risetime analog pulse output at the UTC one minute rollover. This output differs from the time mark pulse in that the time mark pulse is totally asynchronous with respect to UTC, but the cesium reset pulse is forced to occur synchronous with UTC. Accuracy of the cesium reset pulse is approximately 500 nanoseconds. In addition, the UE accepts 1 pulse per second data from the external cesium from which the UE computes a time difference between cesium time and UTC. This time difference is displayed on the CDU as shown in Figure 3.

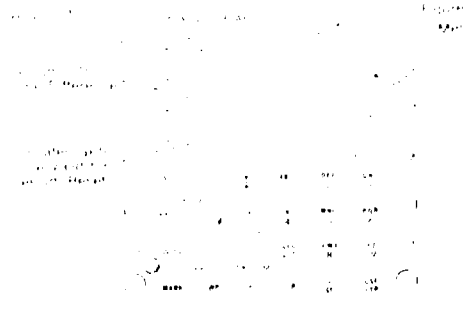


Figure 3. GPS CDU Cesium Time Transfer.

The intent of the display (Figure 3) is to inform the operator as to the cesium/UTC difference with a resolution of 1 microsecond. When the difference exceeds a specified maximum, the operator would manually synchronize the cesium (to 500 nanosecond accuracy) using the reset pulse from the GPS UE.

An extension of the cesium reset capability is possible which could provide a more accurate reset pulse, in the order of 100 nanoseconds, at a higher rate (1 Hz), with a higher resolution difference display on the CDU (100 nanoseconds). This capability will be known as PTTI, precise time and time interval.

GPS Time Dissemination and Time Transfer

The meaning of time determination in GPS ultimately refers to an estimate of UTC via the NAVSTAR GPS. The Control and Space Segments serve as a conduit for transferring UTC to the User. Figure 4 shows this relationship. The User in the end analysis is provided and estimate of UTC for either dissemination or transfer to another clock. The User can also determine GPS system time which may be adequate if only relative time synchronization among Users is required such as in time division multiple axis (TDMA) applications.

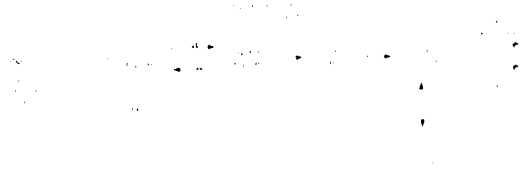


Figure 4. GPS Time Dissemination and Time Transfer.

UE Time Derivation

Control Segment - Time in the GPS system begins with a portable clock that is flown to the GPS Master Control Station. This clock maintains a version of UTC. At the Master Control Station time is transferred from the portable clock (which represents UTC) to GPS time, Figure 5. GPS time is a paper clock maintained by the Master Control Station. Other inputs into GPS time are cesium beam clocks resident at the Control Segment monitor stations.

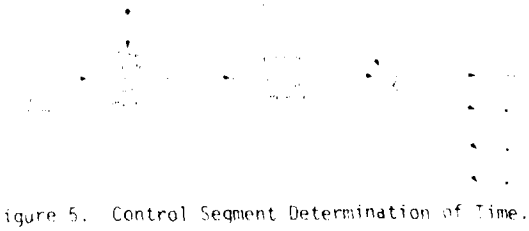


Figure 5. Control Segment Determination of Time.

GPS and UTC will diverge due to relative frequency offset, drift and leap second adjustments made to UTC. However, the difference is continuously estimated and reported to the User via the 50 bit per second data.

In the Space Segment the satellites have their own precision atomic clocks from which all RF and digital signals are derived. The satellite clocks maintain their own version of time which we will call T_{SV}. The Control Segment monitors the satellite transmissions on a once daily basis and is thereby able to estimate the difference in T_{GPS} and T_{SV}. Correction data is loaded into the satellite, combined with the pseudonoise ranging codes and transmitted to the User. UTC/T_{GPS} corrections are also included in the data. Figure 6 shows this relationship.

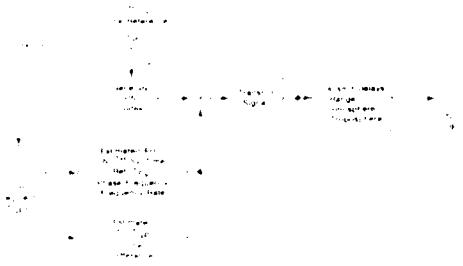


Figure 6. Satellite Segment Determination of Time.

Once transmitted from the satellite the signals are subject to transit delays including ionospheric and tropospheric delays before they reach the User. These transit delays must be appropriately compensated for in the User Equipment.

The User Equipment contains a clock identified in Figure 8 as the frequency standard. The UE frequency standard maintains its own version of time called T_{UE}. The task is to compute the bias in UE local time (5) in order to derive GPS time which then can be related to UTC.

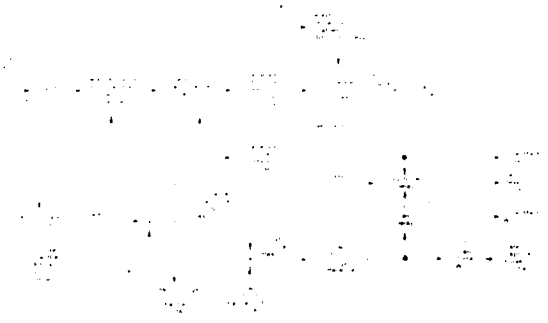


Figure 7. User Equipment.

$$CE(u) = T_{GPS} - T_{UE} \quad (5)$$

Where:

CE(u) = User clock error (i.e., bias)

Time Bias Determination

Determination of the receiver time bias is one of the basic functions of the User Equipment software. To describe this process it is useful to first describe the way in which satellite range is measured by the User Equipment. The following discussion is simplified from the actual UE operation but serves to describe the process.

The phase of the pseudorandom noise code transmitted from the satellite corresponds to the time as kept by the satellite clock. When received by the UE the phase relative to GPS time is derived using (6).

$$T_r = T_{GPS} - T_{PD} + CE(s) \quad (6)$$

Where:

T_r = Received code phase
T_{GPS} = GPS time
T_{PD} = Total path delay
CE(s) = Satellite clock error

The code phase of the signal generated locally in the UE is:

$$T_L = T_{UE} + \Delta P \quad (7)$$

Where:

T_L = Local code phase
T_{UE} = Users time reference
ΔP = Offset between the local code (T_L) and User time reference (T_{UE})

When the User Equipment is tracking the satellite signal:

$$T_L = T_r \quad (8)$$

Therefore:

$$T_{GPS} - T_{PD} + CE(s) = T_{UE} + \Delta P \quad (9)$$

The value of ΔP is the offset in the local code required to make (7) true. Solving (9) for ΔP:

$$\Delta P = T_{GPS} - T_{UE} - T_{PD} + CE(s) \quad (10)$$

Using (5) to simplify (10) yields:

$$\Delta P = CE(u) + CE(s) - T_{PD} \quad (11)$$

Finally, T_{PD} is composed of the true path delay plus various hardware delays in the satellite and in the UE:

$$T_{PD} = (R/c) + T_{ATM} + T_H \quad (12)$$

Where:

T_H = Hardware delays in the satellite and UE

Solving (11) for R yields:

$$R = c \left[CE(u) + CE(s) - \Delta P - T_{ATM} - T_H \right] \quad (13)$$

To solve for the true range, R, each of the right half terms must be known:

a. CE(s) is derived from the satellite clock correction parameters transmitted in the down-link data which include relativistic corrections.
b. ΔP is measured by the receiver controller software.
c. T_{ATM} is derived from the ratio of measurements on L_1 and L_2 frequencies.
d. T_H is determined by Set calibration.
e. CE(u) is determined by the Kalman filter.

Time bias determination by the Set begins by setting the User clock ,T_{UE}, in (10). Absolute accuracy in initially determining T_{UE} is not necessary since it will be further refined by the Kalman estimate of CE(u). Large errors in T_{UE} must be avoided since the calculation of CE(s), satellite position, etc. require the use of T_{GPS} which must be approximated by T_{UE}.

The primary means of determining T_{UE} is by detection of the handover word (HOW) in the downlink data of the satellite. The process is started by performing a code generator reset (i.e., initializing the code generator to the beginning of the GPS week).

If, following the reset, the code VCO is zero and no slew commands are issued, the phase of the code generator would exactly correspond to the number of 50 Hz interrupts since the reset. Non-zero VCO frequency and code slew commands will cause a code phase difference which can be determined by integrating the VCO commands.

Following the code generator reset, the UE receiver controller software will be commanded to acquire a particular satellite on C/A (course acquisition) code. The purpose of the C/A code is to provide necessary timing information to acquire the very long sequence length GPS P code.

Once the C/A code is acquired: T_r modulo 1 millisecond equals T_L modulo 1 millisecond. Since the sequence length of the C/A code is 1 millisecond.

After bit synchronization is attained on the 50 bit per second data. T_r modulo 20 millisecond equals T_L modulo 20 milliseconds. since a GPS data bit is 20 milliseconds long.

Contained in the 50 bit per second data demodulated from the C/A code is the HOW timing word which defines the satellite time of transmission at the leading edge of the next subframe of data. A VCO slew command (T_{SLEW}) with a granularity of 20 milliseconds is sent to the code generator such that:

$$T_{UE} = N - N_r - T_{SLEW} \quad (14)$$

Where:

N = 50 Hz interrupt count
N_R = Interrupt count at reset
T_{SLEW} = Slew value derived from the HOW (in 20 millisecond steps)

Normally the receiver is commanded beyond the expected received code phase, T_r. Subsequent fine granularity back step slew commands are issued until the code loop locks on.

At this point the receiver is making P code pseudorange measurements. It is these pseudorange measurements that are fed into the Kalman filter. The Kalman filter is an optimal estimator that determines $CE_{(u)}$.

Kalman Filter Mechanization

The Kalman filter maintains the statistical and geometric information which is necessary for efficient use of GPS measurements. The position velocity and time (PVT) solution consists of a nominal solution which is corrected by the Kalman filter as shown in Figure 8.

The Kalman filter mechanization is actually composed of four primary functions. These functions are measurement preprocessing, the Kalman filter, solution generation, and receiver aiding. Figure 8 shows a generalized block diagram of these functions and their relation to the rest of the Set software. The nominal solution, $\underline{X}_{NOM}$, and the navigation solution, $\underline{X}_{NAV}$, can be considered the PVT input and output respectively. A brief description of these functions will follow.

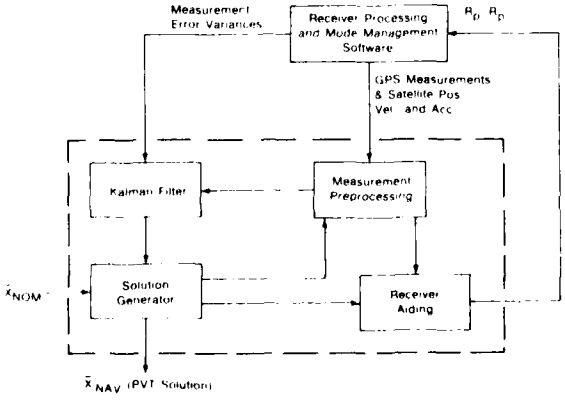


Figure 8. Navigation Software Functions.

Measurement Preprocessing accepts GPS measurements (pseudorange and pseudo-delta-range) from the receiver along with the corresponding satellite position, velocity, and acceleration in ECEF coordinates. The GPS measurements are then linearized prior to processing by the Kalman filter. Also, User to satellite line-of-sight unit vectors are provided for receiver aiding.

The Kalman Filter accepts linearized GPS measurement data and measurement error variances. The filter uses this data to estimate a correction to the present (a priori) navigation solution. The filter also maintains estimate error statistics in the form of an estimate error covariance matrix.

The Solution Generator computes the navigation solution, $\underline{X}_{NAV}$, as required. These navigation solutions are required for measurement preprocessing and receiver aiding, in addition to being the Set PVT solution. The solution is formed by propagating old solutions (possibly with the aid of INS or some other velocity aiding data) and incorporation of the corrections from the Kalman Filter.

There are three state vectors which are maintained by the solution generator. These are the nominal solution, $\underline{X}_{NOM}$ (which contains aiding information), $\underline{X}_{NAV}$ the solution vector, and the Kalman filter generated correction vector $\underline{EX}$. ECEF (earth-centered-earth-fixed) coordinates are used for all position, velocity and acceleration related quantities. External aiding is converted to ECEF coordinates by the interface software. The relation between these three vectors is:

$$\underline{X}_{NAV} = \underline{X}_{NOM} - \underline{EX} \qquad (15)$$

The function of the solution generator is, simply stated, to maintain these vectors in a timely fashion.

The generation of (15) yields a vector which contains the Set PVT solution. In particular, the PVT solution consists of User position $\underline{p}$, velocity, $\underline{v}$, and the GPS time estimate T_{GPS}.

The PVA (position-velocity-acceleration) filter mode is used when there is no position, velocity, or acceleration aiding source available to the Set. An altitude aid may be available. The nominal solution for this mode is presented in (16).

$$\underline{X}_{NOM} = \begin{bmatrix} 0 \\ \underline{0} \\ \underline{0} \\ \delta h_b \\ T_{UE} \\ 0 \end{bmatrix} \begin{array}{l} \text{Position aiding} \\ \text{Velocity aiding} \\ \text{Acceleration aiding} \\ \text{Altitude aid (if available)} \\ \text{Uncorrected UE time} \end{array} \qquad (16)$$

In the PVA filter mode, altitude and Set time are the only navigation aids. The zeros in the state vector are added to correspond to the correction states in $\underline{EX}$. For the PVA filter mode, $\underline{EX}$ contains the "whole value" estimates of position, velocity, and acceleration. The PVA error vector is given in (17).

$$\underline{EX} = \begin{bmatrix} -\underline{p} \\ -\underline{v} \\ -\underline{a} \\ \delta h_b \\ CE_{(u)} \\ -D_{FS} \end{bmatrix} \begin{array}{l} \text{ECEF User position} \\ \text{ECEF User velocity} \\ \text{ECEF User acceleration} \\ \text{Altitude aid bias} \\ \text{Estimated error in } T_{UE} \text{ (i.e., bias)} \\ \text{Frequency standard drift} \end{array} \qquad (17)$$

Notice that the Kalman state vector (17) contains states for UE clock bias and frequency drift. The drift information is important in that CE_u is not calculated continuously, but rather at discrete Kalman filter cycles. Extrapolation of CE_u to the time of signal reception is accomplished through use of the current drift estimate. Referring to Figure 9.

$$\Delta C_{UE} = (1/c)\ (D_{FS})\ (T_{UEN} - T_{UER}) \qquad (18)$$

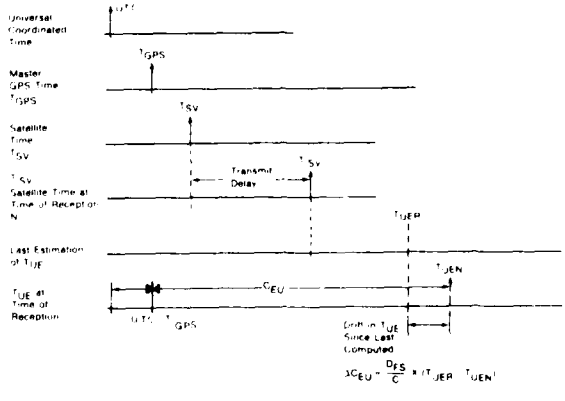


Figure 9. GPS Timing Relationships.

Where:

- ΔC_{UE} = Drift since last Kalman estimate of User bias error, in seconds
- D_{FS} = Current estimate of User clock drift, in meters/seconds
- T_{UER} = User time of last filter cycle for which T_{UE} was calculated, in seconds
- T_{UEN} = User time of signal reception, in seconds.

Once D_{FS} is estimated, it can be assumed to be fairly well behaved over periods of up to 1/2 hour. Consequently underdetermined navigation (less than 4 satellites) can be provided. This is sometimes referred to as "riding the clock". Note that the time output accuracy in this case would be a function of how well the D_{FS} estimate is made.

Another function of the Kalman filter is to properly compensate for all the biases in the

system as shown in Figure 10. The ultimate output of the Kalman process is a correction (CE_u) to User time TU_E.

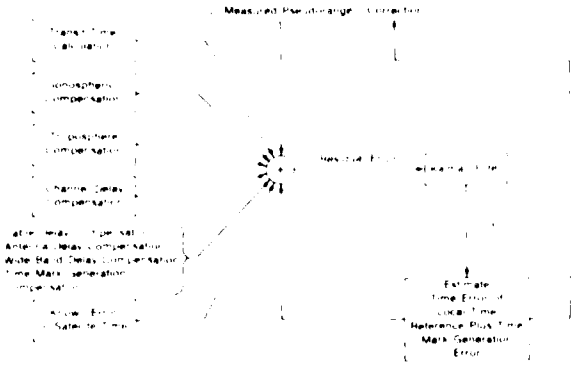


Figure 10. Kalman Filter.

Time Accuracy

Time accuracy of GPS with respect to the User Equipment is a function of the Set design and the inherent error contributed by the other two system segments. Primarily, these are the accuracy with which the Control Segment maintains system time (UTC) and the delays associated with Satellite Segment to UE Set range relationship. This concept of time accuracy can be expressed as a rms value which is a function of the three system segments as shown in (19).

$$A_t = \sqrt{(E_{ue})^2 + (E_{ss})^2 + (E_{cs})^2} \qquad (19)$$

Where:

- A_t = System time accuracy, in nanoseconds.
- E_{ue} = Uncompensated User Set dependent time error, in nanoseconds.
- E_{ss} = Random Time error due to the geometric relationship with the satellite segment and pseudorange measurement innaccuracies, in nanoseconds.
- E_{cs} = Time error inherent to the control segment (estimate of UTC), in nanoseconds (100 nanoseconds, rms).

The error inherent in the control segment, E_{cs}, is 100 nanoseconds rms.

E_{ss} is defined as the time dilution of precision (TDOP) multiplied by the User to satellite range error.[1] Where:

$$E_{ss} = (1/c)(TDOP)(UERE_{sys}) = 27 \text{ nanoseconds} \qquad (20)$$

Where:

- c = Speed of light = 299,792,500 meters/second
- TDOP = Time dilution of precision (1.68)
- $UERE_{sys}$ = User equivalent range error of the GPS system. This relates to the random error associated with the pseudorange measurements, in meters (5.3)

The GPS system specifications (SS-GPS-300 and SS-US-200) define $UERE_{sys}$ to be 5.3 meters and TDOP to be 1.68.

The time bias error attributable to the UE is a function of acceleration sensitivity and hardware delays (19).

$$E_{ue} = E_a + E_h \qquad (21)$$

Where:

- E_a = Uncompensated time error due to Frequency Standard gravity (g) sensitivity
- E_h = Uncompensated time error attributable to hardware (e.g., signal delays)

Hardware Delays in the UE will be determined by analysis and Set calibration. Software routines will bias pseudorange measurements to compensate for these hardware delays.

Acceleration Sensitivity

The User Equipment system specification (SS-US-200) requires the User Equipment to be designed for acceleration levels of:

a. Single Channel Sets - 0.6 g's maximum constant acceleration.
b. Dual Channel Sets - 2.04 g's maximum constant acceleration.
c. Five Channel Sets - 9.17 g's maximum constant acceleration

The User Equipment frequency standard is the primary contributor to the time error resulting from host vehicle acceleration, E_a, as determined by (22).

$$E_a = (D_a)(FS_{as})(HV_a) \qquad (22)$$

Where:

- D_a = Duration of host vehicle acceleration
- FS_{as} = Frequency Standard acceleration sensitivity, per g
- HV_a = Host vehicle acceleration, in g's

As previously discussed, a function of the Kalman filter mechanization is to correct for UE clock bias. Included in this correction factor is the provision to deweight measurements based on host vehicle acceleration data. An available design option is to deterministicly model the frequency standard and apply a deterministic correction for acceleration via the Kalman filter.

In the event that a User is stationary, and the UE noncompensated hardware bias errors, E_h, are

very small, the total system time error can then be approximated by (23):

$$A_t = \sqrt{E_{ss}^2 + E_{cs}^2} = 104 \text{ (nanoseconds)} \qquad (23)$$

For:

$E_{cs} = 100$ nanoseconds

$E_{ss} = 27$ nanoseconds

Hardware Description

Each UE Set is comprised of an antenna subsystem, control display, signal processing, navigation and interface functions. For the purposes of this paper we will highlight the F-16 Set (Figure 11) (i.e., Five Channel Set). The F-16 also represents the Fighter/Attack host vehicle category for GPS Phase IIB.

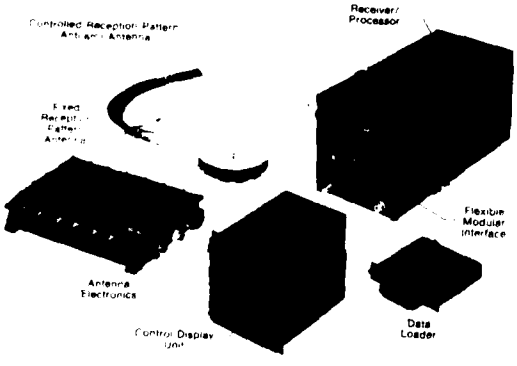


Figure 11. GPS Phase IIB F-16 User Equipment.

The F-16 Set is comprised of:

a. Top mounted controlled reception antenna (CRPA) providing signals to the antenna electronics (AE).

b. Bottom mounted fixed reception antenna (FRPA) providing signals to the antenna electronics (AE).

c. Antenna electronics (AE) providing L_1 and L_2 (L-band signals) IF signals to the receiver group.

d. Receiver group which provide primary information (IF and data) processing and host vehicle interface.

e. Data loader system provides cartridge data entry.

f. Optional control display unit (Standard CDU) provides operator interface and manual data entry.

The UE hardware which primarily affects the time mark output is depicted in Figure 12. A Set component particularly important to the accurate determination of time is the frequency standard.

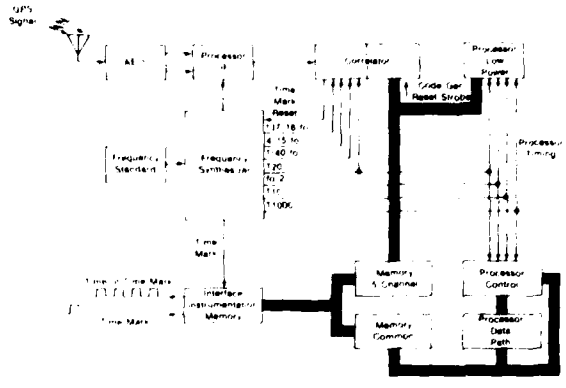


Figure 12. UE Hardware Timing Path.

Frequency Standard

The primary clock or frequency standard (Figure 13) of all UE Sets is an IT cut crystal controlled oscillator with a 10.23 MHz output. This frequency standard is designed for portable and airborne applications with a very rapid warm-up time and low power consumption. Critical parameters of the frequency standard are listed in Table I. A functional block diagram is shown in Figure 14.

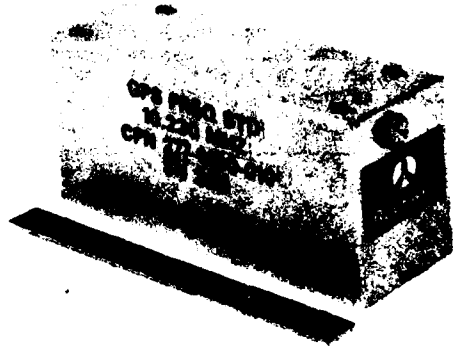


Figure 13. UE Frequency Standard.

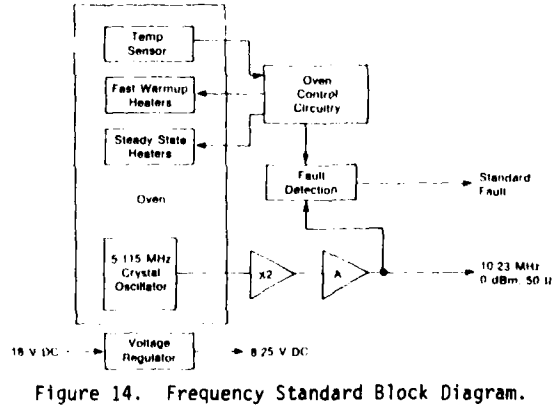


Figure 14. Frequency Standard Block Diagram.

PARAMETER	SPECIFICATION	CONSTRANT
Ambient Stability	$\pm 1 \times 10^{-8}$ Max $\pm 5 \times 10^{-10}$/sec.	-54°C to +85°C + 5°C/min, -54°C to +85°C
Voltage Stability	$\pm 1 \times 10^{-9}$	5 percent supply change
Load Stability	$\pm 1 \times 10^{-9}$	Any load change
Short Term Stability	$\pm 1 \times 10^{-10}$ rms	0.2 sec measurement intervals
Short Term Frequency Drift	$\pm 1 \times 10^{-9}$	For 20 minutes
Warm-up Time	$\pm 1 \times 10^{-8}$ $\Delta f/f$ in 30 minutes	From a 6 hr cold soak at -54°C, after 3 minutes
	$\pm 1 \times 10^{-9}$/sec	Maximum Δf/sec after 3 minutes
Acceleration Sensitivity	3×10^{-10} g 2×10^{-9} g	Long axis (maximum) Short axis (maximum)

TABLE I. FREQUENCY STANDARD PERFORMANCE

Conclusions

The GPS Phase IIB User Equipment can provide a degree of accuracy in time dissemenation and transfer that rivals the portable atomic clock method. It is available as a side benefit to the basic navigation function. Standard output interfaces are provided to accomodate a variety of potential Users. Accuracy is enhanced if the User is stationary. This equipment will be available in the late 1983 time frame.

References

1. Jorgensen, P.S., "Normalized Accuracy Analysis of the NAVSTAR GPS", Aerospace Corporation report TOR-0078(3475-10)-2.

2. Spilker, J.J., Global Positioning System, "Signal Structure and Performance Characteristics", The Institute of Navigation, Washington, D.C., 1980.

3. Hise, T.L., "Receiver Code Control/Time Estimation", Rockwell International Internal Letter, Cedar Rapids, Iowa, December 26, 1981.

4. Sadowsky, J.S., "User Equipment Kalman Filter Description", Rockwell International Internal Letter, Cedar Rapids, Iowa, February 26, 1981.

5. Oscillator, Crystal Controlled, Specification Control Drawing, No. 277-0510, Rockwell International Corporation, Collins Group, February 13, 1981.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

CONSTRUCTION AND PERFORMANCE CHARACTERISTICS OF A PROTOTYPE NBS/GPS RECEIVER

Dick D. Davis, Marc Weiss, Alvin Clements
and David W. Allan

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

Summary

The National Bureau of Standards (NBS) has proposed a particular application of the clear access channel C/A of the Global Positioning System (GPS) signal utilizing the fact that the location of two earth stations may be known. Hence, if one has common-view of a single satellite from these two earth stations, excellent time transfer capability exists. NBS has developed a prototype receiver featuring extremely high time transfer accuracy and low cost. Even though one may not know the absolute delays through the receivers, one can do absolute time transfer by knowing the differential delay between two receivers.

The received satellite signals gave an RMS time fluctuation of the receiver output as good as 3.5 nanoseconds for an omni antenna using 15 second averages. The noise was characterized as white noise phase modulation, which can be averaged below the systematics, which are about 1 nanosecond over a thermal range of several degrees about ambient. The day-to-day time fluctuations, when measuring the time difference between the NBS Boulder and U.S. Naval Observatory (USNO) Washington, D.C. were about 5 ns.

The software and the receiver are configured to be fully automatic with a Z80A microprocessor setting the amplitude for the lock loops of the receiver and setting the synthesizer which corrects for the nominal Doppler shift. The receiver also has a unique feature of using the microprocesser to calibrate a 0.1 ns built-in time interval counter. All that is required on the part of the user is a local 1 pps tick and a 5 MHz signal, plus his local coordinates.

Key Words: Automatic time comparison; deep space network; differential time transfer; frequency transfer; international time comparison; primary frequency standards; SI second; Global Positioning System.

Introduction

The advent of GPS has opened other doors besides navigation for applications using this system. In particular, the staff at the Time and Frequency Division of NBS have capitalized on the idea of using a single GPS satellite signal in common view at two fixed and known coordinate locations on the surface of the earth, e.g., NBS in Boulder, CO, USNO in Washington, DC, National Research Council (NRC) in Ottawa, Canada, Physikalisch-Technische Bundesanstalt (PTB) in Braunschweig, West Germany, or the Bureau International de l'Heure (BIH) in Paris, France. Since navigation is not necessary, this system can be turned into a pure time transfer technique for achieving a significant amount of common-mode cancellation of errors. In particular, any satellite clock errors disappear when the difference is taken between the received times at the two sites; the ephemeris errors for the satellite might be reduced by an order of magnitude or more depending upon the size and direction of the ephemeris errors, and upon the baseline distance between the two earth stations. If convenient times are chosen, there may be significant amounts of common-mode cancellation in the ionosphere of the two paths from the satellite to the two ground stations. A previous paper has been published showing the estimated effects of these various error sources, the results of which gave significant encouragement to pursue this type of approach.[1]

Over the course of about the last 1½ years, a few of the staff of the Time and Frequency Division at NBS have vigorously pursued the design, construction, and testing of a receiver to accomplish these goals. The basic design goals were high accuracy and low cost so that hopefully some industry could capitalize on the research and development effort and make the same available on the open market to a significant number of interested users. On 4 May 1981, the prototype receiver at NBS was locked to a GPS satellite signal on the first attempt. The results of the received data thereafter will be presented in part in the body of the paper.

One of the cost-cutting aspects that has been incorporated is to use the C/A signal only, since preliminary studies showed that it would have adequate accuracy to meet the design goal of better than 10 ns worldwide time transfer using the common-view approach.

Receiver Configuration

Hardware Configuration. The system consists of three rack-mounted chassis, 13.3 cm (5¼ in.) H x 40.6 cm (16 in.) D. A separate low-noise amplifier and down-converter are mounted at the antenna. One rack-mounted chassis contains the receiver processing circuitry, the second contains the the microprocessor-counter and the third is a fail-safe (battery-backup) power supply.

The system is controlled by a Z80A processor operating with a 4 MHz clock. The processor card, which is designed for this application, includes 32 k-bytes of RAM, 16 k of EPROM, 2 serial UARTs (1 for RS-232 and 1 for tape), 9 eight-bit parallel ports, a memory-mapped video display generator and an eight-level priority interrupt structure. Total power consumption of the processor card is eight watts.

Local I/O is provided through a parallel input keyboard and video display. The video display provides "warm feelings" to the user by continuously displaying system status. Local hard copy is available through the RS-232. Two systems can communicate with each other through a dedicated auto-answer/auto-originate modem.

A "micro-cassette" tape drive in the processor chassis is used to load/save programs and data. Each cassette has a capacity of approximately 50 k-bytes.

The time interval counter card in the processor chassis helps keep system cost low with no compromise in performance. The counter is a start-stop interpolator type and is calibrated by the processor before each satellite pass measurement begins. Counter resolution is 0.1 ns. Absolute accuracy is better than 1 ns (3 σ).

Since the ultimate goal is to provide differential time transfer capability with maximum accuracy, we have paid particular attention to those factors in the receiver design that can result in delay variations. For example, receiver bandwidth (RF and IF) is much larger than necessary if the goal were simply to optimize signal/noise performance. Also, special attention was given to the correlation loop servo to minimize the effect of temperature. Although we have not performed exhaustive tests up to this time, it appears that the receiver systematics are less than ±2 ns in a normal room environment where temperature variations are on the order of ±5 °C.

The software and hardware are capable of providing analog azimuth and elevation outputs to steer a high-gain (16 dB) antenna if this option appears desirable to minimize the effects of multipath on timing accuracy. (This option is certainly not necessary from the standpoint of S/N, since in 10 minutes we can average down to the 1-2 ns level of the receiver systematics with the present quad spiral omni antenna wound on a 2.5 cm diameter x 6 cm long cylinder with a 7 cm ground plane. We are able to track satellites down to a 1° elevation angle with the omni antenna.)

Software Configuration. The key element in the performance of this "stand-alone" GPS time transfer system is in the integration of the hardware and software. Since all control and computation is performed by a single Z80A processor, the software is required to be very efficient in execution time, especially the interrupt-driven routines that must perform a return from interrupt within 500 µs. A key "number-crunching" requirement is sufficient numeric accuracy to compute satellite position to 1 cm.

We chose to start from scratch, with the deveopment of a 15 decimal digit floating point package with hex interpreter for number crunching and a unique interrupt implementation in Z80 assembly language for the interrupt-driven programs. The total software package presently stands at approximately 20 k-bytes of object code. This represents about 1 man-year of programming time.

There are 7 interrupt-driven programs, plus the interruptable "number cruncher." The interrupt-driven programs are:

Priority	Name	Interrupt Rate	Function
1	Receiver	50Hz/1kHz	Lock Rcvr-Track
2	Tape	75Hz/200Hz	Read/Write Tape Files
3	RX	150Hz	Receive RS-232 Data
4	Spare	---	---
5	Keyboard	30Hz	Read Kbd-Display-Video
6	Counter	30Hz	Make Counter Measurements
7	TX	150Hz	Transmit RS-232 Data
8	1Hz	1Hz	Run Real Time UTC Clock

The interrupt-driven programs represent a processor overhead time of from 5% to 15%, depending upon what is happening at any instant of time.

One crucial point is that all 7 interrupt-driven programs and the interruptable number cruncher are "alive" at all times (naturally in a time-slice context). It is possible that at a given second, the receiver program is loading the serial data from a satellite and running the GPS clock, the tape program is in the process of writing a data file, the RX program is decoding an operator command, the counter program is making a pseudo-range measurement, the TX program is printing the hard copy results of passes for the previous 10 days, and the 1 Hz program is updating the UTC clock and has determined that this is the second to stop tracking the current satellite. Meanwhile, the interruptable program is in the process of computing the x,y,z coordinates of the satellite now being tracked.

Receiver Results

Shown in figure 1 are the time deviations between the space vehicle clock and the NBS reference clock. The space vehicle clock is a cesium onboard NAVSTAR-5 and, of course, the deviations include propagation fluctuations as well. One observes over the 10½ minute segment of data an rms time fluctuation of 3.5 ns. Each point is a 15 s average. In figure 2 is plotted $\sigma_y(\tau)$ fractional frequency stability and a new method (see Ref. 2) of frequency stability measurement called "modified" for Mod $\sigma_y(\tau)$ vs. the sample time, τ, for the data shown in figure 1. The resulting analysis indicates that the random noise process involved is commonly referred to as white noise phase modulation. The implication of this kind of noise is that one could average the time reading down to an uncertainty of better than 1 ns for averaging times longer than 3 minutes. Shown in figure 3 is a similar analysis for a rubidium standard onboard NAVSTAR-3, denoted space vehicle 6 vs. the reference clock at NBS and one sees similar performance. Also plotted on this $\sigma_y(\tau)$ diagram is the inferred stability of the onboard rubidium clock. The inference here is that one would be able to see the clock instabilities for averaging times beyond about 3000 s (roughly 1 hour).

Shown in figure 4 are the elevation angles vs. UT time for 15 June 1981 as observed from Boulder, Colorado and figure 4b is a similar plot as observed from Washington, DC. It is evident that there are several times over which the common-view, common-mode cancellation technique can be used to do time transfer between Boulder and Washington, DC. Advantage was taken of this to compare the clocks between NBS and USNO. Figure 5 is a plot of those time differences utilizing the average of the two rubidiums and the two cesiums onboard NAVSTAR-3, 4, 5, and 6, respectively. Since the differential delay between the two receivers was not known, a portable clock trip over 25-30 May 1981 was used to calibrate the differential delay before plotting the results in figure 5.

The time difference as obtained across the four satellites as compared to that by portable clock differed by approximately 335 ns. The reason for this difference is not known. It is possible that it is in how time is defined in the pseudo-random code by the designers of the receiver used at USNO vs. the NBS design. The next obvious step, of course, is to build two identical receivers, and compare them side-by-side. It is

fully believed and expected that this difference will disappear, or at least be reduced to the order of a few nanoseconds.

The agreement, when taken across the satellites, in the measure of the time difference UTC(USNO) - UTC(NBS) on any day on the average was 6.5 ns, which would yield a standard deviation of the mean of about 3 ns. This is very close to the clock noise for a sample time of only one day. The frequency difference averaged over 15 days between UTC(USNO) and UTC(NBS) was measured independently using each of the above four satellites and the values were as follows:

NAVSTAR	$y_{UTC(USNO) - UTC(NBS)}$
3	-4.0×10^{-14}
4	-4.0×10^{-14}
5	-3.7×10^{-14}
6	-4.4×10^{-14}
Average	-4.0×10^{-14}

The standard deviation of the mean is 0.14×10^{-14}. This is at least two orders of magnitude better than the results that can be obtained using Loran-C.

Conclusions

We have preliminarily demonstrated that a low-cost, high-accuracy GPS C/A receiver can be built, taking advantage of simultaneous common view of the same satellite. This appears to potentially allow time transfer with accuracies of the order or or less than 10 ns for baselines as large as 3000 kilometers, and frequency can be measured to better than 1 part in 10^{14}. With further studies of the propagation delays, this level of accuracy should be able to be extended on an international basis. It is fortunate that many of the key laboratories involved with accurate time and frequency are at high latitudes. This affords simultaneous common view between such sites as NBS, Boulder and New Delhi, India and USNO, Washington, DC and Radio Research Laboratories (RRL), Tokyo, Japan. This seems suprising at first, but upon calculation becomes obviously doable. This level of national and international time transfer accuracy, of course, affords also excellent frequency transfer accuracy at nominal state-of-the-art levels. These results have potentially significant impact on assisting the monitoring stations for GPS, on syntonizing the tracking stations for the Jet Propulsion Laboratory (JPL) Deep-Space Network, and for doing fundamental time and frequency comparisons for the international time scale and between primary frequency standards laboratories. Any future applications like the above, of course, depend on GPS remaining operationally available at better than the current levels of accuracy. At the present time, we do not know whether this will or will not be true.

Acknowledgments

The authors are deeply indebted to their sponsors, JPL and the U.S. Air Force Space Division. We are also most appreciative of the excellent cooperation of the staff at USNO, in particular, Dr. G. M. R. Winkler and Mr. Ken Putkovich.

References

1. David W. Allan and Marc Weiss, Proc. 34th Annual Symposium on Frequency Control, 334 (1980).
2. David W. Allan and J. A. Barnes, Proc. 35th Annual Symposium on Frequency Control (1981), to be published.

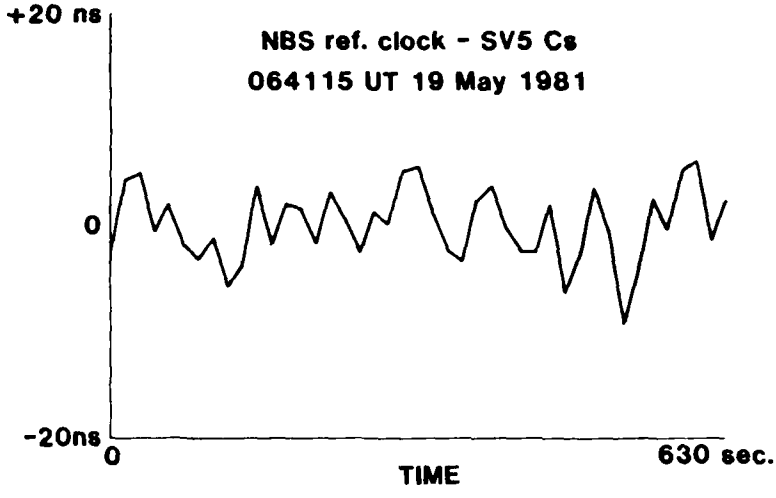


Fig. 1. A plot of the time deviations between a cesium clock onboard the GPS satellite, NAVSTAR-5, vs. the NBS reference clock. Each point is a 15 s average. The rms across the data set was 3.5 ns.

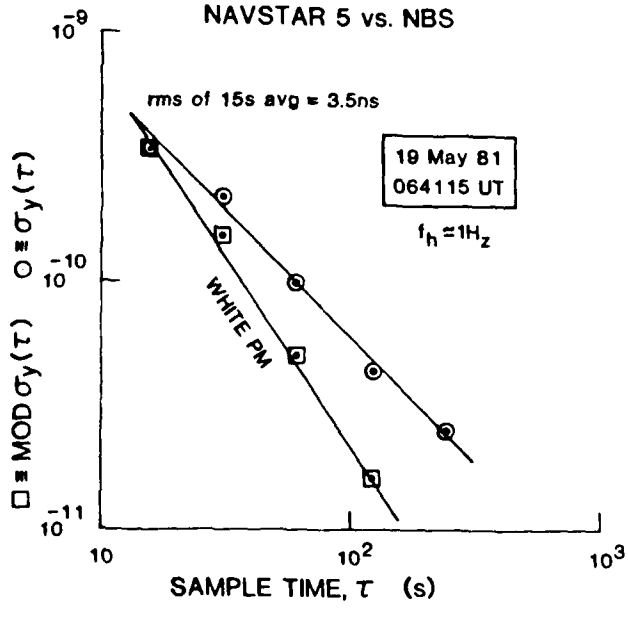


Fig. 2. A plot of the fractional frequency stability of the data shown in Fig. 1. Namely of the cesium in Space Vehicle No. 5 of the GPS constellation. The ordinate is double-valued for both $\sigma_y(\tau)$ and Mod $\sigma_y(\tau)$ as explained in Ref. 2. The noise is well-modeled by white noise PM at a level of 3.5 ns for 15 s averages. This noise is consistent with a meaningful usage of the standard deviation of the mean, which for this data set is 0.5 ns.

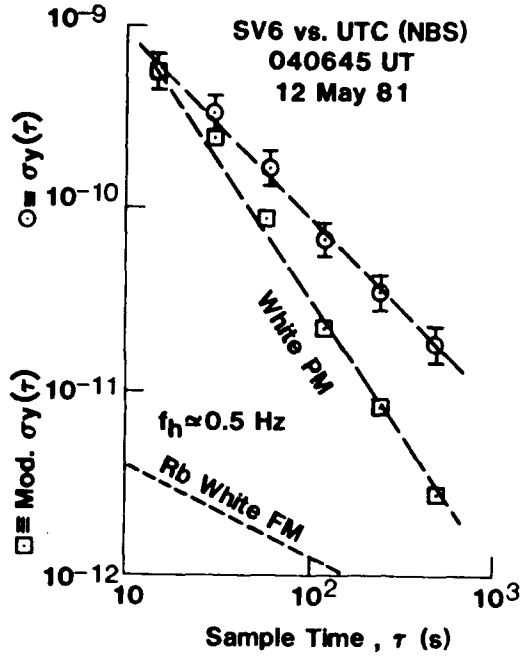


Fig. 3. Fractional frequency stability plot for $\sigma_y(\tau)$ and Mod $\sigma_y(\tau)$ of a rubidium clock in Space Vehicle 6 onboard NAVSTAR-3 as a function of sample time, τ, as explained in Ref. 2. Note also the estimated level of noise of the rubidium clock.

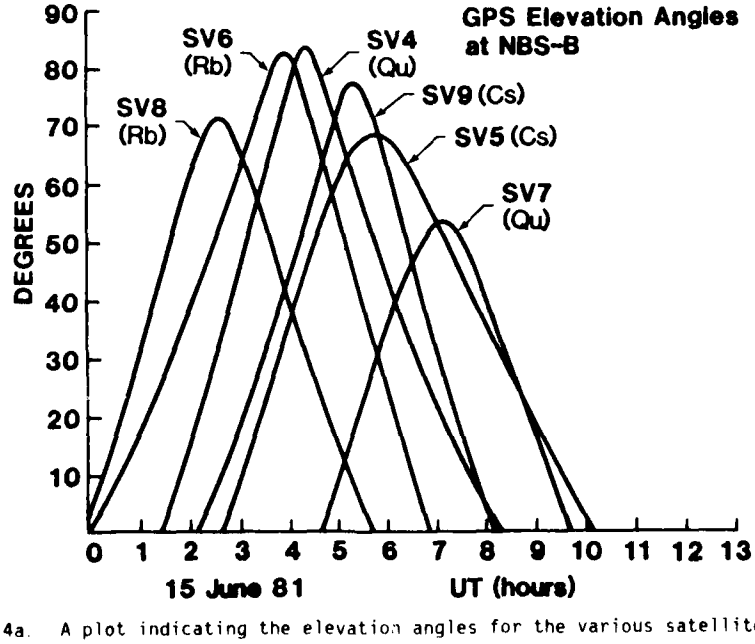


Fig. 4a. A plot indicating the elevation angles for the various satellites in the GPS constellation as a function of hours UT on 15 June 1981 as observed in Boulder, Colorado.

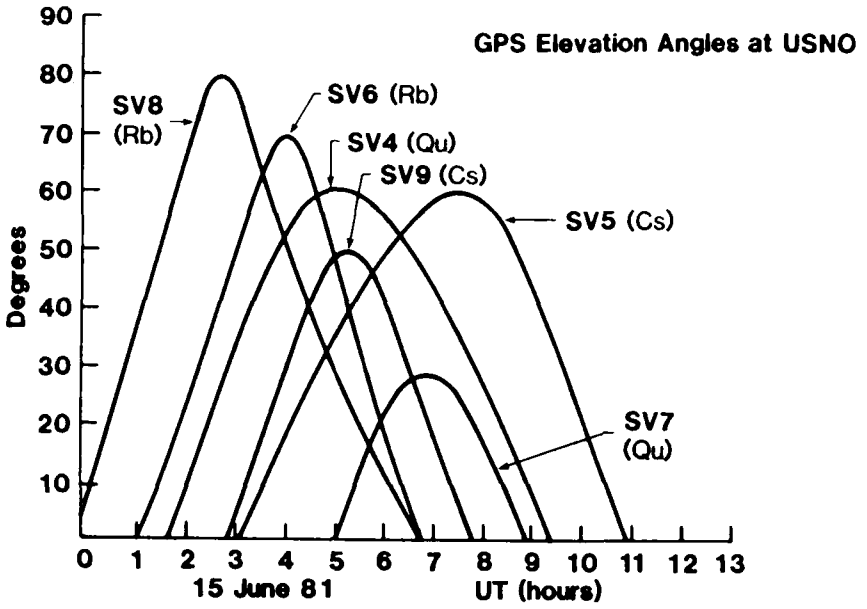


Fig. 4b. A similar plot to Fig. 4a only as observed from Washington, D.C.

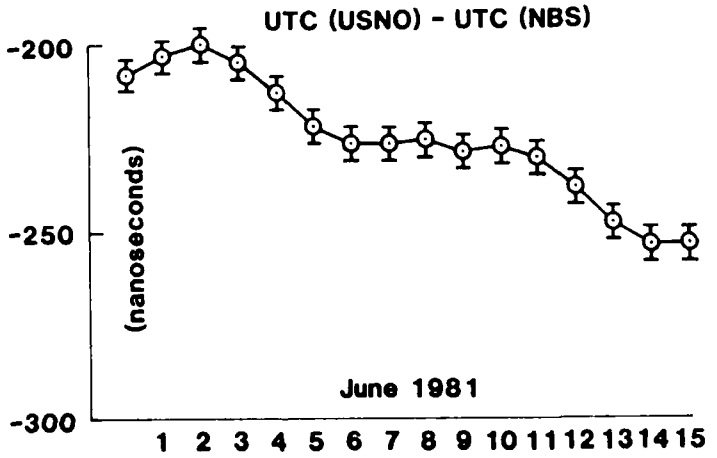


Fig. 5. A plot of the time difference between UTC(USNO) and UTC(NBS) via an average of the four GPS satellites NAVSTAR-3, -4, -5, -6. The error bars are the average standard deviation taken across the four satellites.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

THE NATO III 5 MHZ DISTRIBUTION SYSTEM

A. Vulcan and M. Bloch

Frequency Electronics, Inc.
New Hyde Park, N. Y. 11040

Abstract

This paper describes a high performance 5 MHz distribution system having extremely low phase noise and jitter characteristics and providing multiple buffered outputs. The system is completely redundant with automatic switchover and is self-testing. Faults can be isolated to a modular level by observing front panel status indicators. Since the 5 MHz reference signals distributed by the NATO III distribution system are used for up-conversion and multiplicative functions, a high degree of phase stability and isolation between outputs is necessary. Unique circuit design and packaging concepts are utilized to insure that the isolation between outputs is sufficient to guarantee a phase perturbation of less than 0.001° when other outputs are open circuited, short circuited or terminated in 50 ohms. The circuit design techniques include high isolation cascode amplifiers, the use of negative feedback to stabilize system gain and minimize circuit phase noise contributions, the use of balanced lines in lieu of single ended coaxial transmission media to minimize pickup and degradation of noise floor and the development of simplified fault detection and switchover circuitry to insure continuous operation.

The distribution system is fed by redundant high stability quartz frequency standards which use special crystals with low phase noise and jitter and a daily aging rate better than 5×10^{-11}.

Introduction

A 5 MHz signal generation and distribution system is described which provides the basic reference frequencies for both a transportable and fixed satellite ground station. The system has extremely low vibration induced phase jitter and high isolation between outputs. Both of these characteristics are required for the NATO III satellite terminal mission. The system is modular and self-testing and any modules can be replaced without disturbing system operation.

System Description

The block diagram of Figure 1 shows how a dual frequency standard, Primary Distribution Unit (PDU) and Secondary Distribution Unit (SDU), are interconnected to provide a failsafe system having fifty-six 5 MHz outputs, four 1 MHz outputs and four 100 kHz outputs. Two frequency standards in the Frequency Generation Unit (FGU) provide redundant, stable, low noise signals for the system input. The Electronic Switch Module (ESM) in the PDU accepts the 5 MHz input and feeds it to two Primary Distribution Modules (PDM). The dual electronic switches provide redundancy in the event of faiure of either frequency standard, interconnecting cable, or ESM. The Primary Distribution Module receives the redundant 5 MHz inputs and provides seven balanced outputs which feed various Secondary Distribution Units which can be located up to 500 feet from the PDU. Althouth the PDU, SDU and dual frequency standards comprise the NATO III station distribution system, the SDU can stand alone as a totally independent fifty-six channel low noise distribution system. Interconnection between the PDU and SDU is accomplished on a redundant basis such that the failure of either balanced interconnecting cable or PDM will not effect the final output signal. The requirement for low spurious signals and phase noise necessitates the use of double shielded RG-22B/U twinaxial 95 ohm balanced cable. The entire signal transmission system is isolated from the environmental ground reducing the pickup of unwanted signals due to ground loops and non-common mode noise. Table 1 lists the performance characteristics of the system and Figures 2 and 3 are the detailed block diagrams for the PDU and SDU, respectively.

An amplitude limiting circuit in the SDU Preamplifier Module (PAM) maintains a constant output level of +5 dBm from the SDU over a wide range of input signal levels. This is important so as not to affect the operation of user equipment when interconnecting cables are disconnected or a PDM module in the PDU is removed. Seven independent modules each having eight outputs comprise the balance of the SDU.

The equipment is designed to operate from American or European ac main voltages and frequencies. Dual power regulators ensure that 60 Hz and 120 Hz line related phase modulations are reduced to negligible levels. A high current regulator in each of the dual power supplies and an additional three-terminal

device in each module provides 0.01% line regulation with 150 microvolts of noise. An additional benefit of this system is that radiated susceptibility of the equipment from pickup of extraneous signals from colocated high power RF devices is minimized since the module regulators have the ability to reject interfering input signals by at least 80 dB from 30 Hz to 10 kHz and 40 dB up to 300 kHz. Additional RF filtering is used to meet the EMC requirements up to 18 GHz.

Each active circuit which is critical to system operation is failure-detected. Diode detectors are used for RF alarm generation and micrologic comparators sense the degradation of signal level below a preset value. The alarm signals are summed in each particular subsystem and fed to a common point where an output is provided to external monitoring equipment. Additionally, front panel indicators on the various modules visually indicate the occurrence of a failure or out of spec condition.

The system is designed such that removal or replacement of any module can be effected without causing intolerable phase or amplitude perturbations on other active outputs. This design insures that 5 MHz perturbations which can be multiplied 1000 to 2000 times in subsequent chains of frequency multiplications do not cause system outages. Minimal phase perturbations are insured by the use of high isolation amplifiers and a high degree of shielding between active circuitry.

Circuit Design

The following paragraphs describe two important functional blocks of the the distribution system.

RF Amplifiers

Two amplifier stages used for 5 MHz processing in various parts of the system are cascode circuits with high isolation, low noise, and high dynamic range. This circuit is shown schematically in Figure 4. To meet the isolation requirement of 100 dB, special layout and packaging techniques were utilized. Circuit gain is controlled by the unbypassed emitter resistor which provides ac and dc stabilization of the circuit. Output transformation circuitry converts the collector impedance to 50 ohms with a source VSWR of 1.2 to 1. Good grounding techniques with minimization of base lead inductance and parasitic reactances are necessary to insure optimal performance of the circuit. In the case where a balanced 93 ohm output impedance is required, a small toroidal transformer is utilized to effect the impedance transformation.

Diode CR1 provides bias temperature compensation and CR2 is an RF detector. Capacitor C10 is selected for a detected output voltage of 0.25 volts at the nominal RF level. The dc voltage feeds a voltage comparator whose reference voltage is set to provide an alarm signal at a point approximately 1-1/2 to 2 dB below the minimum acceptable RF level. Thus, temperature effects and other drift parameters will not cause a false alarm.

The various alarm signals are fed into the alarm module in the PDU and preamplifier module in the SDF to generate a composite alarm. These alarm signals which are TTL compatible, are summed with externally generated alarms which are translated to the proper levels and impedances before summation. In the case where the external alarm signals are fed from long twisted shielded pairs or are generated from relay contacts and/or noisy sources which may have high ac ground loop currents, optical couplers are used to isolate the noisy environment from the systems.

Limiting

The SDU must be capable of accepting an input dynamic range of 10 dB and maintain a constant output of +5 dBm $\pm$ 1 dB. Hence, a limiting circuit is necessary which has low noise characteristics and a constant input and output impedance to properly terminate the power splitters and maintain high isolation. Figure 5 shows the limiter used to achieve these objectives. Closely matched back-to-back RF signal diodes limit the RF amplitude to $\pm$0.7 volts and reflected power is absorbed in the hybrid terminations. The quadrature hybrids are implemented with lumped elements as shown in the figure.

Phase Perturbation Measurement

Figure 6 is a block diagram showing how the 0.0016 degree 5 MHz phase perturbation requirement is verified. The 5 MHz input signal feeds two high isolation amplifiers. One amplifier output feeds a X1000 multiplier chain and is used as the unperturbed reference. The other amplifier output feeds the unit under test which then drives an identical multiplier. The 5000 MHz outputs are mixed to yield a dc signal where amplitude is proportional to the phase difference at the mixer inputs. The mixer output is amplified and fed to a high sensitivity oscilloscope. Prior to making a measurement, the variable phase shifter is adjusted for maximum outputs from the EE-6093A Microwave Test Set. This condition determines E_0 in the relationship $\theta = \sin^{-1}(e/E_0) \times 10^{-3}$ where θ is the 5 MHz phase perturbation and e is the change in dc output voltage. The phase shifter is then adjusted for zero volts at the oscilloscope input (point of maximum mixer sensitivity), and the system is perturbed. The value of e is noted and θ is calculated. For example, if E_0 is 1.00 volts and a e of 50 millivolts is recorded, θ equals

0.0015 degrees. This measurement technique is quite versatile and is used for measuring both steady state and transient phase shifts that occur from shock and vibration events, removing and replacing modules, changing load impedances, and varying ac voltage inputs.

Mechanical Packaging

A modular concept has been used for the frequency distribution subsystem in order to enhance maintainability, simplify logistics, and permit economic manufacturing. Figure 7 is a photograph of the FGU showing both quartz frequency standards and a phase comparator. Internal batteries are provided for 24 hours of operation without ac power.

Figures 8 and 9 show the PDU which consists of five different module types. Each module is totally enclosed and has an RFI filter compartment in order to preserve shielding integrity. The modules can be inserted or removed directly from the front panel without disturbing other modules. Each module has a power indicator and various status and fault lights. Front panel test points are provided for monitoring the redundant dc voltages. Figures 10 and 11 shows the mechanical construction of the SDU. This drawer consists of three different module types, preamplifier, distribution amplifier and power supply. Seven distribution amplifiers are used to provide the fifty-six outputs. Figure 12 shows the internal construction of a preamplifier module. Each amplifier section is shielded to maintain an isolation of 100 dB between outputs. The RFI filtering compartment is totally isolated and the input connector directly addresses this compartment. Power, command and alarm signals are RFI filtered at this interface.

Environmental Considerations

The frequency distribution subsystem has been designed to meet the environmental specification shown in Table 1. Conservative component derating ensures that the equipment has a service life of 15 years. Performance specifications are met during high G inputs specifically encountered in transportable vans. Care has been taken to insure that microphonically induced phase and amplitude modulations are reduced to negligible levels. Critical circuit elements are staked in place using resilient adhesives and all RF interconnection cables are likewise encapsulated. The PC Boards are coated with a humidity resistant material. The circuitry has been designed to be essentially broadband in nature to minimize the effect of temperature variations on output levels. Over a range of 0 to +50°C, the level variation is less than $\pm$0.2 dB.

Conclusion

This paper describes a low noise frequency distribution system which is designed for continuous use in severe environments. The design stresses both long term reliability and electrical performance which emphasizes low spurious signals, low cross talk and low phase perturbations. The system consisting of a Frequency Generation Unit, a Primary Distribution Unit and a Secondary Distribution Unit has been qualified to MIL-E-16400 for environment and MIL-STD-461 for electromagnetic compatibility.

TABLE 1

PERFORMANCE CHARACTERISTICS

FREQUENCY DISTRIBUTION SUBSYSTEM

1.	OUTPUTS:	FIFTY-SIX 5 MHz AT + 5 dBm FOUR 1 MHz AT + 13 dBm FOUR 100 KHz AT + 13 dBm ONE 50 KHz AT - 100 dBm
2.	IMPEDANCE:	50 OHMS UNBALANCED, 95 OHMS BALANCED
3.	VSWR:	1.2:1
4.	PHASE PERTURBATIONS:	0.0016°
5.	AMPLITUDE PERTURBATIONS:	0.01 dB
6.	SPECTRAL PURITY:	- 120 dBc FROM 50 Hz TO 2 MHz OFFSET FROM CARRIER
7.	PHASE NOISE:	2 dB ADDITIVE COMPONENT, - 165 dBc/Hz FLOOR
8.	HARMONICS:	40 dB
9.	FREQUENCY STABILITY:	1×10^{-9}/MONTH, 2×10^{-12}/SECOND
10.	MICROPHONICS:	1×10^{-9}/G
11.	MTBF:	30,000 HOURS
12.	EMC:	MIL-STD-461
13.	VIBRATION:	2.5G FROM 2 Hz TO 500 Hz
14.	SHOCK:	15G FOR 11 mSECOND
15.	HUMIDITY:	95% RELATIVE
16.	TEMPERATURE:	- 20°C TO + 65°C

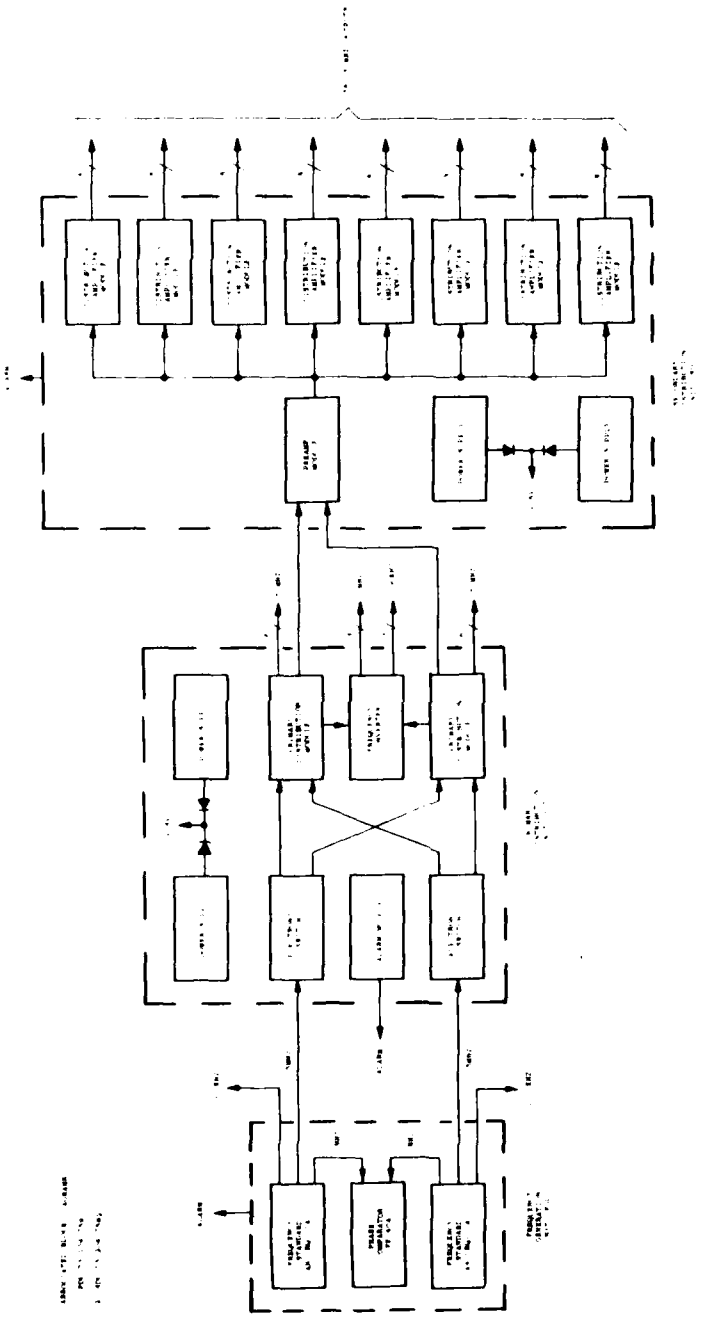


Figure 1. Frequency Distribution Sub-System Block Diagram

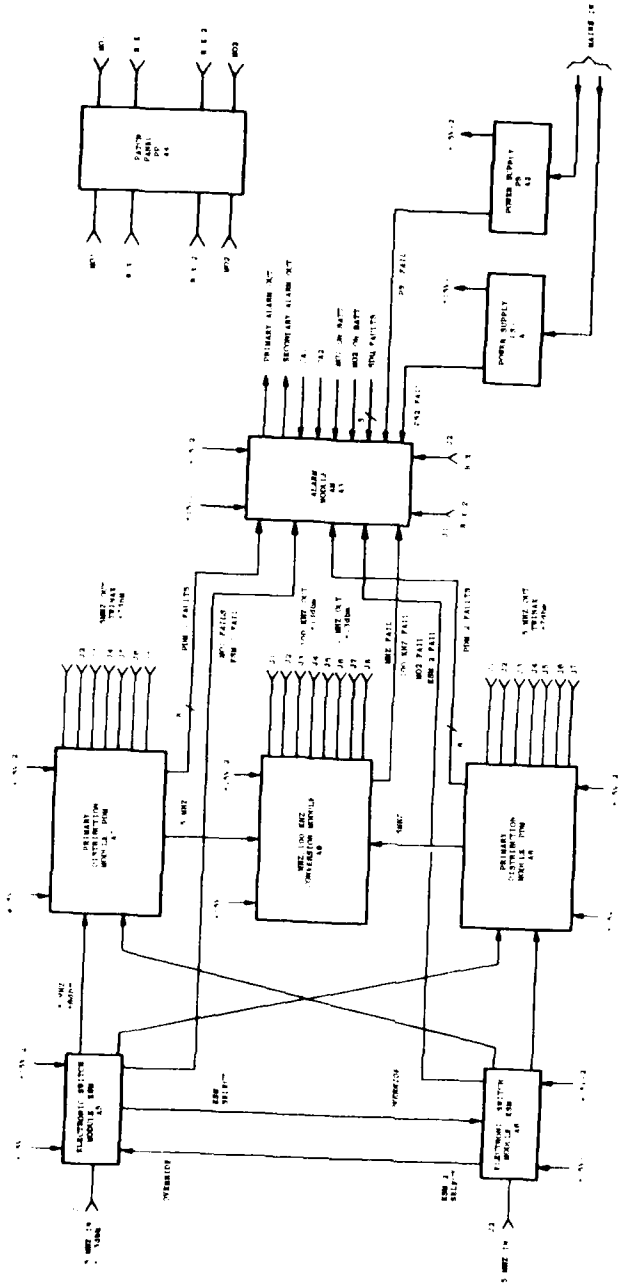


Figure 2. Primary Distribution Unit (PDU), Model FE-798A, Block Diagram

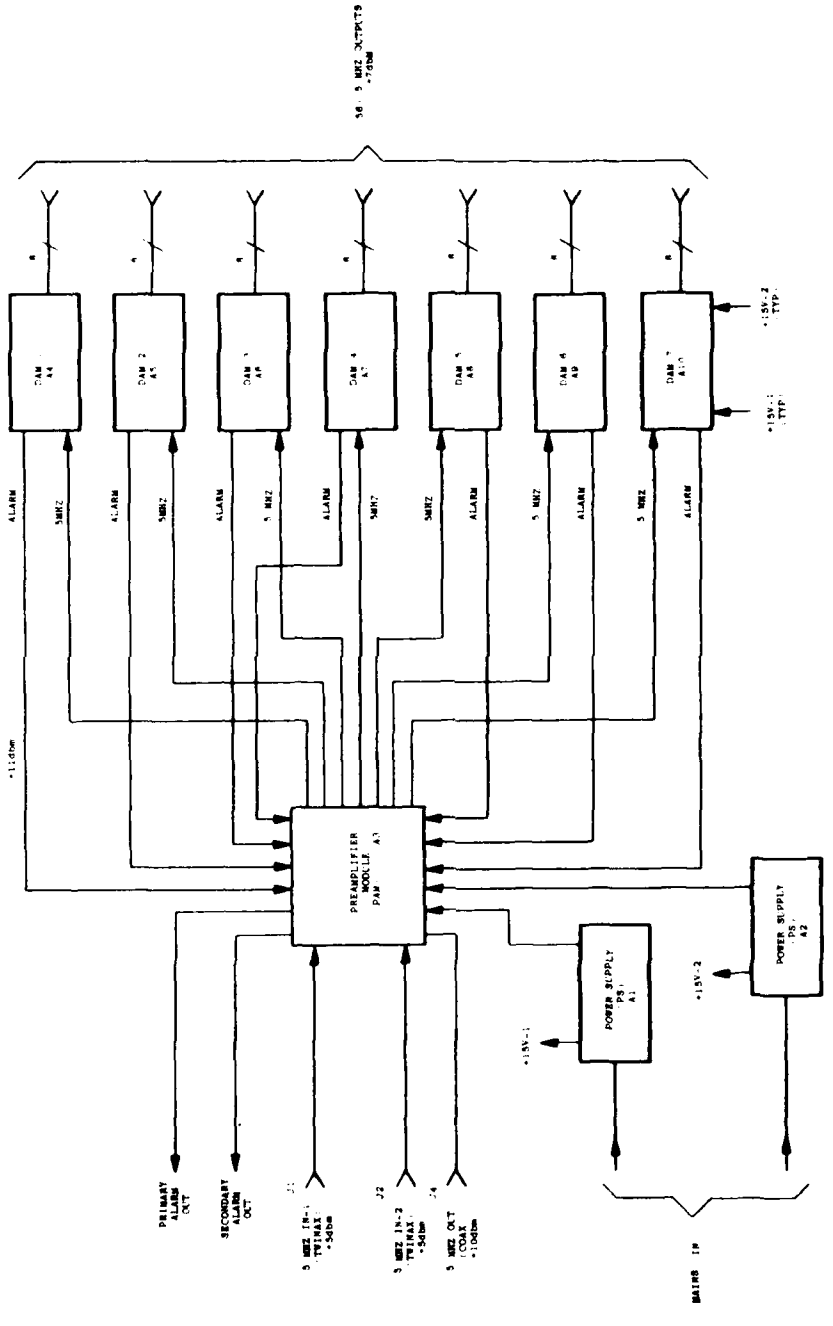


Figure 3. Secondary Distribution Unit (SDU), Model FE-799A

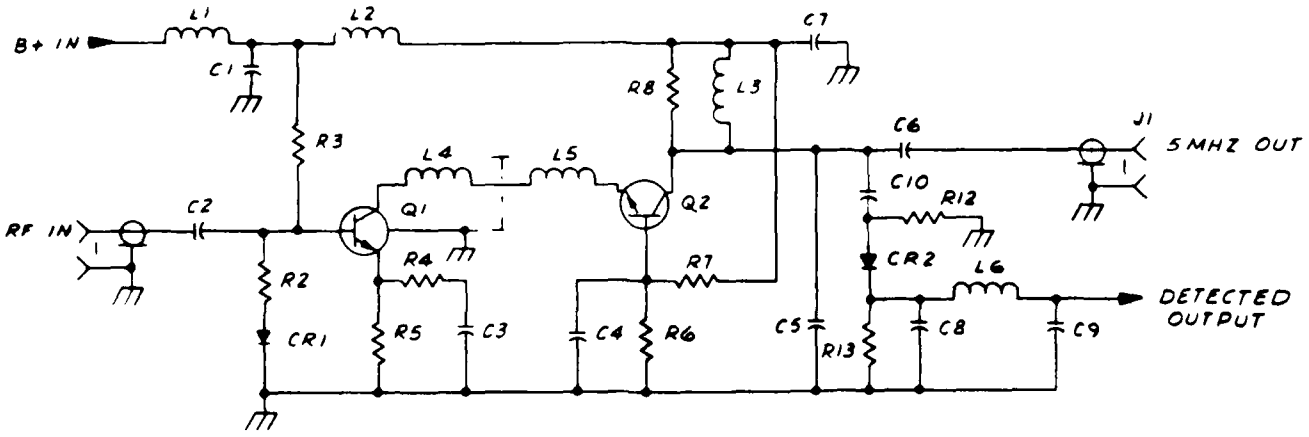


Figure 4. Cascode Amplifier, Schematic Diagram

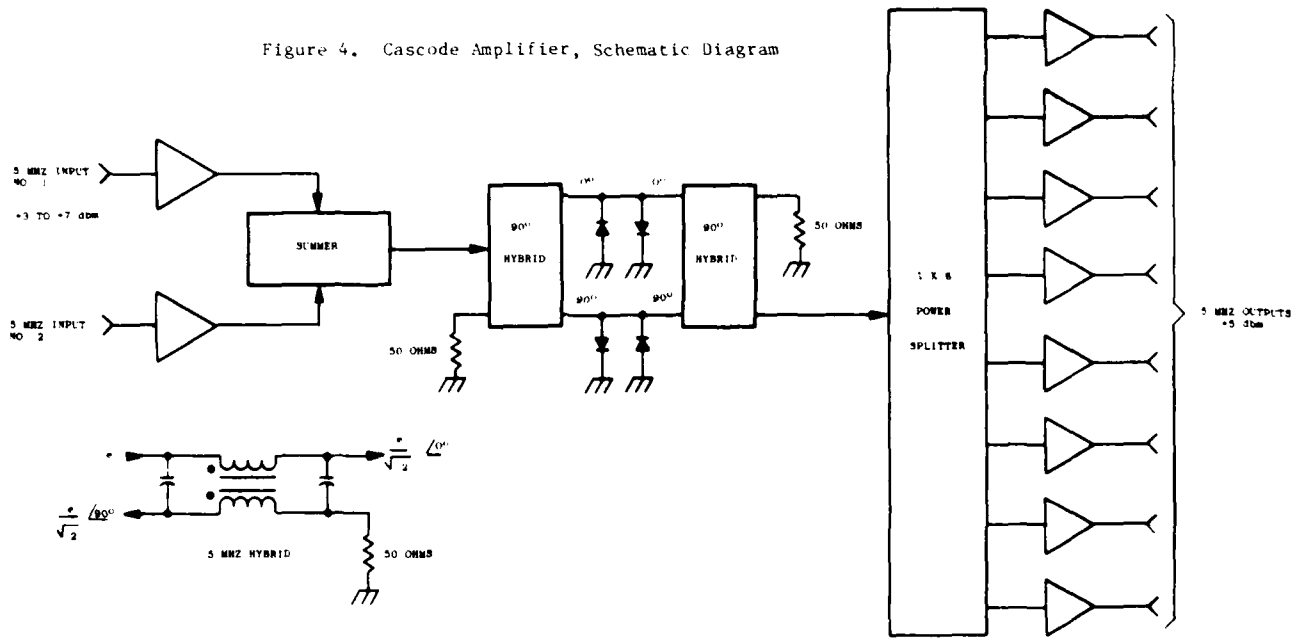


Figure 5. Block Diagram, Constant Impedance Limiter

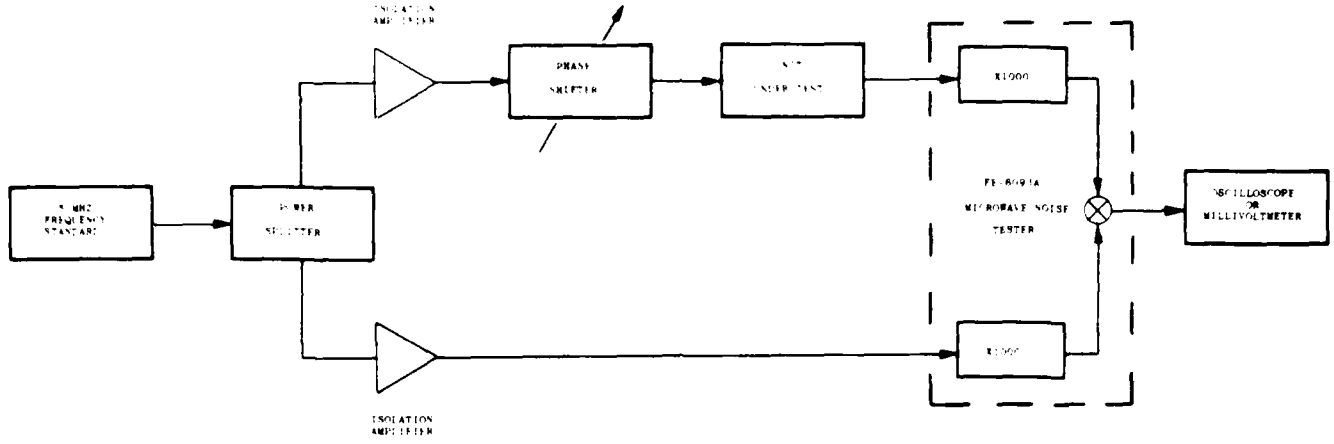


Figure 6. Test Set-Up, Phase Perturbations, Block Diagram

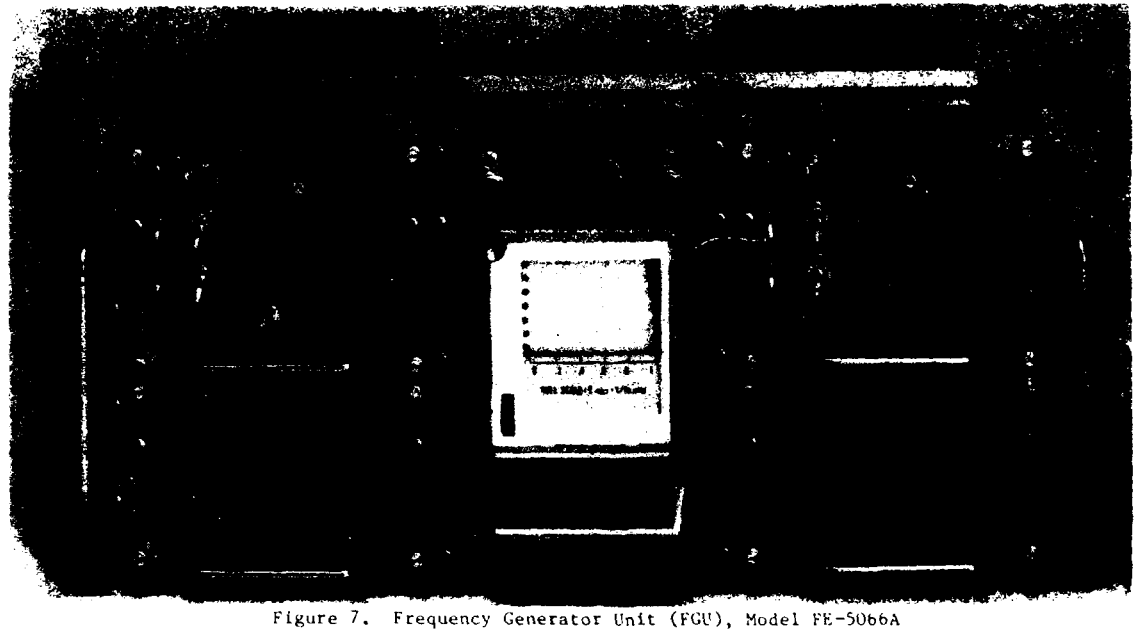


Figure 7. Frequency Generator Unit (FGU), Model FE-5066A

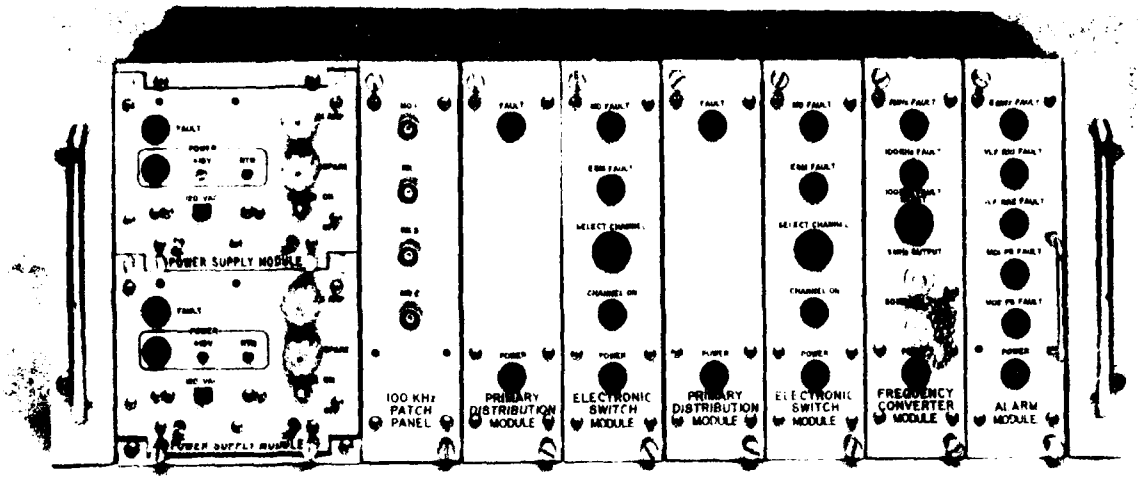


Figure 8. Primary Distribution Unit (PDU), Front View
Model FE-798A



Figure 9. Primary Distribution Unit (PDU), Rear View
Model FE-798A

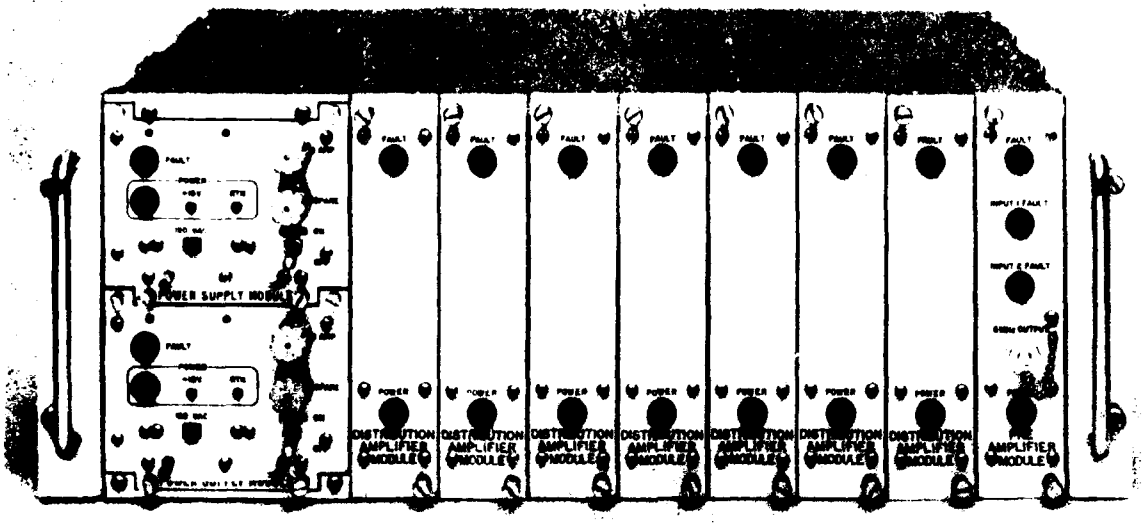


Figure 10. Secondary Distribution Unit (SDU), Front View
Model FE-799A

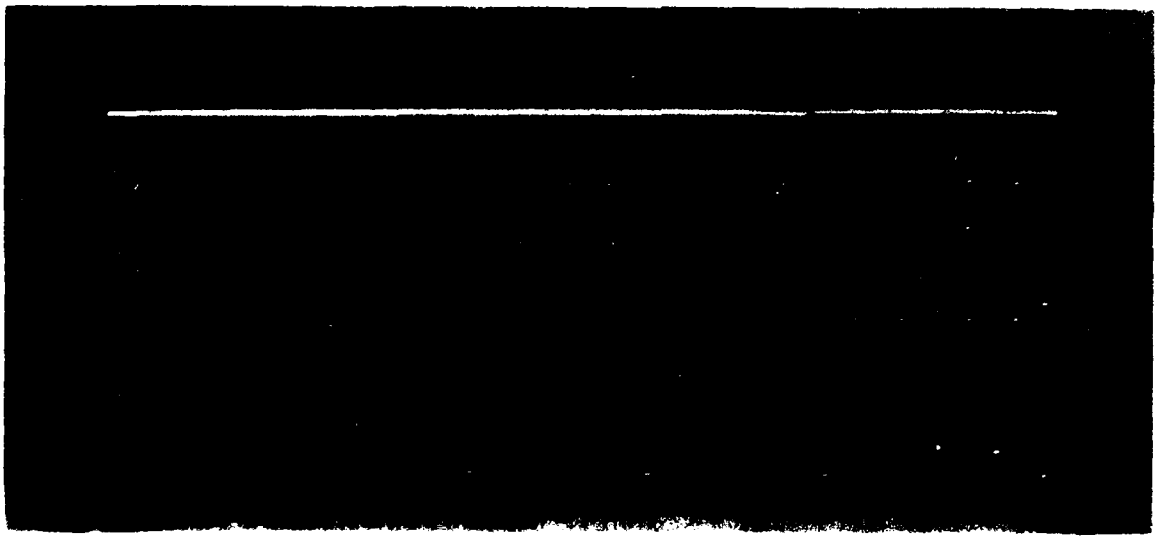


Figure 11. Secondary Distribution Unit (SDU), Rear View
Model FE-799A



Figure 12. Distribution Amplifier Module, Top View

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

LOW NOISE BUFFER AMPLIFIERS AND BUFFERED PHASE COMPARATORS FOR PRECISE TIME AND FREQUENCY MEASUREMENT AND DISTRIBUTION

Richard A. Eichinger

Bendix Field Engineering Corporation

ABSTRACT

Extremely low noise high performance wide band buffer amplifiers and buffered phase comparators have been developed for the NASA Goddard Space Flight Center Atomic Hydrogen Standards Program. These buffer amplifiers are designed to distribute reference frequencies from 30 KHz to 45 MHz from a hydrogen maser without degrading the hydrogen maser's performance. The buffered phase comparators are designed to intercompare the phase of state of the art hydrogen masers without adding any significant measurement system noise. These devices have a 27 femtosecond phase stability floor and are stable to better than one picosecond for long periods of time. Their temperature coefficient is on the order of one picosecond per degree C and they have shown virtually no voltage coefficients. When used in distribution amplifiers and phase comparison systems, these devices have greater than 90 db of isolation.

INTRODUCTION

Bendix Field Engineering Corporation (BFEC) has been providing field engineering support to NASA Goddard Space Flight Center's (GSFC) hydrogen maser frequency standards program for the past ten years. As part of this support, BFEC has recently been building and testing ultra high stability RF buffer amplifiers and phase comparators originally designed by GSFC for distributing and measuring the hydrogen maser frequency outputs. This paper will discuss recent tests by BFEC on the performance of these amplifiers and phase comparators.

GENERAL

In order to provide state of the art performance [illegible] these distribution amplifiers must be able to resolve time steps of a picosecond. Figure 1 [illegible] time [illegible] Samarium Cobalt isolation and buffer stages [illegible] shows 110 dB as the [illegible] required at 5 MHz, which is [illegible] carrier frequency. A five [illegible] signal, corresponding to 90 db at 5 MHz, should be reserved for rare, one time occurrences.

DESCRIPTION OF DEVICES

Figure 2 shows a prototype assembly for precise measurement and distribution. It was selected to show the devices' modular packaging. There are three buffer amplifiers shown on the right. They provide better than 60 db back to front isolation. Gain may be adjusted from 0 to 10 db. These amplifiers are the basic building blocks of the four phase comparators on the left. The phase comparators are made up of two buffers and an RF mixer which are shielded from the filters and zero crossing detector providing a TTL beat signal output.

Buffers in a different configuration are shown in figure 3. This distribution amplifier uses an input driver, an 8-way power splitter and eight buffer amplifiers to provide unity gain output signals from a single input. The splitter increases port to port isolation to 90 db. This configuration was built at BFEC for a NASA hydrogen maser. Figure 4 shows a present distribution amplifier design constructed at BFEC for many applications.

TESTS

Buffer amplifiers, buffered phase comparators, and distribution amplifiers have been tested for:

1. Time domain phase stability
2. Temperature, voltage, shock and AC magnetic field sensitivities
3. Phase noise spectrum out to 100 KHz from carrier
4. Port to port and back to front isolation
5. Bandpass frequency response
6. Harmonic and non-harmonic related distortion out to 1 GHz from a 5 MHz carrier.

Time Domain Phase Measurement

Test results were obtained by an EAS dual mixer test technique for time domain phase measurements. Shown in figure 5, identical beat signals are applied to the start and stop inputs of a time interval counter. The drift of the VCXO's will cancel.

The time interval stability is a direct measurement of phase noise. Note that a hydrogen maser frequency standard is not necessary to resolve maser precision. An order of magnitude better performance can be realized if the time interval is kept smaller than the time constant of the filter. The test system is extremely versatile for modular device testing since test devices are either substituted or added. Phase versus temperature, voltage and shock were recorded using a calculator and digital plotter.

In addition, the calculator was used to make stability calculations. Figure 6 shows an equivalent frequency stability due to random phase changes versus averaging time. The data was taken from an earlier paper entitled "A Modular Multiple Use System for Precise Time and Frequency Measurement and Distribution"[1] where the data was corrected for a single mixer. Data for a dual mixer set-up will be roughly a square root of two times this plot. The plot shows short term data roughly .03 picoseconds per second averaging out to 100 seconds where room temperature cycling began to dominate.

Figure 7 shows a typical phase comparator's temperature performance. The top line is a plot of phase in picoseconds versus the bottom line of temperature variation in degrees centigrade. The change in phase per degree C is less than a picosecond. Note the peak to peak noise fluctuations which are approximately .05 picoseconds. Figure 8 is a typical buffer amplifiers temperature performance. Again, the change in phase per degree C is less than a picosecond. The peak to peak phase noise is up to 0.5 picosecond since the time difference between beats in the dual mixer was not sufficiently short.

In the same test set-up, the buffer was subject to a 5 to 10 G shock. Phase jumps of 10 to 20 picoseconds were recorded. At first, microphonic capacitors were thought to cause the disturbance; however, further testing found that the movement of parts in relation to each other was the cause. The phase jumps were reduced to less than a picosecond when the parts were epoxied in place. BFFC is presently investigating the use of thermally conductive potting material to secure the position of parts relative to one another.

Figure 9 shows the voltage sensitivities of both the buffer and phase comparators. Phase is plotted in picoseconds. The top line shows the change in phase per volt in a typical buffer amplifier. The bottom three lines show changes in phase per volt in each of the three power supplies of the phase comparator. The zero crossing detector of the phase comparator has a one picosecond per volt coefficient. Whereas, buffers in both components have virtually no voltage coefficient since they each contain a separate voltage regulator. Increased channel to channel power supply isolation is achieved in systems of buffer amplifiers.

Spectral Domain Measurements

A spectral phase noise test set-up as shown in figure 10 was used to look at the low frequency spectrum. Identical signals are mixed in quadrature to output small voltage fluctuations proportional to phase. A variable line stretcher was used to measure the phase to voltage conversion ratio. The noise analyzer could be either a lock-in amplifier or low frequency spectrum analyzer where noise power of a given bandwidth is displayed. Any number of test buffers may be added to either signal path.

The phase noise spectrum out to 1 KHz from a 5 MHz carrier is displayed as noise power in a 10 Hz bandwidth. Figure 11 compares the test system shown on the previous slide with two additional buffers. The first peak, *at zero hertz*, is due to the spectrum analyzer's local oscillator. Sixty cycle harmonics are also internally generated, shown in figure 12. Noise levels are roughly 3 db higher for the two additional buffers. Figure 13 shows the phase noise spectrum in a 100 Hz bandwidth out to 100 KHz from a 5 MHz carrier. The 100 Hz bandwidth noise level is higher as expected. Two additonal buffers roughly double the noise power level across the band.

One buffer was introduced to a 10 guass AC magnetic field. The sixty cycle peak rose from an average -138 db to -136 db ± 5 db. The same test was performed on an 8-way distribution amplifier where a 1 db ± 3 db increase was observed.

Figure 12 compares the noise floor of the spectrum analyzer with the noise spectrum of a phase comparator that was rewired to bypass the zero crossing electronics. The band width of the test system was measured at 900 KHz by phase modulating a noise signal on one leg of the test system. A 10 guass AC magnetic field was applied to the phase comparator causing the sixty cycle peak to increase to -80 db. Other mixers tested were equally susceptable to AC magnetic fields and since the buffers are not susceptable, we have initiated procurement of RF mixers with magnetic shielding to reduce this susceptability.

The test set-up used to examine frequency response of buffers and distribution amplifiers is shown in figure 14. It's a basic spectrum analyzer/ tracking generator combination. The input and output signals of the device under test may be reversed to measure back to front or port to port isolation. Figure 15 is a sample group of port to port isolation measurements at 5 MHz. The test results are from a test run on an 8-way distribution amplifier. I had mentioned previously that the splitter increased the port to port isolation. The splitter's design limits the isolation as evidenced by the quadrant grouping of the data. Back to front isolation was measured from all eight ports at greater than 120 db.

Figure 16 shows the frequency response to 100 MHz for both a buffer amplifier without gain and a buffer amplifier with gain. The input level was 0 dbm ± 0.1 db. A typical bandwidth rolls off 3 db at 15 KHz out to 60 MHz. The distribution amplifier's frequency response, shown on figure 17, has a bandpass of 20 KHz to 55 MHz. Figure 17 shows a scan width of 100 MHz and a vertical scale of 10 db per division. Amplifier design is is such that the feedback is trimmed for maximum bandpass response.

Distribution amplifier harmonics are typically greater than 40 db down from the carrier at 5 MHz. Figure 18, typical frequency spectrum, shows the 10 MHz second harmonic at 53 db down from a 13 dbm carrier. There were no spurious signals to a level of 78 db from the carrier out to 1 GHz.

SUMMARY

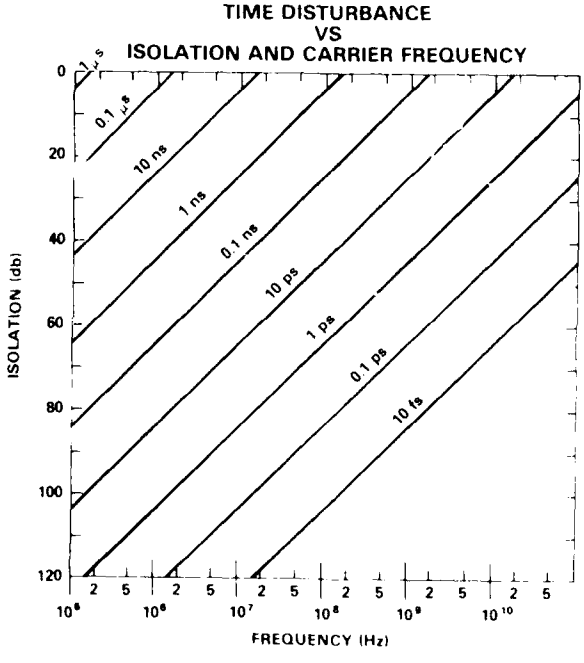
At BFEC, further testing at other carrier frequencies will be persued along with engineering changes to improve existing performance such as:

1. Potting components to reduce shock sensitivity
2. Power supply changes to improve harmonic distortion
3. Magnetic shielding for mixers to reduce AC field susceptability in phase comparators.

Despite the problems just outlined, these devices have already proven extremely useful in improving the specific performance of NASA hydrogen masers. With the engineering changes just described, these devices should prove even more useful for hydrogen maser frequency distribution and measurement and should find application in other precise time and frequency areas.

REFERENCES

1. T. G. Reinhardt, W. A. Adams, G. M. Lee and R. T. Bush, "A Modular Multiple Use System for Precise Time and Frequency Measurement and Distribution", 10th Annual Precise Time and Time Interval Applications and Planning Meeting, Washington, D.C., 1978).

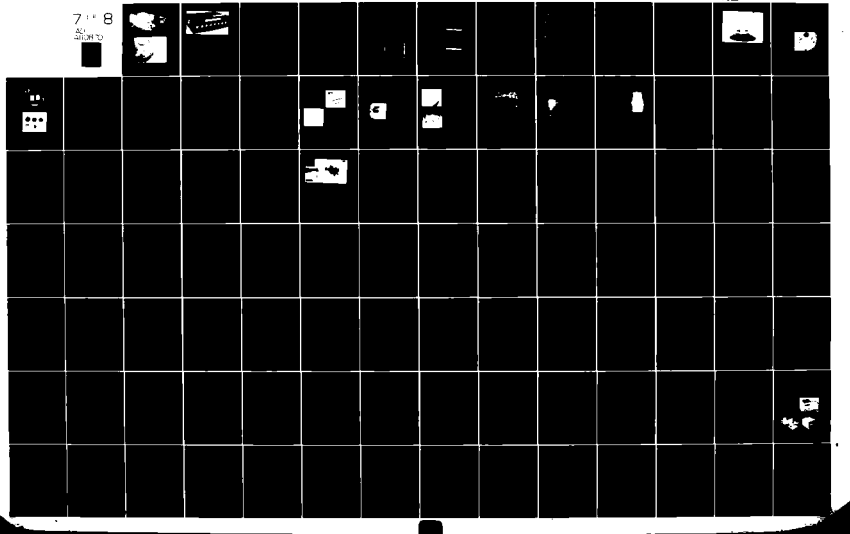


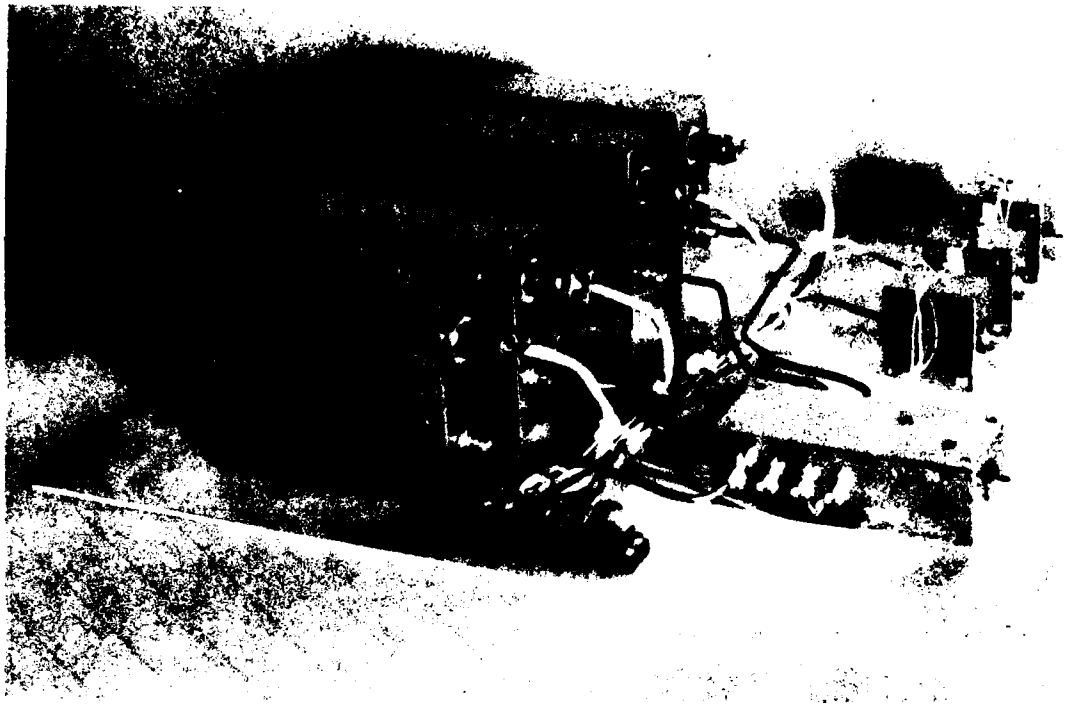
1. Time disturbance versus isolation and carrier frequency

AD-A110 870 ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND AD--ETC F/G 9/5
PROCEEDINGS OF THE 35TH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL, --ETC(U)
1981

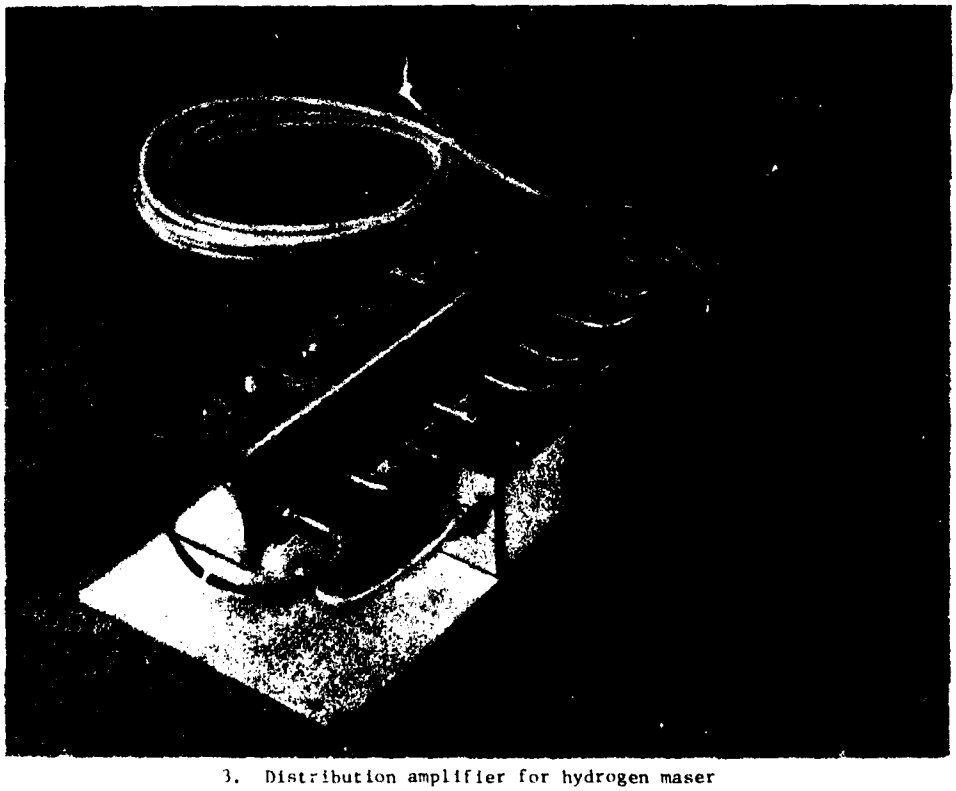
UNCLASSIFIED

NL

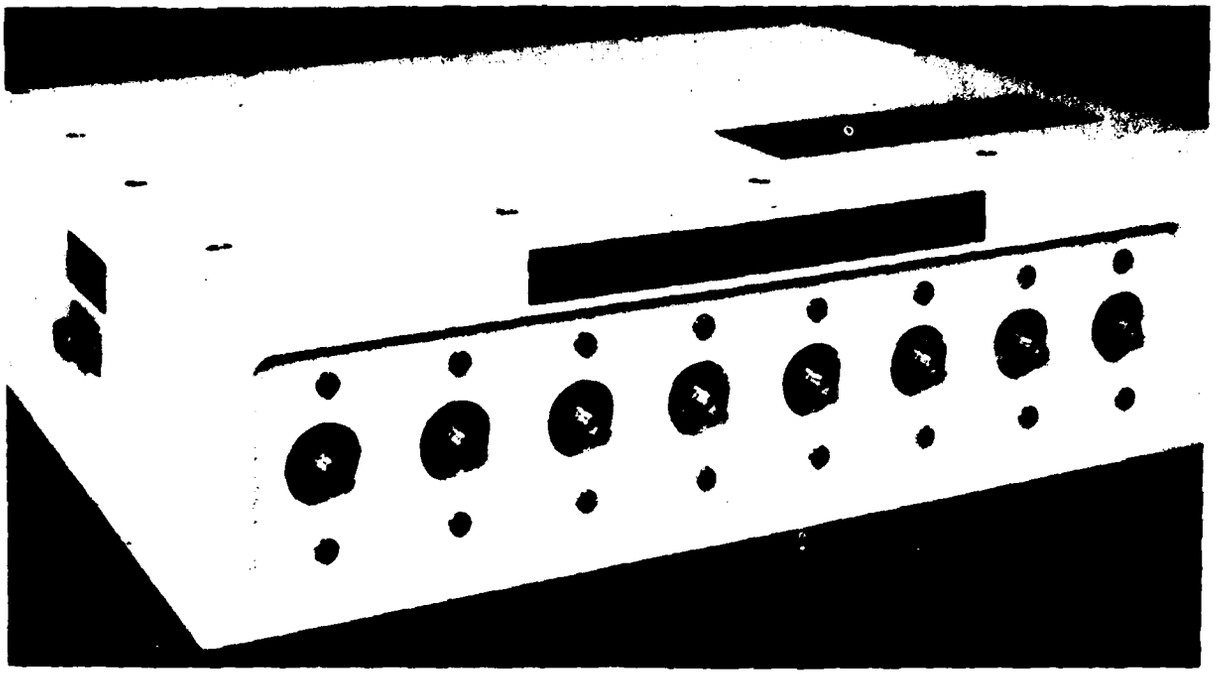




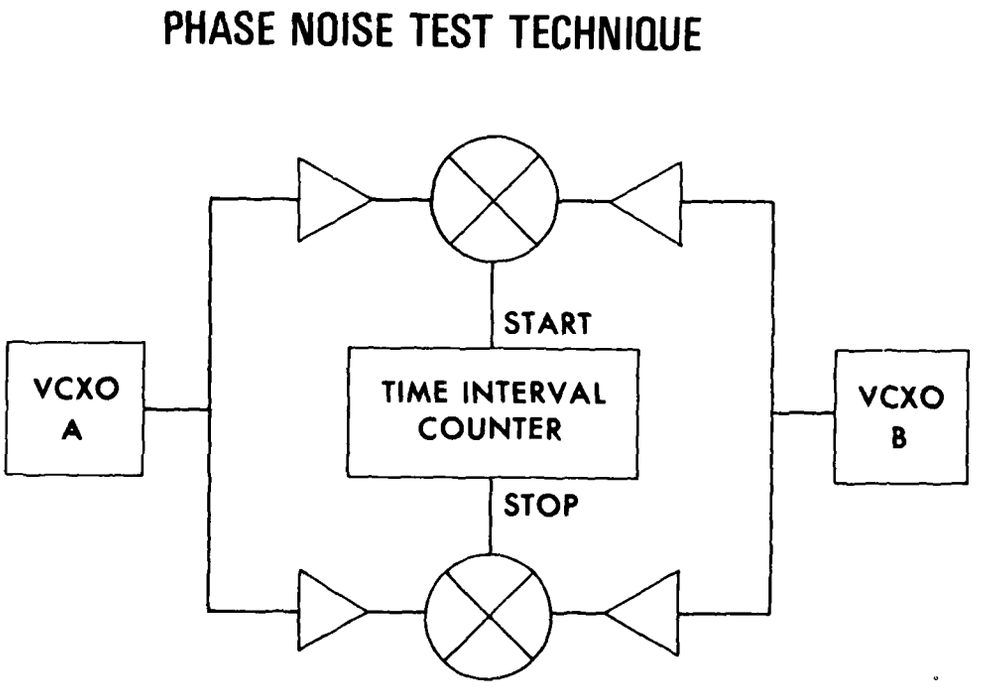
2. Prototype chassis for measurement and distribution



3. Distribution amplifier for hydrogen maser

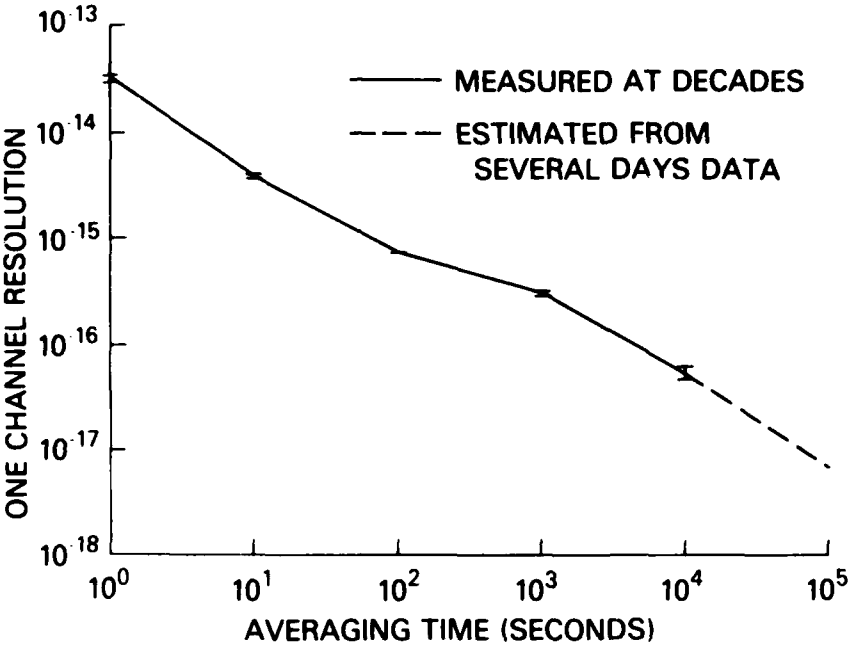


4. Distribution amplifier present design

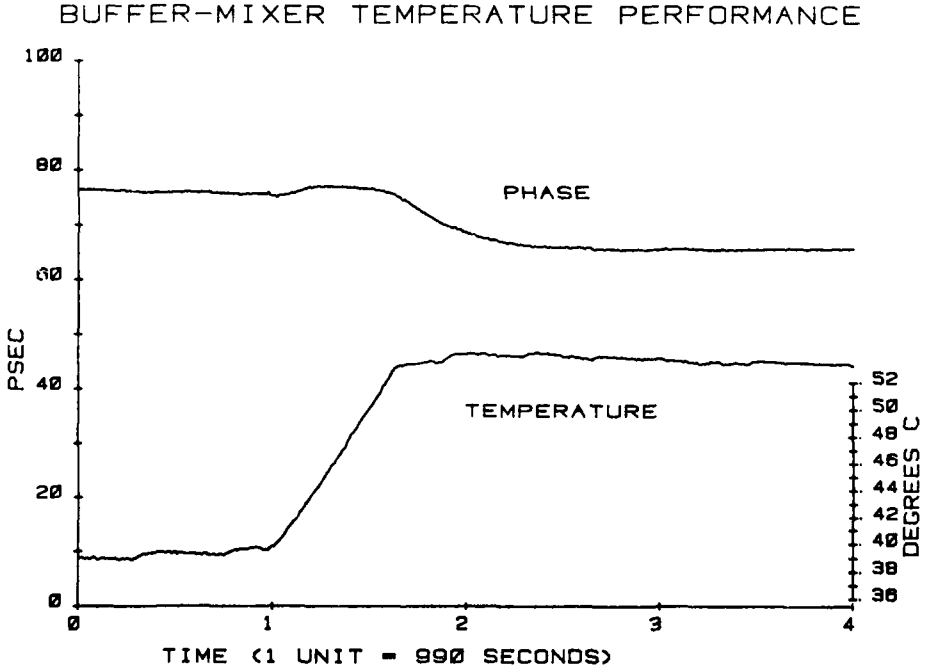


5. Time domain phase measurement

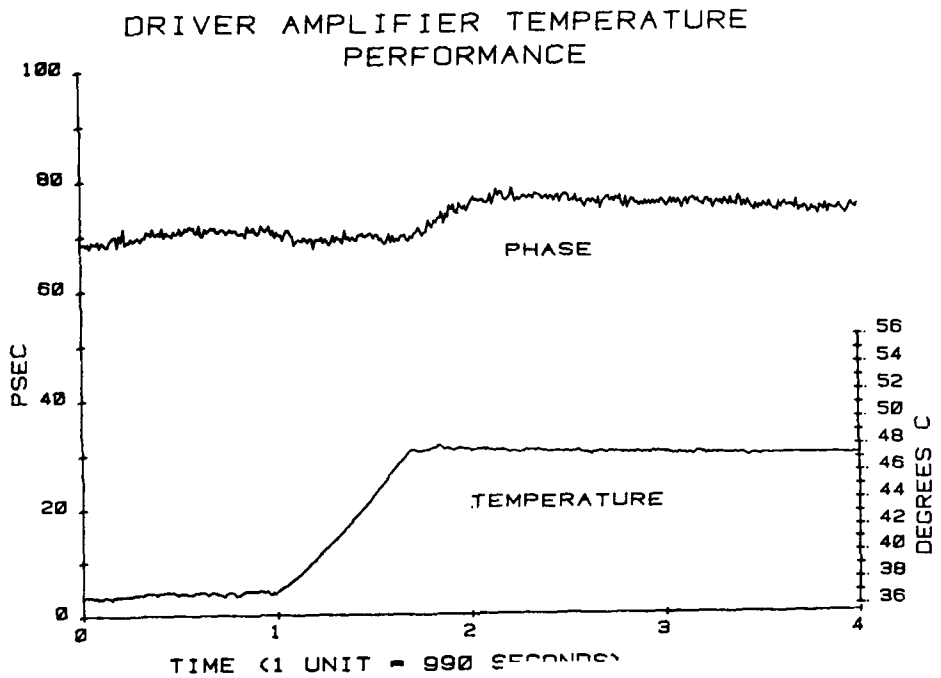
FREQUENCY MEASUREMENT RESOLUTION OF MODULAR SYSTEM



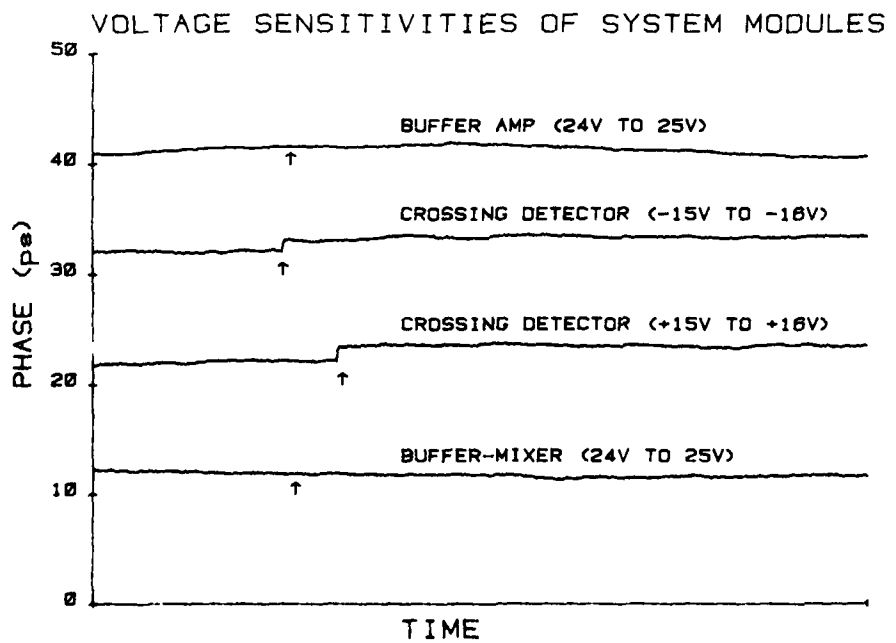
6. Frequency measurement resolution of modular system



7. Phase comparator temperature performance

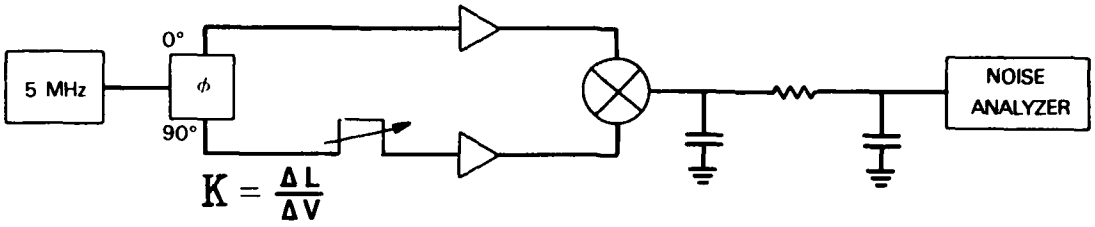


8. Buffer amplifier temperature performance



9. Voltage sensitivities of system modules

SPECTRAL PHASE NOISE MEASUREMENT TECHNIQUE

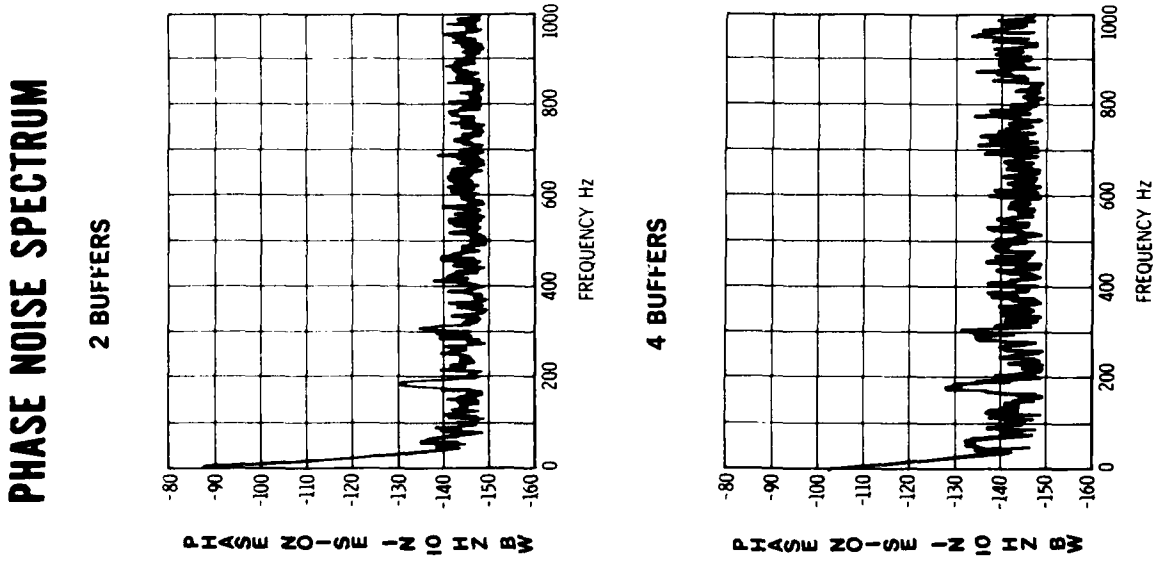


$$\delta\phi \propto \delta V$$

$$\delta\phi = \frac{K\omega}{c}\,\delta V$$

$$\langle \delta\phi^2 \rangle = K^2 \frac{\omega^2}{c^2} \langle \delta V^2 \rangle$$

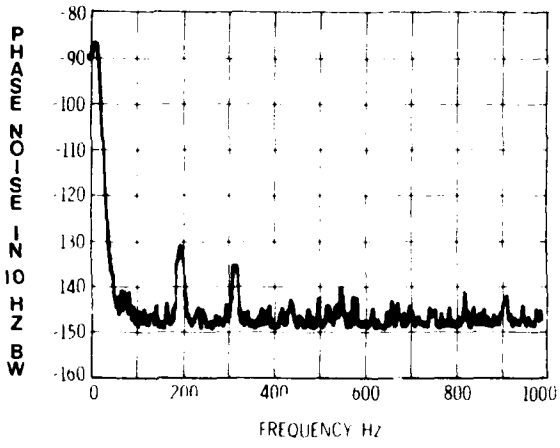
10. Spectral phase noise measurement



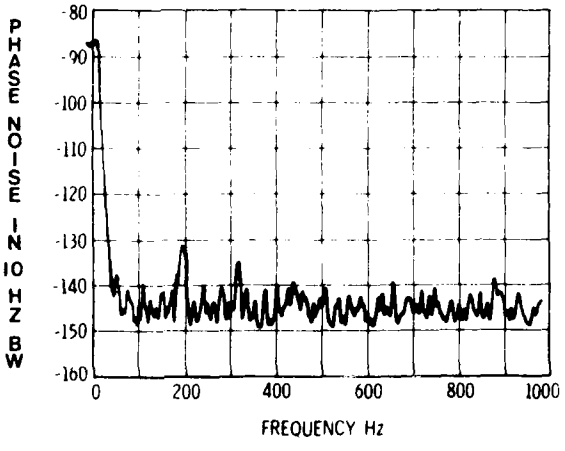
11. Phase noise of buffer amplifier 1 KHz from a 5 MHz carrier

PHASE NOISE SPECTRUM

ANALYZER FLOOR



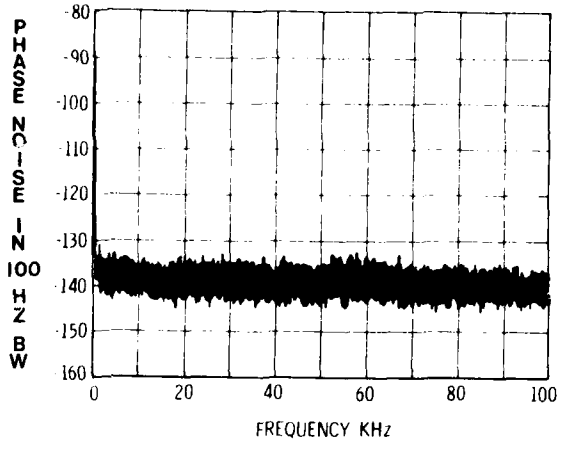
PHASE COMPARATOR



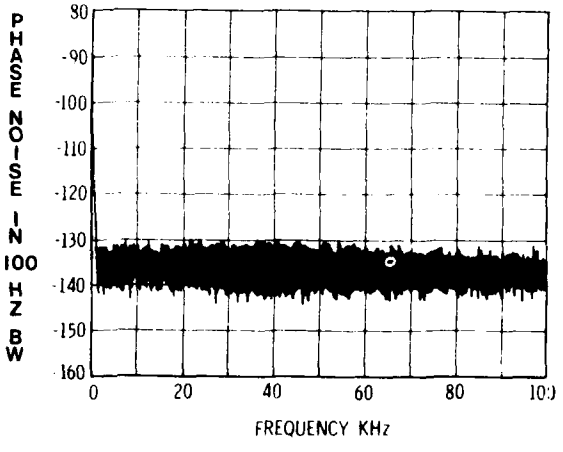
12. Phase noise of phase comparator 1 KHz from a 5 MHz carrier

PHASE NOISE SPECTRUM

2 BUFFERS

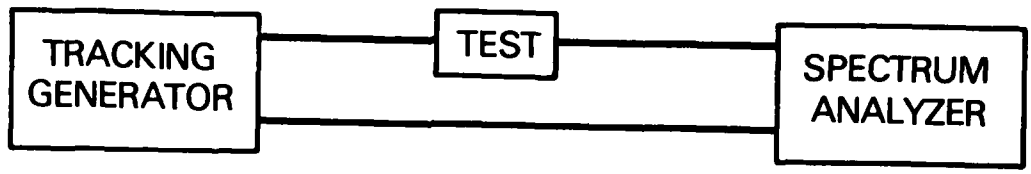


4 BUFFERS



13. Phase noise 100 KHz from a 5 MHz carrier

FREQUENCY RESPONSE MEASUREMENT TECHNIQUE



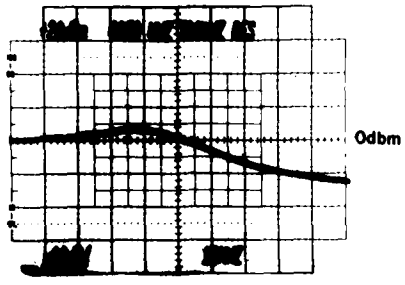
14. Frequency response measurement

Relative Port to Port Isolation (db)

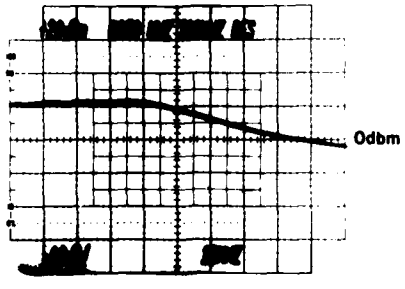
input	output							
	1	2	3	4	5	6	7	8
1	X	98	92	91	109	108	108	108
2	96	X	89	90	106	108	109	108
3	95	95	X	104	108	108	108	110
4	91	91	97	X	108	108	109	108
5	108	108	108	108	X	98	91	92
6	108	108	108	106	98	X	92	92
7	108	108	108	107	91	91	X	98
8	107	108	109	108	93	93	99	X

15. Distribution amplifier port to port isolation at 5 MHz

BUFFER DRIVER

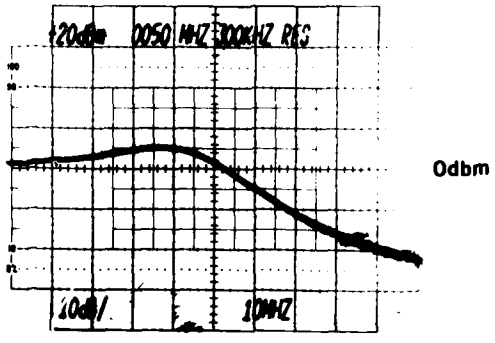


DRIVER AMPLIFIER



16. Buffer amplifier frequency response with and without gain

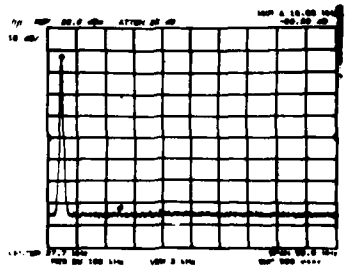
DISTRIBUTION AMPLIFIER



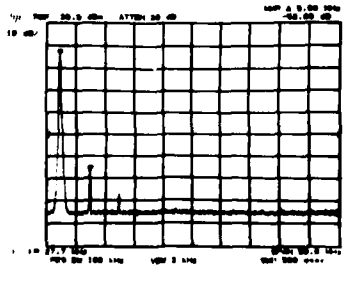
17. Distribution amplifier frequency response

DISTRIBUTION AMPLIFIER HARMONICS

INPUT



OUTPUT



18. Distribution amplifier harmonics

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

THE FUTURE OF THE QUARTZ CRYSTAL INDUSTRY - WORLDWIDE

Panel Discussion Remarks

Arthur Ballato

Electronics Technology & Devices Laboratory
USAERADCOM, Fort Monmouth, New Jersey 07703

My remarks are based on the technological aspects of our subject, and much of the material presented here is taken from the paper of Reference 1. Before attempting any extrapolation to future performance (as seen in a crystal ball), it is instructive, like Janus, to glance backward in time. To this end, Fig. 1 shows worldwide progress in timekeeping for the past millenium.[2]

In the shorter term, a brief look at progression in the technology of crystals is given in Fig. 2 from 1945 onward. This scale is about right for displaying those areas that have the greatest impact at present on our industry - and that might reasonably be expected to influence the future manufacture of units. In the sequel, some of these topics will be touched upon briefly.

Crystal cuts of quartz are displayed in Fig. 3. At present, much attention is being given to the development of doubly rotated thickness mode cuts, specifically the SC cut. As the problem of rapid and cheap X-raying is solved, as will be within the next several years, the SC cut will take over many of the applications of AT cuts; the circuitry problems associated with the SC cut appear already to be overcome. Bulk wave resonators of the grooved and ring-supported variety, reported on elsewhere in these proceedings, can be fashioned with any crystal cut. Ultraminiature strip and GT-cut resonators seem also to have a bright future ahead of them. For frequencies in the upper VHF and UHF, surface wave devices are very attractive. Fast-frequency-hopping sources, and signal processing systems will continue to grow in applications, and SAW devices with them.

Figure 4 traces the history of the processing ambient over the past forty years. The story is a simple one: increasingly high vacua to attain better and better aging rates. The time evolution of vibrator electrode systems is given for the same time period in Figure 5. It would appear that the circle has come full-turn; the difference, of course, is the level of cleanliness, which has increased significantly over the years. The advantages of gold over aluminum for bulk wave resonators are seen in Fig. 6. Deposition of gold requires clean quartz surfaces; to achieve the required level of cleanliness means attention to details. Some of the technology is briefly mentioned in Fig. 2 under the heading "Processing Ambient". This category requires a great deal of attention by our industry. Methods of film deposition in general are listed in Fig. 7.

We introduce the subject of crystal enclosures with Fig. 8 and Fig. 9, each of which is about twenty years old. For comparison, Fig. 10 shows the HC-18 enclosure, with solder seal, and the new ceramic flat-pack with gold gaskets bonded to alumina by thermocompression; the historical evolution seen in Fig. 8 to Fig. 10 is listed in Fig. 11.

One of the most useful areas of quartz resonator technology is listed in Fig. 2 as "Diagnostics"; it is spelled out in more detail in Fig. 12 and Fig. 13. These tools have already taught us much about the quartz, the electrodes and holder, and about the quartz-electrode interface. Diagnostics will play an even greater role in the future.

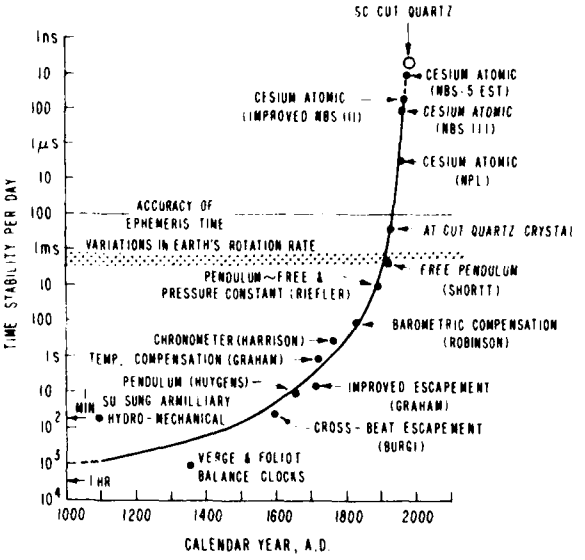
Although not included in the topics of Fig.2, we show, in Fig. 14, some of the developments in temperature control and compensation. A combination of better ovens, use of SC cuts, and, particularly, use of microprocessors, will produce greatly reduced temperature sensitivity. The circuitry problem itself, however, still requires much work. By this is meant designing circuitry around the quartz vibrator in such a manner that the potential of this remarkable component is realized fully. This is far from the case today, and this area represents fertile territory for future work.

In Fig.15 we have sketched frequency stability figures versus time since the appearance of the quartz oscillator. The stability values are for ovenized units sampled at the optimum rate. The lines have been extrapolated to the year 2000. The same presentation is made in Fig. 16, which represents the past and projected performance of commercial oscillators operated over the full environmental ranges of temperature, etc.

Figure 17 shows the aging performance of bulk wave quartz oscillators as function of time. With the use of the technology factors outlined in Fig.2

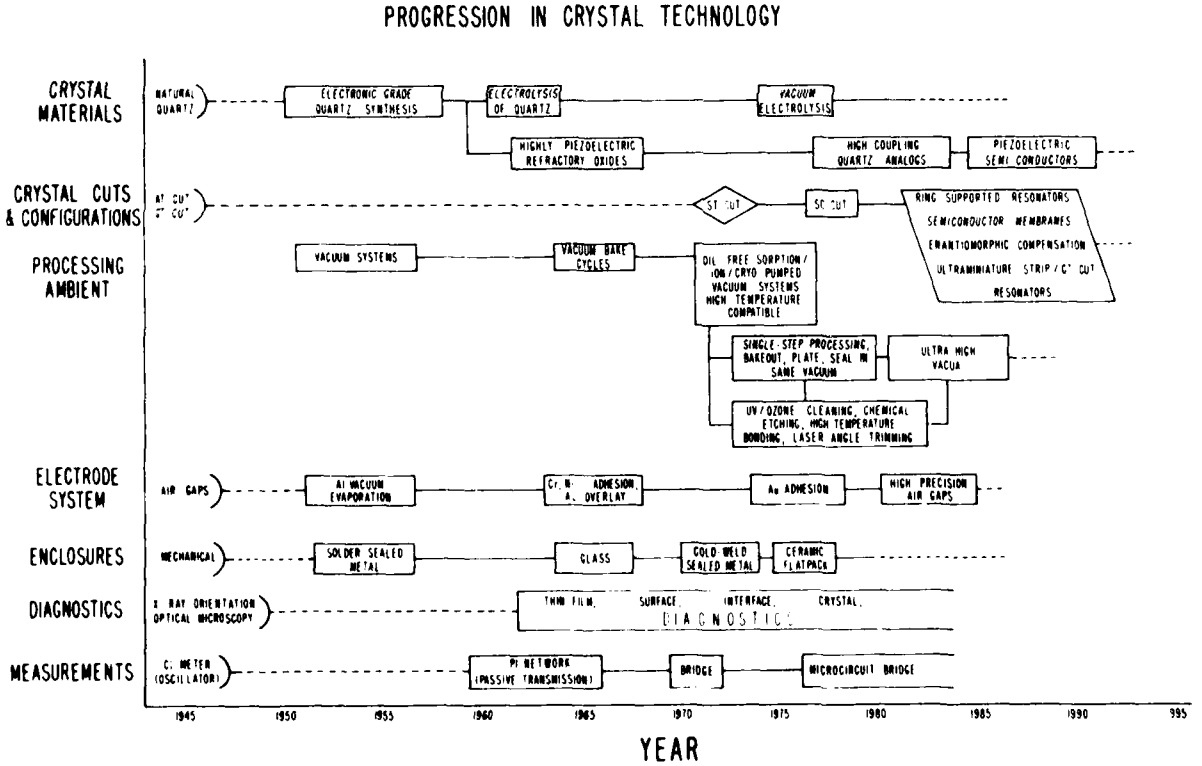
we can look forward to reductions of the aging rate to the 10^{-12} range for small-yield, high precision units by year 2000, and high volume production units in the 10^{-9} range by then. Application of bulk wave technology to the production of SAW devices is an attractive future area of effort.[1]

One remark can be made for certain: the best is yet to come!



PROGRESS IN TIMEKEEPING ACCURACY

1. Progress in timekeeping accuracy.

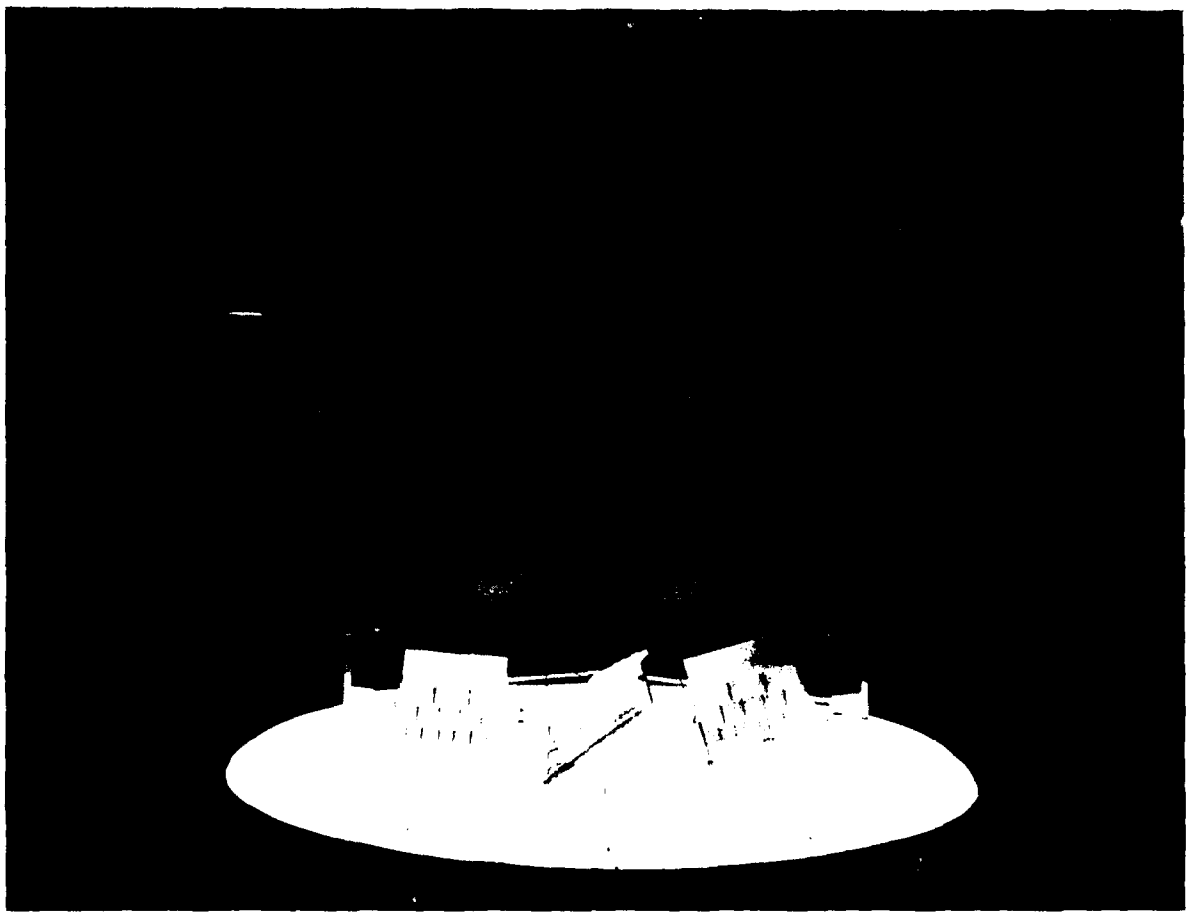


2. Progression in crystal technology.

References

1. T. Lukaszek and A. Ballato, "What SAW Can Learn from BAW: Implications for Future Frequency Control, Selection, and Signal Processing," Proc. IEEE Ultrasonics Symp., November 1980, pp. 173-183.

2. J.A. Barnes, "Basic Concepts of Precise Time and Frequency," in Time and Frequency: Theory and Fundamentals, (B.E. Blair, ed.), NBS Monograph 140, Washington DC 20402, May 1974, Chap. 1, pp. 1-40.



3. Quartz crystal cuts.

AMBIENT IN PRECISION RESONATORS
(1940-1980)

- AIR (ATMOSPHERIC PRESSURE)
- SOFT VACUUM (BACKFILLED WITH DRY N_2 OR He)
- HIGH VACUUM (10^{-6} TORR)
- ULTRA HIGH VACUUM (10^{-9} TORR)

4. Ambient in precision resonators.

EVOLUTION OF QUARTZ VIBRATOR ELECTRODE SYSTEMS
(1940-1980)

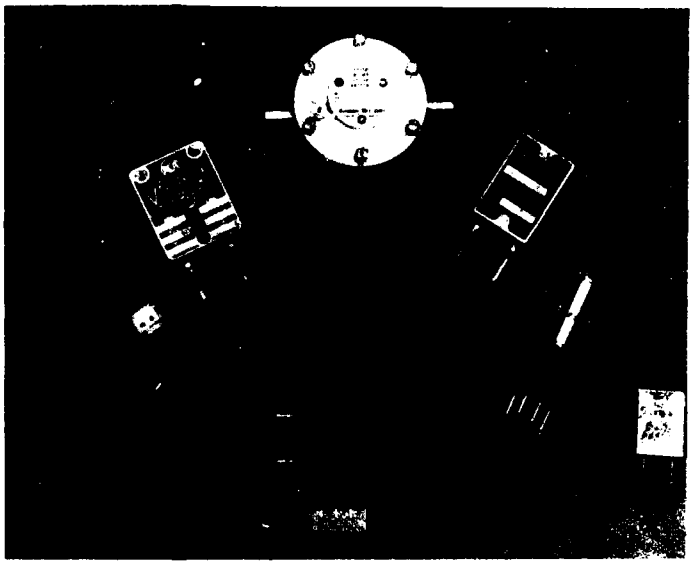
- AIR GAP (ELECTRODES FIXED TO HOLDER OR SPRING-LOADED AGAINST CRYSTAL WITH METAL SPACERS AT EDGE)
- ADHERENT FILM (EVAPORATION OF ALUMINUM, SILVER OR OTHER FILM DIRECTLY ON CRYSTAL SURFACE)
- ADHERENT FILM (EVAPORATION OF ADHESION LAYER OF CHROMIUM, NICKEL, TITANIUM, ETC, PLUS GOLD OVERCOAT)
- ADHERENT FILM (DEPOSITION ON GOLD ON ULTRACLEAN QUARTZ SURFACE)
- AIR GAP (ELECTRODES DEPOSITED ON ENVELOPE SEPARATED FROM VIBRATOR BY QUARTZ SPACERS)

5. Evolution of quartz vibrator electrode systems.

ELECTRODE FILMS ON QUARTZ

GOLD	ALUMINUM
LOW INTRINSIC STRESS	HIGH INTRINSIC STRESS
THERMOELASTIC CONSTANT APPROXIMATES THE AVERAGE FOR QUARTZ (STRESS STAYS LOW WITH TEMPERATURE CHANGE)	THERMOELASTIC CONSTANT DOES NOT APPROXIMATE THE AVERAGE FOR QUARTZ (STRESS LEVEL CHANGES WITH TEMPERATURE)
LOW IN ELECTROMOTIVE SERIES (DOES NOT "RUST")	HIGH IN ELECTROMOTIVE SERIES (SCAVENGES OXYGEN FROM FROM SILICON IN QUARTZ)
GOOD ADHESION TO ATOMICALLY CLEAN QUARTZ SURFACE, POOR ADHESION OTHERWISE	TENACIOUS ADHESION TO QUARTZ
LARGE ACOUSTIC IMPEDANCE MISMATCH TO QUARTZ	VERY GOOD ACOUSTIC IMPEDANCE MATCH (MISMATCH < 5%)

6. Electrode films on quartz.



8. Olde-fashioned quartz crystal enclosures.

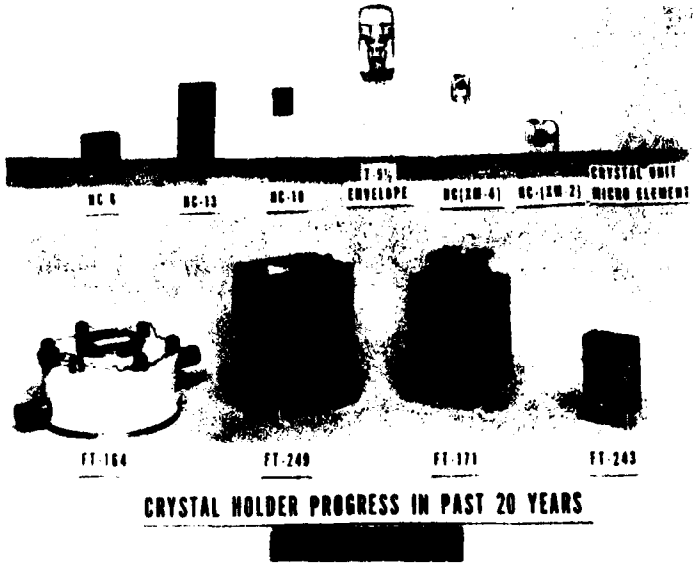
PLASMA PROCESSES

- SPUTTERING
  - DIODE
  - REACTIVE
  - BIAS
  - MAGNETRON
  - ION-BEAM
- PLASMA
  - OXIDATION
  - NITRIDING
  - CARBURIZING
  - ANODIZATION
  - INORGANIC FILM DEPOSITION
  - ORGANIC FILM POLYMERIZATION

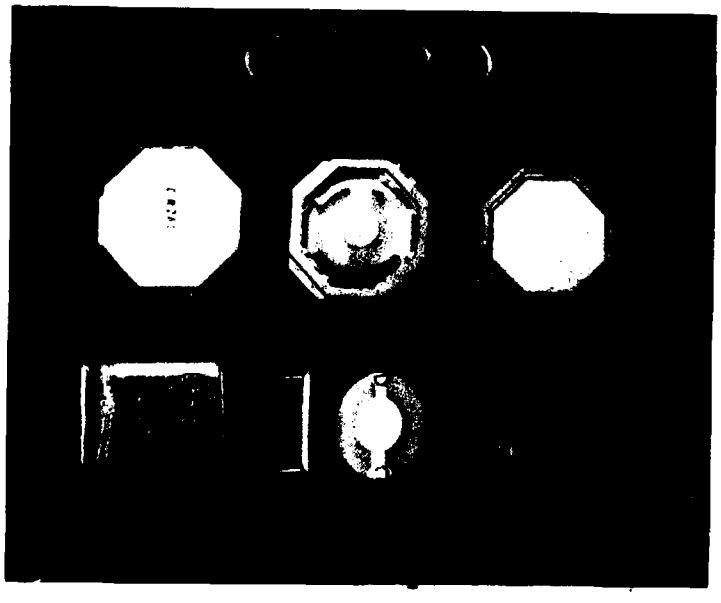
GAS PROCESSES

- CHEMICAL VAPOR DEPOSITION (CVD)
- LASER-ASSISTED CVD
- THERMAL OXIDATION
- GAS-PHASE OXIDATION
- GAS-PHASE PHOTOLYSIS

7. Methods of film deposition.



9. Crystal holder progress from 1940's to 1960's.



10. New ceramic flatpack and old metal enclosure.

EVOLUTION OF QUARTZ VIBRATOR ENCLOSURES (1940-1980)

ENCLOSURE MATERIAL	SEALING MEANS
CERAMIC (PORCELAIN)	SCREWS/GASKETS
PLASTIC	SCREWS/GASKETS
METAL	SOLDER
GLASS	FLAME OR INDUCTION RING
METAL	RESISTANCE OR COLD WELD
CERAMIC (ALUMINA FLATPACK)	THERMOCOMPRESSION BOND/GASKET

11. Evolution of quartz vibrator enclosures.

SMALL ANGLE X-RAY SCATTERING (SAXS)

VERY SMALL ANGLE X-RAY SCATTERING (VSAXS)

IR UV SPECTROSCOPY

OPTICAL MICROSCOPY (OM)

X-RAY DIFFRACTION

X-RAY DIFFRACTION TOPOGRAPHY

REFLECTION INTERFERENCE MICROSCOPY

GLANCING INCIDENCE X-RAY DIFFRACTION

RUTHERFORD BACKSCATTERING; ION BACKSCATTERING

SECONDARY-ION MASS SPECTROSCOPY (SIMS)

ION SCATTERING SPECTROSCOPY (ISS)

ION CHANNELING

13. Diagnostics using photon and ion processes.

DIAGNOSTICS

AUGER ELECTRON SPECTROSCOPY (AES); SCANNING AES

ELECTRON MICROSCOPY (EM); SCANNING EM (SEM)

ELECTRON SPIN RESONANCE (ESR); ELECTRON PARAMAGNETIC RESONANCE (EPR)

ELECTRON SPECTROSCOPY FOR CHEMICAL ANALYSIS (ESCA)

TRANSMISSION ELECTRON MICROSCOPY (TEM); SCANNING TEM

HIGH ENERGY ELECTRON DIFFRACTION (HEED)

LOW ENERGY ELECTRON DIFFRACTION (LEED)

REFLECTION HIGH ENERGY ELECTRON DIFFRACTION (RHEED)

ELECTRON MICROPROBE; ELECTPON PROBE MICROANALYSIS

PHOTOELECTRON SPECTROSCOPY - UV X-RAY

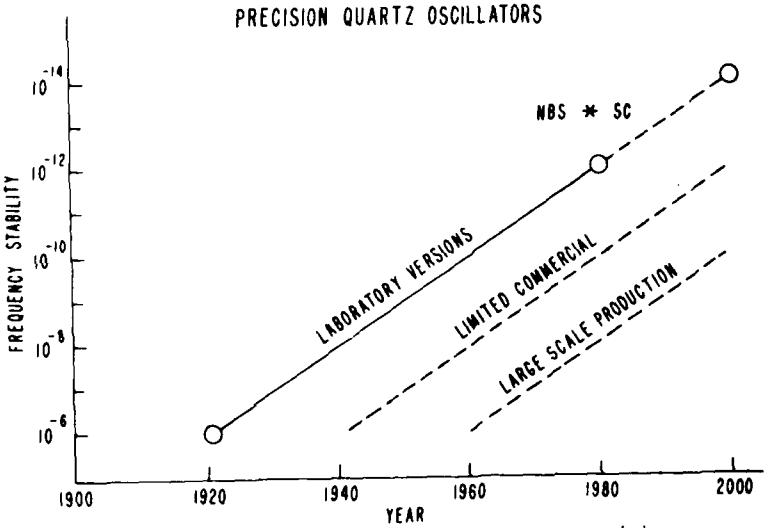
12. Diagnostics using electron processes.

TEMPERATURE CONTROL AND COMPENSATION

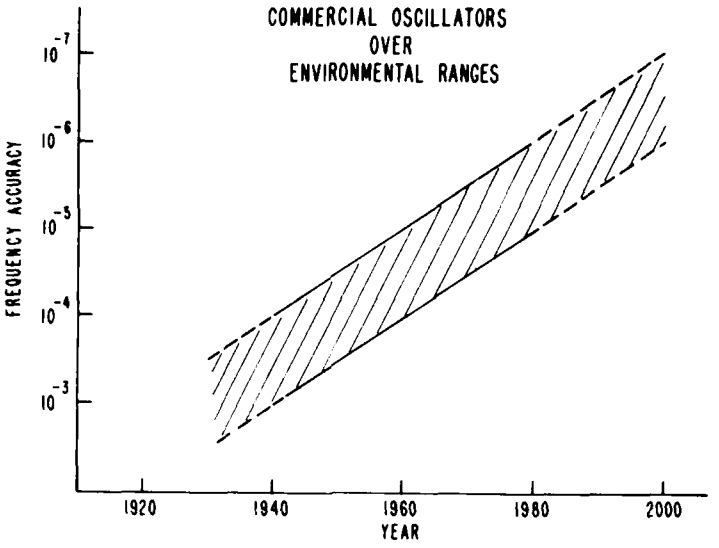
- OVENS (MULTI-STAGE, PROPORTIONAL OVENS, MILLI-KELVIN STABILITY)
- CUT OF CRYSTAL (AT, 1934, SC, 1974, TRANSIENT COMPENSATED)
- CIRCUITRY
  - THERMISTORS SHUNTING OSCILLATOR REACTANCES, C 1955
  - TCXO - (TEMPERATURE COMPENSATED CRYSTAL OSCILLATOR) VARACTOR DRIVEN BY TEMPERATURE SENSING NETWORK, C 1958
  - DIGITAL TCXO - HARD LOGIC AND ROM PLUS D/A CONVERTER TO DRIVE VARACTOR, C 1970
  - DEDICATED MICROPROCESSOR TCXO - IC CHIPS REPLACE HARD LOGIC D/A CONVERTER DRIVES VARACTOR, C 1975

14. Temperature control and compensation.

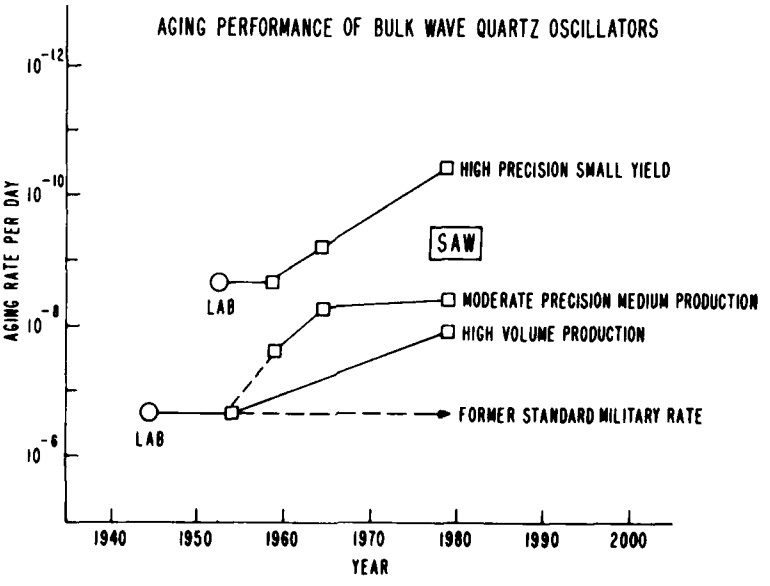
PRECISION QUARTZ OSCILLATORS



15. Frequency stability versus year for precision oscillators.



16. Frequency stability versus year for commercial oscillators over environmental ranges.



17. Aging performance of bulk wave quartz oscillators.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

THE FUTURE OF THE CRYSTAL INDUSTRY - WORLD WIDE

Juergen H. Staudte

Staudte Engineering
Anaheim, California

Introduction

Experience has taught that detailed predictions for the future are not going to be accurate. One thing is for sure, everything will change, but details of how, when, and where cannot be predicted. Nevertheless, businesses must plan ahead in order to prosper or even to survive; therefore, assumptions must be made. One can adjust to the details of how, when, and where as long as the company leadership predicts the general trends and gets the company emotionally, organizationally, technologically, and financially prepared for the future. The following writing will discuss such trends from an overall point of view, giving reasons for the trends, citing numbers to make the point and demonstrating the possible magnitudes of the future trends.

Market Overview

The quartz products industry is presently so diverse and fragmented that any real market data from the past are incomplete and inaccurate. Looking at some past attempts to predict details of the future of the industry were also inaccurate. By talking to several industry leaders, the consensus was that the U.S. sales for quartz products is at least $200 Mill. per year but could be as large as $400 Mill. per year if captive markets are included. The market of U.S., Japan and Western Europe seems to be at least $500 Mill. per year, but can be as large as $1 Bill. per year.

The general prediction of growth in the Electronic Equipment sector, of which Quartz Products have a small part of the value, is a 15% growth per year for at least the next 10 years. There is no reason to believe that the quartz products will grow less than the general electronics growth rate, but there is reason to believe that it will grow faster at a rate of 23% per year. Fig. 1 shows the implication of such a growth. If we assume the pessimistic view point of $500 Mill. present volume and 15% growth, the industry will 4 fold to the $2 Bill. per year rate in the next 10 years. If we take the optimistic view point of presently having $1 Bill. world wide consumption and 23% per year growth the industry will be at $8 Bill. in 10 years. Either way is not bad. This growth definitely will change the operating philosophy and structure of the industry.

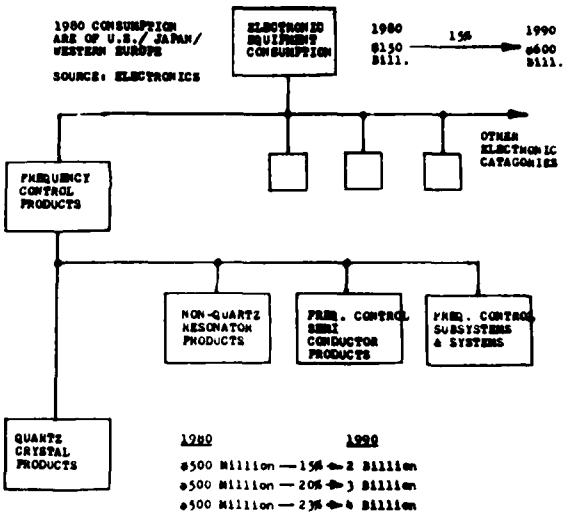


Fig. 1: Quartz crystal products growth over the next ten years and its relation to the electronic equipment market.

Fig. 1 also shows that the quartz products must compete with other ways of frequency controlling. As quartz units become smaller, more rugged and less costly, the non-quartz resonator products such as

the ceramic resonators, and metal reeds and forks will be replaced by quartz devices. The advance and more extensive use of semi-conductors will in some areas reduce the numbers and types of quartz crystals used, but at the same time, will enable new quartz products to be generated such as DCXO and quartz products which perform entire functions (referred to as Functional Quartz Products).

The Quartz Product Market

Fig. 2 shows the possible future catagorization of the quartz products as the industry markets and sells those products. Each catagory is explained in the following paragraphs.

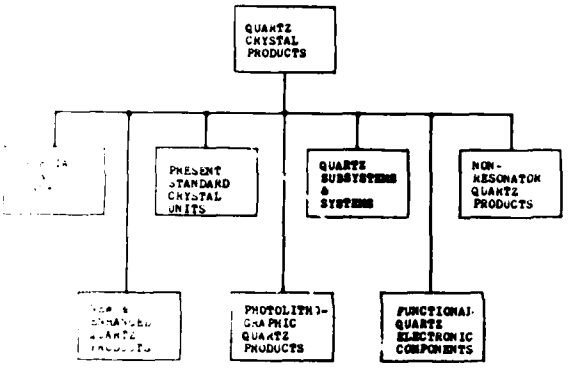


Fig. 2: Catagorization of the future quartz products market.

Quartz Material And Blanks

The trend of many things to come in the quartz crystal industry can be observed by what the semi-conductor industry experienced as it grew. Initially, the large semi-conductor companies did all crystal growing, cutting, and polishing themselves, but speciallized companies emerged generating the raw material, growing the crystals, cutting, and polishing the wafers and selling them to the semi-conductor companies. The same trend is happening in the quartz industry and it will continue. The only exception in the comparison is that the parameters such as the thickness, dimension, angle of the cut, are product parameter determining factors and in some instances may stay under the control of the crystal manufacturers. Many of the new products and manufacturing techniques depend on much better quartz crystal material than what is presently available. Since better material is currently required, those companies that succeed in producing better material will dominate the market. Significant progress in achieving substantially better devices and manufacturing techniques can only be attained through the development of better materials than presently available.

Standard Crystal Products

The present standard products will, in the future, decline in numbers. Relative to the future's total market, standard product's market share will be very small. New products will succeed over the standard products because they will be smaller, more rugged environmentally, better performing and less costly. The quartz contents of newer products will be reduced; glass, ceramic, or quartz crystal material enclosures will become the standard.

Photolithographic Quartz Products

The reason for putting the photolithographic quartz product in a catagory by itself is that it solves many problems which made quartz products undesirable in comparison to other electronic components. Crystals always were considered too large, too sensitive environmentally, too expensive, and too difficult to use relative to other components. As the photolithographic technique can maintain many orders of magnitude better accuracies than conventional techniques, etch any shape, process very large batches, produce very small devices, the handicaps which crystals had in the past are eliminated stimulating wide usage. The crystals always could achieve accuracies and reliability that no other low cost devices can achieve, but soon the cost is low enough that the quartz oscillators are less expensive than LC oscillators, compensated RC circuits, ceramics and metal resonators, and other mechanical systems. Not only can the quartz devices replace many present designs, but, they can enable the designer to change some concepts on how to achieve his goals by using the new quartz products. Fig. 3 shows the catagories of photolithographically produced quartz devices. Presently, only the low frequency devices are mass produced by the photolithographic method in Japan, U.S., and Switzerland. Fig. 4 shows a wafer of tuning forks produced photolithographically. It must be noted that this method can produce devices from 8 KHZ - 100 MHz. The spectrum of devices is single resonators of various types, mechanical coupled filters, monolithic filters, sensors, and SAW devices; however, the actual details of most photolithographical devices of the future must yet be invented.

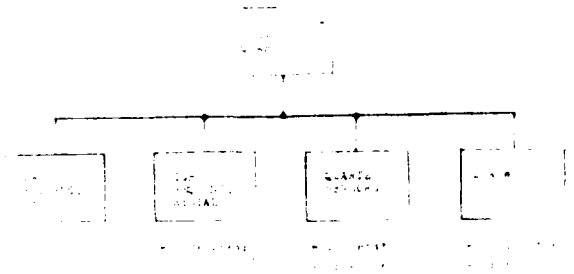


Fig. 3

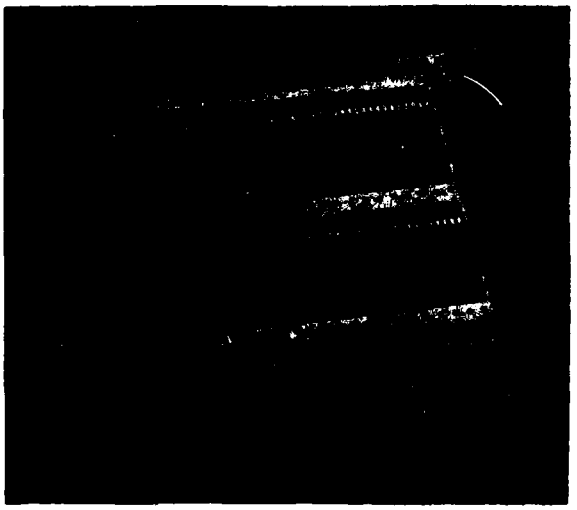


Fig. 4: A .8 inch X .8 inch crystal wafer containing 75 finished, tested, and pretuned crystals. In the future wafers of 4 inch X 4 inch will be processed with several thousand crystals per wafer and of a host of designs covering a frequency range of 5 KHz to over 100 MHz.

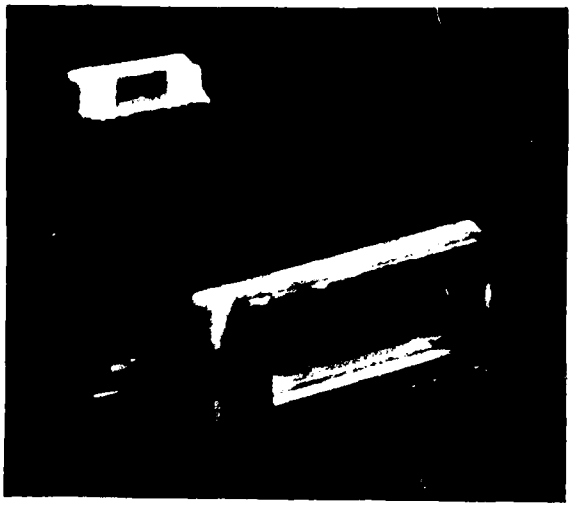


Fig. 5: The two devices shown are NT-tuning fork watch crystals of 1974 vintage produced by the photolithograph process. Presently watch crystals are produced at about the 200 million units per year rate and, in the future, production will further increase and be processed 100% photolithographically.

Fig. 5 shows some already obsolete devices of NT - type tuning forks in a ceramic package but more importantly, it illustrates the capability of the process. The packaging and mounting of the crystal chips produced is still a problem which must be economically solved in a large scale production environment. This again can be done economically with packages made entirely out of single crystal quartz processed by photolithographic methods. The product principle was tried in 1973 and again in 1977 but had to be abandoned for large scale production due to etch channels in cultured quartz which rendered most devices as leakers. The cost of quartz for the entire packaged device can be less than 1¢ if large enough wafers and small enough devices are being fabricated with yield. The processing, testing, and leading can all be performed in large batches and can, therefore, be produced in high labor areas with relatively low capital investment.

New And Enhanced Quartz Products

In the next ten years new and enhanced products will proliferate. The reason for this proliferation is the enourmous growth in the size of the market and the demand the market will impose. The quartz resonator can vibrate in infinite modes, coupled to each other in infinite possibilities, and crystal units can have infinite possibilities in shape, size and angle of cut. Some of those combinations will result in superior devices and generate new products. Packaging, mounting and sealing will evolve rapidly into something less expensive and many times more stable and hermetic than present techniques. A sample of what is to come is shown in Fig. 6.



Fig. 6: The devices shown apply many new designs and manufacturing techniques which are only used by very few manufacturers. The latest construction of a precision crystal in production using an SC cut blank mounted by polyimide organics, enclosed in a ceramic package, sealed by thermo compression bonding, processed in an oxygen plasma or ozone environment and high temperature vacuum environment, using pure gold electrodes on quartz blanks which are deep etch-polished. The small devices are prototypes for a low cost high frequency crystal using identical manufacturing techniques and materials except the crystal is a strip AT photolithographically processed, and the sealing method is by a metal reflow technique.

As new processing knowledge, measuring techniques, and a better understanding of quartz resonator theory is attained or borrowed from other industries, manufacturing techniques will also change radically. To be competitive in this new world, the average manager, engineer, scientist, and manufacturing person has to upgrade their competence and professionalism in order to participate in the future changes or even understand the changes in industry which are presently going on.

Functional Quartz Components Products

The todays' electronic systems engineer really does not want to work with a host of analog components to achieve the function he wishes to perform. He wants to buy the function in a customary package and use it with a customary power supply and layout. For this reason the crystal itself is less popular than the functional electronic building block. Fig. 7 shows the areas where functional building blocks can be achieved today. Most functional building blocks are made today in printed circuits or in hybrid form (see Fig. 8), packaged in odd housings. The future will see more and more of such functional components especially as size and price will shrink rapidly. Fig. 9 shows such products packaged in a dual in line package. Presently, they are available in limited versions but they point the way of things to come. Devices as complex as DCXO's and synthesizers will come in such packages. Since most of such devices are externally frequency programmable, the need for making small quantity production runs of crystals in many different frequencies is eliminated. A set of maybe 30 different programmable functional building blocks may be able to cover 90% of all up to medium precision device needs.

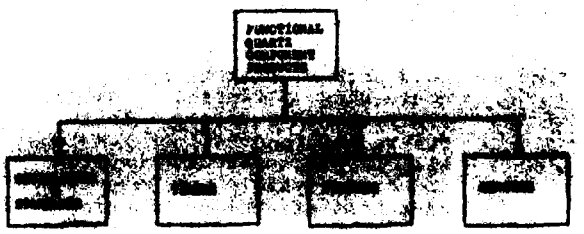


Fig. 7



Fig. 8 This unit performs an entire electronic function having a quartz crystal integrated into the hybrid.



Fig. 9: These dual in line units consist of an IC and a crystal performing an electronic function. The outside units have photolithographically produced crystals of the MT design and the center unit of the tuning fork design. The center unit has a separately processed crystal which later was epoxied into the IC package which provided a cavity for it. Both the IC and crystal can be final tested separately and final laser tuned together as a system. The future will have a host of devices using this principle.

Quartz Sub-Systems and Systems

For the greatest accuracies achievable, a systems approach for making frequency and time references has to be applied. Many products other than the frequency and time references will use quartz devices as a heart of the sub-system or system. These products include DCXO's, accelerameters, pressure gauges, thermometers, synthesizers, etc. and will proliferate. Frequency references with substantial greater accuracy than present devices are needed now in order to implement many of the communication, data transmission and navagational systems already designed. Presently, the most accurate crystal in the laboratory is of electrodeless design, using an SC cut contoured blank with quartz bridges for mounting. Many other innovations, invented or not yet invented, will be applied to the design and fabrication of new devices with lower aging, higher Q, and are environmentally less sensitive. Relatively little theoretical attention has been given to the crystal oscillators external circuitry for achieving its ultimate accuracy. In the future most of the additional accuracies of quartz crystal references will come from the improvement of electronic circuits not devices such as locking together an oscillator designed for short term stability with an oscillator designed for long term stability such as atomic standards manufacturers have done for years. Electronic compensation will also be used for acceleration, vibration, temperature and other environmental conditions. Since the electrodeless design seems to give the ultimate stability, the problem of minor variations in electrode capacitance could be electronically compensated by measuring the variations of the electrode capacitance.

How Will The Industry Change And Operate

As the quartz product industry is similar in technological, management and manufacturing aspects to the semi-conductor industry, studying the management techniques, history and problems of the semi-conductor industry can give us some insight into the future of the crystal industry. The reason the crystal industry has not made as rapid technological advances as the semi-conductor industry is due to its relative small size and lack of growth in the

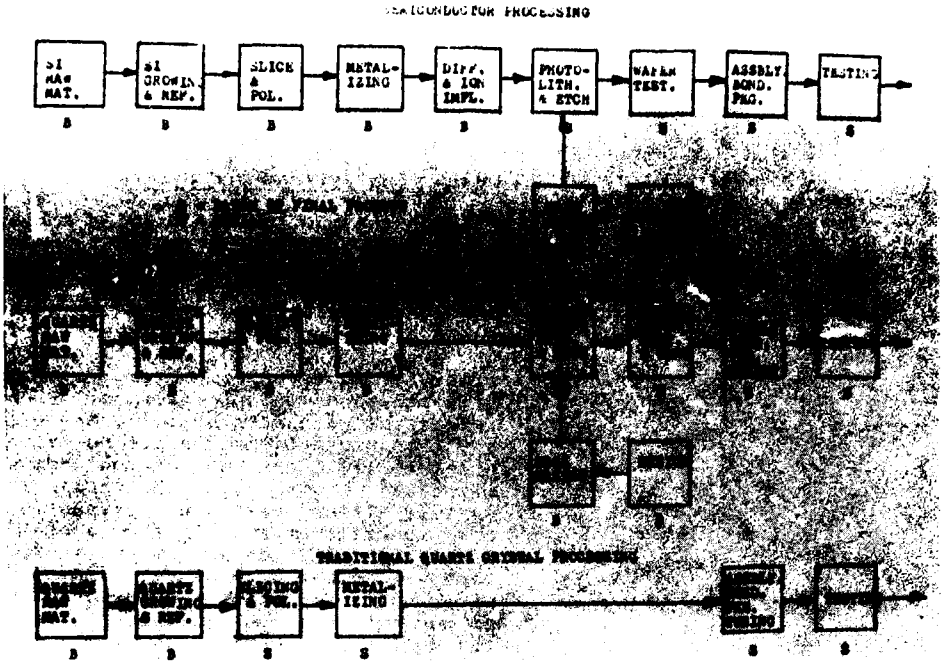


Fig. 10: Some blocks are not arranged in the proper sequence but are shown for better comparison. The crystal industry's similarity to the semi-conductor industry is great especially in the new process technologies. Both industries are material based and share management, process, and manufacturing philosophies. The photolithographic quartz crystal processing was developed by borrowing from the traditional quartz industry's technologies and from the semi-conductor industry's technologies.

past. As the crystal industry has an explosive growth period ahead, technological, management, and manufacturing aspects of the industry will also grow explosively.

Since the crystal industries management techniques, most manufacturing methods, and diagnostic and support technologies are the same as those of the semi-conductor industry, the achievements in all those fields of the semi-conductor industry can be transferred directly making growth in the crystal industry even more explosive in the future. Applying the semi-conductor industries' achievement is just a matter of being able to afford the investments. As the market grows the industry will have companies with significant market share to afford the investments. Fig. 10 shows the similarity of the industries.

The concept of the learning curve which predicts achievements in cost reduction, quality increases, performance increases etc. is a tool being used by the semi-conductor industry. With the learning curve, one can forsee what can be achieved with increased experience which is measured quantitatively by the accumulative volume achieved. This tool can give trends quantitatively, if special conditions are factored into the interpretation of the data. Almost everyone I talked to in the crystal industry does not believe in the learning curve principle or in its validity or applicability to the crystal industry. Nevertheless when historical data of many industries is taken, the accuracy of the principle is substantiated. As the total quartz crystal industries' accumulative volume for the traditional technologies is already very large and the present increases are relatively small, progress will be almost non-existant (except lowering cost by manufacturing in a low labor area). As new technologies, methods, and concepts are emerging those companies possessing the new technologies are in a low accumulative volume position. If those companies are properly

managed, financed, and staffed, they can achieve a steep learning curve. Fig. 11 shows the concept graphically. Accumulative volume is achieved by selling numbers of units. As in any market only a finite number of units can be sold profitably. The company which achieves market share over the competition will achieve the lowest cost, the best performance, and best quality, assuming they can achieve steepness of the learning curve at the same time. Many companies will not achieve as sizeable a market share or steepness of the curve as their competitors and must, therefore, merge, go bankrupt, or restructure and find an insignificant market niche, a precarious position.

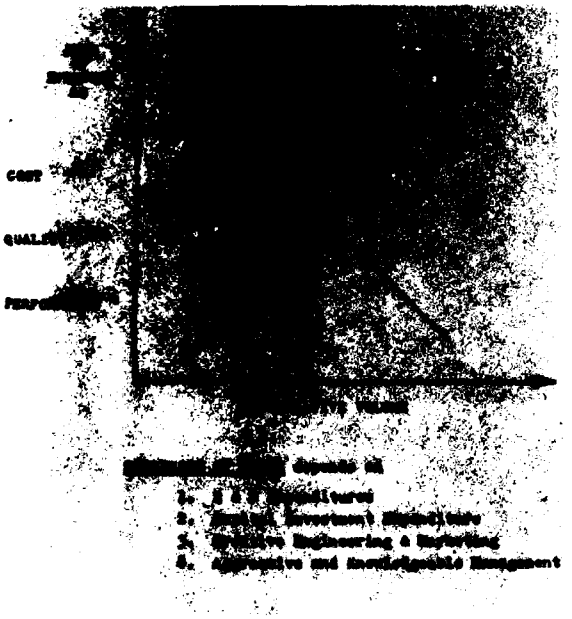


Fig. 11

The next question asked is who is going to be in the game and where does it leave the little guy? The question of course cannot be answered, but looking at the history of other industries, possible answers can be forseen. It is still wide open as to which existing companies or new companies will be cominant and successful in the future because the real growth in volume and technological advances is still to come. Nevertheless these trends already have begun.

Leadership

While good management in the future is an absolute necessity for survival, it is not the only ingredient for success. As companies go through new generations of managements there is a tendency away from the founding enterprises led by an individual or group which understood the product philosophy, its technology and market, toward a management where pure managers lead the enterprise with only shallow knowledge of the industries' technologies and market. Today the different aspects of management have been professionalized to a point where all enterprises can be managed by a team of well trained professional managers. These managers will not succeed in the long term in this highly competitive world, as they cannot lead the enterprise. Fig. 13 shows the elements necessary and process to be followed for future success in the crystal industry. The leadership must have hands on competence in those areas to be able to lead. The technical oriented leadership must, of course, be complemented by managers who are current in such areas as accounting, finance, sales, personnel, QC, general administration, production control, supervision, legal, etc. If on his own the non-technical manager can manage the company, but cannot provide leadership for the company his efforts will ultimately lead to failure.

Large enterprises have a tendency to put managers in charge and let the product specialist be the assistant to the manager. Smaller companies usually are led by a technologist with good market knowledge and, therefore, have an advantage, over larger enterprises despite all other handicaps.

Dedication

In order to succeed in an international competitive environment which the quartz industry is in, a total dedication to the product development, productivity improvement and market expansion must exist. If this dedication in the entire enterprise does not exist, the company will not show staying power. Fig. 12 shows that progress must be made in small steps while volume is accumulating. If an enterprise is not steadily pushing for progress, but reacts only when profitability suffers due to obsolescence or lack of productivity, a crash program to catch up will certainly fail no matter how much money or personnel is thrown at it. This lack of dedication usually exists in every company or department owned by a large enterprise where the crystal products are a sideline and not the mainstay of business of the larger entity. The independent company definitely has the advantage even if it is small in size.

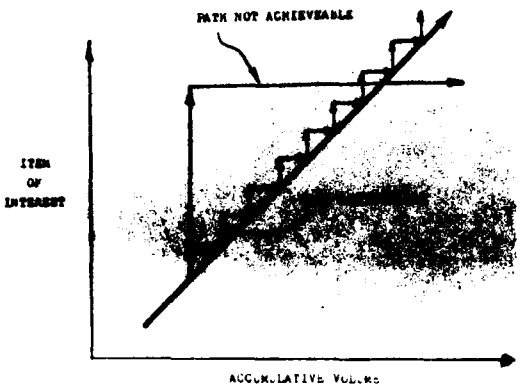


Fig. 12: While the accumulative volume is increasing in the life of an enterprise and if the achieveable path has not been taken by implementing improvement and R & D programs, the lost opportunities cannot be caught up to by a catch up program.

Organization's Structure And Posture

Fig. 13 shows the crystal companies creative process of marketing, development, and engineering. That creative process must be embedded in the enterprises' philosophy. If a company is divided into two main activities, one must divide it into the professional and creative activities. The reason for a companies existence as far as the market (the ultimate provider of money/profit) is concerned are the services or product it delivers. The market does not care about anything else. In order to satisfy the market, the product or service provided must be the best in all aspects and that can only be achieved if the company generates a steady stream of new products of highest quality and at the lowest cost possible. An aggressive posture must be taken in the development of product innovations and increase of productivity. This progress is generated by the creative groups and is implemented by the professional groups in the organization. Such a structure is more natural in a small company but small companies can lack professional power due to the scarcity of resourses.

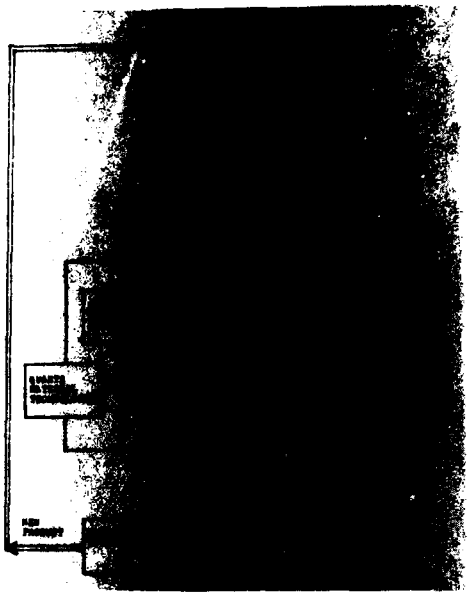


Fig. 13 The crystal industries creative processes are shown. Each company of the industry should either participate in this process or at least be familiar what the industry accomplishes and then apply it into inhouse programs.

Personnel

As the crystal industry grows, it provides many opportunities for motivated and creative new people to enter the field. The work environment and rewards which can be given in a growth industry are best satisfied in a small company. A small company is more likely to concentrate on the creative and professional aspects on hand rather than worry about activities of political, procedural, or bureaucratic natures.

Government

Over the past 15 years the U.S. government actions have virtually shut out the entrepreneural and small company from available investment resources by

its changes in tax laws, and an oppressive amount of regulations. The small enterprise, or individual who in the past have provided virtually all major innovations in America have been made impotent due to the many regulatory restrictions put upon them. Only large enterprises can economically and structurally comply with these restrictions. Presently this trend may be reversing, and will be the biggest determining factor of whether the U.S. enterprises will play a significant part in the world wide crystal business. Investment resources and freedom of action is a condition for successful growth in a competitive world.

Conclusion

The world wide crystal products business will grow explosively over the next ten years to a minimum of $2 Bill. to a maximum of $8 Bill. dollars in sales.

The quality, performance and spectrum of products will significantly increase. The average price will significantly decrease.

Business as usual will lead to the eventual failure of many crystal companies. The number of crystal products companies will decrease significantly by merger and bankruptcies. A few companies will dominate and will have significant market share.

If the U.S. government over the next ten years will provide a favorable business climate by dropping unnecessary regulations and requirements, and by tax reductions, spur investments and bring about resurgence of public offerings, newly formed companies over the next few years and independent small enterprises will grow to dominate the industry. For example the small newly formed semiconductor companies in the 1960's have eventually grown to dominate the industry instead of the existing electronic establishments such as General Electric, Westinghouse, and Bendix. Worldwide the U.S. will have market share over other countries.

If the U.S. government will keep or continue to increase its tax and regulatory burden on the economy, most small companies will be finished over the next ten years in the U.S. and only some large companies will continue some crystal activity. Foreign competition will dominate the crystal industry and the U.S. will only have an insignificant market share which is protected by military needs.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

FUTURE OF THE QUARTZ INDUSTRY - WORLD VIEWS

W. H. Horton

PIEZO TECHNOLOGY INC.
Orlando, Florida 32804

SUMMARY

There is a widespread feeling that the quartz crystal industry, on a global basis, is entering a new phase. First, like many other electronics markets many segments of the market for crystal products are becoming truly international in character. A second aspect, of which the first is a consequence, is the rising importance of high-volume, low-cost crystal products for consumer and commercial applications. The following remarks attempt to present a simplified view of the U.S. industry and its future possibilities.

In order to analyze the future of the quartz crystal industry in the United States it is helpful to understand the present structure of the U.S. industry and markets. Exclusive of companies that manufacture mainly for internal use there are some 60 domestic crystal/crystal filter manufacturers. Of these none has more than 10% of that market and only 3 have more than 5%. There are an additional 30 or so companies that specialize only in oscillators or crystal blanks. In addition there are approximately 7 captive suppliers, representing upwards of one-third of the total U.S. manufacturing capacity.

From these figures it is evident the industry is badly fragmented. Contributing to this, the market is also fragmented. The market divides into broad areas of military, commercial and consumer applications. Within these, a variety of small market segments exist. The large-volume segments are primarily in consumer areas. In these, many of the end items, and an even larger fraction of the crystal products, are manufactured outside of the U.S. At this time there remain only two U.S. companies specializing in the high-volume consumer end-use market.

Until World War II the dominant influence on the quartz crystal industry was the telephone industry. During the next two decades military and telephone requirements shared dominance. Increasingly, since the late '60's these influences have been overtaken by commercial and consumer requirements, with the result that there has been a rapid decline in the technical contributions to the domestic crystal industry from the military laboratories and the telephone industry. In the next decade it appears the consumer area may become dominant.

A recent trend in the industry has been an increasing loss of business in crystals used in high volume in the consumer market to foreign suppliers, with U.S. suppliers taking over the function of distribution and selling. At the same time, most U.S. manufacturers are oriented primarily to supplying the domestic market. The resulting growth in the crystal industrial base outside the U.S. has supported increasing foreign technical innovation. Except for the internal suppliers, individual U.S. manufacturers do not have sufficient size or capital to keep pace with all of this new development effort.

It seems likely that the near term will find drastic readjustments by portions of the industry most affected by the current rapid increase in foreign imports. During these readjustments there will probably be erratic increases in domestic competition as displaced capacity is directed to other market areas.

In the long term, extension of foreign competition to higher technology portions of the U.S. market with adjustment of U.S. industry to a smaller market share will lead to more domestic specialization and fragmentation unless the increasing economic pressure to merge results in significant consolidation. It is likely that the next decade will see greater changes in the crystal industry than in the past 30 years, and only those firms that can adapt to new techniques and markets will have continuing opportunities. Many others will find the future much less attractive.

Worldwide there will be strong growth in the crystal market (now estimated at about $550 Million) at a rate faster than the electronics market as a whole as crystal applications diversify. Much of this growth will be in consumer areas and will be increasingly dominated by foreign suppliers if present trends continue.

In summary there will be excellent worldwide growth in future markets. The nature of these markets is broadening, with increased emphasis on high-volume products which, by and large, will be supplied by the most efficient and aggressive producers.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

"FUTURE OF THE QUARTZ CRYSTAL INDUSTRY WORLD VIEWS"

TOSHIAKI TAKEUCHI

DEPUTY - PRESIDENT OF
NIHON DEMPA KOGYO CO., LTD.
1-21-2, Nishihara Shibuya-ku, Tokyo Japan

Summary

From the management point of view, there are following major factors when we think about the Future of the Industry.

*Future market demands and requirements.

*Future possibilities of technology development.

Finally, reached following conclusions.

*Even smaller size is required.
*Higher reliability is required even for the check crystal units.
*Greater price creates new market demands.
*New type of the crystal units will be realized on production line by out-standing latest technology.
*Broad expansion in the demand will be realized in the future at a result.

Description

First, I wish to enumerate several basic assumptions:

1. When I speak of the "future," I am limiting myself to a discussion of the next ten years or so. Without this definition, we would get too involved in a discussion of a fairytale world of tomorrow, which would not serve our purposes here today.

2. We assume that no major wars will occur in the world over the next ten years to disrupt the present demand structure.

3. We assume that the economy will advance at a predictable and steady rate, with neither a major depression nor a great boom period.

And most importantly, 4. We assume that no material will be created during the next ten years to replace quartz crystals either in performance or cost.

I am neither a scientist nor a fortune-teller. Before speaking of the future, therefore, it would be most appropriate first to take a brief review of the past--examining earlier trends and the problems they involved, as well as taking a look at changes in market needs and in the demand structure.

Since we do not have a great deal of time, let me briefly introduce a basic breakdown of how a review such as this should be carried out.

1. For each application of quartz crystals, it is necessary to have a grasp of at least three factors:

   a. The growth in demand

   b. The level of quality required, in terms of reliability, accuracy, frequency range, physical dimensions, etc.

   c. Trend of cost, both of the quartz crystals and of their final products.

2. The same factors as mentioned above must be considered for competing components such as ceramic crystals.

3. The progress and diffusion of integrated circuits must also be taken into consideration.

4. An accurate evaluation must be made of the striking progress seen in recent years in production technology. And finally, 5. It must also be considered how computers have recently become usable by and accessible to virtually any user at a low price--a vital fact which must not be overlooked for a number of reasons.

After a careful gathering of data on

these five basic points, a proper analysis will lead to the following results:

1. The range of quartz crystal applications is expanding greatly, which in turn indicates a broad expansion in the demand for crystal units. Also, with the progress in ICs and the proliferation of microprocessors, a steady increase can be expected in the demand for products incorporating both ICs and crystal units.

2. The greatest part of the growth in demand is seen in terms of Mass Series items, and the proportion occupied by single-item orders for professional use--such as in telecommunications--grows less and less every year. However, while the proportion of professional-use demand goes down, demand quantity itself continues to rise steadily. Also, in terms of reliability, a far higher level of reliability has been demanded in Mass Series items up till now: but the difference in reliability levels between Mass Series and professional-use items can be expected to disappear.

3. The inroads being made by ceramic crystals into the quartz crystal market are very great, especially in the low-frequency range. As quartz crystal manufacturers, we cannot ignore this threat and must take steps to meet this challenge successfully.

4. As other components proceed to grow ever smaller, quartz crystals too will be required in ever smaller size.

5. It is a true fact that the progress in the computer inductry will expand our own work load. More importantly, though, it will also permit us to make spectacular progress in the technological area. Together with the striking progress in production technologies, I believe that the possibility will arise that new types of crystal units can be created incorporating even more advanced oscillation principles.

6. Until now, progress in general production technologies and steady growth in demand have made investments in mass-production equipment possible. These trends can be expected to continue in the future, leading to ever lower production costs.

From these various points, we can see that the future demands on quartz crystal manufacturers will be a strengthening of development capabilities, execution of heavy facilities investments, acquisition of new production technologies, and provision of an ever higher level of education to employees. Realistically speaking, it would be impossible for all manufacturers to carry out all of these tasks, and if they did, it would merely result in a meaningless competitive race which not only would hinder the prosperity of the inductry but also prevent the user from receiving full benefit. Only meaningful competition can raise the level of technologies and produce full value as needed.

To achieve this meaningful competition, each company should put effort into its own special area, elevating its developmental capabilities and competitive power in that area.

In this way, we will be able to seek out revolutionary new technologies for our quartz crystals, in order to allow the quartz crystal inductry to supply products of higher quality at lower cost, permitting us through our products to serve society on an ever larger scale in the future.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

CURRENT TRENDS AND FUTURE PROJECTIONS IN THE CRYSTAL INDUSTRY WORLDWIDE

Contribution by

Dr. Rudolf Fischer

K V G Germany

Regarding oscillator crystals the CB-boom has built up and faded away due to the frequency synthesis coming up. The range of consumer crystals like those for colour T.V., watches and clocks is still expanding, but has moved to factories especially equipped for high volume mass production and as a last consequence into countries with low manpower cost. The same I think may happen to the microprocessor crystals.

The channel crystal business for FM-communication sets was quite stable in the past years and is now starting become affected by frequency synthesis too. Today the question is not from how many channels up a synthetizer is more economical than a crystal bank. Many modern concepts in Europe are favouring even the one channel set to be equipped with a standard crystal and a synthetizer because of flexibility in frequency setting and to get rid of finite delivery time of the channel crystals. I think the use of exchangeable TCXOs as channel elements will also be only an intermediate solution. High quality TCXOs, low-power ovenized reference sources - both in conjunction with frequency synthesis - will come up for VHF-UHF-communication sets and single-sideband equipment.

Remarkably stable and still increasing in the past was the range of highly professional and military crystals with closely controlled parameters regarding TC-behavior, motional parameters, guaranteed aging performance, and specified shock and vibration resistance. Except for few low frequency crystals the area of the HC-6/U crystal holder is over and has shifted to the HC-18/U. Now a similar shift from HC-18 to the tiny HC-45 will occur together with resistance-welding becoming the standard sealing procedure.

With crystal filters we have a similar evolution: The conventional lattice type multipole-filter in FM-communication sets was replaced completely by miniature monolithic versions mostly built up from dual resonators, the former being left to highly sophisticated filters with special transfer functions or high complexity. Monolithics are coming up for single sideband applications more and more and future concepts are pointing towards higher intermediate frequencies up to 100 MHz where intermodulation performance will be of utmost importance and group delay performance for all filters related to filters in conjunction with data transfer. Of course surface acoustic wave devices will extend the range of broadband-filters up to higher frequencies.

With TCXOs there is still a growing market with a good aspect for the future. Also low-power ovenized and miniaturized frequency standards are becoming of greater interest for the future - all together in conjunction with frequency synthesis.

The main problem of todays crystal industry is the high amount of labour cost especially here in the United States and in Europe.

When regarding the professional and military type crystals and associated filters and oscillators forecasts for the next years are still looking promising. Further impulses might come from the yet not fully evaluated introduction of piezoelectric sensors and may open a new field of application for the products of the quartz crystal industry.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

MONO-ION OSCILLATOR AS POTENTIAL ULTIMATE LASER FREQUENCY STANDARD

Hans Dehmelt

Department of Physics, University of Washington
Seattle, Washington

Abstract

An individual atomic ion localized in the center of a small Paul rf quadrupole trap has potential as an ultimate laser frequency standard because, broadly speaking, the ion may be brought to "A State of Complete Rest in Free Space" by side band cooling. As a consequence, all Doppler shifts vanish. "Free Space" is approximated insofar as the electric trapping field vanishes in the center of the trap and there is no Stark effect. Neither need there be a Zeeman Effect as magnetic fields may be controlled down to the micro-Gauss range. Naturally, there is no transit time broadening. Minute laser powers provided by harmonic generators suffice for saturation of optical transitions as well focused beams may be used. Millionfold atomic amplification of the single-ion fluorescence from a metastable level may bring resolutions of 1 part in 10^{18} within reach.

Introduction

The 1965 experiment[1] of Fortson, Major and Dehmelt in which the 0-0 hyperfine transition in the ground state of $^3He^+$ near 8 GHz was observed with a linewidth of $\sim$ 10 Hz or a resolution of about 1 part in 10^9 marked the advent of high resolution microwave spectroscopy of stored ions proposed earlier by the author.[2] Samples of $\sim 10^7$ ions were used in these experiments. Already then in our 1966 publication we pointed out the potential usefulness of ion storage for laser spectroscopy and laser frequency standards. In related[2,4] studies in 1968 S. Menasian, working in my laboratory, succeeded in detecting Hg^+ ions stored in the $\sim$ 1 meter long linear trapping region of a rf "race track" device.[3] In 1973 Major, now with NASA, reported the record resolution of 2 parts in 10^{14} in an optical pumping experiment[4] on the hyperfine structure transitions near 40 GHz on $\sim 10^5$ stored $^{199}Hg^+$ ions. The pumping light was provided by a ^{202}Hg discharge lamp. Also in 1973 I proposed laser fluorescence spectroscopy on a highly forbidden electronic transition of an electronically refrigerated, individual Tl^+ ion stored in an rf quadrupole trap, "The Tl^+ Mono-Ion Oscillator,"[5] with a projected resolution of 1 part in 10^{14}. In 1975 I added a laser double resonance scheme for effectively amplifying about million-fold the fluorescence intensity used for the detection of the forbidden transition of interest.[6] This scheme also made the study of the similar ions In^+, Ga^+, Al^+ attractive. Incorporation into the Tl^+ experiment of the side band cooling mechanism,[7] searched for by Wineland and Dehmelt in 1974 in experiments on stored electron clouds and found in 1976 by Van Dyck, Ekstrom and Dehmelt in their geonium experiment,[8] was proposed by Wineland and Dehmelt in 1975. Also in 1974 the award of a Humboldt prize enabled me to initiate a Ba^+ mono-ion oscillator[9] experiment at the Universitaet Heidelberg in collaboration with P. Toschek and his group to prepare the ground for the more difficult Tl^+ work. At the same time, I proposed a Sr^+ mono-ion oscillator experiment[10] together with H. Walther, then at the Universitaet Köln.

In 1978 the Heidelberg experiment[11] led to the first isolation and uniform confinement, initially to < 5 μm and eventually to $\leq$ 2000 Å via laser side band cooling to $\leq$ 6 K and eventually[14] to $\sim$ 10 mK as well as continuous visual and photoelectric observation of an individual Ba^+ ion in a small Paul rf quadrupole trap. Simultaneously Wineland and co-workers demonstrated laser cooling to $<$ 40 K of about ten thousand Mg^+ ions[13] confined in a different device, namely a Penning trap incorporating a $\sim$ 1 Tesla magnetic field. Later the potentially very sharp $6^2S_{1/2} - 6^2P_{1/2} - 5^2D_{3/2}$ two-photon transition to the metastable $5^2D_{3/2}$ level was observed by us with a (power-broadened) linewidth of $\sim$ 90 MHz. Applying their technique to an optical cooling and pumping experiment[15] on 300 MHz transitions in 10-100 Mg^+ ions Itano and Wineland improved upon Major's resolution record about 5 fold and also eliminated the 2. order Doppler shift via optical cooling to $<$ 1 K in 1980. Recently Wineland and Itano have also succeded in a Mg^+ mono-ion oscillator experiment[16] and they have established from the residual Doppler effect of the electronic resonance line temperatures in the cyclotron, axial and magnetron motions in their Penning trap of $\sim$ 50 mK, $\sim$ 2.6 K and about minus 0.2 K and a localization of the ion to $\leq$ 15 μm.

Concept of Mono-Ion Oscillator Experiment

The goal of our current preparatory experiment is the observation of copious laser scattering at a very narrow electronic transition at ω_0 on an individual ion stored at the bottom of the potential well formed by a small Paul rf quadrupole trap. The ions of the Group IIIA elements, namely Tl^+, In^+, Ga^+, Al^+, B^+ are excellent candidates for such work. Their metastable lowest $^3P_0$ levels have extraordinary long life times; e.g., the corresponding natural line width of the $3\ ^3P_0$ level of Al^+ is estimated as $\sim$ 10 μHz. However in a mono-ion oscillator such

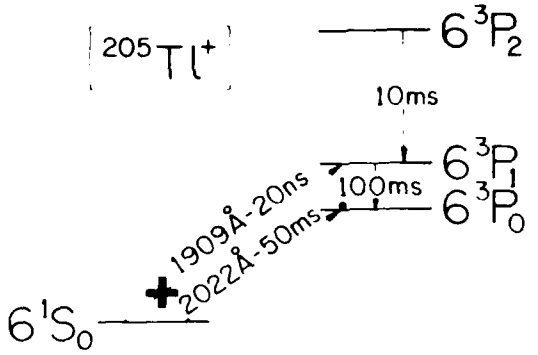


Fig. 1. The [illegible] lowest electronic levels of the Tl^+ ion in the proposed ultrahigh resonance [illegible] transition at 2022 Å of [illegible] the $^3P_0$ level [illegible] natural life[illegible] [illegible] the temporary [illegible] the $^3P_1$ metastable [illegible] [illegible] [illegible] the $^3P_0$ level [illegible] laser radiation at [illegible] Å. [illegible] the detection of about a [illegible] fluorescence is expected [illegible] In^+, Ba^+, Al^+, B^+ [illegible] structure with the width [illegible] [illegible] final [illegible].

[illegible] which can only be explained when a powerful [illegible] mechanism for the corresponding [illegible] scattering is devised. Monitoring of the [illegible] of the optical excitation in the [illegible] ground state via laser fluorescence at the strong $^1S_0$ - $^1P_1$ intercombination line at frequency ω_0 provides [illegible] an atomic amplification mechanism. For Tl^+, see Fig. 1, where the estimated lifetime of the $^3P_0$ level is 50 ms, absorption of a single photon at the forbidden $^1S_0$ - $^3P_0$ transition at 2022 Å suffices to $^1P_1$-laser fluorescence at 1909 Å [illegible] at 50 ms, during which interval otherwise [illegible] photons might have been scattered. Frequency shifts of the sharp 2022 Å line due to the laser excitation are deviated by appropriate pulsing [illegible]. Centering of the ion at the bottom of the trap is achieved via rf and cooling by tuning the 1909 Å laser a few megahertz below the $^1S_0$ - $^1P_1$ resonance at ω_0 - thus that its Doppler side band at $\omega_0 - \omega_y$ is preferentially excited. The vibration frequency in the parabolic trap is [illegible] ω_y. Three-dimensional cooling to [illegible] is achieved by means of a [illegible] strategically directed laser beam.

Apparatus

The heart of the apparatus is the small [illegible] rf trap containing the ion. The design used in the Heidelberg Ba^+ experiment[7,11] is well suited for the future work on Tl^+ etc. and will therefore be sketched here. The trap, see Fig. 2, is formed by appropriately ground wire stubs, the "caps" and a wire "ring." Application of an rf voltage $V_0 \sin \Omega t$ with $\Omega = 2\pi \cdot 16$ MHz created a 3-dimensional well of axial depth [illegible] volt (for Ba^+) in which the ion oscillated at an axial frequency $\omega_z = 2\pi \cdot 2.4$ MHz and perpendicular frequencies $\omega_x, \omega_y = 2\pi \cdot 1.2$ MHz with $|\omega_x - \omega_y| = 2\pi \cdot 10$ kHz. Ideally the cooling laser beam is directed along the body diagonal[7] of the x*,y*,z* = z coordinate system formed by the principal axes of the effective trapping potential. It suffices to approximate these conditions by slightly deforming the ring electrode elliptically and appropriately directing the laser beam through the gaps between ring and cap electrodes. The resonance fluorescence is likewise observed through the ring-cap gap. A vacuum of $\sim 10^{-11}$ Torr is maintained by an ion getter pump in the sealed off, baked glass envelope. The trap was filled by crossing and ionizing a very weak Ba atomic beam with a very weak electron beam in the center of the trap.

Obtaining the ultraviolet power necessary for work with any of the Tl^+ like ions is still a serious experimental problem. For In^+ and Al^+ it should be possible to obtain sufficient laser power to saturate the somewhat allowed $^1S_0$ - $^3P_1$ transition by frequency doubling the output of a stabilized commercial ring dye laser. However scattering rates for these ions are much lower than for Tl^+. The weak laser power sufficient for the forbidden $^1S_0$ - $^3P_0$ transition may probably be obtained by tripling the frequency of a semiconductor diode laser. This latter approach may be the feasible in the future even for the $^1S_0$ - $^1P_1$ laser where resonant harmonic generation schemes have been developed.

Cooling

The following description[1,2,11] brings out the essential features of optical cooling. Due to the Doppler effect, when moving at velocity v opposite to the propagation of a plane light wave of frequency ω, an atom in its rest frame sees an excitation at the frequency $\omega' = \omega(1 + v/c)$ near an atomic resonance at ω_0. Then, again in its rest frame, the atom re-emits photons at ω' in all directions in accordance with its characteristic radiation pattern. Now to an observer in the laboratory due to the Doppler effect these re-emitted photons will exhibit various frequency shifts depending on the angle between the directions of

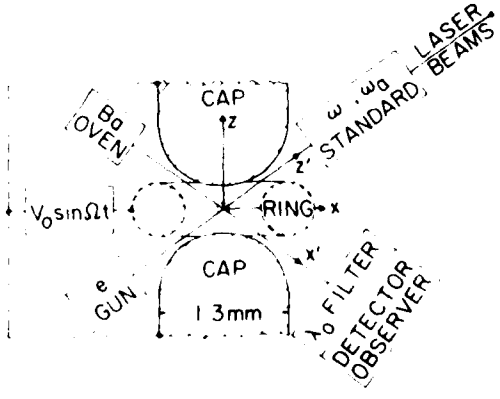


Fig. 2. Apparatus for trapping in ultrahigh vacuum, cooling, and visually observing an individual Ba^+ ion (schematic). Ba oven and electron gun were actually not in the xz plane. Vacuum envelope, pump etc. are not shown. From ref. 11.

re-emission and motion. Nevertheless the average energy of the re-emitted photons will be $\hbar\omega' = \hbar\omega + \hbar\omega(v/c)$. The average energy excess in the energy of the re-emitted over that of the absorbed photons has to come from the kinetic energy in the translational motion which is thereby cooled. This picture is easily adapted to the case of one-dimensional periodic motion at frequency ω_v parallel to the light beam. Now due to the modulated Doppler shift the atom of resonance frequency ω_0, in its rest frame, when irradiated by a laboratory light source at the sharp frequency ω sees a whole excitation spectrum at $\omega'_n = \omega + n\omega_v$, $n = 0,1,2\ldots$. For a sharp spectral line of width $\Delta\omega_0 << \omega_v$ and $\omega'_m = \omega + m\omega_v \approx \omega_0$, $m > 0$, the effect of the m-th side band predominates. Again the average energy of the photons re-emitted in all directions is $\hbar\omega + m\hbar\omega_v$ and the energy excess $m\hbar\omega_v$ has to be provided by the oscillatory motion. For the oscillation amplitude $z_0 << \lambda_0/2\pi$ (in the Lamb-Dicke regime) the spectrum seen by the atom consists only of the carrier at ω of power E plus two weak symmetric side bands at $\omega \pm \omega_v$ of power $E_\pm << E(z_0/2\lambda_0)^2$ and the problem remains manageable even when the requirement $\Delta\omega_0 << \omega_v$ is dropped, see Fig. 3.

The top section of Fig. 3 shows the absorption profile of the stored ion when the laser frequency ω is tuned through the electronic resonance at ω_0. The multiple excitations of the ion in its rest frame at $\omega - \omega_v$, ω, $\omega + \omega_v$ according to the position of these spectral components with respect to the Lorentz profile centered about ω_0, see middle section of Fig. 3, result in the re-emission of these same three components. When viewed from the lab frame each of the 3 re-emitted components will develop its own two Doppler side bands, see bottom section of Fig. 3, without changing the average photon energy. Comparing top and bottom parts of Fig. 3 showing absorption and re-emission as viewed from the lab frame illustrates the cooling process: excitation by an upper-lower or cooling/heating side band photon extracts/adds the energy $h\omega_v$ from/to the vibrational motion. For the case depicted while photons of energy $\hbar\omega$ are absorbed by the ion those re-emitted have a larger average energy $\sim \hbar(\omega + \omega_v)$ for the tuning $\omega = \omega_0 - \omega_v$ shown.

This semiclassical picture indicates strong cooling and an exponential approach in time of absolute zero temperature. Quantum effects establish a finite minimum temperature, however. Obviously, when the ion is in the lowest vibrational state $v = 0$ it cannot be cooled any further and the power in the cooling side band must vanish while that in the heating side band remains finite. This asymmetry in side band strength does not begin with $v = 0$ but is always present, becoming more and more pronounced as v decreases. A similar asymmetry is present in the re-emission side bands. The attainable minimum vibrational temperature and the corresponding vibrational quantum number $\langle v\rangle_{min}$ are determined by the requirement that the heating effect of the stronger but more off-resonant spectral components at $\omega - \omega_v$ and ω is balanced by the cooling effect due to the weaker, more resonant side band $\omega + \omega_v$. Using only familiar elementary quantum mechanics of atomic and molecular spectra I analysed the problem in 1976 and

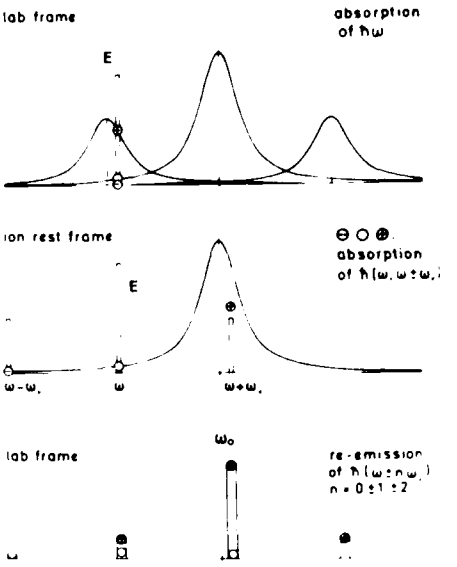


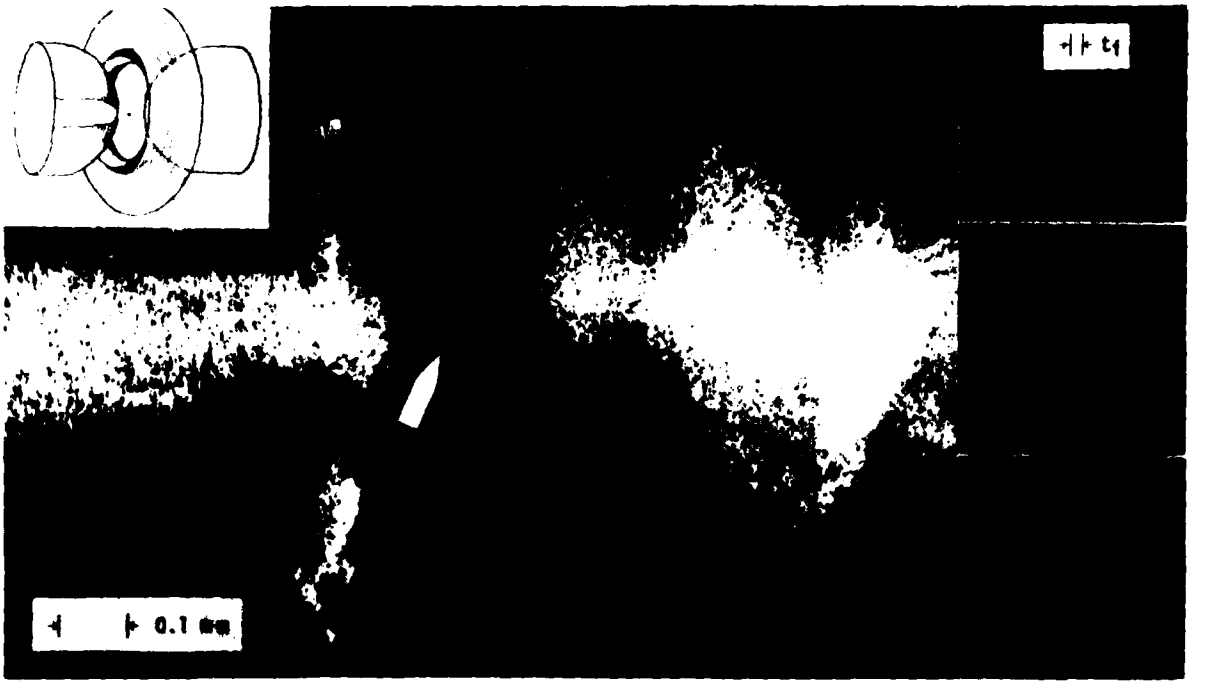
Fig. 3. Spectral components effective in side band cooling (schematic).
As shown in middle section, the ion vibration at ω_v in the parabolic trap parallel to the propagation direction of the exciting electromagnetic wave at ω sees spectral components (of power E_-, E, E_+) at $\omega - \omega_v$, ω, $\omega + \omega_v$. The case of oscillation amplitude $<< c/\omega$ and $E_\pm << E$, the Lamb-Dicke regime, is assumed here. To an observer in the lab-frame the ion appears to have the response spectrum (e.g., for absorption power) shown in top section. In the ion rest frame each of the three exciting components is re-emitted with unshifted frequency and a strength determined by the Lorentzian response profile. When seen from the lab frame (compare bottom section) each re-emitted component develops its own two Doppler side bands. Comparing top and bottom sections shows clearly for the tuning $\omega + \omega_v = \omega_0$ assumed that, while photons of energy $\hbar\omega$ are absorbed by the ion, photons of larger average energy $\sim \hbar(\omega + \omega_v)$ are emitted by it. The balance $\sim \hbar\omega_v$ has to come from the oscillatory motion, which is thereby cooled. After ref. 17, from ref. 12.

applied it to the Tl^+ mono-ion oscillator obtaining a minimum temperature[17,18] of ~ 0.1 mK for cooling with a $\lambda = 1909$ Å laser beam. In this chapter ω_0 refers to the allowed transition used for cooling.

The discussion is easily extended to the realistic 3-dimensional case of an individual stored ion. This requires irradiation along[7] $\vec{i}^* + \vec{j}^* + \vec{k}^*$ where $\vec{i}^*, \vec{j}^*, \vec{k}^*$ refer to the principal axes of the ellipsoidal trapping potential. For the minimal vibrational quantum number we obtained

$$\langle v\rangle_{min} = (q_\omega + q_-)/(q_+ - q_-) \quad ,$$

which applies to vibration along all three principal axes. Here q_ω, q_+, q_- denote the values of the



[illegible] trapped Ba^+ ions. The large photograph [illegible] the [illegible] arrows [illegible] the [illegible] quadrupole trap [illegible] toward the gap between the [illegible] cap electrodes. Trap structure [illegible] illuminated by scattered laser light. A sketch of the whole trap structure [illegible] the same orientation. The three small photos, from top to bottom, show [illegible] the central [illegible] containing [illegible] ions. The trap parameters were [illegible] MHz, ω_z [illegible] MHz, ejection beam off, and [illegible] exposure of 1 min and [illegible] min [illegible] the [illegible] laser [illegible].

[illegible] [illegible] [illegible] [

parameters. Thus much improvement may be expected to occur in the future.

Results of Preparatory Heidelberg and Seattle Experiments

The most important result so far obtained in our mono-ion oscillator work[14] is the 3-dimensional localization in free space to a region of ∿ 2000 Å diameter of an individual Ba^+ ion, see Fig. 4, at Heidelberg. The photographic image of the ion has a diffraction limited diameter of 2 μm. It is possible to de-convolute the much smaller localization range by comparing the 1-ion image with a 3-ion image which, obtained under otherwise identical conditions, has a ∿ 20 times larger area and a ∿ 30 times lower peak brightness. This implies $z_0 \approx \lambda_0/2\pi$ and that the boundary of the Lamb-Dicke regime of the dominant carrier has been reached. The best spectral resolution obtained so far by us[22] is a line-width of ∿ 90 MHz for the $6^2S_{1/2} - 6^2P_{1/2} - 5^2D_{3/2}$ two-photon transition[9] to the highly metastable $5^2D_{3/2}$ state of Ba^+ which has a lifetime of 17 sec, see Fig. 5. This two-photon transition is obviously still strongly power-broadened. In Seattle in work recently made possible by support from the U.S. Office of Naval Research resonance fluorescence from ∿ 300 Mg^+ ions stored in an rf trap in high vacuum has been seen.[23] It has not been possible so far to demonstrate cooling with the weak ≤ 1 μW laser power currently available. This is to be expected from the Heidelberg Ba^+ work[11,12] where for ≤ 50 ions even with laser powers of ∿ 1 mW only ion temperatures of ∿ 100 K could be realized. The presumed cause, rf heating via ion-ion collisions, however, is expected to disappear for a single ion in the trap. The latter expectation is corroborated by the observation of storage times as long as ∿ 30 sec for an uncooled ion.[14] Thus it still may be possible to keep a single Mg^+ ion cold and confined once one has managed the not so mean feat of getting it into the bottom of the parabolic well with the low laser power currently available to us from our somewhat inadequate equipment.

Conclusion

After a fairly difficult start, mono-ion oscillator spectroscopy now appears to be off and running, with at least five groups all over the world actively participating in such studies. The demonstration of electronic linewidths and long-term frequency stabilities in such passive devices of 1 Hz and less is likely to act as a strong stimulus for the future development of laser sources of improved short-term stability, much closer to the fundamental limit near ∿ 10^{-4} Hz and also of new types of laser harmonic generators. Thus the great promise of an atomic line spectral resolution of about 1 part in 10^{18} may be realized in the not too far future. The benefits to physics, metrology, cosmology and technology resulting from such an advance in the accuracy of measurement of an observable as fundamental as time need not be detailed here.

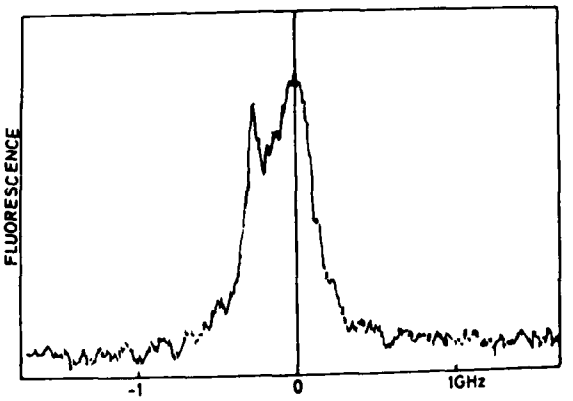


Fig. 5. Resonance fluorescence near 493 nm from an individual barium ion vs. scanned frequency ω_a of 650 nm laser. The 493 nm laser frequency ω was adjusted to $\omega_{sp} - \omega = 2\pi \cdot 280$ MHz to effect side band cooling. The frequency coordinate calibration is given for $\omega_a - \omega_{pd}/2\pi$. The sharp two-photon transition $6^2S_{1/2} - 6^2P_{1/2} - 5^2D_{3/2}$ to the metastable $5^2D_{3/2}$ level occuring for $\omega - \omega_a = \omega_{sp} - \omega_{pd}$ is visible at $\omega_a - \omega_{pd} = -2\pi \cdot 280$ MHz. Here ω_{sp}, ω_{pd} are the respective single photon resonance frequencies. After ref. 22.

Acknowledgements

I am indebted to Gary Janik for reading the manuscript and valuable comments. Thanks also to Linda Chapel for typing the manuscript and to Dave Azose for photographic work.

The Heidelberg work is supported by a grant from the Deutsche Forschungs-Gemeinschaft. A contract of the U.S. Office of Naval Research supports the Seattle work.

References

1. E.N. Fortson, F.G. Major and H.G. Dehmelt, Phys. Rev. Lett. 16, 221 (1966).
2. H.G. Dehmelt, Phys. Rev. 103, 1125 (1956).
3. S. Menasian, see H.G. Dehmelt, in "Atomic Physics," Plerum, 1969.
4. F.G. Major and G. Werth, Phys. Rev. Lett. 30, 1155 (1973).
5. H.G. Dehmelt, Bull. Am. Phys. Soc. 18, 1571 (1973).
6. H.G. Dehmelt, Bull. Am. Phys. Soc. 20, 60 (1975).
7. D. Wineland and H. Dehmelt, Bull. Am. Phys. Soc. 20, 637 (1975).
8. R. Van Dyck, Jr., P. Ekstrom and H. Dehmelt, Nature 262, 776 (1976).
9. H. Dehmelt and P. Toschek, Bulletin APS 20, 61 (1975).

10. H. Dehmelt and H. Walther, Bulletin APS 20, 61 (1975).
11. W. Neuhauser, M. Hohenstatt, P. Toschek and H. Dehmelt, Phys. Rev. Lett. 41, 233 (1978).
12. W. Neuhauser, M. Hohenstatt, P. Toschek and H. Dehmelt, Appl. Phys. 17, 123 (1978).
13. D.J. Wineland, R.E. Drullinger and F.L. Walls, Phys. Rev. Lett. 40, 1639 (1978); D.J. Wineland, Bull. Am. Phys. Soc. 24, 634 (1979).
14. W. Neuhauser, M. Hohenstatt, P.E. Toschek and H. Dehmelt, Phys. Rev. A22, 1137 (1980).
15. W.M. Itano and D.J. Wineland, Phys. Rev. (A), in press.
16. D.J. Wineland and W.M. Itano, Physics Letters 82A, 75 (1981).
17. H. Dehmelt, private communication (1976); Morris Loeb Lectures in Physics, Harvard University (1977).
18. H. Dehmelt, Nature 262, 777 (1976); Bull. Am. Phys. Soc. 24, 634 (1979).
19. D.J. Wineland and W.M. Itano, Phys. Rev. A20, 1521 (1979).
20. J. Javanainen, Appl. Phys. 23, 175 (1980).
20a. A neighboring ion at the equilibrium distance in the well[14] of ~ 5 μm would cause an analogous quadrupole shift of ~ 200 Hz which would not be averaged out by the motion of the ions.
21. J.L. Hall in "Atomic Physics 3," Plenum (1973).
22. W. Neuhauser, M. Hohenstatt, P.E. Toschek and H. G. Dehmelt, in "Spectral Line Shapes," B. Wende, editor. Walter de Gruyter & Co., Berlin, New York (1981).
23. W. Nagourney and H. Dehmelt, Bulletin APS 26 (1981).
24. Detection was by absorption of a beam of resonance light.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

PROPOSED STORED $^{201}Hg^{+}$ ION FREQUENCY STANDARDS*

D. J. Wineland, Wayne M. Itano, J. C. Bergquist,
and F. L. Walls

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

Summary

In this paper, we discuss the performance potential and the problems of implementing a microwave frequency (and time) standard and an optical frequency standard utilizing $^{201}Hg^{+}$ ions stored in a Penning trap. Many of the discussions apply to ion storage-based frequency standards in general. Laser cooling, optical pumping, and optical detection of the microwave or optical clock transition could be achieved using narrowband radiation at the 194.2 nm $6p\ ^2P_{\frac{1}{2}} \leftarrow 6s\ ^2S_{\frac{1}{2}}$ transition, while selectively mixing the ground-state hyperfine levels with appropriate microwave radiation. A first-order field-independent microwave clock transition, which is particularly well-suited to the use of the Penning ion trap is the 25.9 GHz $(F,M_F) = (2,1) \leftrightarrow (1,1)$ hyperfine transition at a magnetic field of 0.534 T. The two-photon Doppler-free $5d^9\ 6s^2\ ^2D_{5/2} \leftrightarrow 5d^{10}\ 6s\ ^2S_{\frac{1}{2}}$ transition at 563 nm is a possible candidate for an optical frequency standard. Both standards have the potential of achieving absolute accuracies of better than one part in 10^{15} and frequency stabilities of less than 10^{-16}.

Introduction

In this paper, a specific proposal is made for a $^{201}Hg^{+}$ stored ion microwave frequency (and time) standard which could have an absolute fractional uncertainty of less than 10^{-15}. We also discuss the possibilities for a $^{201}Hg^{+}$ optical frequency standard. A future stored ion frequency standard may not take the exact form described here; nevertheless, it is useful to investigate a specific proposal, since many of the same generic problems will be encountered in any standard based on stored ions.

Since the pioneering work of Dehmelt and co-workers, who first developed the stored ion method for high-resolution spectroscopy,[1] it has been clear that these techniques provide the basis for an excellent time and frequency standard.[2-10] This conjecture is based primarily on the ability to confine the ions for long periods of time without the usual perturbations associated with confinement (e.g., "wall shift" as in the hydrogen maser). Starting with the work of Major and Werth[5] reported in 1973, groups at Mainz[7] and Orsay[10] and at least one commercial company[11] have sought to develop a frequency standard based on $^{199}Hg^{+}$ ions stored in an rf trap. The choice of Hg^{+} ions for a microwave stored ion frequency standard is a natural one because its ground-state hyperfine structure is the largest of any ion which might easily be used in a frequency standard, and its relatively large mass gives a small second-order Doppler shift at a given temperature. This work has been developed to a fairly high level; the group at Orsay[10] has made a working standard whose stability compares favorably to that of a commercial cesium beam frequency standard. However, the full potential of the stored ion techniques has yet to be realized; this appears to be due to two problems: (1) Historically, it has been difficult to cool the ions below the ambient temperature; this is made more difficult in the rf trap by "rf heating"[1] —a process not clearly understood, but one that makes it difficult to cool even to the ambient temperature.[12] For both the rf and Penning traps, the inability to cool below the ambient temperature means that one must contend with the frequency shift from the second-order Doppler or time-dilation effect. Although it is possible to calculate this shift from the measured Doppler (sideband) spectra, to

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

do so with the required accuracy may be difficult for ions near room temperature. (2) A second problem is that the number of ions that can be stored in a restricted volume (dimensions $\lesssim$ 1 cm) is typically rather small ($\lesssim 10^6$). This, coupled with the somewhat poor signal-to-noise ratios realized with conventional lamp sources, causes the short-term stability in a frequency standard based on ions to be degraded, even though the Q's realized are quite high.

In the past two or three years, both of these problems have been addressed in experiments and the results suggest viable solutions. In 1978, groups at the National Bureau of Standards (NBS) and Heidelberg demonstrated[13,14] that radiation pressure from lasers could be used to cool ions to temperatures < 1 K, thereby reducing the second-order Doppler shifts by 2-3 orders of magnitude below the room temperature case. As discussed below, the cooling is most favorable for very small numbers of ions (down to one ion), so that there is a trade-off between the maximum number of ions we can use and the minimum second-order Doppler shift that can be achieved.

To improve signal to noise, we note that in certain optical-pumping, double-resonance experiments, it is possible to scatter many optical photons from each ion for each microwave photon absorbed.[9] This can allow one to make up for losses in detection efficiency due to small solid angle, small quantum efficiency in the photon detector, etc., so that the transition probability for each ion can be measured with unity detection efficiency. This means that the signal-to-noise ratio need be limited only by the statistical fluctuations in the number of ions that have made the transition.[15] This is discussed in a simple example in Appendix A.

More recently, the narrow linewidths anticipated for the stored ion technique have been observed. A resonance linewidth of about 0.012 Hz at 292 MHz has been observed for the $(m_I, m_J) = (-3/2, 1/2) \leftrightarrow (-1/2, 1/2)$ hyperfine transition of $^{25}Mg^+$ at a magnetic field of about 1.24 T where the first derivative of the transition frequency with respect to magnetic field is zero.[16] (The Ramsey interference method was implemented by applying two rf pulses of 1 s duration separated by 41 s). These narrow linewidths should be preserved with hyperfine transitions of higher frequency, such as in $^{201}Hg^+$, but, of course, more attention must be paid to field homogeneity and stability.

These results have encouraged us to begin studies on the $^{201}Hg^+$ system and although this ion may not provide the "final answer," it appears to provide a case where inaccuracies significantly smaller than 10^{-13} can be achieved. The discussion here is largely devoted to a microwave frequency standard with a design goal of accuracy better than 10^{-15}; however, the possibilities for a stored ion optical frequency standard in $^{201}Hg^+$ are also briefly included.[17,32] For a given interaction time, the Q of a transition will scale with the frequency. Therefore, in principle, an optical frequency standard would have clear advantages over a microwave frequency standard. Our decision to work on a microwave frequency standard (as well as an optical standard) is motivated largely by practical considerations: (1) Before the full potential of an ion-storage optical frequency standard could be realized, tunable lasers with suitable spectral purity must become available. This problem may be nearing solution.[20] (2) If an optical frequency standard is to provide time, the phase of the radiation must be measured. This appears to present a much more formidable problem.[18,19] Both of these problems are already solved in the microwave region of the spectrum—thus, the attraction for investigating a microwave frequency standard.

$^{201}Hg^+$ Stored Ion Microwave Frequency Standard

$^{201}Hg^+$ ions will be stored in a Penning trap. The choice of the Penning trap over the rf trap is motivated primarily because it appears easier to cool a cloud of ions in a Penning trap than in an rf trap. Residual heating mechanisms in the Penning trap are quite small[13] whereas in the rf trap (where "rf heating" occurred), they can be substantial.[14] Of course, this problem does not

exist for the rf trap if single ions are used, but for a microwave frequency standard, it is desirable to use as many ions as possible in order to increase signal-to-noise ratio. Use of a Penning trap means that one must use transitions that are independent of magnetic field to first order[8,9,16]; this limits the number of transitions available, but need not be an absolute restriction. The $^{199}Hg^{+}$ isotope is therefore not considered, since there are no frequency extrema at practical fields.

The energy level structure[21] of the $^{201}Hg^{+}$ ion vs. magnetic field is shown in Fig. 1. The

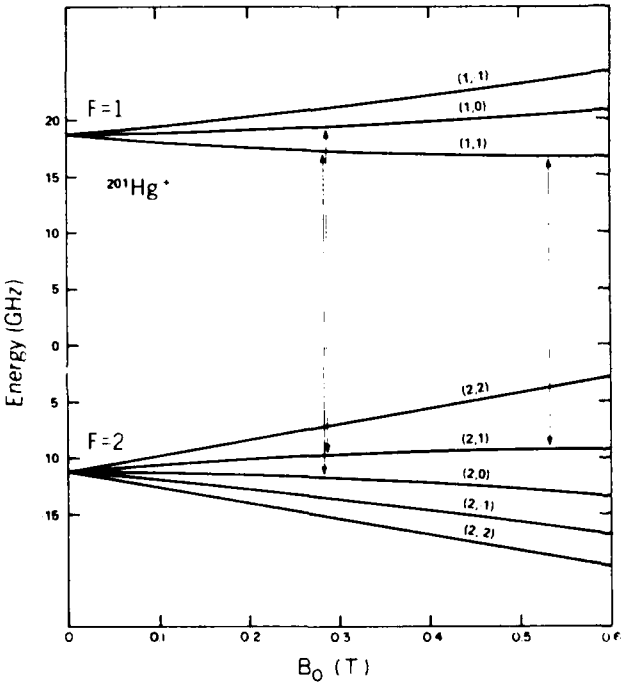


Figure 1. Ground state hyperfine energy levels of $^{201}Hg^{+}$ vs. magnetic field. States are designated by the (F, M_F) representation. Three transitions are indicated at the fields where the transition frequencies are independent of magnetic field to first order.

possible field independent "clock transitions" are the $(F, m_F) = (2,0) \leftrightarrow (1,1)$ and $(2,1) \leftrightarrow (1,0)$ transitions at about 0.29 T, the $(2,1) \leftrightarrow (1,1)$ transition at 0.534 T, the $(1,1) \leftrightarrow (1,0)$ transition at 3.91 T and the $(1,0) \leftrightarrow (1,-1)$ transition at 28.1 T. In principle, one desires to work at the highest microwave frequency possible (for highest Q) at the highest attainable magnetic field (to maximize the number of ions). The choice seems to be between the 25.9 GHz $(2,1) \leftrightarrow (1,1)$ transition at 0.534 T and the 7.72 GHz $(1,1) \leftrightarrow (1,0)$ transition at 3.91 T. At these magnetic fields, where the transition frequency is independent of magnetic field to first order, the second-order field dependence is given by $\Delta\nu/\nu_0$ $((2,1) \leftrightarrow (1,1)) = {}^1/_6\,(\Delta H/H_0)^2$ and $\Delta\nu/\nu_0$ $((1,1) \leftrightarrow 1,0)) \cong .04(\Delta H/H_0)^2$. The remaining part of the proposal is modeled around the $(2,1) \leftrightarrow (1,1)$ transition at 0.534 T because of its higher frequency. From the second-order field dependence noted above, it is necessary to control the field stability and homogeneity over the ion cloud to better than 10^{-7} in order to achieve 10^{-15} accuracy. This is an important problem, of course, but not insurmountable—we note that the free-running stability of the magnetic field must be better than 10^{-7} over the clock transition time; in longer term it can be stabilized to this level by locking the field to a Zeeman transition in the ions.[16]

We will assume that a laser can be tuned to the $6p\ ^2P_{\frac{1}{2}} \leftrightarrow 6s\ ^2S_{\frac{1}{2}}$ transition at 194.2 nm with sufficient power to provide laser cooling, optical pumping, and fluorescence detection. A specific scheme for observing the clock transition might be the following: (we will assume that when the ions are cold, the widths of the Doppler-broadened optical lines are approximately equal to the natural width (~ 70 MHz)).

(1) We tune the laser about 35 MHz below the $6p\ ^2P_{\frac{1}{2}}$ (1,-1) $\leftrightarrow$ $6s\ ^2S_{\frac{1}{2}}$ (2,-2) optical transition. This laser tuning gives the maximum cooling possible,[22] but rapidly pumps the ions out of the (2,-2) ground state. (For example, the (1,-1) excited state decays to the (1,-1) ground state with a probability equal to 0.329.) For cooling and detection, we require each ion to scatter many photons. Unfortunately, the simple schemes[15,16,23] for multiple scattering that have been realized with ions like $^{25}Mg^{+}$ do not exist for Hg^{+}. There-

fore, for the initial cooling, we must mix all the ground states with microwave radiation. This effectively reduces the cooling by a factor of 8 from the two-level ion case, since only $^1/_8$ of the ions are in the (2,-2) ground state, but should not be a problem given enough laser power (see below).

(2) Once cooling is accomplished, we can pump all the ions into one of the clock levels by, for example, mixing all the ground-state levels except the (1,1) level. This pumping occurs because we are exciting the ions in the wings of other optical transitions; for example, we are pumping in the wings of the (1,0) ↔ (2,-1) optical transition (about 3.5 GHz away) and the (1,0) excited state decays into the (1,1) ground state with a probability of 0.039. (Note that the (1,-1) excited state cannot decay to the (1,1) ground state.) We remark that we have neglected the decay of the ions from the $^2P_{\frac{1}{2}}$ excited state to either the $^2D_{3/2}$ or $^2D_{5/2}$ states. That this is a good approximation is shown in Appendix B. The ratio of the scattering rate on the (1,-1) ↔ (2,-2) optical transition to the pumping rate into the (1,1) ground-state level is about 6×10^4, i.e., about 6×10^4 photons are scattered by each ion before it is pumped into the (1,1) ground-state clock level. This number can be reduced by tuning the laser to another optical transition; this may be necessary if the laser power is too small. As discussed below, it is desirable for detection purposes for this number to be large.

(3) After the pumping is achieved, the laser and ground-state mixing rf are turned off (to avoid ac Stark shifts and relaxation) and the clock transition is driven. If we use the Ramsey interference method as was done for $^{25}Mg^+$,[16] the resulting linewidth is about one-half as wide as if we use continuous excitation (the Rabi method). Therefore, the Ramsey method will be assumed.

(4) After the rf cycle is complete, the laser is turned back on as well as the mixing rf (excluding the (1,1) ground-state level) and the fluorescence scattering is observed. It should not be too difficult to detect the scattered photons with about 10^{-3} overall efficiency. This would mean that about 60 photons would be collected for each ion that had made the transition (before it is repumped to the (1,1) level) insuring that the noise would only be due to the statistical fluctuations in the number of ions that had made the transition (see Appendix A).

In order to lock a local oscillator to the center of the clock transition, one would first complete one of the above cycles with the local oscillator tuned to the half-intensity point (where the transition probability for each ion is one half) on, say, the low side of the central Ramsey peak. These photon counts could then be stored and the cycle repeated with the local oscillator tuned to the high-frequency side of the line. These resulting counts could then be subtracted from the first to give an error signal which can be used to force the mean frequency of the local oscillator to be at the center of the clock transition. We will assume that the time for fluorescence observation and repumping is much less than the clock transition time.

Systematic Frequency Shifts

We have already mentioned the stringent requirements on magnetic field stability and homogeneity. Using a superconducting magnet, it is possible to stabilize the magnetic field to better than 10^{-7}. (This may require locking the field to an NMR probe adjacent to the trap over the time of the clock transition). The field homogeneity requirements will be more difficult to satisfy (assuming a 1 cm diameter spherical working volume inside the trap), but are still feasible.

The electric fields from the applied trap potentials and from Coulomb interactions between ions can cause second-order Stark shifts, but the resulting fractional frequency shifts are estimated to be much less than 10^{-15}. The black body ac Zeeman shift[38] is estimated to be $\Delta\nu/\nu_0 \cong 1.3 \times 10^{-17}\ (T/300)^2$ and is therefore neglected. The black body ac Stark shift[39] is estimated to be $\Delta\nu/\nu_0 \cong 2 \times 10^{-16}\ (T/300)^4$ and therefore must be accounted for or reduced environmental temperatures are required.

In spite of the laser cooling that has been achieved, we must still be concerned with the second-order Doppler frequency shift. In the Penning trap, the cyclotron-axial temperature of $^{201}Hg^{+}$ ions needs to be cooled to below 1.45 K to insure that the second-order Doppler shifts (on these degrees of freedom) is less than 10^{-15}. These low temperatures should be easy to obtain. A more serious problem exists for the magnetron rotation of the cloud; the kinetic energy in this degree of freedom will probably limit how small the second-order Doppler shift can be.[24] In the limit of very small numbers of ions, the magnetron kinetic energy should be negligible,[25] but in the case discussed here, we would like to use the maximum number of ions possible and this will cause problems as described below.

The "magnetron" rotation in a Penning trap is simply a form of circular $\vec{E} \times \vec{B}$ drift; that is, the radial electric fields in the trap from the applied potentials on the electrodes and from space charge act in a direction perpendicular to the magnetic field. This causes the ions to drift in circular "magnetron" orbits about the axis of the trap. If we assume that we must keep the ions inside a 1 cm diameter spherical working volume, then qualitatively, the nature of the problem is as follows: if we add more ions to this volume, then the magnetron frequency increases due to two effects. First, we must increase the applied trap potentials to overcome the increased space charge repulsion along the z axis, which tends to elongate the cloud in this direction. Consequently, the magnetron frequency increases due to the increased potentials and the increased space charge fields in the radial direction. In Appendix C, we estimate the maximum number of ions contained in a 1 cm diameter sphere assuming that the second-order Doppler shift is 10^{-15} for ions on the perimeter of the cloud at z = 0. We obtain $N_{max} = 8.2 \times 10^{4}$ and note that the applied trap voltage is only 71 mV for $z_0 = r_0/1.64 = 0.8$ cm.[1]

Frequency Stability

At optimum power (transition probability is equal to one at line center), we can closely approximate the number of detected photon counts for each experimental cycle as

$$N = N_i\, n_d \left(\frac{1 + \cos(\omega - \omega_0)\, T}{2} \right) \qquad (4)$$

where, as in Appendix A, N_i is the number of ions in the trap and n_d is the average number of detected photons for each ion that has made the transition. ω and ω_0 are the frequency of the applied rf and "clock" center frequency; T is the time between the rf pulses at the beginning and end of the rf period. (We assume that the time of the rf pulses is much less than T.) As described above, the interrogating oscillator is switched between $\omega - \pi/2T$ and $\omega + \pi/2T$ (where we assume $|\omega - \omega_0| \ll \pi/2T$), and the resulting counts subtracted to give an error signal. The sensitivity to mistuning of ω is given by calculating the slope of the signal in Eq. 4 at the half-intensity points ($|\omega - \omega_0| = \pi/2T$). We have

$$\left|\frac{dN}{d\omega}\right|_{|\omega-\omega_0| = \frac{\pi}{2T}} = \frac{N_i n_d T}{2}$$

After one full switching cycle (taking the difference of the counts from both sides of the line), the error signal is

$$\delta N = 2 \left|\frac{dN}{d\omega}\right|_{|\omega-\omega_0| = \pi/2T} \times \delta\omega$$

where $\delta\omega = \omega - \omega_0$. Since N fluctuates statistically, these fluctuations (δN) give rise to frequency fluctuations in the locked local oscillator:

$$(\delta\omega)^2_{rms} = \frac{\delta N^2}{4 \left|\frac{dN}{d\omega}\right|^2_{|\omega-\omega_0| = \pi/2T}}$$

Maximum frequency stability is thus given by

$$\sigma_y^2(2T) = \frac{1}{2} \frac{(\delta\omega)^2_{rms}}{\omega_0^2}$$

Assuming $n_d \gg 2$ (Appendix A), we have[26]

$$\sigma_y(\tau) = \frac{1}{2T\omega_0 \sqrt{N_i}} \sqrt{\frac{2T}{\tau}} \qquad \tau > 2T$$

For $N_i = 8.2 \times 10^4$ and $T = 50$ s,

$$\sigma_y(\tau) = 2 \times 10^{-15}\, \tau^{-\frac{1}{2}} \qquad \tau > 100 \text{ s}$$

and

$$Q = \frac{\omega_0 T}{\pi} \sim 2.6 \times 10^{12}$$

Lasers

Perhaps the single reason that such a proposal has not been made previously is that the required narrowband tunable laser has not been available at 194.2 nm. However, it appears that two possible approaches lend themselves to initial experiments in this system. Briefly, the first approach might be to use an externally narrowband filtered (~ 100 MHz) pulsed ArF excimer laser. We estimate that it should be possible to achieve saturating intensity from such a filtered laser. However, the pulse lengths of these lasers are quite short ($\lesssim$ 10 ns), so that only about two photons per ion will be effective in each laser shot to drive the optical transition. (The lifetime of the upper $^2P_{\frac{1}{2}}$ (1,-1) state is about 2.3 ns and decays with 0.46 probability to the (2,-2) ground state.) The potential advantage of excimer laser systems is that the repetition rates can be quite high (KrF lasers have been built with 1 kHz repetition rates). However, at the present time, it is probably only feasible to realize 150 Hz repetition rates for ArF lasers. This should allow reasonable signal to noise in the above scheme when the ions are cold; however, we estimate cooling times of order 20 minutes, which is uncomfortably long.

A second scheme using frequency mixing of cw lasers in nonlinear crystals is presently being pursued at NBS. Tunable coherent radiation in the 194 nm region has previously been produced by phase-matched sum frequency mixing of pulsed lasers in a potassium pentaborate (KB5) crystal.[27] The second harmonic of the 514 nm Ar^+ single frequency cw laser line, when mixed with radiation near 790 nm from a cw dye laser, will generate single frequency cw radiation at 194 nm. The estimated efficiency of this process for a 3 cm crystal is given by

$$P_3 \simeq 8 \times 10^{-5}\, P_1\, P_2,$$

where P_3 is the output power at 194 nm, P_1 is the input power at 257 nm from the doubled Ar^+ laser, and P_2 is the input power at 790 nm.[28,29] All powers are expressed in watts. The second harmonic of 514 nm radiation can be generated in 90° phase matched temperature tuned KDP or ADP crystals with an efficiency (for a 5 cm crystal) given by

$$P_1 \simeq 2.5 \times 10^{-3}\, P_0^2,$$

where P_1 is the output power at 257 nm and P_0 is the input power at 514 nm. Thus, about 10 mW can be produced with a 2 W input. The output power can be increased considerably by using a cavity to build up the circulating power. As much as 300 mW of cw power at 257 nm has been produced in this way.[30] Assuming that 200 mW at 257 nm and 500 mW at 790 nm are available, about 8 μW at 194 nm could be produced. From the experience with Mg^+ ions,[13,15,16] this should be adequate power for initial experiments. The output could be further increased by building a cavity around the KB5 crystal to increase the circulating power at 790 nm, 257 nm, or both. Finally, the output power at the ions could be increased by building a ring cavity around the trap. The frequency of the Ar^+ laser can be stabilized to an I_2 absorption. The dye laser can be stabilized and tuned using standard techniques. The temperature of the KB5 may have to be shifted slightly from room temperature in order to satisfy the phase-matching conditions. We estimate from the observed temperature tuning at 201.6 nm[31] that the required shift is less than 25 °C.

Two-Photon Optical Frequency Standard in $^{201}Hg^+$

We will briefly describe the properties of a $^{201}Hg^+$ optical frequency standard that has been suggested previously.[32] ($^{199}Hg^+$ should, of course, also be considered because of its simpler

structure.) Using two-photon Doppler-free spectroscopy,[33] it should be possible to excite $^{201}Hg^{+}$ ions from a ground-state sublevel to a particular $^{2}D_{5/2}$ sublevel using a dye laser tuned to approximately 563.2 nm. An excited state and magnetic field could be chosen so that the transition frequency could be independent of magnetic field to first order. For the case of this optical transition, the second-order dependence of fractional frequency offset due to magnetic field would be reduced by approximately the ratio of the optical frequency (5.33×10^{14} Hz) to the ground-state hyperfine frequency. This represents a reduction in sensitivity of about four orders of magnitude which would greatly relax the constraints on magnetic field. The ground and excited states could be chosen such that the ground-state level would be depopulated by the two-photon transition; therefore, detection could be accomplished in essentially the same manner as described for the microwave case above. The lifetime of the excited states[35] is about 0.11 s, so that the Q of this transition is about 7.4×10^{14}!

A drawback to this scheme is the accompanying ac Stark shift[33]; this shift is formally equivalent to the "light shift" in rubidium frequency standards. We have roughly estimated the transition probability per unit time to be $W \cong 0.3\ I^2/\delta\nu$ where $\delta\nu$ is the larger of the (doubled) laser linewidth or the natural linewidth (1.4 Hz) and I is the laser intensity in each direction in W/cm^2. We have also estimated the accompanying light shift to be $\Delta\nu \cong -I$ (Hz). If the (doubled) laser linewidth is less than the natural width and if we drive the transition near saturating intensity ($W \cong 1/s$), then $I \simeq 2.2$ W/cm^2, which implies a fractional frequency shift of 2×10^{-15}.

The black body ac Stark shift[38] is estimated to be $\Delta\nu/\nu_0 \simeq 10^{-16}\ (T/300)^4$. The frequency shift[40] due to the interaction of the quadrupole moment of the atomic D levels with the applied quadrupole field and fields due to ion-ion collision is estimated to be less than 10^{-16} for the conditions described here.

In initial experiments, it will very likely not be possible to obtain (doubled) laser linewidths less than 1.4 Hz; however, laser linewidths of a few tens of hertz should be achievable. Therefore, the above projections may not be too optimistic.

Assuming the same conditions as for the microwave case, predicted stabilities are also quite dramatic (assuming laser linewidths are sufficiently narrow). If we assume that the (Rabi) interaction time is 1 s and that the detection and repumping time is much less than 1 s, this would imply (for $N_i = 8.2 \times 10^4$)

$$\sigma_y(\tau) \cong 2 \times 10^{-18}\ \tau^{-\frac{1}{2}} \qquad \tau \gtrsim 2 \text{ s}$$

For $N_i = 1$, $\sigma_y(\tau) \cong 6 \times 10^{-16}\ \tau^{-\frac{1}{2}}$.

Conclusions

Some of the problems associated with a stored-ion frequency standard have been addressed by making a specific proposal around the $^{201}Hg^{+}$ ion. Although other interesting candidates exist, this system appears feasible enough that experimental work has begun at NBS. Current efforts are aimed at producing the 194.2 nm laser light, producing narrowband 563.2 nm laser light for the two-photon transition, studying the ion cloud dynamics in order to produce the rather diffuse, spatially stable ion clouds and making ion traps with significantly increased collection efficiency.

Acknowledgments

The authors wish to acknowledge the continued support from the Air Force Office of Scientific Research and the Office of Naval Research. We thank H. Dehmelt, C. Pollock, and H. Robinson for critical comments and we thank R. Ray and B. Barrett for manuscript preparation.

APPENDIX A

Statistical Fluctuations in Detected Photon Counts

For simplicity, we will assume the conditions shown in Fig. 2. The actual conditions discussed in this paper and in Refs. 15 and 16 are somewhat

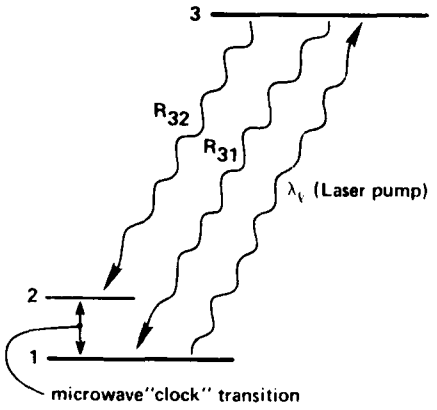


Figure 2. "Model" 3 level system to show the effects of statistical fluctuations in an optical pumping, double resonance experiment.

more complicated, but the basic result still applies. We assume that a laser is tuned to the 3 ← 1 transition wavelength and that we can neglect excitation in the wings of the 3 ← 2 transition by this laser. We also assume that the decay branching ratios R_{32} and R_{31} are such that $R_{31} >> R_{32}$.

The basic scheme for an optical-pumping double-resonance experiment in this simple example is the following: (1) If the number of scattering events (for each ion) is sufficiently large, then essentially all the ions are pumped into the 2 level. (2) The laser is now turned off and the "clock" (1 ← 2) transition is driven. (3) The ions that have made the transition are then detected by turning the laser back on and observing the fluorescence scattering by collecting the light in a phototube.

For each ion that has made the clock transition, let the average number of detected photons be n_d; the rms fluctuation in the number of detected photons per ion is $\sqrt{n_d}$. Because of inefficiencies in collection and detection, typically $n_d << R_{31}/R_{32}$, but the interesting case will be when n_d is still significantly greater than one. We will assume that the clock transition is driven with optimum power, so that when we are tuned to the half-intensity point on the line (for maximum frequency sensitivity), the probability, p, that each ion has made a transition is 0.5. In this case, if the total number of ions is N_i, then the average number that have made the transition is $pN_i = N_i/2$ and the rms fluctuations in the number of ions that have made the transition on each experimental cycle is

$$\Delta N_i = \sqrt{N_i\, p(1-p)} = \sqrt{N_i}/2.$$

For one photon counting cycle, the average number of photons detected is $N_{TOT} = n_d\, N_i/2$. The fluctuations in the number of detected photons (ΔN_{TOT}) is due to two causes: (1) The fluctuations in the number of photons counted $(\Delta n_{TOT})_d$ due to fluctuations in the counted photons for each ion. Since these are statistically independent:

$$(\Delta n_{TOT})_d = \sqrt{\sum^{N_i/2} (\Delta n_d)^2} = \sqrt{n_d\, N_i/2}$$

(2) The fluctuation in the number of photons counted $(\Delta n_{TOT})_i$ due to the fluctuations in the number of ions that have made the transition:

$$(\Delta n_{TOT})_i = \Delta N_i\, n_d = n_d \sqrt{N_i}/2$$

Since these two processes are also statistically independent, the total fluctuations in N_{TOT} are given by:

$$\Delta N_{TOT} = \sqrt{(\Delta n_{TOT})_d^2 + (\Delta n_{TOT})_i^2}$$

$$= \frac{n_d \sqrt{N_i}}{2} \sqrt{1 + \frac{2}{n_d}}$$

Therefore when $n_d >> 2$, the fluctuations in the photon counts are given by the statistical fluctuations in the number of ions that have made the transition. We must note, of course, that the above arguments assume that the laser intensity and cloud of ions is stable, and therefore, fluctuations in signal due to changes in laser-cloud spatial overlap are negligible.

Appendix B

$^2P_{1/2} \rightarrow {}^2D_{3/2,5/2}$ Decay Rates

Decay from the $5d^{10}\ 6p\ ^2P_J$ levels to the $5d^9\ 6s^2\ ^2D_J$ levels is allowed through configuration interaction, which can mix some $5d^{10}\ 6d$ amplitude into the D states and some $5d^9\ 6s\ 6p$ amplitude into the P states.[35] Crandell, et al.[36] have determined that the decay rate from the $^2P_{3/2}$ state to the $^2S_{1/2}$ ground state is 350 ± 30% times greater than the decay rate to the $^2D_{5/2}$ state. If LS coupling is valid, the $^2P_{1/2}$ state decays to the $^2D_{3/2}$ with probability 3×10^{-7}. This decay is highly suppressed because the energy difference is only 933 cm^{-1}, which corresponds to a wavelength of 11 μm. (The decay rate is proportional to the cube of the energy difference.) Decay of the $^2P_{1/2}$ state to the $^2D_{5/2}$ state is forbidden by the electric dipole selection rules, since it requires J to change by 2. Hyperfine and Zeeman interactions mix different J states, making this decay slightly allowed. However, we estimate this probability to be less than 10^{-11} at a magnetic field of 0.5 T. The $^2D_{3/2}$ state decays to the $^2D_{5/2}$ state at a rate of about 54 s^{-1} and to the $^2S_{1/2}$ state at a rate of about 42 s^{-1}. The $^2D_{5/2}$ state decays to the $^2S_{1/2}$ at a rate of about 9.5 s^{-1}.[35] If the laser-induced $^2S_{1/2}$ to $^2P_{1/2}$ transition is denoted by γ_ρ, the ions reach the $^2D_{3/2}$ state at a rate of $3 \times 10^{-7} \cdot \gamma_\ell$ from which they decay, with probability about 0.56, to the $^2D_{5/2}$ state, where they stay for an average of about 0.11 s before decaying back to the $^2S_{1/2}$ state. If γ_ρ is less than 6×10^6 s^{-1}, the ions spend less than 10% of their time in the metastable D states, so that this trapping does not cause a problem for the cooling. Transitions to the D states occur at a rate which is much less than the optical pumping rates between ground-state sublevels, so their neglect in the previous discussion on optical pumping is justified.

Appendix C

Maximum Number of Stored Ions

Since the cyclotron and axial kinetic energies are assumed cold, we will assume a uniform charge distribution for the ions.[37] For a spherical ion cloud, the magnetron rotation frequency of the cloud (ω_m) is given from the equations of motion as:

$$\omega_m = \frac{\omega_c}{2} - \sqrt{\left(\frac{\omega_c}{2}\right)^2 - \left(\frac{\omega_z^2}{2} + 4\pi\frac{e\rho}{3M}\right)} \qquad (C.1)$$

where ω_c is the unperturbed ion cyclotron frequency $\omega_c = eB/Mc$, ω_z is the axial oscillation frequency derived from the applied trap voltage (V_o) as:[1]

$$\omega_z^2 = 4e\ V_o/M(r_o^2 + 2z_o^2)$$

and ρ is the space charge density. We have for the space charge potential from the ions (inside the cloud)

$$\phi_i = \frac{2}{3}\pi\rho\ (r^2 + z^2)$$

and the trap potential may be expressed as

$$\phi_T = \frac{V_o\ (r^2 - 2z^2)}{r_o^2 + 2z_o^2}$$

When the axial-cyclotron motion is cold, we have $\phi_i(z) = \phi_T(z) = 0$; this implies

$$\rho = \frac{3}{4\pi}\ \frac{M}{e}\ \omega_z^2 \qquad (C.2)$$

If we assume we want the maximum second-order Doppler shift $\left(\frac{1}{2}\left(\frac{v}{c}\right)^2\right)$ less than ε, then we require $\omega_m = c\sqrt{2\varepsilon}/r_{c\ell}$ where $r_{c\ell}$ is the radius of the cloud. If we want to maximize the number of ions $N = 4\pi\rho(r_{c\ell})^3/(3e)$, then we want to maximize this expression subject to the above contraints on ω_m. Substituting Eqs. (C.1) and (C.2) into this expression for N we have:

$$N = 2\ \frac{\sqrt{2}\ Mc}{3e^2}\ r_{c\ell}\ \sqrt{\varepsilon}\left(\omega_c\ r_{c\ell} - c\sqrt{2\varepsilon}\right) \qquad (C.4)$$

(Values of $r_{c\ell}$ where this expression is negative are unphysical because we assume that ε is a fixed value on the perimeter of the cloud (at $z = 0$) no matter what its size is. For very small clouds, this requires V_o large enough that the ions are unbounded—the case for $N \lesssim 0$.)

From (C.2) and the expression for for ω_z^2,

$$V_o = \pi(r_o^2 + 2z_o^2)\,\rho/3 \qquad (C.5)$$

For the conditions assumed in the text, $r_{c\ell} = 0.5$ cm, $\omega_c = 40.5$ kHz, and $\varepsilon = 10^{-15}$ we have $N \simeq 8.2 \times 10^4$. Assuming $z_o = r_o/1.64 = 0.8$ cm, then $V_o = 0.071$ V.

References

1. H. G. Dehmelt, Advan. Atomic and Mol. Physics 3, 53 (1967) and 5, 109 (1969).
2. F. G. Major, NASA Report X-521.69.167, Goddard Space Flight Center (1969).
3. H. A. Schuessler, E. N. Fortson, and H. G. Dehmelt, Phys. Rev. 187, 5 (1969).
4. H. A. Schuessler, Metrologia 7, 103 (1971).
5. F. G. Major and G. Werth, Phys. Rev. Lett. 30, 1155 (1973) and Appl. Phys. 15, 201 (1978).
6. R. Ifflander and G. Werth, Metrologia 13, 167 (1977).
7. M. D. McGuire, R. Petsch, and G. Werth, Phys. Rev. A 17, 1999 (1978).
8. F. Strumia, Proc. 32nd Ann. Symp. on Freq. Control (Fort Monmouth, NJ, May 1978) p. 444.
9. F. L. Walls, D. J. Wineland, and R. E. Drullinger, ibid. p. 453.
10. M. Jardino, M. Desaintfuscien, R. Barillet, J. Viennet, P. Petit, and C. Audoin, Proc. 34th Ann. Symp. on Freq. Control (Fort Monmouth, NJ, June 1980) p. 353 and Appl. Phys. 24, 1 (1981).
11. M. D. McGuire, Bull. Am. Phys. Soc. 26, 615 (1981).
12. D. A. Church and H. G. Dehmelt, J. Appl. Phys. 40, 3421 (1969).
13. D. J. Wineland, R. E. Drullinger, and F. L. Walls, Phys. Rev. Lett. 40, 1639 (1978).
14. W. Neuhauser, M. Hohenstatt, P. Toschek, and H. Dehmelt, Phys. Rev. Lett. 41, 233 (1978).
15. D. J. Wineland, J. C. Bergquist, Wayne M. Itano, and R. E. Drullinger, Opt. Lett. 5, 245 (1980).
16. Wayne M. Itano and D. J. Wineland, Phys. Rev. A, to be published.
17. See also: H. G. Dehmelt, Bull. Am. Phys. Soc. 18, 1521 (1973) and 20, 60 (1975), and D. J. Wineland and H. G. Dehmelt, Bull. Am. Phys. Soc. 20, 637 (1975).
18. J. L. Hall, Science 202, 147 (1978).
19. D. J. Wineland, Proc. 11th PTTI, (NASA Publ. 2129, Nov., 1979) p. 81.
20. J. L. Hall, Proc. Int. Conf. on Atomic Phys., Boston, Mass. (Aug. 1980).
21. The ground-state hyperfine separation was taken from the measurements of Y. Guern, A. Bideau-Méhu, R. Abjean, and A. Johannin-Gilles, Physica Scripta 14, 273 (1977).
22. D. J. Wineland and Wayne M. Itano, Phys. Rev. A 20, 1521 (1979).
23. Wayne M. Itano and D. J. Wineland, Bull. Am. Phys. Soc. 24, 1185 (1979).
24. A similar limitation for the rf trap is imposed by the kinetic energy in the forced "micromotion" of ions not at the center of the trap. See D. J. Wineland, Proc. 2nd Int. Conf. on Precision Meas. and Fundamental Constants, Gaithersburg, MD, June 1981, to be published.
25. D. J. Wineland and Wayne M. Itano, Phys. Lett. 82A, 75 (1981).
26. The basic scheme proposed for $^{137}Ba^+$ in Ref. 9 should be possible; however, two important errors in this paper should be noted: first, the number of scattered photons before repumping was overestimated because pumping in the wings of the other optical lines was not considered. Second, although the improvement in signal-to-noise ratio due to multiple scattering was noted, it was overestimated because it was not assumed that the fluctuations would be limited to the statistical variations in the number of ions that made the clock transition. (See Appendix A).
27. R. E. Stickel, Jr. and F. B. Dunning, Appl. Opt. 17, 981 (1978).
28. G. D. Boyd and D. A. Kleinman, J. Appl. Phys. 39, 3597 (1968).
29. F. B. Dunning and R. E. Stickel, Jr., Appl. Opt. 15, 3131 (1976).
30. P. Huber, Opt. Commun. 15, 196 (1975).
31. R. E. Stickel, Jr. and F. B. Dunning, Appl. Opt. 16, 2356 (1977).
32. P. L. Bender, J. L. Hall, R. H. Garstang, F. M. J. Pichanick, W. W. Smith, R. L. Barger, and J. B. West, Bull. Am. Phys. Soc. 21, 599 (1976).
33. See for example: V. S. Letokhov and V. P. Chebotayev, Nonlinear Laser Spectroscopy, (Springer-Verlag, New York) 1977.
34. Ye. V. Baklanov, V. P. Chebotayev, and B. Ya. Dubetskii, Appl. Phys. 11, 201 (1976).
35. R. H. Garstang, National Bureau of Standards J. Research 68A, 61 (1964).
36. D. H. Crandall, R. A. Phaneuf, and G. H. Dunn, Phys. Rev. A 11, 1223 (1975).
37. See for example: S. A. Prasad and T. M. O'Neil, Phys. Fluids 22, 278 (1979). T. M. O'Neil and C. F. Driscoll, Phys. Fluids 22, 260 (1979).
38. T. F. Gallagher and W. E. Cooke, Phys. Rev. Lett. 42, 835 (1979).
39. L. L. Lewis, Wayne M. Itano, and D. J. Wineland, to be published.
40. H. G. Dehmelt, paper in these prodeedings.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ, 07703, May 1981

OPTICAL PUMPING BY LASERS IN ATOMIC FREQUENCY STANDARDS*

L. L. Lewis and M. Feldman

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

Summary

Single-mode, near-infrared diode lasers may improve the performance of atomic frequency standards. In the case of rubidium standards, the short-term stability may be improved by using laser diodes for optical pumping in place of conventional rf-excited lamps. In cesium beam standards, the lasers may replace both sets of state selection magnets, resulting in greater signal-to-noise, more reliable beam detection, easily reversed beam direction for cavity phase shift measurement, reduced Majorana transitions, and a smaller, more easily regulated C-field. The degree to which these improvements are realized depends upon the characteristics of available lasers. In this paper, we report measurements of laser intensity and frequency noise and their effects on clock performance. The light shift in a laser-pumped Rb clock is given, as well as the stability curve for that clock. Preliminary work on optical pumping in a cesium beam is also reported.

Key Words. Laser Diode, Atomic Frequency Standard, Optical Pumping, Light Shift, Laser Stabilization.

Introduction

The development of inexpensive, reliable, single-mode laser diodes within the communications industry has made the use of lasers in atomic frequency standards much more attractive. In particular, the use of these devices in vapor cells, atomic beams, and laboratory primary frequency standards is possible. The use of optical pumping by lamps and lasers in other types of standards, such as ion storage traps[1,2] will not be discussed in this article.

Table 1 gives typical stability and accuracy performance of existing commercial and laboratory atomic frequency standards. There are a number of ways in which laser diodes may help improve these specifications. In the case of Rb standards, the greater efficiency in optical pumping afforded by laser diodes over conventional lamps should improve signal-to-noise, resulting in better short-term stability. While long-term stability would still be limited by buffer gas or wall shifts, light shifts might be reduced by switching the laser diode and interrogating the microwave resonance while the light source is off.[7] In addition, since pumping by laser diodes is not dependent upon coincidence of hyperfine lines in different isotopes as conventional rubidium lamps are, cesium may be used in the vapor cell instead of rubidium. Cesium offers advantages of greater microwave resonance frequency, slower atoms (because of greater atomic mass) and natural isotopic purity.

Anticipated benefits to commercial cesium standards also include increased stability. Since atoms are converted to the desired hyperfine state instead of rejected by a state-selection magnet, more atoms contribute to the signal for a given beam intensity. Laser diodes can also be used for fluorescence detection of atoms, eliminating the second state-selection magnet, and relaxing requirements on alignment. Such a detection method should have good long-term stability, sensing nearly all atoms in the beam regardless of velocity and position. Using laser diodes for both state preparation and state detection results in a highly symmetric device in which the atomic beam direction can be rapidly reversed, or even maintained continuously in both directions.[8] This would permit correction for cavity phase shifts, which may limit long-term stability of commercial standards. Simplification of beam tube design could result in some weight savings as well.

The accuracy of laboratory primary frequency standards may be increased through the use of

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

laser diodes.[10] Elimination of state-selection magnets permits the extension of the C-field region to include both the optical pumping and the optical detection regions of a standard, which should prevent Majorana transitions among magnetic sublevels. These transitions may produce frequency offsets of unknown magnitude in present cesium primary standard designs.[9] A more symmetric beam reversal is possible with optical pumping methods, since all atoms may be included regardless of velocity and position within the beam. This should permit a better cancellation of cavity phase shifts. Angle tuning of the laser may result in velocity selection of atoms in the cesium beam, favoring slower velocity atoms, which would give a narrower microwave resonance. Finally, two lasers can be used to pump all of the cesium atoms into a single magnetic sublevel,[11] giving better control of Majorana transitions, freedom from frequency pulling by adjacent Zeeman levels, still greater numbers of usable atoms for a given beam intensity, and permitting operation at a lower C-field and a lower magnetic shielding coefficient than would otherwise be possible.

Laser diodes also introduce new difficulties into any atomic standard. They must be cooled to a highly regulated temperature, and the electronics required to perform this function, as well as provide a regulated injection current to the laser, is appreciable. The long-term reliability and useful lifetime of laser diodes is unknown. While manufacturers' projections of life range as great as 100 years, this does not include an estimate of wavelength stability. Single-mode, cw lasers are available today for as little as $100 each, but this price does not guarantee the exact wavelength required for optical pumping of a given atom. Finally, use of such a narrow spectrum source complicates the problem of light shifts in both vapor cells and atomic beams, and may introduce short-term noise through frequency and amplitude noise of the laser.

General Characteristics of Laser Diodes

A recent review of single-mode GaAlAs laser diodes is given by Botez.[12] A more extensive treatment is found in the books by Kressel and Butler[13] and by Thompson.[14] We have made measurements with transverse junction stripe (TJS) lasers,[15] channeled-substrate-planar (CSP) lasers,[16] and other stripe geometry, double-heterostructure lasers. The most promising results have been obtained with the first two types and consequently the discussion here will concentrate on them.

The very small active region in these devices (0.4μm x 2μm for the TJS laser, and 2μm x 10μm for the CSP laser) results in a diffraction-limited beam which must be collimated for most applications. We found that large numerical aperture microscope objectives gave the best collimation, although a number of measurements were made using single aspheric lenses. In either case, reflected light is a problem, as it can seriously affect the mode structure of the laser.

Both types operate in nearly single transverse and longitudinal modes, although the CSP type tends to stay longer in a given longitudinal mode as the injection current and temperature of the diode are changed. Longitudinal mode spacing is about 0.35 nm in both devices. Laser linewidth is about 150 MHz for the TJS laser (about 3 mW typical output), and less than 30 MHz for the CSP laser (about 10 mW typical output) as measured both with the Fabry-Perot spectrum analyzer, and with atomic resonance fluorescence curves. A detailed study of line broadening in CSP lasers has been made by Fleming,[17] who concluded that the spectral width was not in agreement with the Schawlow-Townes expression[18] based on quantum phase fluctuations.

Varying the amount of aluminum and dopants in the semiconductor alloy produces laser diodes of different wavelengths, ranging from about 780 nm to 905 nm in commercially available devices. Coarse tuning of the wavelength is achieved by changing the temperature (about 0.2 nm/°C) and fine tuning is accomplished through adjustment of the injection current (about 0.02 nm/mA).

Circuits have been designed and constructed at NBS to control the laser injection current to one part in 10^6, and the laser temperature to

about 1 millidegree centigrade for times in excess of one day. Further laser frequency stability has been obtained by locking the laser to an atomic absorption line for periods of days. This method has given long-term frequency stability of better than 100 kHz.

Although a given laser diode may operate near a desired wavelength at a reasonable value of current and temperature, it may not be possible to tune to an exact wavelength. As current and temperature are changed, the laser may hop to an adjacent cavity mode, leaving sizable gaps in the tunable spectrum. About half of the ten TJS diodes without package windows which we tested could be tuned to the Rb D_2 resonance at 780 nm. Only one of the ten TJS lasers with windows, which we studied, could be tuned exactly to that wavelength. It may be that the windows reflected enough light back to the laser to interfere with tuning. When a particular laser is tuned to the Rb resonance, the laser may change to a new cavity mode later for reasons which are not well understood. Even though manufacturers project lifetimes for laser diodes in excess of 100 years, this mode-hopping problem may limit their usefulness in atomic frequency standards. Measurements are in progress at the National Bureau of Standards (NBS) to determine how long these lasers can be locked to an atomic absorption line.

With typical operating currents of about 40 mA and 100 mA repectively, the TJS and CSP lasers themselves consume about 80 mW and 200 mW each. However, a system which includes temperature regulation, as well as current control, could require as much as 5 watts.

Stability of a Laser Pumped Rb Clock

The short-term stability of a conventional Rb clock is usually limited by the shot noise associated with the light detected by the photocell after the Rb vapor cell.[19] A typical value of signal to background of as little as 0.1% means that large improvements in short-term stability might be obtained by increasing the fraction of detected light which contributes to useful signal. As long as other noise sources do not become relatively large, the stability of the clock should improve as the square root of signal/background.

We have taken a commercial Rb clock and replaced the rf-excited rubidium lamp with a laser diode. Various sources of noise have been examined, and the short-term stability measured. Our most interesting results were obtained with a CSP-type laser operating at the D_2 transition (780 nm). Since the commercial Rb vapor cell contains both ^{85}Rb and ^{87}Rb, the best measurements were obtained for the F=1 transition in ^{87}Rb, which is well isolated from the ^{85}Rb lines. The output from the laser diode was roughly collimated to a 1 cm diameter beam, where the power density increased by about a factor of two from the perimeter to the center of the beam. The average laser power was varied from about 0.1 mW/cm^2 to 1.0 mW/cm^2 by inserting neutral density filters. The frequency of the laser was controlled both by temperature and injection current regulation, and by locking the laser to the center of the Rb absorption line.

Figure 1 shows the microwave resonance obtained by sweeping the commcercial clock's crystal oscillator over a small range, while pumping the cell with about 0.23 mW/cm^2 of D_2 light. While the sweep is somewhat nonlinear, it is clear that the linewidth of the resonance is approximately 1300 Hz, which gives a Q of only 5×10^6. It is believed that this large linewidth is due to broadening associated with the rather high laser power density. Figure 2 shows the dependence of microwave resonance linewidth on laser power density. The increase in linewidth due to an input laser light of 1 mW/cm^2 is about 1400 Hz. A rough estimate of the effective laser power in the Rb cell may be made from that linewidth. The change in microwave linewidth is given by

$$\delta(\Delta\nu_\mu) = \frac{1}{2\pi}\,\gamma_\ell\,\frac{\Delta\nu_\ell}{\Delta\nu_{Rb}} \qquad (1)$$

where γ_ℓ is the laser pumping rate on resonance, and $\Delta\nu_\ell/\Delta\nu_{Rb}$ is the ratio of the laser linewidth to the rubidium optical resonance linewidth in the

cell. The ratio of the laser pumping rate to the spontaneous emission rate γ_n is just

$$\frac{\gamma_\ell}{\gamma_n} = \frac{B_{21}\, u(E)}{1/\tau_n} = \frac{\lambda^3}{8\pi h}\, u(E) = 2.9 \times 10^{19}\, u(E) \quad (2)$$

where B_{21} is the Einstein B coefficient, τ_n is the natural decay time of the excited state (~ 30 ns), u(E) is the laser energy density per unit bandwidth, in joule-s/cm³, and λ is the wavelength of the optical transition. Combining this expression with the previous equation , and assuming a laser linewidth of ~ 30 MHz and a Rb optical linewidth of 1 GHz, gives a laser power density of

$$\rho = u(E)\cdot c\cdot \Delta\nu_\ell \quad (3)$$

$$= \frac{2\pi\, c\delta(\Delta\nu_\mu)\, \Delta\nu_{Rb}}{\gamma_n\, (2.9 \times 10^{19})} = 0.27 \text{ mW/cm}^2 \,. \quad (4)$$

Note that the final expression is independent of the actual laser linewidth, as long as $\gamma_n/2\pi \ll \Delta\nu_\ell \ll \Delta\nu_{Rb}$.

It is reasonable that the effective laser power in pumping the Rb should be less than the input laser power because of absorption and variation in microwave power through the cavity. The above result is therefore taken as support of this mechanism of line broadening.

The peak change in detected current across the microwave resonance was about 10^{-6} A. The average detected current was 10^{-5} A. This gives a maximum signal-to-noise in a one hertz bandwidth of

$$\frac{S}{N} \sim \frac{I_{sig}}{\sqrt{2e\, I_{bkg}}} = 5.6 \times 10^5 = 115 \text{ dB} \,. \quad (5)$$

Taking into account the relatively small microwave modulation amplitude, the expected signal to shot noise was greater than 100 dB in a one hertz bandwidth. This should result in an Allan variance stability of

$$\sigma(\tau) \sim \frac{0.2}{Q\,(S/N)}\, \tau^{-\frac{1}{2}} = 4 \times 10^{-13}\, \tau^{-\frac{1}{2}} \,. \quad (6)$$

Figure 3 gives the actual performance of the clock from 10^{-2} to 10 seconds, using a laser diode locked to the center of the ^{87}Rb, F=1 transition, as seen in the mixed isotope cell of the clock. At times much greater than 10 seconds, the clock frequency stability leveled out at the 10^{-12} level. The stability at one second is about 1.5×10^{-11}— comparable to the clock performance with the original Rb lamp installed, but considerably worse than the performance limited by shot noise should be.

The probable cause of the limit in short-term stability, and the 10^{-12} floor on stability at longer times, is frequency noise on the laser, mediated through the light shift.[20] The light shift in this particular device was measured by tuning the laser diode in steps of 100 MHz across the F=1 resonance, while recording the clock's output frequency. The light shift for an input laser power of 0.23 mW/cm² is given in figure 4. This curve gives a peak-to-peak shift on the dispersion curve of about 550 Hz, in rough agreement with the predictions of Mathur,[21] if one uses the reduction factor for light intensity obtained from the microwave resonance broadening calculations above. An accurate measurement of the light shift in rubidium would require thorough knowledge of the laser beam profile within the vapor cell as well as information about the microwave modes in the cavity, neither of which was available. It is probably more accurate to use the light shift measurements as an indication of the effective laser power than to estimate the light shift from measured laser power.

Bearing the above caveats, figure 5 gives the light shift dependence upon input light intensity in this particular clock measured at the laser frequency of maximum light shift. The curve is linear, with slope ~ 1.2 kHz/mW/cm². A similar curve was obtained for the F=2 transition, but with a slope minus one-half that of the F=1 curve. The magnitudes of the two slopes should be equal, but differ because of the presence of ^{85}Rb, which reduces the laser light intensity at the F=2 transition.

Similar light shift measurements have been

performed in cesium by Arditi and Picqué,[22] also using a laser diode, with results of the same magnitude.

The clock stability in figure 3 was obtained with a laser input power of 0.23 mW/cm². At this light level, the change in microwave frequency with laser frequency at the optical resonance peak is about -0.67 Hz/MHz, or a fractional frequency change of 10^{-10}/MHz. The long-term stability of the clock could thus be accounted for by a 10 kHz floor on the lock of the laser frequency to line center. (Which means the laser frequency is locked to a part in 10^5 of the resonance linewidth. A better lock might be obtained by using either saturated absorption or atomic beam techniques).

Figure 6 gives the frequency noise spectrum of the CSP laser diode. The TJS lasers tested gave a factor of three larger frequency noise, with a similar spectral dependence. These figures should be treated as an upper limit to the noise in these lasers. Since the cavity modes of the devices are very sensitive to reflected light, it is possible that a different testing procedure may give better results. However, Dandridge,[23] has obtained results with an interferometric testing method which agree within a factor of two with these measurements.

Our frequency noise curves were obtained by tuning the laser diode to the side of a Rb absorption line, and interpreting the intensity noise on the transmitted light in terms of frequency noise. The intensity noise observed, both away from the absorption line and on the peak of the resonance, was much smaller than that seen on the side of the line. The frequency noise at 127 Hz of ~ 100 kHz/$\sqrt{\text{Hz}}$ would limit the clock's frequency stability at one second to about 2×10^{-11}, in agreement with figure 3. The attack time on the servo loop which locked the laser diode to line center was not fast enough to significantly improve the stability of the laser at one second. In order to remove frequency noise as a limiting factor in short-term stability, a tighter laser lock is required. Also, if a lower temperature vapor cell is used, the laser intensity can be reduced with a proportional reduction in sensitivity to light shifts, without los. of signal.

No attempt is planned at this time to reduce the effect of laser frequency noise in this particular system because the crystal oscillator itself reaches a minimum stability of ~ 10^{-11} at 0.3 second. Consequently, very little improvement could be expected over the results of figure 3 without other modifications of the clock.

Another source of noise in the laser-pumped Rb clock is intensity noise in the laser light. Figure 7 gives the CSP laser diode intensity noise spectrum; this spectrum was limited by the noise in the detection electronics and must be considered an upper limit only. Other measurements in CSP lasers have been made by Dandridge[24] who obtained fractional intensity noise as small as -145 dB/$\sqrt{\text{Hz}}$ at 100 Hz.

Using our limit of -120 dB/$\sqrt{\text{Hz}}$ at 127 Hz, and a signal-to-background ratio of 2%, the limit to stability due to laser intensity noise in this system would be about $\sigma(\tau) \sim 2 \times 10^{-12}/\sqrt{\tau}$. Using Dandridge's results, with a signal-to-background of 10%, and a microwave Q of 10^7 gives $\sigma(\tau) \sim 1 \times 10^{-14}/\sqrt{\tau}$. In this case, other sources of noise will limit the clock stability.

These results suggest that using a laser diode pump source with proper adjustment of Rb clock parameters, it may be possible to obtain stabilities of a few parts in 10^{13} at one second. A clock with such excellent short-term stability might find application as a local oscillator for a stored ion standard.[1] The clock's long-term stability would still be limited by the problems which plague conventional Rb clocks[7]—light shift changes due to light source intensity changes (which could be minimized either by adequate laser locking or by chopping the light[7]), buffer gas or wall shifts, and changes in the microwave spectrum. However, since laser pumping can be used in a system with higher Q than that of conventional standards, some relief from even these problems is expected.

Laser Pumped Atomic Beam Frequency Standards

Laser diodes may be useful as both a pumping source and a state detector in atomic beam fre-

quency standards. Arditi and Picqué have demonstrated the use of a laser diode for these two purposes in a Ramsey-structure cesium beam clock. The work at NBS has proceeded along somewhat different lines, leading toward implementation of two or more lasers in a single clock, as is described below. While the preliminary Cs work at NBS has been done with dye lasers, the above results for laser diodes may be used in order to estimate noise in planned atomic beam devices.

Figure 8 is a sketch of a beam apparatus which has been used to measure optical pumping, detection, and microwave transitions in cesium. In the early experiments, one krypton ion laser pumped dye laser (HITC dye @ 852 nm) was used to pump the D_2 transition of atomic cesium. A second laser, pumped by the same krypton ion laser, was used to monitor the state of atoms which passed to the second region of the beam tube. The fluorescence signal in this region is an indication of the number of cesium atoms in a particular ground-state hyperfine level (figure 9). For example, figure 10 is a dispersion curve of the downstream cesium beam fluorescence detected by a frequency modulated laser swept across the $F=4 \rightarrow F'=5,4,3$ lines. Total data acquisition for this curve was one minute. The fact that the $F=4 \rightarrow F'=5$ transition is much larger than the other two lines is due to optical pumping on the two lesser transitions. The selection rule $\Delta F=\pm 1,0$ prevents atoms from entering the F=3 ground state, which results in fluorescence of many photon per atoms. The $F=4 \rightarrow F'=4,3$ transitions move atoms to the F=3 state, where they no longer contribute to the fluorescence signal.

This detection laser can be left on the peak of the $F=4 \rightarrow F'=5$ dispersion curve, while a second laser is tuned to a pumping transition, the fluorescence signal downstream decreases. This is demonstrated in figure 11, where the $F=4 \rightarrow F'=4$ transition is pumped. Use of a cycling transition, such as $F=4 \rightarrow F'=5$, for fluorescence detection can result in 100% quantum efficiency, even when the light collection efficiency is small. The disadvantage of this method is that intensity noise in the laser appears in the detected signal. If the detection laser is tuned to a pumping transition, such as $F=4 \rightarrow F'=4$, the quantum efficiency may decrease, but every atom contributes an equal number of photons to the fluorescence signal, regardless of the atom's velocity, or of the laser's intensity (above a minimum value required for complete pumping).

If a microwave resonance between F=4 and F=3 of the ground state is now induced between the two laser regions, the fluorescence signal in the second region will be partially restored. This signal as a function of microwave frequency is shown for a Rabi-type resonance structure in figure 12 for two values of C-field. The peak which did not change frequency with change in C-field is the $m_F=0 \rightarrow m_F=0$ clock transition. The adjacent lines are π transitions present because the C-field is not precisely parallel to the magnetic field vector, $\vec{H}$, of the microwaves. As the C-field is increased, its direction becomes more nearly parallel to $\vec{H}$, enforcing the selection rule $\Delta m_F=0$. The outmost microwave resonances are the $m_F=\pm 1 \rightarrow m_F=\pm 1$ transitions.

In a cesium standard with a microwave Q of 10^7 and an effective beam current of 10 pA, the expected stability would be about $4 \times 10^{-12}\ \tau^{-\frac{1}{2}}$, if the limiting noise is shot noise associated with the atomic beam. If a cycling transition is used for fluorescence detection, the intensity noise in the laser light must be less than about $10^{-5}/\sqrt{Hz}$ in order that shot noise dominate. As seen above, this condition holds with the CSP-type laser diodes.

A more difficult problem arises with detector noise. Typical best performance at room temperature for a silicon PIN diode is a NEP ~ 10^{-14} watt/$\sqrt{Hz}$. At one photon/atom, the 10 pA beam current used above would emit shot noise of about 2×10^{-15} watt into 4π steradians. The use of a cycling transition with perhaps 10^3 photons/atom would bring this noise level above the detector noise level without requiring great collection efficiency or the use of photomultiplier tubes.

A saturating laser intensity (~ 10 mW/cm^2) over a 1 cm length of the atomic beam should produce such a large number of photons. If the

collection efficiency is increased, the interaction region may be correspondingly reduced in length.

An additional feature of the cycling fluorescence is that is tends to weight slower atoms more heavily, since they remain within the laser beam longer. This should result in a somewhat higher microwave Q for the system.

It is anticipated that frequency noise in the laser diode light will not introduce additional noise into an atomic beam standard, since the microwave region is separate from the pumping and detection regions. However, a systematic offset will be produced due to the light shift induced by fluorescent radiation entering the microwave region.[26] Using a value of 20 Hz $cm^2/\mu W$ for the light shift in Cs as an upper limit,[22] and 2×10^{-8} watt/4π steradian as the fluorescence emitted from a cycling detection region 1 cm from the microwave region, the resultant light shift would be about 3×10^{-2} Hz, or a fractional shift of 3×10^{-12}. Since the intensity of the laser light should not change over a period of a year, this probably does not present a problem for commercial applications, but it is a serious consideration for primary standards, where an accuracy of 10^{-14} is sought. A detailed calculation is in progress at NBS.

Further increase in both signal-to-noise and immunity from certain systematics should be obtained by pumping the cesium atoms into a single magnetic sublevel. One way of accomplishing this rearrangement of energy level populations is through the use of two lasers.[11] One laser is tuned to the F=3 → F'=4 transition (figure 9), pumping atoms into the F=4 hyperfine level. A second laser, with electric field polarization parallel to a weak magnetic field in the same pumping region, is adjusted to the F=4 → F'=4 line. In this case, the atoms from the F=4 level are pumped into F=3 level, with the exception of the F=4, m_F=0 magnetic sublevel, which remains unaffected by either laser. This may be seen by considering the Clebsch-Gordon coefficient which connects the F=4, m_F=0 level with the excited state:

$$\langle j_1 j_2 m_1 m_2 | JM \rangle = \langle 4100 | 40 \rangle = 0, \qquad (7)$$

where the notation used is the same as that of Messiah.[27] In this case, j_1=F, m_1=m_F, J=F', and M=$m_{F'}$. The photon's quantum numbers are given by j_2=1 and m_2=0. Ultimately, all atoms are pumped into the F=4, m_F=0 sublevel. This results in a factor of sixteen increase in signal over conventional state selection, and removes adjacent transitions from the microwave spectrum. A similar arrangement may be used to pump atoms into the F=3, m_F=0 sublevel, since Eq. 7 holds for any two levels with j_1=J and m_1=m_2=M=0.

Experiments are in progress at NBS to demonstrate this new pumping technique. At present, the individual selection rules involved for the method have been demonstrated using dye lasers. Figure 13 illustrates the most critical rule. In this example, the detection region laser is tuned to the F=4 → F'=5 cycling transition. The upstream, pumping laser is tuned to the F=4 → F'=4 line (852 nm). A microwave spectrum is displayed as in figure 12. In curve A, the electric field of the laser radiation is parallel to the static magnetic field in the pumping region. Some small signal is present at the m_F=0 → m_F=0 microwave frequency because the optical pumping from the other magnetic sublevels changes the population of the F=3, m_F=0 level. In curve B, however, the polarization of the laser is rotated by about 6° from the magnetic field orientation. This removes the m_F=0 selection rule, and results in a marked increase in the population of the F=3, m_F=0 sublevel relative to the F=4, m_F=0 level. This produces a larger microwave transition signal at m_F=0 → m_F=0. The next effort to actually pump all of the cesium atoms into a single magnetic sublevel will be made with diode lasers.

Conclusions

The properties of some commercial laser diodes have been studied. The frequency and intensity noise characteristics of these lasers place some restrictions on the short-term stabilty of both vapor cell clocks and atomic beam standards. These difficulties probably can be overcome and laser diodes may be used to improve the performance of atomic frequency standards. The

frequency stability and light shift offsets in a laser-diode-pumped rubidium clock have been measured, and the short-term stability compared to that of a conventional Rb clock. A preliminary investigation has been made of optical pumping in a cesium beam using dye lasers and the selection rules necessary to pump all of the cesium atoms into a single magnetic sublevel have been verified.

Acknowledgments

The authors acknowledge the support of the Frequency and Time Standards Group, Time and Frequency Division, in this work. Especially valuable were the contributions of A. Derbyshire and H. Hemmati. This research was funded in part by the Department of the Air Force. We are grateful to A. Aiki for providing CSP laser diodes.

References

1. D. J. Wineland, W. M. Itano, J. C. Bergquist, and F. L. Walls, 35th Annual Symp. on Freq. Control (1981).

2. M. Jardino, M. Desaintfuscien, R. Barillet, J. Viennet, P. Petit, and C. Audoin, Proc. 34th Annual Symp. on Freq. Control, U.S. Electronics Command, Ft. Monmouth, NJ (1980).

3. Efratom California, Inc., Irvine, CA.

4. Frequency and Time Systems, Inc., Danvers, MA.

5. D. J. Wineland, D. W. Allan, D. J. Glaze, H. W. Hellwig, and S. Jarvis, Jr., IEEE Trans. Instr. and Measurement IM-25, 453-458 (1976).

6. F. L. Walls, measurement of cesium primary standards versus hydrogen masers.

7. T. C. English, E. Jechart, and T. M. Kwon, Proc. 10th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 147-168 (1978).

8. M. Feldman, J. C. Bergquist, L. L. Lewis, and F. L. Walls, 35th Annual Symp. on Freq. Control (1981).

9. D. W. Allan, H. Hellwig, S. Jarvis, Jr., D. A. Howe, and R. M. Garvey, Proc. 31st Annual Symp. on Freq. Control, U.S. Army Electronics Command, Ft. Monmouth, NJ, 555 (1977); G. Becker, IEEE Trans. on Instrumentation and Measurement IM-27, 319 (1978).

10. L. Lewis, F. L. Walls, D. J. Glaze, and D. J. Wineland, to be published.

11. This method was called to our attention by L. Cutler. See also H. J. Gerritsen and G. Nienhuis, Appl. Phys. Lett 26, 347 (1975).

12. D. Botez, J. Opt. Comm. 1, 42 (1980).

13. H. Kressel and J. K. Butler, Semiconductor Lasers and Heterostructure LED's (Academic Press, 1977).

14. G. H. B. Thompson, Physics of Semiconductor Laser Devices (John Wiley and Sons, Somerset, NJ, 1980).

15. H. Namizaki, IEEE J. Quantum Electron. QE-11, 427 (1975); W. Susaki, T. Tanaka, H. Kan, and Makoto Ishii, IEEE J. Quantum Electron. QE-13, 587 (1977).

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

17. M. W. Fleming and A. Mooradian, Appl. Phys. Lett. 38, 511 (1981).

18. Melvin Lax, in Physics of Quantum Electronics, edited by P. L. Kelly, B. Lax, P. E. Tannenwald (McGraw-Hill, New York, 1966) p. 735.

19. C. Audoin and J. Vanier, J. Phys. E: Scient. Inst. 9, 697 (1976).

20. J. P. Barrat and C. Cohen-Tannoudji, J. Phys. (Paris) 22, 329 (1961); J. Phys. (Paris) 22, 443 (1961); M. Arditi and T. R. Carver, Phys. Rev. 124, 800 (1961).

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

22. M. Arditi and J.-L. Picqué, J. Phys. B: Atom. Molec. Phys. 8, L331-L335 (1975).

23. A. Dandridge, A. B. Tveten, R. O. Miles, D. A. Jackson, and T. G. Giallorenzi, Appl. Phys. Lett. 38, 77 (1981).

24. A. Dandridge, A. B. Tveten, R. O. Miles, and T. G. Giallorenzi, Appl. Phys. Lett. 37, 526 (1980).

25. M. Arditi and J.-L. Picqué, J. Phys. Lett. (Paris) 41, L379-L381 (1980).

26. We are grateful to A. Brillet for emphasizing the seriousness of this problem in an optically-pumped cesium primary standard.

27. A. Messiah, Quantum Mechanics (John Wiley and Sons, Inc., New York, 1958).

TABLE 1. Performance of Atomic Frequency Standards

	Short-Term Stability	Long-Term Stability	Accuracy
Commercial Rb Clock (3)	$2 \times 10^{-11}/\sqrt{\tau}$	3×10^{-11}/month	---
Commercial Cs Clock (4)	$10^{-11}/\sqrt{\tau}$	3×10^{-12} (2×10^{-13} @ 1 day)	---
Laboratory	$5 \times 10^{-13}/\sqrt{\tau}$ (5)	10^{-14} @ 1 week (6)	10^{-13} (5)

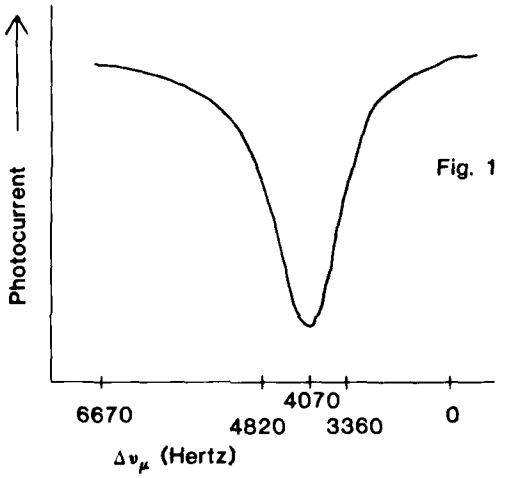


Figure 1. Rubidium $m_F=0 \rightarrow m_F=0$ microwave resonance in a laser pumped cell. The laser is tuned to the F=1, D_2 transition in ^{87}Rb.

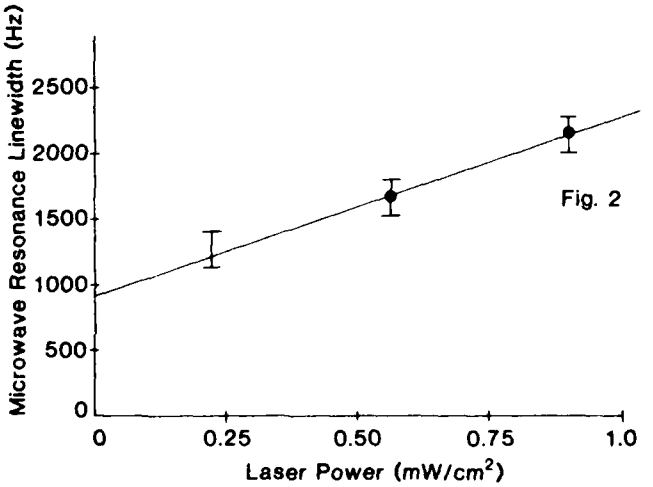


Figure 2. Rubidium microwave resonance linewidth as a function of input laser power density.

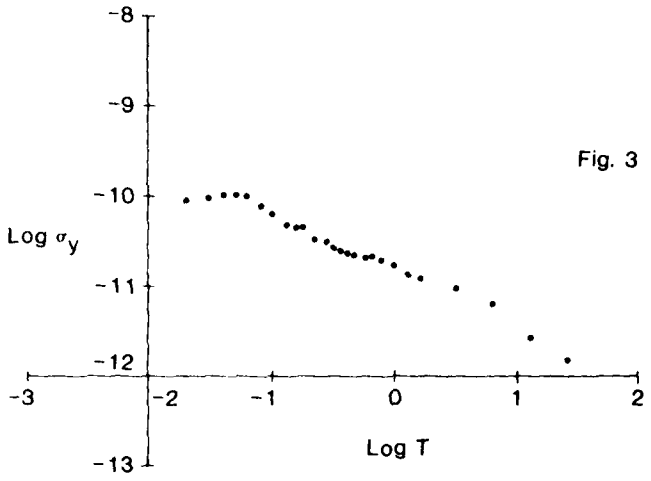


Fig. 3

Figure 3. Allan variance stability curve for a laser-pumped rubidium clock. The laser is locked to the F=1, D_2 transition in ^{87}Rb.

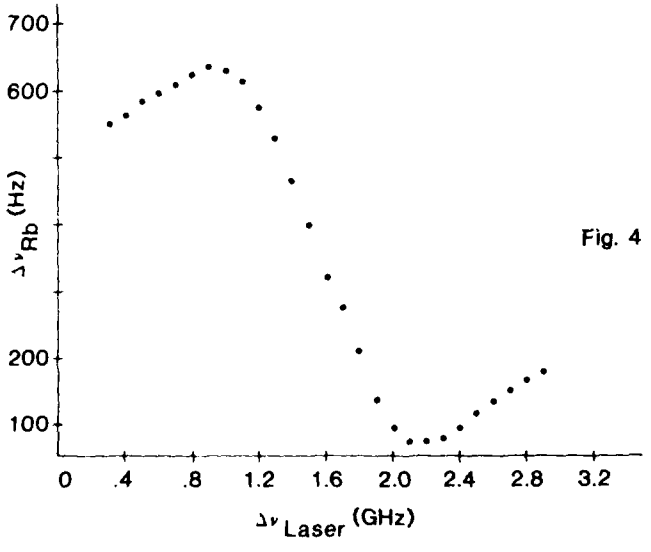


Fig. 4

Figure 4. Light shift in ^{87}Rb clock. Input laser light of 0.28 mW/cm^2 is tuned across the F=1, D_2 transition.

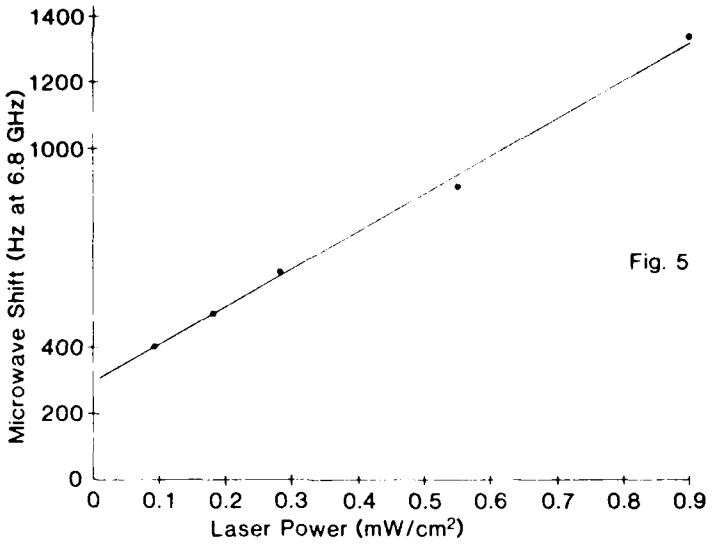


Figure 5. Maximum light shift in a ^{87}Rb clock as a function of laser power density. The laser is tuned to the low frequency dispersion peak of the F=1, D_2 transition.

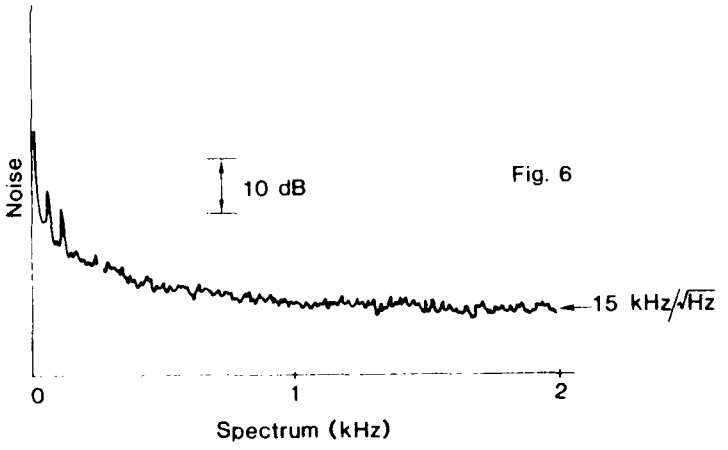


Figure 6. Frequency noise spectrum of a CSP laser diode. The curve was obtained by tuning the laser to the side of a Rb absorption line.

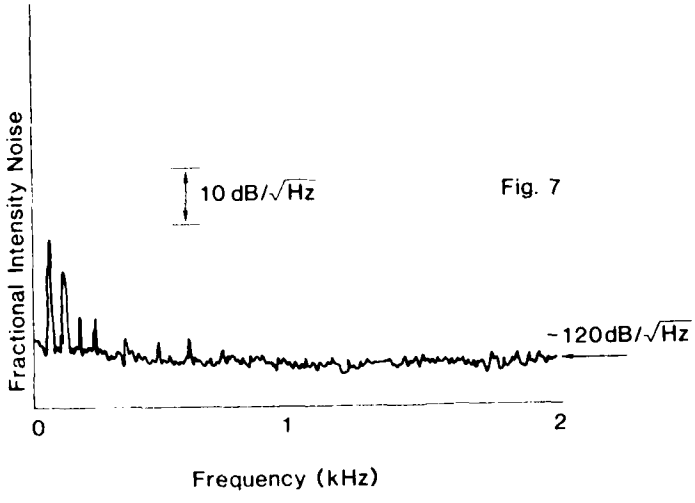


Fig. 7

Figure 7. Intensity noise spectrum of a CSP laser diode. The curve is only slightly greater than the noise level of the measuring system and must be regarded as an upper limit only.

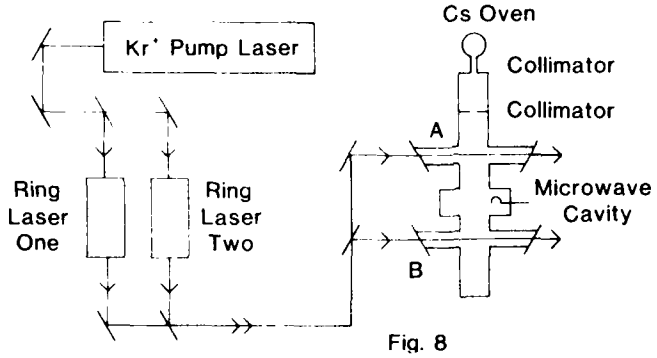


Fig. 8

Figure 8. Laser pumped/detected atomic beam.

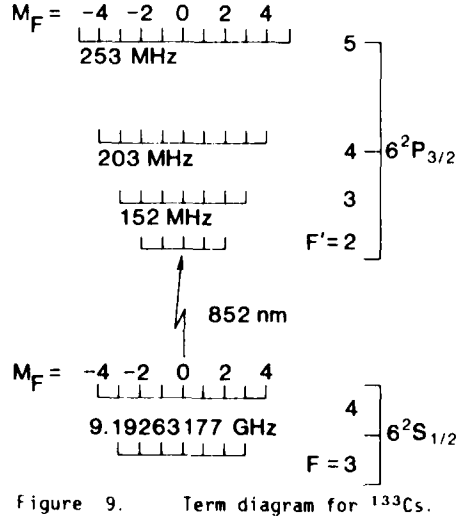
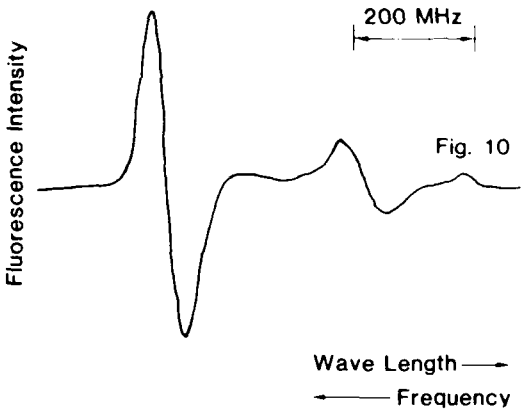
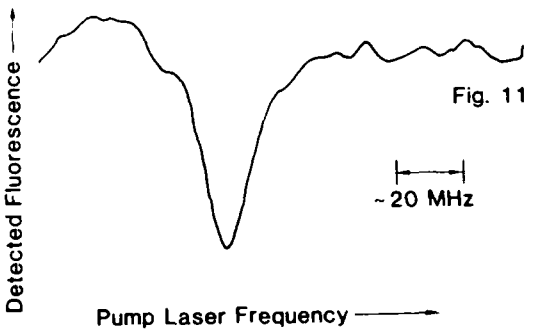


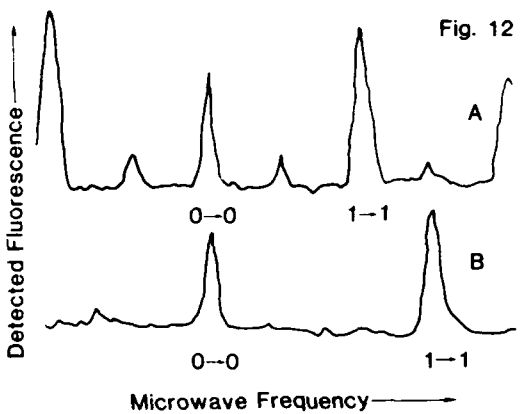
Figure 9. Term diagram for ^{133}Cs.



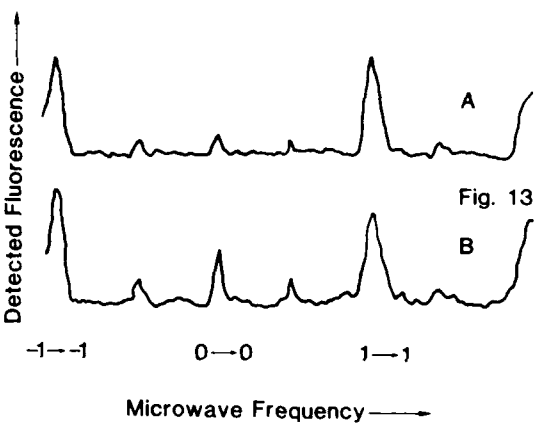
Fluorescence dispersion curve in cesium atomic beam. From left to right, the peaks are F=4 → F'=5,4,3.



Optical pumping upstream of the F=4 → F'=4, D_2 cesium line with fluorescence detection downstream of the *F=4 → F'=5 line.*



Microwave Rabi spectra for laser pumped cesium beam. The C-field is greater in Curve B.



Microwave Rabi spectra for laser pumped cesium beam. The F=4 → F'=4 pump laser electric polarization is parallel to the C-field in Curve A and rotated by 6 degrees from parallel in Curve B.

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

PRELIMINARY INVESTIGATION OF A NEW OPTICALLY PUMPED ATOMIC RUBIDIUM STANDARD

Mark Feldman, J. C. Bergquist, L L. Lewis,
and F. L. Walls

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

Summary

We are studying two types of optically pumped glass cells which do not contain a buffer gas and have no wall coating in which beam-like properties are exploited. The first device is a sealed glass tube of about 1 cm diameter and 20 cm length. A small amount of ^{87}Rb metal is localized at one end by temperature gradients which also control the vapor pressure. The cell has the properties of a broad atomic beam for the transport of optically pumped atoms from one end to the other with collimation given by the aspect ratio of the tube. At each end the Rb "beam" is crossed by a laser. In each interaction region, the laser optically pumps atoms into one of the 5 $^2S_{\frac{1}{2}}$ hyperfine levels, as well as detects population changes between the hyperfine levels. In a second device, graphite inserts are included in the glass tube. The graphite strongly getters Rb, thereby providing collimation and significantly reducing scattering of laser light from background Rb atoms. A TE_{011} microwave cavity is positioned between the two laser interaction regions.

In the broad beam device we have observed the transport of optically pumped atoms between the ends with a S/N ratio of 300:1 in 1 s. A tunable dye laser was used for the pumping and signal detection. In the collimated beam device, we have observed microwave transitions using a diode laser for pumping and fluorescence detection.

Key words. Atomic Frequency Standard; Laser Frequency Standard; Optical Pumping; Rubidium Beam; Rubidium Cell; Rubidium Frequency Standard.

Introduction

Portable atomic rubidium standards are in widespread use because of their relatively low cost, good short term stability, low power requirements, and physical package (small size, light weight). However, presently available devices are limited in their long term accuracy and reproducibility by two large systematics: light shifts and buffer gas pressure shifts. We are in the preliminary stages of investigating two different physical packages for Rb (or Cs) which show promise of significant reduction in these offsets, while still maintaining the advantages of a small atomic standard. The initial goal is a device with 10^{-10} to 10^{-11} accuracy and rapid turn-on capability compared to available standards.

The two packages are described separately below and their potential for portable standards are discussed. Briefly, these devices have the advantages for accuracy and long-term stability of an atomic beam: no wall coatings, no buffer gases, and an optical interaction region which is separate from the microwave region. These general features are illustrated in fig. 1. However, they differ significantly from commercial portable beam standards in their simplicity of construction and use of optical pumping to eliminate state selection magnets. For practical implementation, both schemes do depend on the availability of reliable diode lasers. So far, we have performed all experiments using Rb, though some advantages of Cs are pointed out below.

^{87}Rb Beam/Cell

General Description. Fig. 2a illustrates a simple glass beam/cell[1] which we have investigated. The atomic source consists of ^{87}Rb localized at one end by temperature gradients and heated to 40-50 °C. The atomic vapor density at this temperature presents about one optical absorption length to the intersecting laser beams which are tuned to the 5 $^2S_{\frac{1}{2}}$ F=2 → 5 $^2P_{3/2}$ F=1,2,3 Doppler broadened D2 resonance (see fig. 3). At this operating temperature and pressure, the mean free path for ground-state spin-exchange collisions remains much greater than the cell dimensions, so that Rb atom-atom collisions are unimportant. Thus, attenuation of the laser light and not the atomic

mean free path length determines the highest useful temperature in the beam/cell.

If it is assumed that a wall collision effectively destroys any difference between ground-state populations, we may then think of this device as producing a broad, large source area atomic beam for the transport of optically pumped atoms from one optical region to the other. The beam's effective collimation is given approximately by the aspect ratio, a/r, of the tube, where a is the tube radius, and r is the separation between optical regions.

It should be noted that if a single laser is used for both interaction regions and significant net mass transport of Rb does not take place from one end of the tube to the other, then the device is also inherently symmetrical: each optical interaction region may be considered to be both a pump and a detection region. In this case, we have dual opposed atomic beams which may be used to cancel out phase shifts associated with the microwave cavity. For clarity in the discussion which follows, however, one laser will be referred to as a pump and the other as a detector.

With reference to fig. 2b, we now consider in detail the transport of pumped atoms between the optical interaction regions in terms of atomic densities. If the laser beam is expanded to cover the entire cross section of the cell, it will interact with the atoms in a disc-shaped region of radius a. Then from simple geometrical considerations,[2] the density of atoms at r which originated in the disc and have not suffered a wall collision is given by

$$\rho_o(r) = \left(\frac{a}{2r}\right)^2 n \quad , \tag{1}$$

where n = background atomic density. For our experimental situation, a = 0.5 cm, r = 8 cm, and $\rho_o(r) = 10^{-3}n$.

In general, not all of the atoms contributing to $\rho_o(r)$ will be optically pumped if the laser spectral width $\Delta\nu_\ell$ is narrower than the transverse beam Doppler width $\Delta\nu_D'$. In this case, $\Delta\nu_D' \sim \Delta\nu_D (2a/r) \sim 75$ MHz, where $\Delta\nu_D$ is the full Doppler width of ~ 600 MHz. In addition, the number of atoms pumped from a hyperfine level depends upon the number of magnetic substates in that manifold. This may be expressed in terms of a weighting factor, g_F, where $g_1 = 3/8$ for F = 1 and $g_2 = 5/8$ for F = 2 of ^{87}Rb. Thus, for $\Delta\nu_\ell < \Delta\nu_D'$, the density of atoms at r which were pumped into a single hyperfine manifold at the originating disc is

$$\rho_p(r) = g_F \frac{\Delta\nu_\ell}{\Delta\nu_D'} \left(\frac{a}{2r}\right)^2 n \quad . \tag{2}$$

We have assumed that the product of laser intensity and optical interaction time is sufficient to pump all atoms absorbing within $\Delta\nu_\ell$. If $\Delta\nu_\ell \gtrsim \Delta\nu_D'$, the linewidth factor becomes unity for an appropriately tuned pump laser.

The density of background absorbers ρ_B is simply determined by the ratio of $\Delta\nu_\ell$ to the full Doppler width $\Delta\nu_D$ (for $\Delta\nu_\ell < \Delta\nu_D$) times the weighting factor g_F of the ground state level involved in the optical transition:

$$\rho_B = g_F \frac{\Delta\nu_\ell}{\Delta\nu_D} n \quad . \tag{3}$$

The ratio ρ_p/ρ_B is thus independent of $\Delta\nu_\ell$:

$$\frac{\rho_p}{\rho_B} = \frac{\Delta\nu_D}{\Delta\nu_D'} \left(\frac{a}{2r}\right)^2 = \frac{1}{8} \frac{a}{r} \tag{4}$$

We next consider the effect of saturating the microwave clock transition ($M_F = 0 \rightarrow M_F = 0$) in the region between the two lasers. For both laser beams at the same frequency, this will increase the absorption in the detection region by increasing the density of atoms in one of the $FM_F = 0$ states. For ^{87}Rb the change in density which contributes to this absorption is given by

$$\rho_s(r) = \frac{1}{16 g_F} \rho_p + \frac{1}{16(1-g_F)} \rho_p = \frac{1}{16 g_F (1-g_F)} \rho_p \tag{5}$$

where we have made the approximation that the pumped atoms are distributed equally among the receiving hyperfine magnetic sublevels. (For example, if we pump out of F=2, then $g_F = {}^5/_8$ and the F=1 populations are increased.)

The ratio of signal density to background density is then given by

$$\frac{\rho_s}{\rho_B} = \frac{1}{128} \frac{a}{r} \frac{1}{g_f(1-g_f)} \qquad (6)$$

Under our experimental conditions, $\rho_s/\rho_B \sim 2 \times 10^{-3}$

Signal. For a simple linear absorption model (neglecting saturation effects), the transmitted light intensity will be

$$I_t = I_0 e^{-\sigma(\rho_B - \rho_p)\ell} \qquad (7)$$

$$= I_0 e^{-\sigma\rho_B\ell}(1 + \sigma\rho_p\ell), \qquad (8)$$

where σ is the optical absorption cross section and ℓ is the path length through the vapor. Microwave modulation with phase sensitive detection of the absorption signal then results in a signal

$$I_{PSD} \sim I_0 e^{-\sigma\rho_B\ell} \rho_s \sigma\ell \qquad (9)$$

$$\sim I_0 e^{-1} \frac{\rho_s}{\rho_B} \qquad (10)$$

$$\sim 7 \times 10^{-4} I_0 , \qquad (11)$$

where we have chosen $P_B\sigma\ell = 1$ (one optical depth), and the same experimental conditions as used above are invoked. The approximate expressions obtained above are valid for laser powers below saturation, but they must be modified if I_0 approaches the saturation intensity[3].

Noise. Below, we identify several sources of noise and briefly discuss some aspects of absorption and fluorescence detection with regard to these noise terms.

(1) Laser intensity noise. For the intensity (and frequency) stabilized dye laser used in these experiments, fractional intensity fluctuations have been reduced to less than $5 \times 10^{-7}/\sqrt{Hz}$ beyond 500 Hz. For the diode lasers which we have used, fractional intensity noise on a single mode (these "single mode" diodes generally exhibit multimode operation, with one strong mode and a few weak auxiliary modes) ranged from $10^{-5}/\sqrt{Hz}$ to less than $10^{-6}/\sqrt{Hz}$. The photon shot noise for laser powers in the range of 10^{-5} to 10^{-4} watts at 780 nm is 0.5 to $1.5 \times 10^{-7}/\sqrt{Hz}$. These power levels are typical in these experiments.

(2) Laser frequency noise. The Doppler broadened absorption profile converts frequency noise to amplitude noise. Near line center, frequency to amplitude noise conversion is reduced. To a good approximation, we can represent the Doppler broadened absorption profile in Gaussian form,

$$P(\omega(t)) \propto P_0 \exp[(\omega_L(t) - \omega_0)/\Delta\omega_0]^2 ,$$

where $\omega_L(t)$ is the instantaneous laser frequency (in angular units), ω_0 is the line center, and $\Delta\omega_0 = \frac{\pi}{\sqrt{\ell n 2}} \Delta\nu_D$. If we write

$$\omega_L(t) = \omega_0 + \varepsilon_1 + \varepsilon_2(t)$$

where ε_1 is a fixed frequency offset from line center and $\varepsilon_2(t)$ is a time dependent change in the laser frequency and if

$$\varepsilon_1, \varepsilon_2(t) \ll \Delta\omega_0 ,$$

then the fractional frequency to amplitude noise term is given by

$$\frac{\varepsilon_2^2(t) + 2\varepsilon_2(t)\varepsilon_1}{(\Delta\omega_0)^2}$$

Empirically, we can expect in a 1 Hz bandwidth that

$$\varepsilon_1 \cong 2\pi \times 75 \text{ kHz}$$

$$\varepsilon_2(t) \cong 2\pi \times 100 \text{ kHz}$$

and $$\Delta\omega_o \cong \frac{\pi}{\sqrt{\ell n 2}} \times 560 \text{ MHz}$$

Thus, the fractional intensity fluctuations due to FM to AM conversion are less than $2 \times 10^{-7}/\sqrt{\text{Hz}}$. For diode lasers we have measured ~ 0.1 MHz/$\sqrt{\text{Hz}}$ at 500 Hz which also gives a fractional intensity of $\leq 2 \times 10^{-7}/\sqrt{\text{Hz}}$. (For $\Delta\nu_D$ = 75 MHz, we obtain $1.2 \times 10^{-5}/\sqrt{\text{Hz}}$.)

(3) Detector noise. Solid state photodiodes are available with dark currents of $\cong 10^{-14}$ A and noise equivalent powers (NEP) of approximately 1×10^{-14} W/$\sqrt{\text{Hz}}$. This value is negligible for the absorption experiments but may dominate the noise for weak fluorescence signals.

(4) Background absorption noise. The atomic absorption process introduces another noise contribution associated with the number of atoms which have made an optical transition. If we let P_A represent the power absorbed by the background atoms out of the incident laser beam, then this noise contribution is proportional to $\sqrt{P_A}/h\nu/\sqrt{\text{Hz}}$, where $h\nu$ is the optical quantum of energy. For densities which give one optical absorption depth, $P_A \cong 0.6\ P_{\text{incident}}$, and this contribution is of the same order as the laser shot noise contribution.

(5) Signal noise. The signal noise term, which is proportional to the square root of the number of atoms which have made the microwave clock transition, is the fundamentally limiting noise term. In practice, one is usually limited in signal-to-noise ratio by one of the previously listed noise sources, and not by the signal shot noise term.

In absorption and fluorescence experiments, all of the noise terms are present and they must be individually examined for each experimental situation. In the collimated and well-gettered Rb beam tube of our second design, the background noise is virtually eliminated. Furthermore, fluorescence detection in this scheme has an additional advantage: if the laser intensity-interaction time product is sufficient to optically pump all atoms then, in principle, it is possible to eliminate the noise due to laser intensity fluctuations for times longer than the average transit time through the laser beam.

Experiment. Preliminary to the application of microwaves, we have observed the transport of optically pumped atoms between the interaction regions in an absorption experiment. An intensity stabilized, single-mode, tunable jet stream dye laser (5 MHz spectral linewidth, intensity noise $\cong 10^{-7}/\sqrt{\text{Hz}}$ beyond 500 Hz, power attenuated to 50 microwatts) was used for pumping and downstream detection. By chopping the pump beam either mechanically or with an electro-optic crystal and synchronously detecting the downstream absorption, we observed a maximum signal-to-noise (S/N) ratio of $\cong$ 300:1 (τ = 1 s). This signal is the change in the detected intensity synchronous with the chopped pump beam. It is due to the change in absorption of the atomic beam when the sublevel populations in one of the lower F levels are redistributed.

To compare with our description, we expect that without microwaves

$$I_{PSD} = I_o e^{-1} \frac{\rho}{\rho_B} \cong 3 \times 10^{-3}\ I_o$$

The largest noise contribution arose from laser intensity fluctuations at the 5×10^{-7} level. With this level of noise, we expect a S/N of about 5800:1 in 1 s, roughly a factor of 20 better than the 300:1 which we have observed. Stray light from window scattering precluded the possibility of fluorescence detection. We note that a frequency stability of 1×10^{-10} in 1 s. requires a S/N ratio of about 600 at 1 s on the microwave resonance.

Standards Potential. Additional work is required to evaluate experimentally the maximum S/N of this device. We now have a diode laser with spectral width of 150 MHz which should increase signal because the factor $\Delta\nu_\ell/\Delta\nu_D'$ will be approximately

unity. Also, fluorescence detection (easily possible with microwave modulation as opposed to light modulation) will give a smaller signal, but should also decrease the noise, and we expect an improved signal-to-noise ratio.

One systematic error of this device, which will eventually need to be investigated, is the light shift produced by fluorescent light originating in the pumping and detection regions. We note that because of the symmetry of the beam /cell, light from one region will be blue shifted while light from the other region will be red shifted, so that the atom will be subject to a symmetric frequency distribution. Furthermore, since the cell is optically thick, the intensity scattered into the microwave region will be somewhat reduced by multiple scatterings (this will also contribute to depumping atoms in the beam).

Rb Beam Tube

General Description. Fig. 4 illustrates a glass beam tube. Rb (natural isotopic abundance) is localized in one end by temperature gradients. However, a glass capillary and graphite inserts provide a collimation ratio of about 10:1 in order to form a directed atomic beam of Rb. The present tube is actively pumped by a small ion appendage pump, though it is anticipated that the tube will operate in a sealed -off mode due to the excellent gettering properties of the graphite. The absence of any appreciable density of background Rb atoms makes this device much more like a simplified, conventional atomic beam than the previously discussed beam/cell.

It is expected that one gram of natural Rb in the source will last more than a year when operated at 90 °C, providing an atomic beam flux of 3×10^{10} ^{87}Rb atoms/s into a 3 mm^2 area at the detector. Again, the configuration can be made symmetric by using two sources to provide opposing beams.

Signal. To estimate the microwave signal, we must know what population difference can be obtained between the clock transition sublevels by laser pumping. The solution of the optical pumping equations for a single laser is outlined in the Appendix, with results presented in figs. (5), (6), and (7). The calculation assumes that both the laser width and residual Doppler width of the atomic beam are less than the excited state hyperfine splittings, so that absorption takes place on a single F transition. It is clear that the $F = 2 \rightarrow F' = 2$ line pumped by σ-polarized light maximizes the population difference, δ_{oo}, between the $F = 1$, $m_F = 0$ and $F = 2$, $m_F = 0$ levels. (The same line pumped with π light results in $\delta_{oo} = 0$). If the optimum δ_{oo} of 0.4 is attained, we expect the microwaves to induce a 20% change in the total beam fluorescence in the detection region.

We will use this figure to estimate the S/N though the actual change in detected fluorescence may be less when the angular distribution of the light emitted during the detection cycle is taken into account.

For a beam current of 3×10^{10} ^{87}Rb atoms/s, a collection solid angle factor of 1%, and detector efficiency of 70%, we expect 5×10^8 detected photons per s, (~ 3 photons/atom), and a synchronously detected signal of about 1×10^8 photons/s (1.6×10^{-11} A) fluorescence.

With 100 μW pump power ($\cong 4 \times 10^{14}$ photons/s), we would expect a fractional absorption signal of about 3×10^{-5}, about 3 times larger than the microwave absorption signal expected from the beam/cell using the 5 MHz wide laser, and somewhat smaller than might be expected from the beam/cell absorption if a broader linewidth laser were used.

Noise: The noise sources are the same as those previously discussed. With fluorescence, we have found experimentally that the major noise contribution is due to the detector dark current at the level of 1×10^{-14} W/$\sqrt{\text{Hz}}$, implying a theoretical S/N of better than 1000:1 in 1 sec for a 1% solid angle (fractional shot noise limit is 10^4:1). In absorption, the shot noise limit to S/N ratio is about 600:1 for 100 μW of laser power.

Experiment: Fig. 8 shows the laser induced fluorescence observed in the detection region as laser injection current is slowly swept. The ^{85}Rb lines

are strongest due to the 3:1 natural abundance ratio of ^{85}Rb:^{87}Rb. The largest components of each array occur on cycling transitions such as $F = 2 \cdot F' = 3$ in ^{87}Rb. Fig. 9 shows a derivative fluorescence detection signal with improved resolution of the ^{87}Rb, $F = 2 \cdot \cdot F' = 1, 2, 3$ spectrum. The arrow indicates the discriminator point used to lock the laser to the $F = 2 \cdot F' = 2$ transition.

With the laser locked, microwave power was applied in the intermediate region using a TE_{011} cavity. The glass beam tube fits axially along the center of the cavity in order to avoid field reversals and to maintain inherent symmetry of the TE_{011} mode. Figs. 10 and 11 show the microwave resonances we have observed by phase sensitive detection with the microwave frequency modulated ±2 kHz at a 400 Hz rate.

Theoretically the clock transition signal is strongest for $F = 2 \cdot F' = 2$ σ pump light, but gives no signal for π pump light. The fact that we see the clock transition for both polarizations indicates poor definition of the magnetic field within the cavity. We attribute this problem to inadequate magnetic shielding of the earth's field in all of the interaction regions. At present, the observed S/N is far below what we have estimated should be possible.

Nevertheless, from the data in fig. 11, we conclude that the observed resonance is already adequate for obtaining a frequency stability of about 3×10^{-8} in 1 second.

Standards Potential. In addition to improving the field definition, we feel there are a number of changes which would produce a Rb beam with fractional stability at 1 second of less than 1×10^{-10}. The largest improvement would come by improving the collection efficiency of the detected light from the present 1% to better than 20% and by replacing the Rabi cavity with a Ramsey cavity. Another improvement would be obtained by using a more intense laser to increase the population difference δ_{00}. At present, $\sigma_A Nt = 26$, while from fig. 6 a value of at least 100 is desirable.

It is clear that the simplified beam device should operate at least as well with Cs as with Rb. The benefit of a single Cs isotope, a more widely spaced excited state hyperfine structure, and a somewhat higher atomic Q will enhance the signal and make it easier to implement diode laser auto-locking schemes. We expect to have laser diodes operating at the Cs wavelength within a few months.

At present, the beam tube appears to be a much more promising device than the beam/cell, though both warrant further investigation.

Both devices clearly eliminate many of the accuracy and long-term stability limitations of present portable atomic frequency standards which have fractional stabilities of 10^{-10} - 10^{-12}, but, at this time, the improvement has come at the expense of reduced S/N ratio. We expect to improve the S/N ratio, and note that the simplified optically pumped Rb (Cs) beam tube offers the potential of at least an order of magnitude improvement in long-term stability and accuracy as compared to the presently available Rb portable standards.

Appendix

The atomic absorption probability per second, P_{fi}, to the level $|f\rangle = |n'IJ'F;M_{F'}\rangle$ from the level $|i\rangle = |nIJFM_F\rangle$ may be written as (MKS units):

$$P_{fi} = \frac{e^2\pi}{\varepsilon_0 \hbar^2 c} \frac{dI}{d\omega} |\langle f | \hat{e} \cdot \hat{r} | i \rangle|^2$$

where

$\frac{dI}{d\omega}$ = intensity per unit angular bandwidth of the pumping light with polarization unit vector $\hat{e}$.

Expanding the matrix element and eliminating the reduced matrix element by considering the analogous expression for spontaneous emission, we obtain

$$P_{fi} = \sigma_A R_{fi} N$$

where $\sigma_A = 3\lambda^2$ (λ = optical transition wavelength)

$$N = \frac{1}{\hbar\omega\Gamma}\frac{dI}{d\omega}$$

= number of photons/m^2 - s within the atomic absorption linewidth

($\tau = \frac{1}{\Gamma} = \frac{1}{\Delta\omega_A}$ = upper state lifetime).

and for light that connects the levels F and F' which is linearly polarized parallel (π) or perpendicular (σ) to an external field we have

$$R_{fi}(\pi) = \frac{(2J'+1)(2F+1)(2F'+1)}{4}\begin{pmatrix} F' & 1 & F \\ -M_F' & 0 & M_F\end{pmatrix}^2 \times \begin{Bmatrix} J' & F' & I \\ F & J & 1\end{Bmatrix}^2$$

$$R_{fi}(\sigma) = \frac{(2J'+1)(2F+1)(2F'+1)}{8}\left[\begin{pmatrix} F' & 1 & F \\ -M_F' & 1 & M_F\end{pmatrix}^2 + \begin{pmatrix} F' & 1 & F \\ -M_F' & -1 & M_F\end{pmatrix}^2\right]\begin{Bmatrix} J' & F' & I \\ F & J & 1\end{Bmatrix}^2$$

where ( ) and { } are 3-j and 6-j symbols,[4] respectively. With the same assumptions as in Ref. 5, the optical pumping rate equations are then

$$\frac{dn_i}{dt} = -\sigma_A N \sum_f R_{fi}\, n_i(t) + \sigma_A N \sum_{fi'} B_{if} R_{fi'}\, n_{i'}(t)$$

where

$n_i(t)$, $n_{i'}(t)$ are lower level populations

B_{if} is the spontaneous emission branching ratio from f to i, normalized to unity ($\sum_i B_{if}=1$; these are listed in Ref. 5).

This set of equations is easily solved by iteration on a computer. For convenience, the iteration has been done in increments of $\sigma_A Nt$, which appears as the abscissa in figs. 5, 6, and 7. At t=0, we assume the atom to be equally distributed among the lower levels (n_i = 0.125 for ^{87}Rb).

We can relate $\sigma_A N$ to the initial scattering rate per atom as follows. Let P be the rate of absorpton (photons/s or transitions/s) for an atom known to be in the level F with equal distribution among the M_F sublevels. Then P is obtained from P_{fi} by summing over excited sublevels and averaging over initial sublevels:

$$P = \frac{1}{2F+1}\sum_{M_F M_F'} P_{fi} = \frac{\sigma_A N}{2F+1}\sum_{M_F M_F'} R_{fi}$$

Using a sum rule on 3-j symbols[4], we find

$$P = \sigma_A N K_{FF'}$$

where $$K_{FF'} = \frac{(2F'+1)(2J'+1)}{12}\begin{Bmatrix} J' & F' & I \\ F & J & 1\end{Bmatrix}^2$$

We see that $\sigma_A N$ is related to P by the factor $K_{FF'}$. $K_{FF'}$ is given in Table 1 for ^{87}Rb D2 light. For a saturating laser beam (absorption rate = spontaneous emission rate), we have $P = 1/\tau$ or $N_{SAT} = 1/\tau\sigma_A K_{FF'}$. Since $K_{FF'}$ can vary by as much as a factor of 10, we see that the saturation intensity is strongly dependent on the F→F' route.

References

1. This idea was originally suggested to us by S. R. Stein and H. Hellwig. Some features have been described in: P. Cerez, M. Arditi, and A. Kastler, C. R. Acad. Sc Paris 267, 282 (1968).
2. A. C. Tam and W. Happer, Phys. Rev. Lett. A38, 278 (1977); N. Ramsey, Molecular Beams (Oxford Univ. Press, London, 1969).
3. P. G. Papas, M. M. Burns, D. D. Hinshelwood, M. S. Feld, and D. E. Murnick, Phys. Rev. A 21, 1955 (1980).
4. A. R. Edmonds, Angular Momentum In Quantum Meachnics (Princeton Univ. Press, Princeton, N.J., 1960).
5. M. Arditi, I. Hirano, and P. Tougne, J. Phys. D: Appl. Phys. 11, 2465 (1978).
6. P. Cerez and F. Hartmann, IEEE Journ. Quantum Elec. QE-13, 344 (1977).

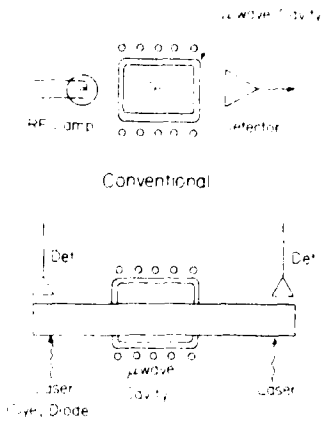


Figure 1. General comparison of physics packages for present Rb standards and the schemes being investigated here.

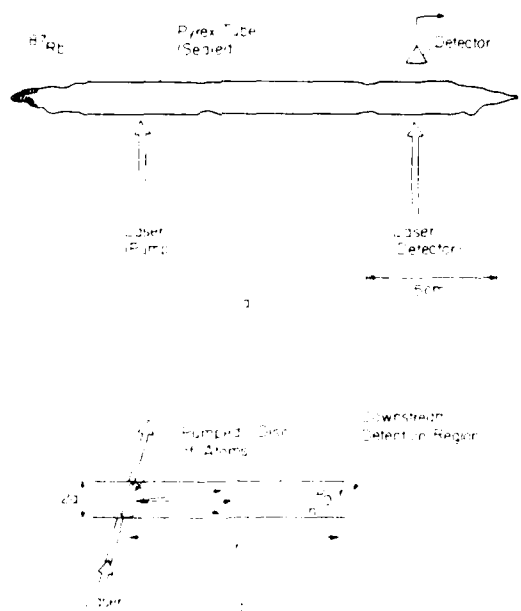


Figure 2. (a) Pyrex ^{87}Rb beam/cell. The sections intercepted by the laser beam incorporate pyrex spectrophotometer cells to obtain flat, high optical quality sides. Detection is by transmission monitoring. (b) Schematic illustration of the optical interaction region.

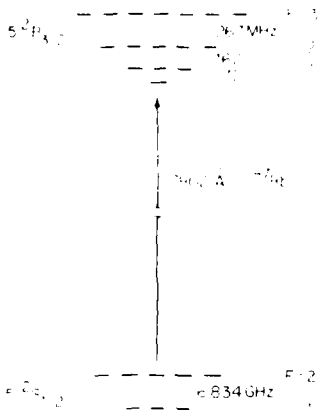


Figure 3. Energy levels relevant to D2 optical pumping in ^{87}Rb (not drawn to scale).

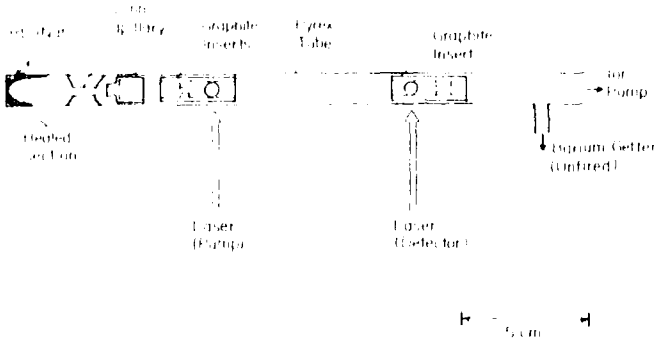


Figure 4. Pyrex atomic beam tube with graphite inserts. Spectrophotometer cells are used as in the beam/cell. However, detection is by fluorescence monitoring at right angles to the laser beam. A TE_{011} microwave cavity with cylindrical axis along the beam fits between the two optical interaction regions.

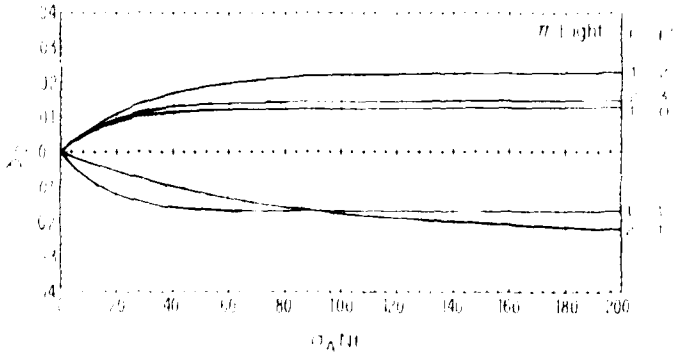


Figure 5 The nonzero population differences $\delta_{00} = n(F = 2, M_F = 0) - n(F = 1, M_F = 0)$ are plotted vs. σNt for ^{87}Rb D2 (780 mm) light of π polarization. For a 100 μW laser beam of 3mm dia, 150 MHz spectral width at 780 nm, $\sigma_A N = 3.87 \times 10^6$ (see Appendix for details). Transit time through the beam then corresponds to $\sigma_A Nt \sim 26$

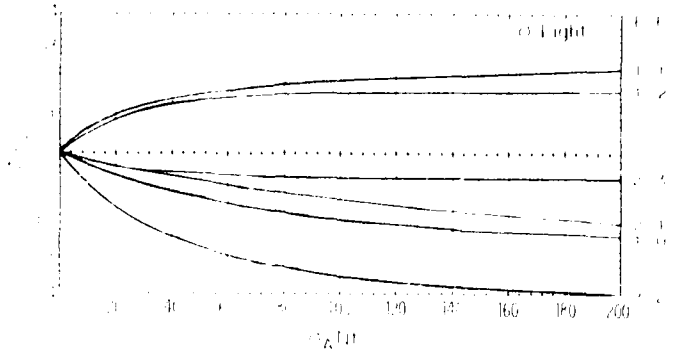


Figure 6. Same as Figure 5, but for σ polarization.

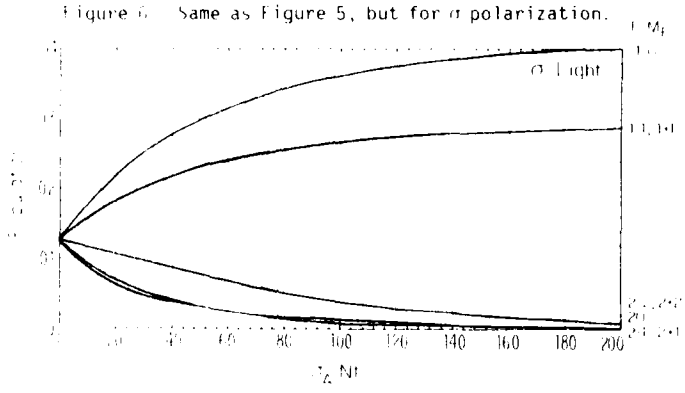


Figure 7. Ground state sublevel populations vs. $\sigma_A Nt$ for ^{87}Rb D2 (780 nm). The pumping is done on the 5 $S_{\frac{1}{2}}$ F = 2 → 5 $^2P_{3/2}$ F' = 2 transition with σ polarized light (see Appendix for details). This transition results in the largest possible population difference between the M_F = 0 clock sublevels for pumping with a single linearly polarized laser.

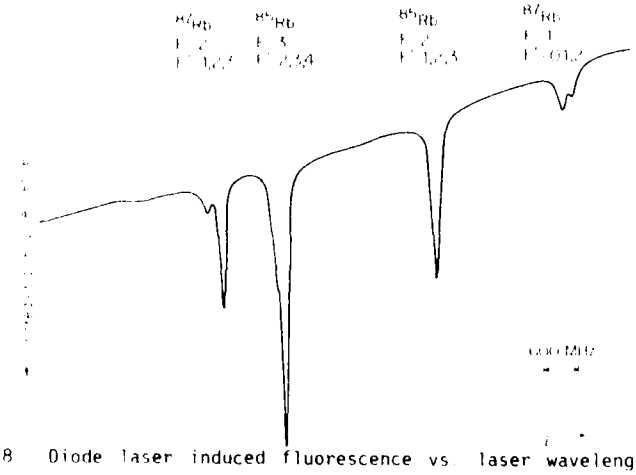


Figure 8. Diode laser induced fluorescence vs. laser wavelength observed downstream in the Rb beam tube (pump beam blocked). Sloping baseline is due to the change in intensity of the stray laser light as the injection current is swept. Diode Laser: type TJS, 90 μW power, 150 MHz spectral width, 3 mm beam diameter. Source temperature = 100 °C.

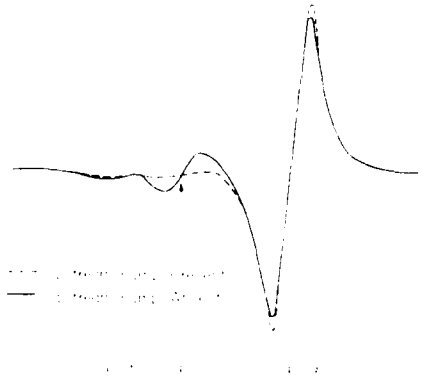


Figure 9. Diode induced fluorescence vs. laser wavelength observed downstream for the ^{87}Rb $5\ ^2S_{1/2}$ F = 2 → $5\ ^2P_{3/2}$ F' = 1, 2, 3 transitions. Phase sensitive detection with laser fm modulated over ~ 40 MHz at 200 Hz rate, time constant = 400 ms. To obtain the microwave data in figs. 10 and 11, the laser was locked to the F' = 2 fluorescence peak (arrow) observed in the pumping region.

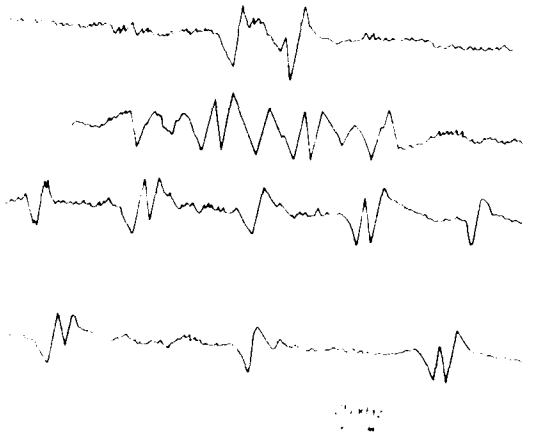


Figure 10 Microwave resonances in the fluorescent light for several values of an applied longitudinal magnetic field (field increases from top to bottom) Horizontal axis is a 400 KHz sweep of the microwave frequency about the center value 6.834627GHz while the diode laser is locked to the $5\ ^2S_{1/2}$ F = 2 → $5\ ^2P_{3/2}$ F' = 2 transition. Pumping is done with π polarized light.

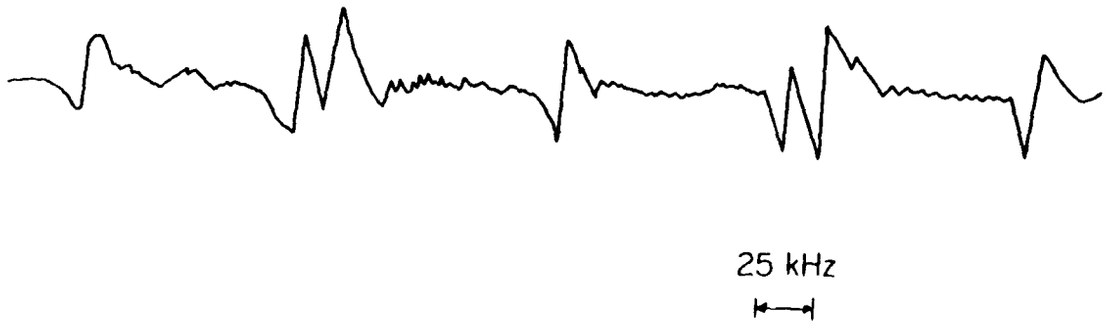


Figure 11. Same as fig. 10, but with σ polarized light.

F ↓ \ F' →	0	1	2	3
1	.028	.069	.069	0.0
2	0.0	.008	.042	.117

Table 1. Values of $K_{FF'}$ for ^{87}Rb D2 light (see Appendix). The same values apply for π and σ polarization.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ, 27-29 May 1981

[illegible] SAPPHIRE LAMPS [illegible] IN SATELLITE-BORNE ATOMIC RUBIDIUM CLOCKS

Thomas C. English and Ernst Jechart

Efratom Systems Corporation
18851 Bardeen Avenue
Irvine, California 92715

Summary

The apparent failure of state-of-the-art, all-glass, rubidium spectral lamps in the vacuum environment of GPS satellite-borne atomic clocks has given rise to a program to develop a new type of lamp with an order of magnitude greater lifetime than a glass lamp.

Terrestrial experience with sapphire and sapphire-like materials strongly suggests that a lamp made from these materials would satisfy this requirement. Accordingly, several developmental lamps of this type have been built. Preliminary results indicate that their operating characteristics are very similar to those of glass lamps in almost all respects, with two exceptions: First, the ignition properties are somewhat different. Second, there is a significant UV output that is accompanied by ozone production. This problem has been eliminated by coating a thin, UV absorbing coating to the outside of the lamp.

Finally, it is concluded that the technology exists to make one-year accelerated life tests of these lamps is presently available.

Key Words: Rubidium, spectral lamp, sapphire, atomic clock, GPS.

Introduction

[illegible] Rb [illegible] and [illegible] a key part in the development of the Navstar Global Positioning System (GPS). Although the navigation performance of these clocks has greatly exceeded the specifications,[1,2] there have been reliability problems. Some of these problems have been traced to the clock electronics or associated equipment and are not fundamental in nature. The most serious reliability problem is the apparent failure in orbit of some of the electrodeless discharge, rubidium spectral lamps that are used for optical pumping.

Telemetry and other data suggest that the failure mode is loss of free metallic Rb in the lamp, with the most likely loss mechanisms being chemical reaction of free Rb with the glass envelope and/or diffusion of free Rb into the glass at the elevated temperatures encountered in the space (vacuum) environment. An extensive investigation,[4] begun in late 1979, is presently studying these and other questions. In the meantime, an interim approach has been adopted for GPS Phase II. This interim approach consists of increasing the Rb fills of glass lamps and using a different type of glass (Corning 1720) for the envelope. Accelerated life test data taken by Efratom and Rockwell suggest a ground-based lifetime in excess of 10 years under these conditions.

In addition to this interim approach, Efratom has initiated an internally funded program to develop a lamp having a lifetime that is an order of magnitude greater than that of a glass lamp by using sapphire and sapphire-like materials instead of glass. This program is the subject of this paper.

Lamp Characteristics

State-of-the-art rubidium spectral lamps that are used in present-day commercial rubidium frequency standards generally have the following characteristics:

1. Operating temperature: 150° C.
2. Electrodeless discharge operation: [illegible] of frequency 100 MHz, rf power to discharge [illegible] W.
3. Lamp envelope material: glass of a type that is not readily attacked by rubidium.

In the sapphire lamp development program described in this paper, characteristics (1) and (2) are retained, and the goal is to build an extremely long-lived sapphire-type lamp that has spectral characteristics that are similar to those of present state-of-the-art glass lamps.

Why Sapphire?

Sapphire has many desirable properties that make it suitable for use in Rb spectral lamps of the type used in optically pumped rubidium frequency standards. Some of these properties are:

1. 100% Al_2O_3 single crystal that is optically transparent when polished.

2. Chemically inert except under extreme conditions.
3. Contains no silicates, water or other impurities that could react with rubidium.
4. High density (3.9 g/cc typ. [illegible] g/cc for most alumina), pore-free material that make it virtually impermeable.
5. Excellent optical, thermal and electrical properties.
6. Refractory material with melting point of [illegible] °C.

The primary disadvantage of sapphire is that it is difficult to work during the lamp fabrication process. That sapphire is more expensive than glass is of secondary importance because materials [illegible] a small fraction of the total [illegible].

Practical Experience with Sapphire

A good indication of the suitability of sapphire and sapphire-like materials for use in rubidium spectral lamps can be obtained by considering two very practical examples where such materials have been used for containment of alkali metals under extreme conditions of temperature and pressure.

High Pressure Sodium Lamps

These devices have been available for many years and are commonly used for street lighting and general commercial illumination.[5] A typical lamp of this type has the characteristics[6-8] given in Table 1. The envelope is made of Lucalox which is General Electric Company's trade name for its proprietary, sintered, polycrystalline alluminum oxide (Al_2O_3) ceramic. Lucalox is similar to sapphire in that it is high purity (99.9% Al_2O_3) and has a density approaching that of sapphire; however, it is translucent rather than transparent, and is much less expensive than sapphire. These lamps are operated by direct passage of ac current through the lamp thereby forming an arc that heats the center of the Lucalox tube to [illegible]100° C. The partial pressure of the sodium in these lamps is a significant fraction of one atmosphere. (The purpose of the mercury -- See Table I -- is to increase the voltage gradient of the arc which reduces the ballast requirements.) Even with these high temperatures and pressures, these lamps are still sufficiently long-lived that after 24,000 hours (2.7 years), 97% are still operating.[8] As the lamp is operated, free sodium is gradually lost, thereby increasing the mercury pressure and and lamp operating voltage. End of life occurs when the voltage needed to maintain the arc increases beyond what can be supplied by the power source. The mechanisms for partial sodium loss are varied and include reaction of sodium with the Lucalox, and with impurities in the Lucalox and the electrodes.

Space Laser Arc Lamp

Experience with the effect of potassium on sapphire at temperatures of [illegible]°C has been obtained as part of a program[9,10] to build a 250 W, CW, potassium vapor arc lamp for use in pumping a Nd:YAG space communications laser. Table 2 gives the characteristics of this arc lamp.

In the initial phase of this program, difficulty was experienced in two areas: Pitting[9,11] of the inside of the sapphire envelope, and development and growth of cracks[11,12] inside the sapphire envelope, eventually resulting in lamp failure. It was suggested that both these effects were a result of free potassium reacting with residual oxygen in the lamp, and the sapphire envelope, to form potassium aluminate ($KAlO_2$) and/or potassium beta alumina ($K_2O \cdot Al_2O_3$). These problems were solved by eliminating free oxygen from the interior of the lamp. This was accomplished by: Reducing the initial oxygen content of the lamp by improved processing and filling procedures; and use of a suitable oxygen getter in the lamp. Subsequent lamps that were built using these procedures showed no frosting and cracking, and gave the results shown[9,13] in Table 2 where the failure mode was degradation of the metal-to-sapphire seals on the ends of the lamp.

Comparison with Present Application

The operating conditions for the above two examples (high pressure sodium lamps and potassium arc lamp) are rigorous indeed. By comparison, for the spectral lamps of interest here:

1. Operating temperatures are lower by [illegible] °C.
2. Alkali vapor pressures are $\sim 10^5$ times lower.
3. Since the alkalis are less reactive in the order Li, Na, K, Rb, Cs, rubidium is less reactive than either sodium or potassium.
4. No electrodes or metal seals are used.

In view of these large and very important ameliorating factors, it is likely that the reaction of Rb with sapphire and sapphire-like materials (such as Lucalox) is nil for a spectral lamp of the type considered here.

Approach

Our overall approach to the realization of a fully operational, long-lived, sapphire, Rb spectral lamp can be divided into two phases, as indicated in Fig. 1. At the present time, Phase I is still in progress.

Lamp Design

The philosophy adopted for lamp design was to use as basic materials either sapphire, sapphire-like materials or a combination of the two. Unfortunately, these materials cannot be worked using the (very convenient) fabrication techniques that are normally used for glass: Sapphire has a very sharply defined melting point where it suddenly changes from solid to liquid. By way of comparison,

glass exhibits plastic behavior over a significant temperature range (usually ~200° C) that makes it easy to form using a gas-oxygen torch. Because of this "shortcoming" of sapphire and sapphire-like materials, fabrication methods that are different than those used for manufacture of glass lamps must be devised. In particular, what is needed is:

1. A method of joining sapphire and/or sapphire-like parts together.
2. A method whereby the lamps can be hermetically sealed after filling with Rb metal and buffer gas.

A single method was chosen to accomplish both the above objectives. This method involved using a special alkali-resistant glass to "glue" the sapphire and sapphire-like parts together, and also to provide a connection to a glass vacuum system by means of which a lamp could be filled. The lamps were designed with a glass fill tube on the one end that could be "tipped off" with a gas-oxygen torch subsequent to filling. The cross section of a lamp filled in this manner is shown in Fig. 2. The window and end caps are each "glued" to the body by an annular layer of the special alkali-resistant glass that is ~0.25 mm thick (not shown in drawing). Similar lamps have also been made out of sapphire alone (no Vistal) -- See Tables 3 and 4.

For purposes of comparison with the sapphire-Vistal lamp shown in Fig. 2, two glass lamps are pictured in Fig. 3: A regular commercial Rb lamp and a GPS [illegible] Rb lamp. The main difference in geometry is that the all-glass lamps have hemispherical ends, whereas the sapphire-type lamps have ends that are capped, right circular cylinders.

Anticipated Lifetime

Table 4 allows a direct comparison of developmental sapphire lamps and all-glass lamps. The important parameter is the interior surface area. If it is assumed that the reaction (and/or diffusion) rate of Rb with sapphire-Vistal is negligible, and that the reaction (and/or diffusion) rate of the Rb with the glass, per square centimeter of glass, is the same as for an all-glass lamp, then an estimate of the operating life of a sapphire lamp can be made. For a given glass lamp and a given sapphire lamp, the ratio of the operating lifetimes should equal the inverse ratio of the interior glass surface areas. For the 10 mm O.D. sapphire lamp, the operating lifetime is therefore expected to be about an order of magnitude greater than that of an all-glass lamp, all other factors (such as Rb fill) being equal. For the sapphire-Vistal lamp, the ratio is not as favorable but the interior area of glass could easily be reduced to give a lifetime factor of 10 in the next generation of these lamps.

The operating lifetime of a lamp will also depend on its operating temperature. For an all-glass lamp operating in a vacuum, the temperature at the front of the lamp, where the light emerges, can exceed the temperature of the thermostatically controlled base by nearly 60° C. This large a temperature gradient across the lamp is undesirable because it is expected to increase significantly the rate of chemical reaction of free Rb with glass at the front of the lamp (any diffusion of Rb into the glass would also be increased). For a sapphire lamp the corresponding temperature gradient would be much less, mainly because the thermal conductivity of sapphire is approximately 20 times larger than that of glass.

Results to Date

In this section we present some preliminary results from Phase I of our development program.

Performance

Measurements of light intensity for the D_1 (794.7 nm) and D_2 (780.0 nm) resonance lines of Rb were made for a sapphire-Vistal lamp and compared to a GPS 1720 glass lamp having the same O.D. It was found that the D_1 intensities for the two lamps were the same to within 2%; similarly for the D_2 line. These small differences in D_1 and D_2 intensities are of no practical significance for rubidium frequency standard operation.

The sapphire-Vistal lamp was also compared with the GPS all-glass lamp by placing each, sequentially, in the same operating rubidium frequency standard (Efratom, Model FRK). The Rb resonance signal for the sapphire-Vistal lamp was found to be 6% larger than for the GPS glass lamp. The sapphire-Vistal lamp produced a fractional frequency offset in the unit, relative to the GPS lamp, of $+ 1 \times 10^{-10}$. These variations are of the order of those normally encountered for different lamps from the same production run and are of no practical importance.

The short-term frequency stability (Allan Variance) of both sapphire and sapphire-Vistal lamps was measured for averaging times of 4.2 and 32 sec and compared to glass lamps in the same frequency standards. The differences are small although the larger diameter sapphire lamp (10 mm O.D.) does give slightly better results than the sapphire-Vistal lamp (9 mm O.D.), presumably because of the larger window area of the former.

Long-term drift was also measured for both a glass and a coated (see below) sapphire-Vistal lamp. The results are shown in Figs. 4 and 5. Judging from these plots, coated, sapphire-like lamps exhibit no serious long-term drift.

Areas of Concern

As regards the operational and functional characteristics of sapphire lamps, two areas of concern have surfaced during our developmental program. The first is the generation of ozone and the concurrent formation of an unwanted thin film on the outside of the lamp. The second is sometimes sluggish lamp ignition.

Ozone Generation. Experimentally, it is observed that significant quantities of ozone (O_3) are present outside a sapphire lamp during operation and that the presence of ozone is associated

with the plasma discharge inside the lamp. The most likely mechanism for production of ozone is ultraviolet (UV) photodissociation of oxygen in the surrounding air by UV light from the lamp according to the reaction equation[14]

$$3\,O_2 + h\nu\ (\lambda \leq 240\ \text{nm}) \rightarrow 2\,O_3.$$

The ultraviolet light produced in the plasma discharge will be readily transmitted by the sapphire envelope (sapphire has 86% transmission at 240 nm). On the other hand, the absence[15] of ozone and (by inference) UV emission from a glass lamp may be attributed to the nearly complete opacity of glass at 240 nm and below.

Film Formation. After operation of a sapphire lamp for ~ 2 days, it is observed that a thin film forms on the outside of the sapphire window (and also on the sides of the lamp, whether they be sapphire or Vistal). This film has the following properties: (1) Thin (interference colors are present); (2) Transparent - translucent; (3) Granular; (4) Hard; (5) Does not damage sapphire in any way; (6) Can be scrapped off window with a sharp knife; (7) Can be removed with hydrofluoric acid (which does not attack sapphire); (8) After removal, film reforms again during subsequent lamp operation. The exact mechanism by which this film is produced is unknown but in view of (5), above, it is clear that it must be due to material transport from the lamp surroundings to the lamp itself. The most likely possibilities are reaction of ozone/UV with materials in the immediate vicinity of the lamp (e.g. Delrin AF of rf coil form, mica window on lamp oven) and subsequent vapor transport of the reaction product to the lamp surfaces.

Possible Solutions. At least two different solutions to the above problems are possible:

1. Operate the lamp in a vacuum and use only UV resistant materials near the lamp.
2. Put a coating on the outside of the sapphire envelope that absorbs UV radiation.

For lamp operation in space (that is, in a high vacuum environment) ozone is absent and only UV radiation remains as a potential problem. In this case, part of solution (1), above, is automatically implemented. The problem of finding materials that exhibit long-term stability in a UV environment remains (actually, the extent to which UV emitted from the lamp is a problem has not yet been established), but the real drawback of solution (1) is that any testing or use of the lamp in air is ruled out completely. For these reasons we have implemented solution (2) which we feel to be a much more satisfactory overall solution to the ozone/UV problem.

Results for Coated Lamp. We have developed a method of bonding a thin, stable, inorganic coating to a sapphire lamp. This coating material is transparent to near infra red and visible radiation but virtually opaque to UV radiation (below 240 nm). So far, we have applied this technique to one sapphire-Vistal lamp and have carried out various tests to determine the efficacy of this coating. The results of these tests may be summarized as follows: (1) No ozone is produced[15] by the coated lamp; (2) After six weeks of lamp operation there is no evidence of thin-film formation of the type observed for an uncoated lamp; (3) The long-term drift data for this lamp (Fig. 4) are quite acceptable relative to glass lamps (Fig. 5); (4) The short-term stability of a clock using this lamp is comparable to that for a GPS glass lamp in the same clock.

Sluggish Ignition. After filling, and upon operating for the first time, both sapphire and sapphire-Vistal lamps ignite instantly from a cold (room temperature) start. However, when a sapphire-Vistal lamp is operated for awhile at normal operating temperature (~ 115° C), subsequent reignition from a cold start is sluggish and up to several minutes are required after lamp oscillator and lamp oven turn on. (For an all-glass lamp, the ignition time under the same conditions is a few seconds.) The "all-sapphire" lamp (entry No. 2 in Table 3) has also been sluggish on some occasions and under certain circumstances but usually this lamp ignites instantly from a cold start. In no case has a sapphire or sapphire-Vistal lamp either failed to ignite or ignited in the wrong mode.

A separate investigation has shown that a sapphire lamp can lower the Q of the lamp oscillator tank circuit when the lamp is cold. This, of course, will act to inhibit ignition (see below). Since sapphire, Vistal, and glass by themselves do not measurably affect the Q, this result suggests that a conductive (invisible) Rb film may be formed under certain conditions on the inside surface of a sapphire-Vistal or sapphire lamp that makes the lamp lossy in the rf field. Rubidium film formation could reasonably be expected to depend on the thermal design, environment and previous thermal history of a lamp, thereby possibly explaining the diversity of observed ignition times for different lamps under different circumstances. The reason a sluggish lamp ignites after several minutes is due to heating of the lamp both by direct, unintentional rf dielectric heating of the envelope (lossy film) and indirect heating by conduction via the thermostatically controlled lamp oven. This heating can conceivably produce ignition in either (or both) of two ways:

1. Since the ionization potential of Rb is only 4.2 eV, compared to ~ 15 eV for the buffer gas, vaporization of Rb metal lowers the value of the critical rf field strength needed for ignition.
2. Heating of the lamp envelope may vaporize the conducting film on the inside surface, thereby increasing the Q of the lamp oscillator tank circuit. This would then act to increase the rf electric field available for ignition inside the lamp.

It should be noted that these phenomena are not yet well understood. More work needs to be done in this area.

The only disadvantage of sluggishness that we see is that it might place a limit on how rapidly a clock can be warmed up from a cold start. (Fast warm-up is of no concern for the GPS Program.) Even this may not be a problem, though, because the warm-up time of a glass lamp is considerably longer than its ignition time. For both glass and sapphire lamps, the lamp is not warmed up until it is hot enough to give a vapor pressure of Rb that can produce sufficient D_1 and D_2 radiation for optical pumping. Whether or not a lamp ignites prior to this is of little importance. Also, it should be mentioned that the test fixtures that we used in our ignition tests were not designed for rapid warm-up. If special rapid warm-up test fixtures had been used, it is likely that sapphire lamp ignition would have occured sooner.

Life Testing

It is not an easy matter to conduct meaningful accelerated life tests on a device with an anticipated lifetime in excess of 100 years. This is especially true for electrodeless discharge, Rb spectral lamps because any attempt to stress the device can change the operating conditions to the point where the failure mechanism is changed, thereby invalidating the testing process. For rubidium spectral lamps, where failure occurs due to loss of free Rb, an ideal testing method would allow non-destructive measurements of free Rb to be made as the lamp was operated under normal conditions. Fortunately, such a method (calorimetric method) is currently available (see below).

Based on life test results obtained for glass lamps, we anticipate a rubidium consumption of 5 µg during the first year of operation of a sapphire lamp. This is in addition to the 20 - 30 µg of Rb consumed during the first few days of operation by contaminants such as water vapor that are initially present in the lamp ("dirt effect").

The calorimetric method, suggested by E.S. & T. and presently in use at The Aerospace Corporation, nondestructively determines the amount of free Rb metal in a lamp by measuring the heat of fusion of Rb metal at the melting point (38.5° C). So far, these measurements have been carried out using two different models of Perkin-Elmer Differential Scanning Calorimeters (DSC). The Model DSC II allows measurements to ±(5 + 1 µg), typically, while the Model DSC IV allows ±(1% + 1 µg), typically.[4] Using this type of instrument, a typical (simplified) life-test scenario might proceed as follows:

1. Short-fill sapphire lamp with 40 to 50 µg of Rb.
2. Measure initial free-Rb fill using DSC II after tip-off of lamp but before operation.
3. Begin lamp operation under normal conditions.
4. After 100 hrs. of operation, remeasure free Rb to determine amount of Rb consumed by dirt effect (expected to be 20 - 30 µg.)
5. Depending on initial Rb fill quantity and the exact amount of Rb consumed by the dirt effect, 10 - 50 µg of free Rb should be left for the actual life test (one year duration). This quantity of free Rb should allow measurement uncertainties of ± 1½ µg or less. This would allow measurement of a 5 µg/year consumption rate to ±3% or better.

Use of the Perkin-Elmer DSC II should therefore give sufficient resolution for measurement of the true rate of free Rb consumption in a sapphire lamp (exclusive of dirt effect) provided that the initial Rb fill can be kept in the range of 40 - 70 µg. This will be limited by precision of vacuum technique in controlling the Rb fill of DSC lamps.

Future Effort

Phase I future effort should be directed toward operational testing of additional sapphire lamps to increase the present data base, especially for coated lamps. Additional investigation of the sapphire lamp ignition process should also be conducted. Once an operationally satisfactory lamp is chosen, accelerated life testing, including the use of glass lamps as controls, should begin. The method employed should make use of short-fill techniques and differential scanning calorimetry for nondestructive measurement of free Rb.

Conclusions and Summary

Based on the practical experience of others with sapphire and sapphire-like materials, and our own experimental results on sapphire lamps for use in Rb frequency standards, we can draw the following preliminary conclusions:

1. Reaction (including diffusion) of free Rb with sapphire and sapphire-like materials (Lucalox, Vistal) is expected to be nil for the operating conditions encountered in Rb spectral lamps of the type used in rubidium atomic clocks.

2. Developmental lamps have been built using special alkali-resistant glass "glue" to join the sapphire and sapphire-like parts together, and to provide a convenient means of filling these lamps with Rb metal and buffer gas.

3. These lamps (both sapphire and sapphire-Vistal) are operationally very similar to state-of-the-art, all-glass lamps, and provide comparable performance when used in a rubidium frequency standard in place of a glass lamp.

4. Based on simple arguments, the lifetime of these lamps is anticipated to be an order of magnitude greater than that of a glass lamp. This corresponds to a lifetime in excess of 100 years.

5. These lamps emit UV radiation in sufficient amounts that ozone is generated when a lamp is operated in air. This ozone/UV interacting with materials surrounding the lamp causes a thin film to form by vapor transport on the outside of the lamp after 2 - 3 days of operation. This problem can be

eliminated by bonding a thin inorganic coating to the outside of the lamp that is opaque to UV but transparent to the Rb resonance radiation (D_1 and D_2 lines).

7. Lamp ignition from a cold start for the sapphire-Vistal lamp is sluggish, amounting to as much as several minutes (compared to seconds for a glass lamp). On the other hand, the "all-sapphire" lamp usually ignites instantly. No lamp has ever failed to ignite or operated in the wrong mode, and both types of lamp ignite instantly when hot.

8. Proper technology (calorimetric and short-fill) is presently available to conduct one-year accelerated life tests on these lamps.

9. Production models of sapphire lamps are possible by 1982.

Acknowledgements

We would like to thank Jeff Hayner for his able technical support of this program. One of us (T.C.L.) would also like to express thanks to Dr. John Farrar of Rockwell International for many helpful ideas, discussions and suggestions. The gift tubes of Mr. James Knight of Coors Ceramics in supplying Vistal samples is gratefully acknowledged. This work was supported entirely by Efratom internal company funding.

References

1. These clocks are a space-qualified version of the Efratom Model FRK-H commercial clock (more precisely, rubidium frequency standard) built by Rockwell International using Efratom-supplied physics packages.

2. Frank E. Koide and Darrell J. Dedrick, "Rubidium Frequency Standard Test Program for Navstar GPS," Proc. 10th Precise Time and Time Interval (PTTI) Applications and Planning Meeting, November 1978 (NASA Technical Memorandum 80250) pp. 379-392.

3. D. M. Tennant, "Navstar Global Positioning System (GPS) Clock Program: Present and Future," Proc. 12th PTTI, 1980 (NASA Conference Publication 2175) pp. 703-718.

4. This investigation is being conducted by The Aerospace Corporation, El Segundo, California.

5. J. I. McVey, "High-Pressure Sodium Lamp Technology," IEE Proc. 127A, 158 (Apr 1980).

6. Kurt Schmidt, "High Pressure Sodium Vapor Lamp," U.S. Patent 3,248,590 (26 Apr 1966).

7. Rodney E. Hanneman, Paul J. Jorgensen, and Dimitrios M. Speros, "Alumina-Ceramic Sodium Vapor Lamp," U.S. Patent 3,493,477 (3 Jul 1969).

8. General Electric Co., "Lucalox Life, Mortality and Maintenance," Lamp Specification Bulletin # LX-400-P1 (Dec 1969).

9. Norman C. Anderson, "Laser Pump Lamps," Air Force Avionics Laboratory Tech. Report, AFAL-TR-79-1002, Feb. 1979 (Available from NTIS as ADA 072704).

10. Norman C. Anderson, "K-Rb Laser Pump Lamp," Air Force Avionics Laboratory Tech. Report AFAL-TR-75-112, Nov. 1975 (Available from NTIS as ADB 009019).

11. Norman C. Anderson, R. M. Linford and James M. Heitman, "Potassium Lamp for Space Laser Communications," SPIE (Laser & Fiber Optics Communications) 150, 48 (1978).

12. Norman C. Anderson, "Basal Plane Cleavage Cracking of Synthetic Sapphire Arc Lamp Envelopes," J. Amer. Ceram. Soc. 62, 108 (1979).

13. Norman C. Anderson, IIC Technology, Inc., Sunnyvale, CA private communication, 1980.

14. Gerhard Herzberg, Molecular Spectra and Molecular Structure, Vol. I Spectra of Diatomic Molecules Second Edition (Van Nostrand, 1950) p. 483.

15. Ozone concentrations in air as low as 15 ppb can be detected by smell. See for example, N. Irving Sax, Dangerous Properties of Industrial Materials, Fourth Ed. (Van Nostrand Reinhold, New York, 1975), p. 989.

16. Dr. Charles Volk, The Aerospace Corporation, El Segundo, CA, private communication, 1981.

CHARACTERISTICS OF COMMERCIAL HIGH PRESSURE SODIUM LAMPS

- Construction: - GE Lucalox envelope
  - tungsten electrodes
  - niobium end seals
  - fill ~ several mg sodium + mercury + starting gas
- Operating temperature: ~ 1100 °C at center
- Sodium pressure: fraction of 1 atm
- Operating power: 50 W to 1000 W
- Mean operating lifetime: 28,500 hrs ($3\frac{1}{4}$ yrs)
- Failure mode: partial loss of sodium

TABLE 1. Characteristics of commercial high pressure sodium lamps.

CHARACTERISTICS OF A CW POTASSIUM VAPOR ARC LAMP

- Construction: - sapphire envelope
  - tungsten electrodes
  - Kovar alloy end seals
  - fill ~ several mg potassium + inert starting gas
- Operating temperature: ~ 1100 °C at center
- Potassium pressure: fraction of 1 atm
- Operating power: 250 W
- Operating lifetime: - no light output decrease after 3000 hrs (4 mo)
  - some still operating beyond 10,000 hrs (> 1 yr)
- Failure mode: oxidation of metal end seals causes air leak

TABLE 2. Characteristics of a CW, potassium vapor arc lamp.

MATERIALS OF CONSTRUCTION

Lamp Type	Construction Material		
	Window	Body	End Cap
Sapphire - Vistal	sapphire (optically flat, polished)	Vistal	Vistal
Sapphire	same as above	sapphire (transparent, as grown)	sapphire (unpolished -- frosted)

NOTE: All lamps use special alkali-resistant glass to "glue" sapphire & sapphire-like parts together, and to form "tip-off" at rear of lamp.

TABLE 3. Materials used for construction of sapphire and sapphire-Vistal lamps.

LAMP GEOMETRY

Lamp	O. D. (mm)	Volume (cc)	Interior Surface Area (cm^2)		
			Glass	Sapphire/Vistal	Total
Commercial Glass	10	.96	5.1	0	5.1
GPS 1720 Glass	9	.75	3.8	0	3.8
Sapphire	10	.76	.4	5.6	6.0
Sapphire-Vistal	9	.61	.6	3.9	4.5

TABLE 4. Geometrical data for glass and sapphire lamps.

SAPPHIRE LAMP DEVELOPMENT PROGRAM

Phase I

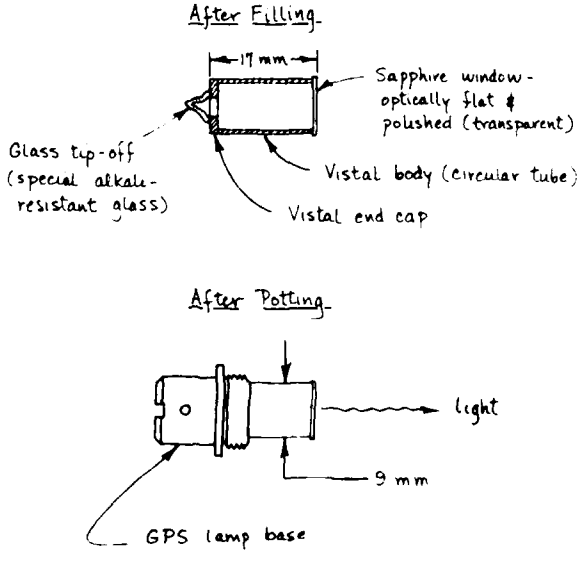
- Build lamps of different designs
- Test performance and compare with glass lamps
- Modify designs as required and retest

Phase II

- Select final design(s)
- Build lamps for life testing
- Conduct one-year accelerated life tests

FIGURE 1. Methodology for sapphire lamp development program.

SAPPHIRE-VISTAL* LAMP CONSTRUCTION



*Vistal = Coors Ceramics' version of GE's Lucalox, a translucent, polycrystalline, sapphire-like material

FIGURE 2. Construction details for sapphire-Vistal lamp.

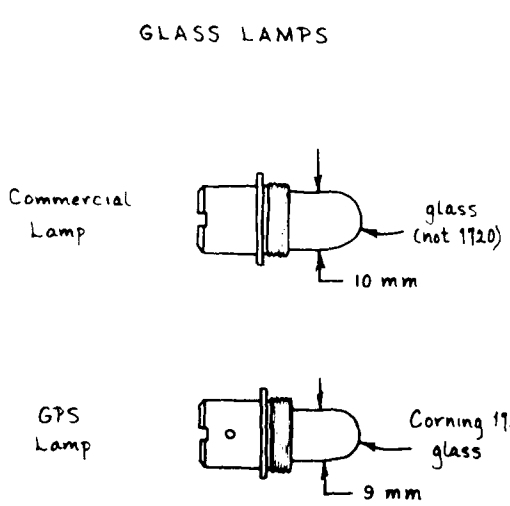


FIGURE 3. Glass lamps.

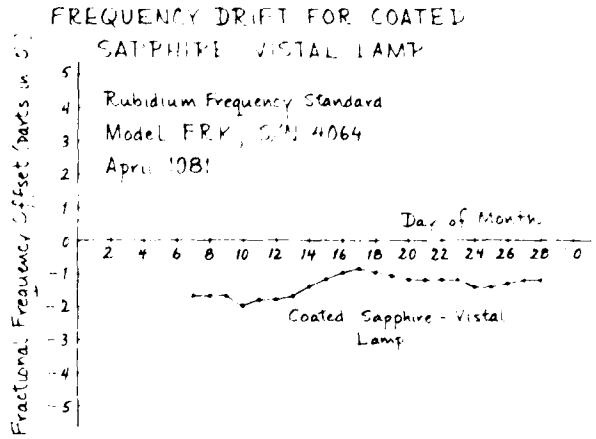


FIGURE 4. Long-term drift plot for rubidium frequency standard using coated sapphire-Vistal lamp.

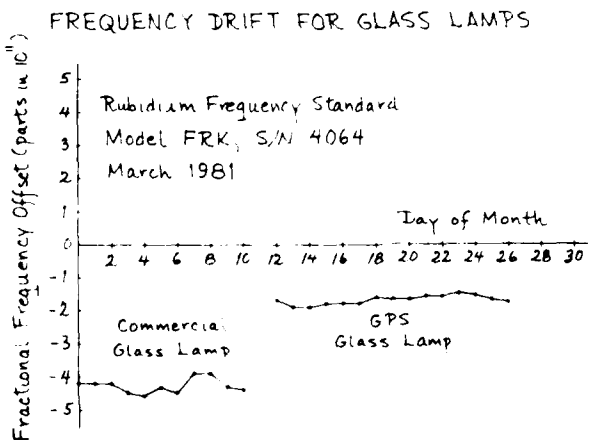


FIGURE 5. Long-term drift plot for rubidium frequency standard using glass lamps.

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

A MINIATURE, HIGH-PERFORMANCE RUBIDIUM FREQUENCY STANDARD

Toshio Hashi, Kazuharu Chiba, and Chikao Takeuchi

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

Summary

A miniature, general-purpose rubidium gas cell frequency standard has been developed to meet the ever-increasing needs for precision oscillators in a wide variety of applications. Presented in this paper are the design concepts and performance characteristics of this newly-developed frequency standard.

The new rubidium frequency standard embodies new electrical and structural design features and improvements, such as a miniaturized 2-cell type optical-microwave unit without a filter cell, advanced servo-electronics employing a number of integrated circuits, and simplified, fabrication-oriented structure with a mother-board and plug-in subassemblies. The miniaturization of the optical-microwave unit has been accomplished by eliminating the filter cell through the use of natural rubidium in the resonance cell instead of isotope Rb-87.

As a result of these features, the new rubidium frequency standard has small volume (2 liters), light weight (2.9 kg), and wide operating temperature range (-20°C to +60°C).

A series of measurements has been made to characterize the performance of the new rubidium frequency standard. In spite of the substantial reduction in size, it has exhibited excellent short- and medium-term stability comparable to that of the larger rubidium frequency standards intended for laboratory use, and relatively high environmental stability under extended operating conditions.

Introduction

Atomic frequency standards are no longer used exclusively in research laboratories. They have already found many applications in the field and systems involving precision frequencies. In Japan, rubidium frequency standards were developed to meet the demands for precision oscillators in communications, broadcasting, and precision measurement systems.[1] For a decade, FUJITSU has been producing rubidium frequency standards in several models, which are now in practical use as frequency and clock sources in systems such as digital data networks, and FDM communications systems.[2,3]

Most of today's applications for atomic frequency standards require a high degree of frequency stability and compactness for ease of transportation and installation. Our continuous effort has been directed toward volume not much greater than that of high stability crystal oscillators, and high performance level, which are generally incompatible.

In order to achieve a compact design of the rubidium frequency standard, it was necessary to reduce the size of the conventional optical-microwave unit, which accounts for a large part of the volume. The use of a 2-cell type optical-microwave unit without a filter cell was expected to serve this purpose. Preliminary experiments on the optical-microwave unit suggested the feasibility of using inexpensive natural rubidium in the resonance cell to eliminate the filter cell without degradation in the signal-to-noise ratio of the atomic resonance signal.

Along with this, electrical and mechanical design approaches were also taken to achieve a relatively low parts count and a simplified, cost-saving design without sacrificing performance.

In the following, the design and performance characteristics of the new rubidium frequency standard (designated model 5407A) are discussed in detail.

Design

Figure 1 is a simplified block diagram of the model 5407A rubidium frequency standard. The system consists of 5 subassemblies:

1. optical-microwave unit
2. voltage controlled crystal oscillator (VCXO)
3. frequency multiplier and synthesizer
4. servo control circuits
5. voltage regulator

The interconnections among them are, for the most part, made through the mother-board into which they plug. The mechanical structure was

also simplified to achieve light weight and ease of construction. This simplified structure greatly facilitated the assembly and test of the circuits. Figure 2 is an exploded view of the model 5407A showing the subassemblies and the basic mechanical structure.

Optical-microwave unit

The reduction in size of the bulky optical-microwave unit is indispensable for the miniaturization of the rubidium frequency standard. We developed a miniature, 2-cell type optical-microwave unit which employs a resonance cell containing natural rubidium, i.e., isotopes Rb-87 and Rb-85 in their naturally occurring proportion.

Problems posed by the 2-cell scheme and miniaturization were high temperature-coefficient of the resonance cell, and undue heating due to power dissipated within the optical-microwave unit.

(A) Temperature-coefficient of resonance cell

The temprature-coefficient of the resonance cell is dependent on the pressure ratio of the buffer gases. However, the optimum pressure ratio for the conventional 1-cell scheme is not suitable for the 2-cell scheme because of the effect of Rb-85 contained in the resonance cell. The resonance cell containing natural rubidium acts as a distributed filter/resonance cell with the part facing the rubidium lamp filter-like, and the opposite side resonance cell-like. Therefore, the filtering function is heavily dependent on the temperature of the integrated cell, causing light intensity change at the part acting as the resonance cell. Hence, a combined effect of buffer gases and light shift is observed.

At an early stage of development, our investigation focused on determination of the optimum ratio of buffer gases to obtain the lowest possible temperature-coefficient. The ratio of the buffer gases was selected on the basis of a series of experiments. Figure 3 and Fig. 4 show the temperature dependences of the resonance cell and of the lamp cell. These experimental results bear out the above assertion that the temperature-coefficient of the resonance cell is heavily dependent on the lamp cell temperature, i.e., light intensity, and the temperature-coefficient of the lamp cell is also dependent on the resonance cell temperature. Temperature-coefficients of less than $\pm 5 \times 10^{-11}$/deg. have been obtained at reasonable temperatures. The operating temperature of the lamp cell can be set at the point where the resonance signal level is at maximum, with improved short-term stability, but with possible adverse effects on the standard's sensitivity to ambient temperature changes.

(B) Operating temperature range

An additional important advantage of the 2-cell scheme is the fact that the optimum operating temperature of the integrated cell is higher by 10 - 20 degrees than that of the conventional resonance cell containing isotope Rb-87. (Fig. 3) This serves the purpose of attaining a higher operating temperature limit for the rubidium frequency standard, since its upper temperature limit is determined by the operating temperature of the resonance cell. Hence, the integrated cell containing natural rubidium allows a wide operating temperature range to be achieved without degradation in the signal-to-noise ratio of the error signal, i.e., short-term stability.

(C) Magnetic, thermal, and physical aspects

An important concern for a portable standard is its susceptibility to the ambient magnetic field, especially the earth's magnetic field. To reduce this effect to a negligible value for a worst-case reorientation in the earth's field, i.e., for about 1 gauss change, the optical-microwave unit is provided with double magnetic shields of mu-metal, and is mounted with the C-field direction at right angles to the horizontal surface.

Solving the thermal problems is also important to achieve wide operating temperature range, low power consumption, and fast warmup. The optical-microwave unit was designed to increase the thermal isolation of the cavity resonator from the lamp house and the external environment.

The volume of the new optical-microwave unit is about 0.5 liters, exclusive of the second magnetic shield. Figure 5 shows an exterior view of this unit.

VCXO and other circuits

The 5 MHz slave oscillator is a compact high stability crystal oscillator using a 5 MHz, 5th overtone AT-cut crystal unit with Q 2.5 x 10', and a single stage oven. It exhibits an aging rate of typically less than $1 \times 10^{-}$/year, and temperature stability of less than $1 \times 10^{-}$ from -20°C to +60°C. Fast warmup to $5 \times 10^{-}$ in 25 minutes is achieved with an input power of 5.5 watts at +25°C. The electronic control range is > $\pm 2 \times 10^{-}$. The volume is 40 x 40 x 80 (mm^{3}).

The circuits for frequency synthesis and servo control also have been made compact by the use of a relatively large number of integrated circuits, and by efficient mounting. Advanced circuit techniques have resulted in a relatively low parts count and elimination of bulky inductors and transformers from the servo control circuits. In particular, the phase detector and the integrator were carefully designed and adjusted to produce minimum offsets and drifts, because they would adversely affect the long-term stability and temperature stability.

The frequency synthesizer uses a sampling phase lock loop, and permits adjustment of the standard's output frequency over a range of 1.5×10^{-7} with incremental steps of about $2 \times 10^{-}$. This, combined with a C-field range

of 2 x 10^{-}, allows precise frequency adjustments to be made over a wide range.

Figure 6 shows the exterior view of the model 5407A rubidium frequency standard. The 5 MHz output is on the rear panel. The built-in meter on the front panel monitors internal circuit functions. Lock or loss of lock is indicated on the front panel. C-field adjustment is accessible from the front panel.

Performance characteristics

Long-term stability

Long-term stability is determined mainly by the stability of the pumping light and by the chemical processes which can occur in the resonance cell. Since these processes may produce a positive or a negative frequency drift, units may have different drift rates.

Figure 7 shows the long-term stability plots of two model 5407A's for about six months. Neither of them exhibited noticeable drifts. Aging rates are far better than 2 x 10^{-11}/month, and compare favorably with that of larger rubidium frequency standards for laboratory use.

Short-term stability

Figure 8 shows the short-term stability of the model 5407A. The stability was measured by referring to a low-noise crystal oscillator for 's of 10^{-} to 10 seconds, and by comparing two of the model 5407A's to each other for 's longer than 10 seconds. The measurement was made in a benign environment, because temperature changes and other environmental changes would influence the stability for averaging times longer than about $10^{}$ seconds. For averaging times smaller than the servo time constant (about 2 sec), the frequency fluctuations are determined by the crystal oscillator. For averaging times longer than the time constant, they are mainly determined by the signal-to-noise ratio of the output of the optical-microwave unit, resulting in a change in the slope. The stability improves as τ^{-1} from $=10^{}$ to about 1 sec, as $\tau^{-1/2}$ from $=1$ to $10^{}$ sec, and levels off at about 4 x 10^{-1}.

Environmental stability

Rubidium frequency standards are more or less sensitive to environmental changes. This sensitivity must be minimized to take advantage of the intrinsic stability. The model 5407A was subjected to changes in ambient temperature and magnetic field.

(A) Temperature

Figure 9 shows the effect of ambient temperature on the frequency of the model 5407A. The ambient temperature was held at +25°C for 3 hours, lowered to -20°C, and then was stepped up in 10 degree increments to +60°C, maintaining each intermediate temperature for 3 hours.

Frequency changes of the 2 samples were both less than 1 x 10^{-10} with respect to the frequency at +25°C over the range from -20°C to +60°C. The similar temperature coefficients of the 2 samples, without erratic behavior, suggest a good possibility of further improvement.

(B) Magnetic field

Figure 10 shows the effect of the ambient magnetic field on the frequency of the model 5407A. The model 5407A was subjected to reorientation in the earth's field. The frequency change was less than 1 x 10^{-11} for any orientation, i.e., for about 1 gauss change.

Warmup characteristic

Figure 11 shows the result of warmup measurement of the model 5407A after 24 hours off-time at +25°C. The frequency control loop is established 25 minutes after turn-on. Forty minutes after turn-on, the frequency is within 5 x 10^{-11} of the original value. The retrace is better than 1 x 10^{-11}.

Supply voltage characteristic

Figure 12 shows the effect of changes in supply voltage on the frequency. The model 5407A has a voltage regulator to reduce this effect, and no discernible frequency change was observed for ±10% changes in supply voltage.

Technical data

A summary of the technical data on model 5407A is shown in Table 1.

Conclusion

The details of the design of a miniature rubidium frequency standard have been presented as well as the performance test results.

Many electrical and structural features and improvements have been incorporated into the design, which have resulted in extreme compactness, light weight, wide operating temperature range, and low power consumption.

The long-term stability has been found to be better than 2 x 10^{-11}/month. Short-term stability measurement has shown that $\sigma_y(\tau) = 1 \times 10^{-1}$ for τ = 100 seconds. Sensitivity to environmental changes was also minimized in order to utilize the intrinsic frequency stability under extended environmental conditions. The frequency change caused by ambient temperature is less than 1 x 10^{-10} with respect to the frequency at +25°C for the range from -20°C to +60°C. The effect of the ambient magnetic field is less than 1 x 10^{-11} for any orientation in the earth's magnetic field.

These results demonstrate that no performance

has been sacrificed for the small volume, and that the new rubidium frequency standard is promising for use in a wide range of applications including use in the field.

Acknowledgement

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

References

1. Committee on Atomic Frequency Standard, "Principle and Practice of Atomic Frequency Standards", Technical Report of the Institute of Electrical Engineers of Japan, No. 15 (II) (1971).

2. K. Yakuwa, et al., "Rubidium Frequency Standard", FUJITSU, vol. 26 No. 6 (1975).

3. M. Otsuki, et al., "Digital Data Transmission Equipment", FUJITSU, vol. 29 No. 5 (1978).

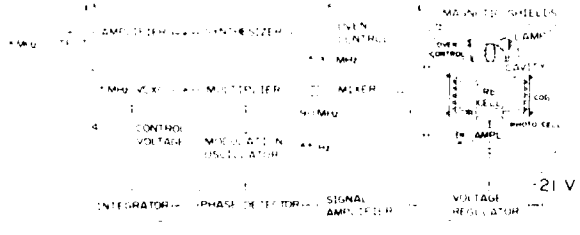


Fig. 1 Block diagram of model 5407A

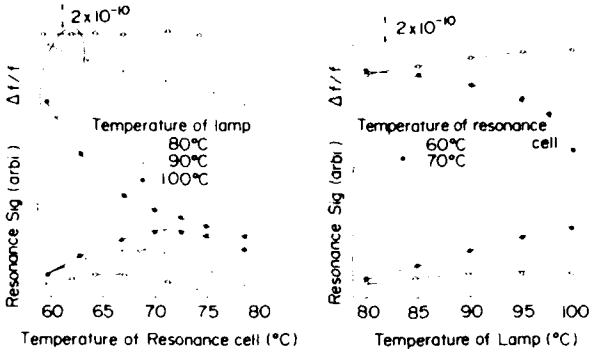


Fig. 3, Fig. 4 Temperature dependence of resonance cell and lamp cell

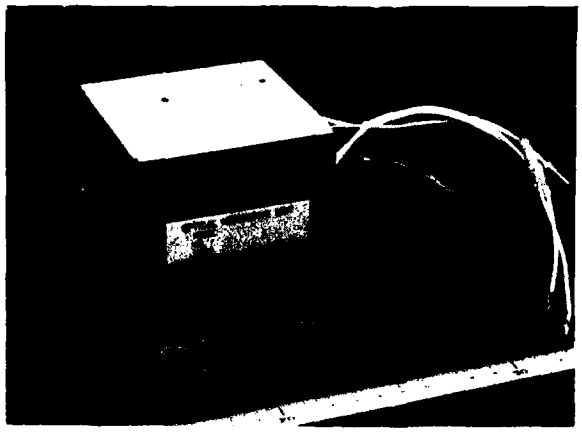


Fig. 5 Optical-microwave unit used in model 5407A

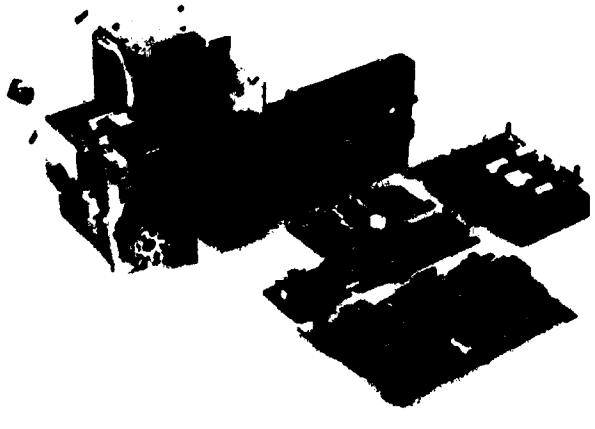


Fig. 7 Exploded view of model 5407A showing the subassemblies

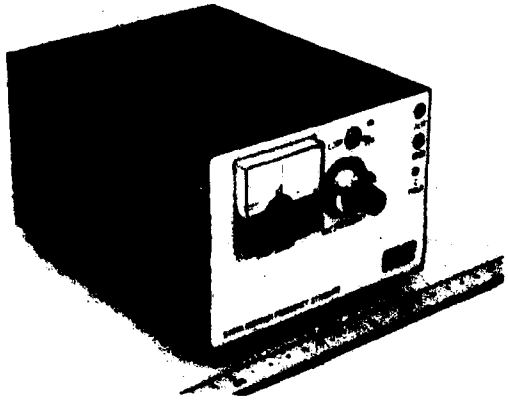


Fig. 6 Model 5407A rubidium frequency standard

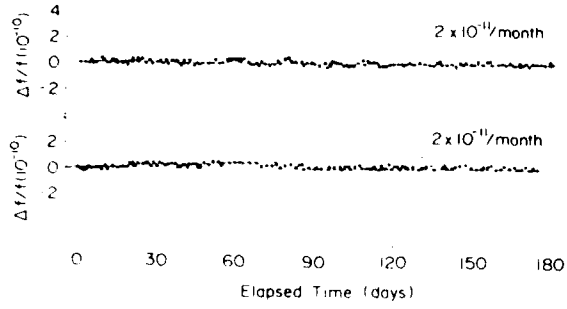


Fig. 7 Long-term stability

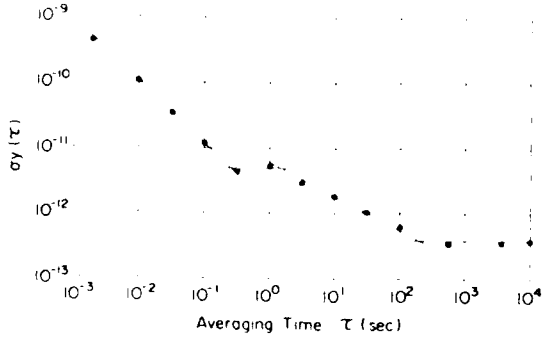


Fig. 8 Short-term stability

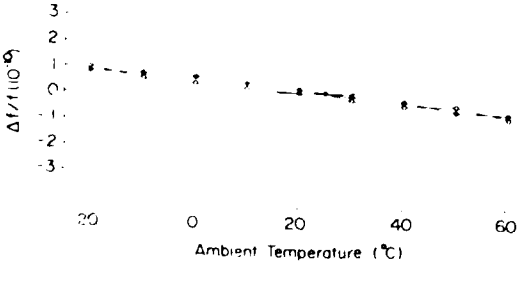


Fig. 9 Frequency versus temperature (for two 5407A's)

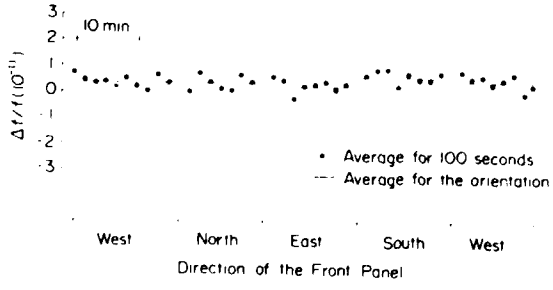


Fig. 10 Frequency versus orientation

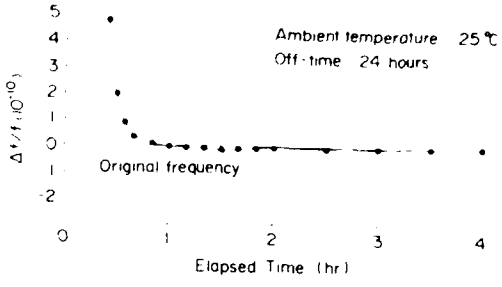


Fig. 11 Warmup characteristic

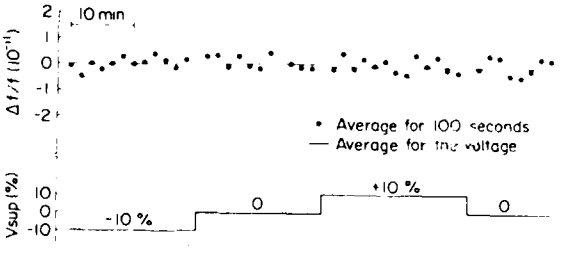


Fig. 12 Frequency versus supply voltage

Table 1. Technical data on model 5407A

Output
- Frequency: 5 MHz
- Level: 1 V_{rms} into 50 ohms
- Signal-to-noise ratio: 90 dB at rated output (BW 30 kHz)
- Connector: OSM

Input
- Voltage: -21 V
- Power: 17 W (steady @ +25°C)
  28 W (during warmup)

Dimensions: 95 mm (3.7") high, 130 mm (5.1") wide, 166 mm (6.5") deep

Weight: 2.9 kg (6.4 lbs)

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

PERFORMANCE OF THE GPS CESIUM BEAM FREQUENCY STANDARD IN ORBIT

Martin W. Levine

Frequency and Time Systems, Inc.
Beverly, Massachusetts 01915

Summary

The navigation-system clocks for the GPS NAVSTAR 5 and NAVSTAR 6 satellites are FTS-developed and manufactured cesium beam frequency standards (CFS). The cesium clocks have been in continuous operation since 6 April 1981 and 6 May 1980, respectively, and are the first cesium standards to provide satellite navigation and time-transfer signals over an extended period of time. The NAVSTAR cesium standards are preproduction models, predecessors to the GPS Phase II standards being manufactured by FTS for GPS space vehicles 9 through 12. The design and performance details of the PPM CFS was described at the 10th PTTI Planning Meeting.[1]

The on-orbit time and frequency stability of the navigation signals is computed from data disseminated by the U.S. Naval Observatory. The performance of the on-board cesium clocks, as recorded at USNO, exceeds the expectations of GPS Phase I and in most respects approaches the projected requirements for GPS Phase III.

Introduction

The preproduction model (PPM) series of cesium beam frequency standards is an evolutionary development from the Prototype Model flown aboard Navigation Technology Satellite II (NTS-2)[2]. The PPM standards were developed and manufactured by Frequency and Time Systems (FTS) under contract to the Naval Research Laboratory for the NAVSTAR 5 through NAVSTAR 8 satellites for the Global Positioning System constellation.

The performance specification and physical characteristics of the PPM cesium beam frequency standards are summarized in Table 1. The measured frequency stability of the PPM qualification unit, expressed as the Allan Variance of the frequency over averaging intervals from 10 to 10^5 seconds, is shown in Figure 1.

NAVSTAR 5 (Space Vehicle 5) was launched in February 1980 and the navigation signal was initially controlled by a rubidium frequency standard. The on-board cesium standard, PPM Serial Number 10, remained in passive storage for approximately 14 months. On 6 April 1981, the rubidium standard was turned off and the cesium standard was turned on to provide the navigation system master clock signal. NAVSTAR 6 (Space Vehicle 9) was launched on 26 April 1980 and the on-board cesium clock, Serial Number 11, was powered up 10 days later, on 6 May 1980. Tables 2 and 3 summarize the salient characteristics and the flight histories of PPM Serial Number 10 and 11, respectively. Both cesium standards have operated continuously since turn-on, with performance as described below.

Measurement Philosophy

The characterization of a high-performance atomic frequency standard can be a difficult and challenging task in a laboratory environment. Installation of the frequency standard in a NAVSTAR satellite orbiting 20,000 km above the earth's surface introduces interesting new metrology problems. In general these problems fall into three classifications:

a. Doppler Shift. The time-rate-of-change of the transmission delay (phase) from the satellite to the ground measurement point, the Doppler shift, is generated by the geometric motion of the transmitter with respect to the receiver and the changes in ionospheric and tropospheric group delays. The variations in ionospheric and tropospheric delays, in turn are the result of both relatively short-term density fluctuations and of periodic and secular terms due to the changing elevation of the satellite as observed from a given ground station.

b. Relativistic Effects. Corrections for non-trivial general relativistic effects must be applied to the received frequency data. Again, periodic terms appear in the relativistic corrections as a result of orbital eccentricity. Secular terms may also be present as a result of solar and lunar perturbations, radiation pressure and zonal harmonics of the earth's gravitational field[3].

c. Observation Time. A particular satellite is not continuously visible from a given ground receiver, so that an unbroken phase tracking of an on-board frequency standard is not possible. Furthermore, the satellite, even if visible, will be observed at low elevation angles for relatively long periods of time, seriously degrading the quality of the received data.

A two-way transmission link between the satellite and the ground receiver is the optimum solution to the Doppler shift measurement error. The transponder provides a direct measurement of the total transmission delay; and therefore permits estimation of the on-board clock performance independent of orbital and atmospheric perturbations. It should be noted that the two-way transmission does not account for the relativistic effects and corrections for relative velocity and for gravitational potential of the on- board clock must be applied to the received data[4]. The GPS system, however, only transmits signals - therefore the clock frequency errors, the satellite orbital errors, and the atmospheric group delays are coupled, and the performance of the clock can be extracted only by accurately modeling all of the perturbations which affect the phase of the received signal.

Measurement Systems

The most obvious source of GPS performance data is the GPS Control Segment, which is comprised of four Monitor Stations, an Upload Station and a Master Control Station. The on-line computation of satellite ephemeris corrections and explicit clock corrections is performed at the Master Control Station, which also controls the uploading of the satellite navigation message. Frequency standard data obtained from the Control Segment is presented elsewhere in these Proceedings by D.E. Tennant[5]. The mission of the Control Segment, however, is position determination, not the characterization of frequency standards. The Control Segment allocates overall system errors so as to optimize the navigational function of the System and the data may not explicitly demonstrate the full time-keeping capability of the on-board clocks.

The United States Naval Observatory has developed the capability to generate and disseminate GPS time, relative to the Observatory Master Clock, and all data presented in this paper is based on measurements made at the Observatory.[6] GPS data and information is readily available through the USNO Automated Data Service and can be accessed by teleprinter through the commercial telephone network.

The USNO time-transfer capability is based on a Stanford Telecommunications, Inc. (STI) single-channel C/A-code receiver; a block diagram of the system is shown in Figure 2. The time-transfer receiver locks onto and tracks the GPS signal. The decoded navigation message is routed to the processor which computes:

a. the satellite-to-receiver range from the known location of the antenna and the received ephemeris data

b. the ionospheric group delay from the known elevation and slant-range of the satellite and the ionospheric model transmitted as part of the navigation message

c. the tropospheric group delay from the known elevation and slant-range of the satellite

d. the clock correction from the model transmitted as part of Data Block 1 of the satellite navigation message. The estimated difference between the satellite transmission time (t_{SAT}) and the GPS system time (t_{GPS}) is represented by the second order polynomial[3]

$$t_{SAT}-t_{GPS} = a_2(t_{SAT}-t_o)^2 + a_1(t_{SAT}-t_o) + a_0 \quad (1)$$

where t_o is a system reference time. It should be noted that all relativistic effects; secular and periodic, are absorbed into the parameters a_0, a_1 and a_2.

The time transfer receiver generates a received-epoch pulse, once for every six-second frame of the navigation message. The pseudo range counter measures the time difference, with 10 nanosecond resolution, between the received epoch pulse and an epoch pulse derived from, and synchronous with, the USNO MC. The processor applies the four corrections noted above to the pseudo range measurement to compute the difference between the USNO MC and GPS system time, $t_{MC} - t_{GPS}$.

The USNO GPS data file includes, in addition to $t_{MC} - t_{GPS}$, the time-of-day at which an observation was made, the elevation angle of the satellite, the data age (time since the last updating of the navigation message) and other information which serves as a guide to the statistical reliability of the measurement. USNO also provides the difference between USNO MC time and the satellite transmission time, $t_{MC} - t_{SAT}$, which is computed by adding the GPS clock correction, $t_{SAT} - t_{GPS}$ to the MC versus GPS data which is the normal output of the time transfer unit. The $t_{MC} - t_{SAT}$ difference is independent of the explicit GPS clock correction model but is, of course, coupled to the GPS ephemeris through the computation of the satellite-to-receiver geometric range.

Data Editing

A relatively large amount of data was processed in order to evaluate the performance of the cesium frequency standards aboard space vehicles 5 and 9 and specific, a priori, guidelines were derived to filter this data as follows:

a. No measurements taken prior to October 1980 were used. Major discrepancies between the USNO MC and GPS system time were observed on time-transfer measurements taken from April 1980 through September 1980.

b. The uncorrected time difference $t_{MC}-t_{SA}$ was found to be more representative of the performance of the on-board clocks than the modeled difference $t_{MC} - t_{GPS}$.[7] This was found to true even though the uncorrected time difference does not take into account the frequency and phase errors introduced by relativistic effects. Therefore, the NAVSTAR cesium clock performance must be at least as good as indicated by the $t_{MC}-t_{SA}$ measurements, and probably significantly better in the long term.

c. Within the above constraints, all points were included which satisfied these a priori criteria:

Upload data age less than two hours
One (major) pass per day
Slope less than 50 ps/s
Number of six-second samples per measurement greater than 50
Satellite elevation angle greater than 15°

In certain cases, data points have been included which did not satisfy one or more of the above criteria, but no points were excluded!

Stability Data

Space Vehicle 5

The transition between rubidium-clock operation and cesium-clock control aboard space vehicle 5 is shown in Figure 3. The cesium standard was turned on on 6 April 1981 (MJD 44702) and immediately stabilized as shown on the right-hand side of the figure. Independent least-square linear fits over approximately 30 days for each clock have been applied to the data to remove the frequency offset, to the extent feasible, between the on-board satellite clocks and the USNO MC. The time error scale (ordinate) is 100 nanoseconds per minor division.

Figure 4 is a plot of the time deviation of PPM cesium frequency standard number 10 aboard space vehicle 5 from 12 April 1981 to 12 May 1981 (MJD 44705 to 44735). Note that the time error scale is reduced to 5 nanoseconds per minor division. The data sequence starts three days after the CFS was powered-up. The least-squares linear fit shows an offset of -106 ns/day (1.2×10^{-12}) with respect to the Naval Observatory and a standard deviation of the fit of less than 35 ns. A trial quadratic fit yields a drift of less than 0.5 ns/day squared, which is not statistically significant. The functional stability of the time fluctuations is approximately 1.6×10^{-13} for a one-day averaging interval. The remarkable time-keeping performance of this clock is evidenced by the fact that the (uncorrected) time stayed within ±65 ns of the USNO Master Clock for 30 days.

Space Vehicle 9

The performance of PPM cesium frequency standard number 11 aboard space vehicle 9 is shown in Figure 5 over the same thirty-day interval that the performance of PPM number 10 has been measured. The cesium standard had been operation for just under one year at the beginning of this data sequence. The least-square fit shows an offset of less than 45 ns/day (5×10^{-13}) with respect the USNO MC and a standard deviation of less than 30 ns. There is no detectable frequency drift term. The functional stability of the time fluctuations for one-day averaging is approximately 1.4×10^{-13} day. The (uncorrected) time remained within ±65 ns of the USNO Master Clock over the 30 day interval, virtually duplicating the exceptional performance of PPM number 10.

For comparison purposes, the difference between the USNO Master Clock and GPS system time is plotted in Figure 6, using the same set of satellite observations used to derive Figure 5. The much larger, and atypical, time deviations are evident from even a cursory comparison of Figure 6 with Figure 5.

The performance of the space vehicle 9 cesium clock over a much longer interval and at an earlier stage in its mission is shown in Figure 7. The data sequence extends for 100 days, from 24 October 1980 to 31 January 1981 (MJD 44535 to 44635). The time error plotted is the deviation from a least-squares linear fit with an offset of approximately 15 ns/day (-1.7×10^{-13}) and a standard deviation of less than 65 ns. The computed quadratic drift term is less than 0.1 ns/day^2 (a fractional frequency drift of less than 1×10^{-15}/day). The peak-to-peak time deviation of the satellite clock, with respect to the USNO MC, is less than ±140 ns over 100 days.

The fractional stability of the time fluctuations for one-day averaging intervals is approximately 2.1×10^{-13}. The flight of space vehicle 9 provides an almost unique opportunity to obtain an estimate of the very-long-term stability of a cesium frequency standard. An Allan Variance for five-day averaging is slightly greater than 5×10^{-14} and for ten-day averaging less than 4×10^{-14}, over the 100-day period. The calculated decreases at a rate somewhat greater than $\tau^{-1/2}$ (where τ is the averaging interval)

suggesting that system phase noise; ephemeris errors, atmospheric and relativity corrections, etc., is a contributor to the observed time dispersion, even for ten-day averaging intervals, and that the flicker floor of the cesium standard is less than 4×10^{-14}.

Conclusions

The frequency stability of the orbiting cesium frequency standards is summarized in Table 4, demonstrating performance levels significantly exceeding the expectations of just a few years ago. The ultimate noise floor appears to be in the order of 1×10^{-14}, with overall noise performance surpassing the clock requirements of the GPS Phase III.

Acknowledgement

The author wishes to thank G. Winkler and K. Putkovich of the U.S. Naval Observatory for their assistance in providing and analyzing the GPS timing data, and A. J. Van Dierendonck of Stanford Telecommunications, Inc. for help in interpretation of the GPS navigation message. The PPM cesium frequency standards were designed and built under Contract N00014-74-C-0061 for the Naval Research Laboratory.

References

1) M. Levine, "Performance of Preproduction Model Cesium Beam Frequency Standards for Spacecraft Applications", Proceedings of the 10th Annual Precise Time and Time Interval Applications and Planning Meeting, November 1978, pp 169-193.

2) G. A. Bartholomew, "Satellite Frequency Standards", in Global Positioning System, The Institute of Navigation, Washington, DC, 1980, pp 21-28.

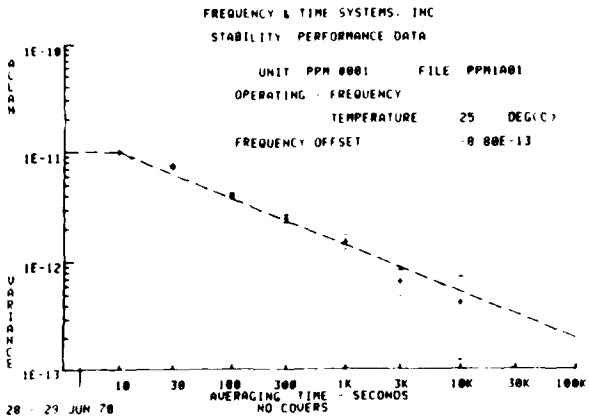
3) A. J. Van Dierendonck, S. S. Russell, E. R. Kopitzke and M. Birnbaum, "The GPS Navigation Message, in Global Positioning System, The Institute of Navigation, Washington, DC, 1980, pp 55-73.

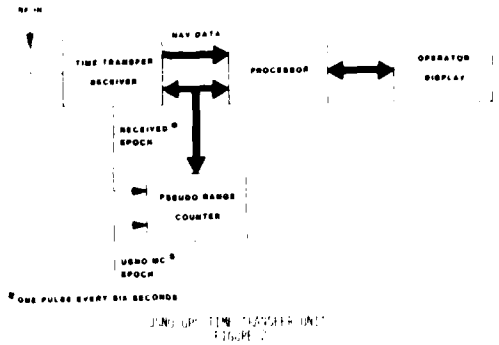
4) R. Vessot and M. Levine, "Preliminary Report on the Gravitational Redshift Rocket-Probe Experiment", Proceedings of the 2nd Frequency Standards and Metrology Symposium, July 1976, pp 659-673.

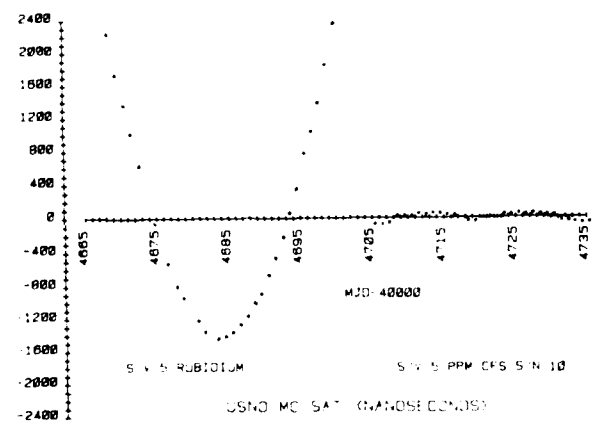
5) D. M. Tennant, These proceedings.

6) K. Putkovich, "Initial Test Results of USNO GPS Time Transfer Unit", Proceedings fo the 34th Annual Symposium on Frequency Control, May 1980, pp 326-333.

7) K. Putkovich, "USNO GPS Program", Proceedings of the 12th Annual Precise Time and Time Interval Applications and Planning Meeting, December 1980, pp 387-413.







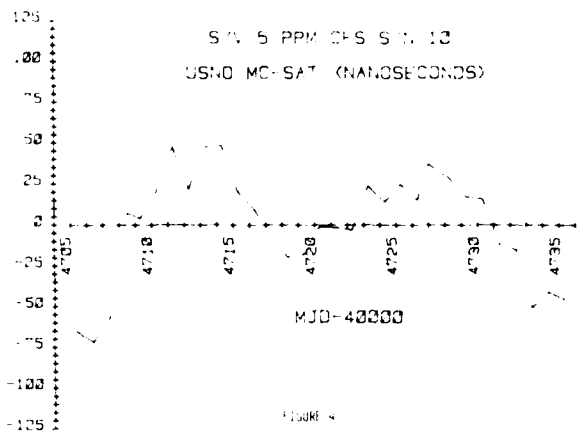
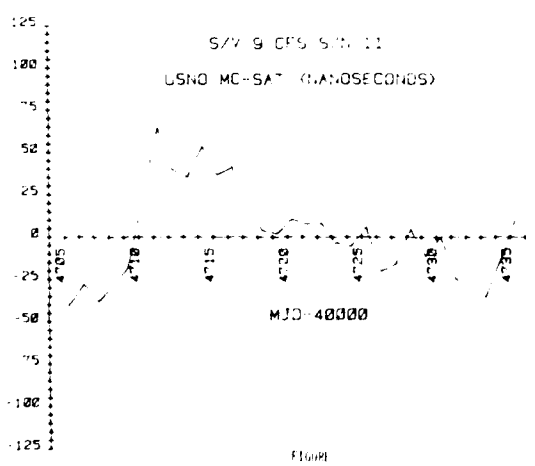


FIGURE 4



FIGURE

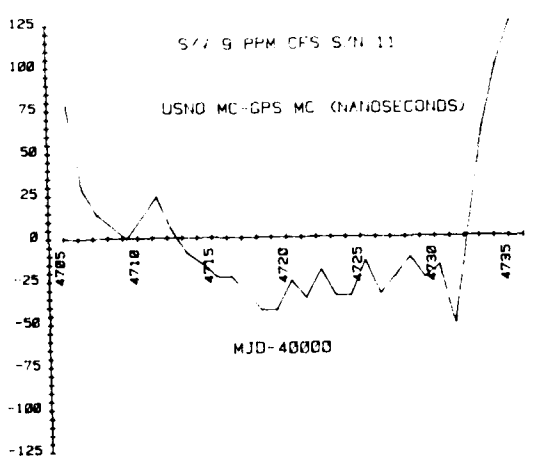


FIGURE 5

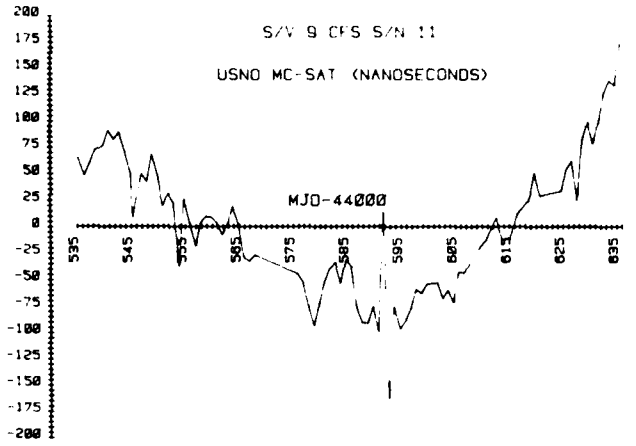

S/V 9 CFS S/N 11
USNO MC-SAT (NANOSECONDS)
MJD-44000
200
175
150
125
100
75
50
25
0
-25
-50
-75
-100
-125
-150
-175
-200
535
545
555
565
575
585
595
605
615
625
635

FIGURE 7

Proc. 35th Ann. Freq Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

Long Term Performance of VLG-11 Masers

J. White and K. McDonald

Naval Research Laboratory
Washington, D. C. 20375

SUMMARY

There has recently been increased interest in active hydrogen masers for use in timekeeping applications for space and military programs. Most of the published data on masers has related to their short term stability. The VLG-11 maser, one of two generally available, has been produced since 1977 but no long term stability or reliability information has been available to potential users. The U.S. Naval Research Laboratory (NRL) has over three years experience with Smithsonian Astrophysical Observatory's (SAO) VLG-11 masers. NRL purchased two of the first three masers built in this series in the fall of 1977. An additional maser was delivered in January, 1980. This paper documents the performance of these three masers and also an NRL owned VLG-10, the predecessor of the VLG-11 series.

Introduction

With the advent of increasingly tighter timing requirements in a variety of space, military, and private communication and navigation programs there is greater interest in very high performance clocks such as the hydrogen maser. While hydrogen masers were available commercially over a decade ago, only a few have been sold. The lack of widespread use has been due to several factors, including cost, size, weight, a reputation for poor reliability, and, probably most importantly, the availability of cesium standards that did not suffer as much from these problems. However, recent developments in programs such as the Global Positioning System and various spread spectrum communication systems have shown the need for performance exceeding that of current design cesium clocks.

There are presently two sources of active hydrogen masers in the United States. They are the Smithsonian Astrophysical Laboratory (SAO), which builds the VLG-11 maser and the Applied Physics Laboratory (APL), which builds the NR masers. There is also increasing interest in the passive technology masers. At present there is no commercial or institutional source for these masers although development work is taking place in several laboratories [1,2,3].

The U.S. Naval Research Laboratory has supported hydrogen maser development at SAO for over ten years beginning with the first SAO maser, the VLG-10 and continuing through the present series. NRL now owns one VLG-10 and three VLG-11 masers. There are maintenance records documenting failures and events which limit performance along with tuning and frequency offset data.

Reliability

One of the most commonly cited concerns relating to the application of hydrogen masers has been reliability. Early hydrogen masers suffered from a number of problems. There were not only the problems of low volume production but also physics problems relating to wall coatings, dissociators, pumps and cavity tuning. The experience gained through the SAO masers at NRL indicates that there has been a great improvement in reliability from the time of delivery of the first VLG-10 through recent VLG-11 deliveries.

Figure 1 shows a summary of the failures and significant events in operation of the NRL owned masers. This table covers the one VLG-10, serial number P0, and the three VLG-11 masers, P9, P10, and P12. P0 was the prototype VLG-10 maser delivered in 1972. P9 and P10 were the first two VLG-11 masers. They were delivered to NRL in the fall of 1977. Maser P12 is one of the most recent VLG-11 masers. It was delivered in January, 1980. Since that time two additional VLG-11 masers have been completed. P13 is the SAO house maser which is used as a testbed for new development ideas as well as the SAO house reference. P14 was delivered to the Jet Propulsion Laboratory. It should be noted at this time that both the VLG-10 and VLG-11 maser models vary in some respects from unit to unit depending on customer requirements and design changes made to improve performance. The data presented here is technically only applicable to the particular masers owned by NRL.

To set a reference for evaluating the current masers it is helpful to examine the

performance of earlier models. Based on experience with the first VLG-10 and also with two pre H-10 masers, manufactured by another source, in 1963, it is not hard to understand how masers got a bad name. The data on P0 show that there were significant problems with that maser. All but the most recent ion pump problems occurred during the first four years of its life. The most serious problem was with the initial dissociator design. From delivery until a change in dissociator design in 1975 there were constant problems with low signals and white hydrogen discharges. There were also problems with the maser receiver. These primarily related to the LC oscillator at 1.4 Ghz which was phase locked to the crystal oscillator to provide the first local oscillator in the receiver. Receiver design changes have eliminated this problem in later masers. For the period beginning in 1976 to the present, maser P0 has performed reliably. It has had some ion pump problems and the receiver problem has recurred from time to time.

As Figure 1 shows, the three VLG-11 masers have been significantly better. The first two masers, P9 and P10, both had some initial problems related to mechanical and electrical design. The physics problems of the earlier masers are not present. The newest unit, P12, has had no failures in the sixteen months it has been at NRL. The histories of P9 and P10 are documented in figures 2 and 3 respectively.

The problems with maser P9, figure 2, have all been electrical. Two of the three relate more to the design of the wiring harness than to the electronic design of the maser. The remaining problem initially appeared to be intermittent tuning of the cavity. This was later found to be due to failure to reset the maser from a test mode.

Looking next at figure 3 the nature of the problems with maser P10 has primarily been ion pump failure. In the first two years of life the ion pump power was cut off by the circuit breaker five times due apparently to excessive current being drawn in the pump. The pump plates were replaced and no further problems have been noted. P10 also had one failure related to the mechanical design. An intermittent condition developed where the tuning of the maser would jump. The problem was traced to the mounting of the tuning diode, which was subsequently corrected.

Maser P12 has had no failures or loss of operation.

Stability

The excellent short term stability of hydrogen masers for averaging time in the range of 1 to 10000 seconds has been well documented [4,5]. Users requiring good long term stability, however, have less information available. Long term stability is affected by frequency drift, barometric pressure sensitivity, temperature sensitivity and magnetic field sensitivity. The environmental factors have been measured and reported previously [4]. This report will show the long term frequency changes for masers P9 and P12. Data for maser P10 are not included due to the breaks in the data caused by ion pump and diode repairs.

Figure 4 shows the frequency offset between masers P9 and P12 for the year 1980. The short breaks are missing data. The large frequency jump is from retuning the maser cavities. The initial drift rate is about 3×10^{-14}/day. Tapering off slowly over the period. In order to determine how much of the total drift can be assigned to each maser the offsets of each with respect to a cesium standard, HP 5061/opt 004, are given in figures 5 and 6. It can be seen that both masers show positive frequency drift with respect to time. Maser P9 shows a drift of less than 1×10^{-14}/day while P12 has a rate starting around 5×10^{-14}/day. As a crosscheck the frequency offset of the reference cesium standard with respect to the U.S. Naval Observatory time scale is shown in figure 7. This comparison was made via TV line ten time transfer using WTTG in Washington. While this method produced noisy data, it does show that the drift is quite low.

A further reduction of the P12 data is possible. Figure 8 is a plot of the drift rate determined by making linear fits to the frequency for short segments of the data in figure 6. Figure 6 also shows the results of two cavity tunings. Since the frequency tends to return to the same point after each tuning it is reasonable to assume that most of the drift is due to a time dependent change in cavity tuning. The x axis for figure 8 is labeled to indicate the elapsed time since the final assembly and mechanical loading of the maser cavity.

The cavity tuning data from the logbooks can be used to independently show changes in cavity tuning to correlate the rate and source of the drift in the output frequency. This has been done for masers P9 and P12. Both masers have been tuned periodically. The cavity tuning was done by the conventional method of adjusting the cavity frequency at high and low source pressures to locate the point where the output frequency is pressure independent. Using the measured cavity tuning curves, figure 9, the approximate change in cavity frequency can be found for each tuning. An equation was fit to each tuning curve to convert the applied diode tuning voltage to cavity frequency. Using these relations the logged results of each cavity tuning were changed to sets of cavity frequencies. The starting point for each curve was the last tuning point at SAO prior to shipment. This was arbitrarily assigned as the reference tuning point. Figure 10 shows the relative changes in cavity tuning. As in figure 8 the x axis is scaled in time since final assembly. The points shown indicate the changes in frequency relative to the 1.4 GHz center frequency. The corresponding change in the maser output is about 1.7×10^{-12} for every

10^{-13} Hz of cavity drift. This allows a more direct comparison in the rates for the two masers. The drift rates calculated by this method are approximately the same as that measured for the output frequencies.

There has been some discussion on the possible causes of the cavity tuning drift. Based on input from SAO and JPL, the most probable theory centers on the mechanical settling of the joints in Cervit material used in the VLG-11 cavity. The frequency changes are all positive, indicating an increase in cavity frequency and so a decrease in the size of the cavity. Physically this could be explained by a compression of the cavity material. Such a compression would be greatest at the time of initial mechanical loading and should decrease over time.

There is no obvious explanation for the difference in tuning drift rates for these two masers. The P12 data is comparable to measurements from SAO on maser P13 [6] and JPL on maser P14 [7]. NRL maser P10 was evaluated at JPL in 1980 [8]. The drift rate at that time was on the order of 1×10^{-14}/day.

Within the measurement constraints of a cesium reference and a phase resolution of .1 nanosecond there were no observable effects which correlated with environmental factors, nor were any found when comparing the masers to each other through the same measurement system. This may be in part due to correlation of effects between masers.

SUMMARY

The VLG-11 masers at NRL have shown significantly improved reliability over older designs. Observed problems have been of two basic types, electronic and ion pump. The electronic problems should decrease with increasing production experience and the performance of the most recently acquired maser bears this out. The ion pump problem is a more difficult matter, common not only to hydrogen masers but also to cesium standards. A solution may come from the space passive maser programs at NRL and Hughes Research Laboratory [9,10]. A getter pump has been developed for the space program which has high hydrogen capacity while requiring no high voltage power. A small ion pump would still be required to pump residual gases, but should have a long life since the problems of flaking associated with pumping hydrogen would be eliminated. There is also the additional benefit of reduced stray magnetic fields with the removal of the large ion pump. NRL and SAO are engaged in the design of a pump retrofit modification for the VLG-10/11 masers to use the getter pump. SAO has prior experience with getter pumps from the red shift space probe maser program [11]. Preliminary tests have been done on maser P13 at SAO.

The long term frequency drift is of concern to timekeepers. Since this is a smooth process it should be possible to model the tuning change. The model could then be used to generate a periodic tuning correction to offset the effects of the drift. This would have the advantage of reducing changes in the output frequency without the need for a full tuning procedure at short intervals.

The authors wish to thank R.F.C. Vessot and E. Mattison of SAO and P. Kuhnle of JPL for their cooperation and stimulating discussion.

References

1. F. L. Walls, D. A. Howe, "A Passive Hydrogen Maser Frequency Standard", Proceedings, 32nd Annual Symposium on Frequency Control, U. S. Army Electronics Command, Ft. Monmouth, N.J. pp 492-498 (1978).
2. J. D. White, A. Frank, V. Folen, "Passive Maser Development at NRL", Proceedings, 12th Annual Precise Time and Time Interval Applications and Planning Meeting, pp 495-514 (1980).
3. E. M. Mattison, E. L. Blomberg, G. N. Nystrom, R. F. C. Vessot, "Design and Testing of a Small Passive Hydrogen Maser", Proceedings, 31st Annual Symposium on Frequency Control, U. S. Army Electronics Command, Ft. Monmouth, N. J., pp 549-553 (1977).
4. M. W. Levine, R. F. C. Vessot, E. L. Mattison, "Performance Evaluation of the VLG-11 Atomic Hydrogen Maser", Proceedings, 32nd Annual Symposium on Frequency Control, U. S. Army Electronics Command, Ft. Monmouth, N. J. pp 477-485 (1978).
5. V. S. Reinhardt, "Recent Progress in the NASA - Goddard Space Flight Center Atomic Hydrogen Standards Program", Proceedings, 12th Annual Precise Time and Time Interval Applications and Planning Meeting pp 463-494, (1980).
6. R. F. C. Vessot, private communication.
7. P. F. Kuhnle, private communication.
8. "Operating Parameters Characterization of SAO Hydrogen Maser, VLG-11, S/N 10", Jet Propulsion Laboratory report (1980).
9. S. A. Wolf, D. U. Gubser, L. D. Jones, "Vacuum Pumping System for Spaceborne Passive Hydrogen Masers", Proceedings, 12th Annual Precise Time and Time Interval Applications and Planning Meeting, pp 581-590 (1980).
10. H. T. M. Wang, "An Oscillating Compact Hydrogen Maser", Proceedings, 34th Annual Symposium on Frequency Control, U. S. Army Electronics Command, Ft. Monmouth, N. J., pp 364-369 (1980).
11. R. F. C. Vessot, M. W. Levine, "Performance Data of Space and Ground Hydrogen Masers and Ionospheric Studies for High Accuracy Comparisons Between Space and Ground Clocks", Proceedings, 28th Annual Symposium on Frequency Control, U. S. Army Electronics Command, Ft. Monmouth, N. J., pp 408-414 (1974).

SUMMARY OF FAILURES/EVENTS

SOURCE	P0	P9	P10	P12
CAVITY/SHIELDS	NONE	NONE	1	NONE
PUMPS	2	NONE	5	NONE
DISSOCIATOR	6	NONE	NONE	NONE
VALVE	1	NONE	NONE	NONE
RECEIVER	14	NONE	NONE	NONE
POWER SUPPLY/ WIRING	NONE	4	2	NONE

1. Summary of VLG-11 Maser Failures/Events

P9 FAILURES/EVENTS

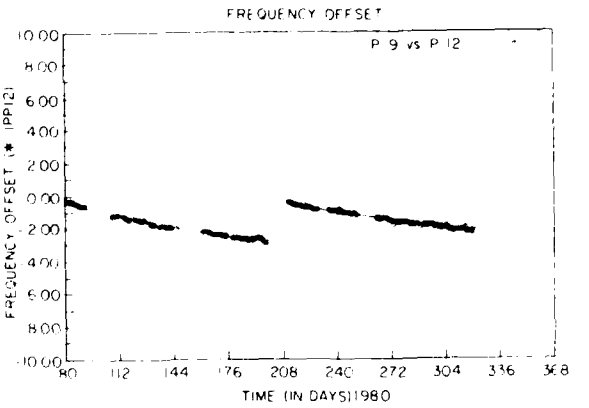
DATE	CIRCUIT	PROBLEM
5/23/78	RECEIVER OVEN PREAMP	OP AMP LOCK UP CHANGED RESISTOR
4/2/79	CAVITY TUNING	NO TUNING HUMAN ERROR
2/9/81	TANK DOME HEATER CONTROL	NO OUTPUT– INTERMITTANT OPEN THERMISTOR/WIRING
2/26/81	10 VOLT POWER SUPPLY	NO OUTPUT CONNECTOR PROBLEM

2. List of Maser P9 Failures/Events

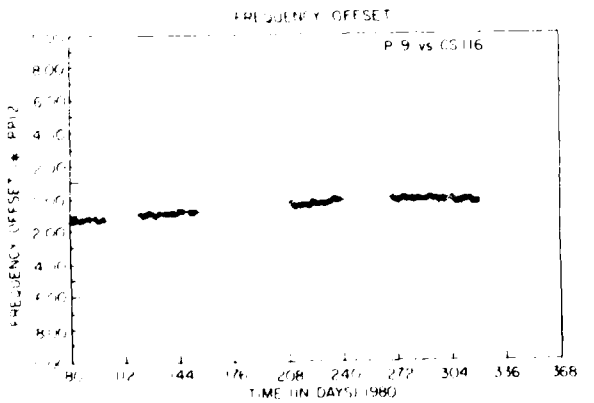
P10 FAILURES/EVENTS

DATE	CIRCUIT	PROBLEM
11/14/77	ION PUMP/POWER SUPPLY	BREAKER POPPED RESET
5/30/78	ION PUMP/POWER SUPPLY	BREAKER POPPED RESET
6/8/78	ION PUMP/POWER SUPPLY	BREAKER POPPED RESET
6/29/78	CAVITY TUNING	INTERMITTENT DIODE MOUNT REPAIRED
1/8/79	ION PUMP	SHORT RESTARTED
5/17/80	ION PUMP	SHORT REPLACED PUMP

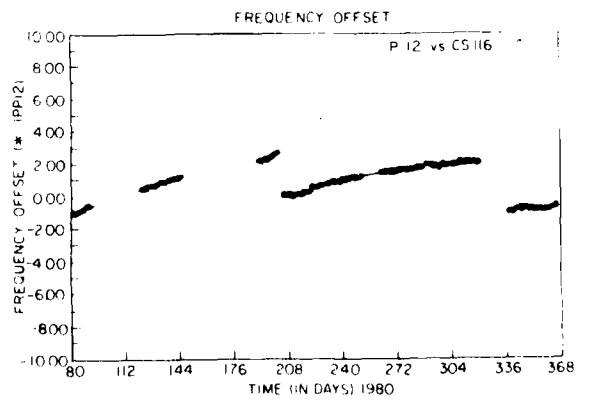
3. List of Maser P10 Failures/Events



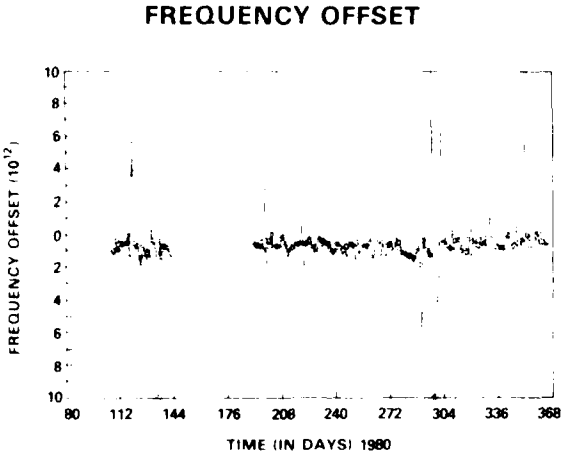
4. Frequency Offset of Masers P9 and P12



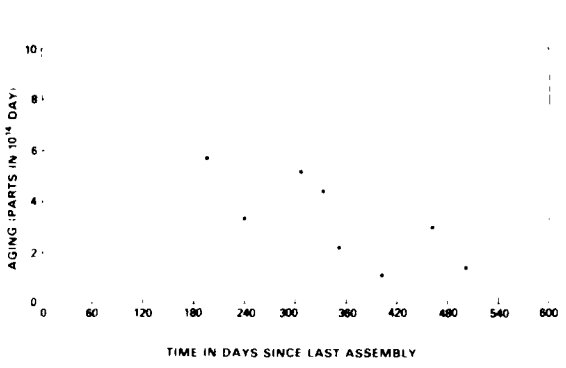
5. Frequency Offset of Maser P9 vs Cesium



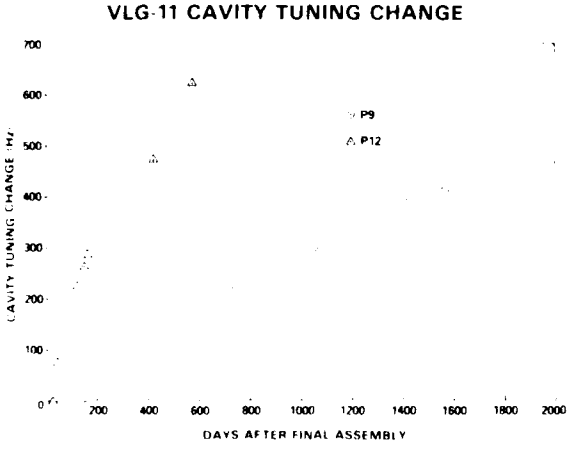
6. Frequency Offset of Maser P12 vs Cesium



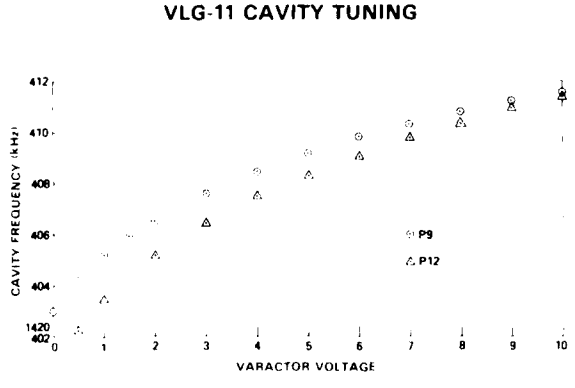
7. Frequency Offset of Cesium Reference vs USNO via WTTC



8. Measured Aging Rate for Maser P12



9. VLG-11 Cavity Tuning



10. Cavity Tuning Change for Masers P9 and P12

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703, May 1981

FEASIBILITY OF EXTREMELY SMALL HYDROGEN MASERS

H. E. Peters

Sigma Tau Standards Corporation
Tuscaloosa, Alabama

Abstract

The feasibility of achieving much smaller and lighter hydrogen masers than have been built to the present was pointed out in a paper presented at this symposium in 1978.[1] Experimental work to establish the feasibility of a "Small Hydrogen Maser," with physics unit only 7.0" inches in diameter and 10.5" inches long, has since been undertaken with support of the United States Air Force,[2] and progress with this work was presented in a paper, "New Hydrogen Maser Results," which was presented at the 34th Annual Symposium on Frequency Control in 1980.[3]

In the following sections progress with the "Small Hydrogen Maser" prototype which is currently being assembled is presented. In addition, the effect of cavity pulling on stability is discussed, and a feasible design for an "Extremely Small Hydrogen Maser," with a volume of [illegible] liters in diameter by [illegible] long, is cited.

Key words: Atomic oscillator; Hydrogen Maser; Small Hydrogen Standard; Frequency standard; Frequency stability.

Introduction

The active hydrogen maser oscillator has exceptionally good stability for all measurement intervals longer than one second; however it has found its greatest usefulness in specialized applications at standards laboratories, tracking stations, and radio astronomy observatories due to its remarkably superior stability for the intermediate measurement intervals between one second and one day. Limited from wider applications by its high cost, large size, and heavy weight, a very significant effort has been expended at many research and development laboratories in the last few years to overcome these restrictions without degrading the stability.

The present paper reports on the progress [illegible] in size and weight reduction by use of [illegible] and discusses the trade-offs involved in the storage bulb. The use of a new approach [illegible] with a minimum cost to achieve oscillation is presented and the effect of cavity pulling on the maser stability is considered. At the extreme size reduction of the "Extremely Small Hydrogen Maser," the performance achievable and the corresponding size and weight benefits are considered in the last section.

Small Hydrogen Maser

Figure 1 is a picture of the "Small Hydrogen Maser" which is presently being assembled. The instrumentation, controls and electronics subsystems have been completed. The processing of physical sub-systems such as source, storage bulb, cavity and vacuum enclosure is in progress at this writing; these fit within the blue cylinder in the center of the framework.

The electronics sub-systems have been packaged in functionally separate modules, each of which may be uncovered for operational testing without disconnection of power. They may also be removed, repaired, or replaced as units in case of malfunction. The modules on the front panel are: 1. Vac-Ion pump supply; this is a DC-DC converter which provides 3,000 volts for the pump, 2. The source pressure control module which automatically regulates hydrogen flow, 3. The receiver synthesizer; this supplies the 5MHz reference frequency for the receiver phase lock loop - there are 11 decades of control which give a resolution of 5×10^{-15} for the output frequencies, 4. The receiver VCO and output buffer amplifiers, 5. The receiver local oscillator multiplier and IF amplifier module, 6. The module containing the magnetic field and cavity frequency controls, 7. The instrumentation readout module. There are 16 readout channels which are selected by 4 binary coded switches to provide visual indication of variables on a 4½ decade digital panel meter.

The power supply has been placed in a module mounted at the rear of the frame. The cover may be removed for changing connections or trouble shooting without disconnecting the power. Batteries for uninterruptable standby operation are placed in a separate external battery pack for long term operation without AC power, a 12 Ah [illegible] battery is used which will last for [illegible] years.

Breadboard Maser Tests

First tests of a breadboard hydrogen maser which had cavity and bulb dimensions similar to the prototype maser presently under construction indicated that the maser did not oscillate at the hydrogen flux available. The beam was observed using pulsed stimulated emission. After optimizing the source collimator and realigning the beam optics, the situation was improved, but the maser still would not quite oscillate. While the breadboard maser oscillation parameters were not as ideal as might be achieved with improved techniques of bulb and cavity processing, it was decided to try a new method of cavity Q enhancement as an alternative design approach to assure efficient operation and to facilitate tests. This was successful, and using this method the breadboard maser oscillated. The effect on maser stability and the method used are discussed in the following sections.

The minimum cavity Q for oscillation was approximately 20,000, the hydrogen flux was 5×10^{-5} Torr-liter/second, and the line Q was 9×10^{8}. Use of a smaller line Q in the prototype maser, and larger bulb aperture, will lower the oscillation threshold and lessen the effect of spin-exchange in reducing oscillation amplitude.

Cavity Pulling

Factors which relate to the resonant cavity surrounding the maser storage bulb are most critical in determining whether the maser will oscillate, how high the hydrogen beam flux must be, and to what extent the maser output frequency is changed by cavity pulling. If the cavity changes by an amount Δf_c, the oscillation frequency will change by an amount $\Delta f_m = (Q_c/Q_l)\,\Delta f_c$. Thus we would like to use a low cavity Q. However, the flux required for oscillation, as well as the spin-exchange parameter, vary as $(1/Q_c)$, so too low a cavity Q results in unreliable operation, low signal to noise ratio, a requirement for high flux, and early pump saturation. So one is forced to strike a balance between achieving a practical oscillating maser and attainment of the best long term stability.

In the small hydrogen maser, without Q enhancement, the ratio of cavity Q to line Q is approximately 1.2×10^{-5} which is comparable to the value attained in conventional large hydrogen masers. If active Q enhancement by some factor is used, the maser will be that much more sensitive to cavity frequency changes. Also, depending upon the technique used, there may be severe electronic or environmental perturbations of the maser output frequency.

While it is clearly desirable to minimize the amount of Q enhancement used, the small ratio of Q_c to Q_l, as well as the low cavity thermal coefficient, permits Q_c to be raised significantly while still maintaining good long term stability, providing some of the problems inherent in previous attempts to use Q enhancement can be avoided.

Active Cavity Q Enhancement

Figure 2 illustrates two approaches to hydrogen maser cavity Q enhancement. Figure 2a illustrates the methods described in references (4) and (5). Two coupling loops are used and coaxial cable connects these to a variable gain amplifier, a band-pass filter, and a phase matching circuit so that a signal of proper amplitude and phase is returned to the cavity to reduce the losses. The amplifier, filter, phase matching network, and coaxial connections are outside the thermally controlled and shielded region of the maser cavity. The inherent mechanical, thermal, and electronic perturbations inherent in this method introduce unaceptable frequency perturbations for use as a standard. Wang[5] has added a cavity stabilization servo to avoid these problems, however, this entails additional complexity, as well as new perturbations.

Figure 2b illustrates the method developed for use in the small hydrogen maser. In this case, a transistor amplifier is coupled directly to a single loop within the maser cavity. The phase and gain parameters are adjusted by varying the bias voltage. The circuit is mounted on the cavity adjacent to the magnetic shield. Using swept frequency methods, the enhanced cavity resonance is measured just as in a conventional maser cavity without gain. The signal is detected through a separate, independent, loop using very light coupling. It should be emphasized that the only real difference between these two figures is that there are no RF components, other than the output connection, outside the maser inner magnetic shield in Figure 2b. Thus one may achieve the degree of thermal and mechanical stability typical of a single cavity structure.

Extremely Small Hydrogen Maser

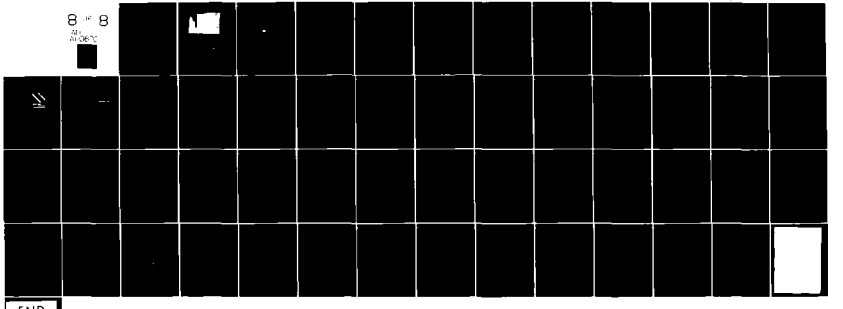
If active cavity Q enhancement is to be used, the usual restrictions on cavity size are greatly reduced. In fact, by reducing cavity size, one of the most difficult problems can be eliminated. In a large cavity there are numerous high Q resonant modes. To avoid spurious oscillations one must couple only to the proper mode, or assure that the gain at extraneous resonance frequencies is small. This can be accomplished by a band-pass filter in the conventional method of figure 2a, or by proper configuration of the cavity and judicious orientation of the coupling loop in the method of figure 2b.

A much better way is to make the cavity small enough that the only mode within the band-pass of the cavity amplifier is the one desired at the hydrogen frequency. A cavity 4 inches long by 2 inches diameter, with bulb 1 inch in diameter having electrodes configured as in reference (1),

AD-A110 870 ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND AD--ETC F/G 9/5
PROCEEDINGS OF THE 35TH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL, --ETC(U)
1981

UNCLASSIFIED

NL



END
DATE FILMED
3-82
DTIC

may be made resonant in the proper mode, and extraneous resonances are an octave or more higher in frequency. A test cavity of copper, with a quartz tube, which closely approximated the desired physical shape and electrical parameters, has been measured. The loaded Q ranged from 5,000 with no active gain, to over 50,000 with gain, and could be controlled by bias voltage variation within this range. There were no other resonances within the 1,800 MHz maximum frequency of the instrumentation.

The line Q of a maser with a one inch diameter bulb would be about 4×10^8, extrapolating from the 2 inch diameter bulb of the SHM or the 5 inch diameter bulb typical of large hydrogen masers. With an enhanced cavity Q of 24,000, the pulling factor is only 6×10^{-5}. The threshold for oscillation will go down directly as the bulb diameter, so less hydrogen flux will be required. The spin exchange parameter will remain essentially unchanged from that of larger masers. Of course, the intrinsic accuracy will be about a factor of 5 worse than that of large hydrogen masers, so use as a fundamental standard is not anticipated, but where the requirement is for exceptional stability for periods up to a few days or a few weeks, the ESHM may be very useful.

Figure 3 is a drawing which illustrates the remarkable size reduction that may be accomplished. This shows the "physics package" of a design which is only 5 inches in diameter and 10 inches in length. This is also, most likely, the smallest practical size, since the source assembly is about the size of the cavity, and may not be reduced significantly without encountering dissociator problems. The pump and the required electronics are also comparable in weight and size to the physics package, thus there is little motivation to further reduce the physics package size.

Stability Goals

Figure 4 illustrates the stability goals for the "Extremely Small Hydrogen Maser," as well as the expected performance of the "Small Hydrogen Maser." For reference, the stability currently realized with typical good conventional large hydrogen masers is shown.

Though the stability of the SHM (or the ESHM) has not yet been measured, there is a good basis for predicting what it should be if we achieve cavity stability similar to that which is obtained with analogous electro-mechanical oscillators under optimum environmental conditions. It should be emphasized that the hydrogen maser may be viewed as a device which improves upon the stability of a cavity resonator by the ratio of line Q to cavity Q - typically about a factor of 10^5.

Theoretical noise limits do not provide a good basis for estimating the stability of a gain-enhanced cavity system because long term systematic variables or the degree of environmental isolation achieved are the predominant considerations. Consideration of the performance of devices such as super-conducting cavity stabilized oscillators, or crystal oscillators, which depend directly on gross physical dimensions just as a microwave cavity does, give a very optimistic view of the stability potential for a maser with enhanced cavity Q. For example, improvement by a factor of 6×10^{-5} over the stability of a very good crystal oscillator would give stabilities in the 10^{-17} range (other limits, such as perturbing noise or additive noise, would actually be encountered before this level is achieved.)

Thus the simple addition of an amplifier in a cavity loop does not inherently make the system unstable in frequency. There is, therefore, a very good basis for predicting the stabilities illustrated in figure 4 for the "Extremely Small Hydrogen Maser" as well as for the "Small Hydrogen Maser."

Conclusion

The work reported herein has been undertaken to demonstrate that the excellent stability properties of a good hydrogen maser may be achieved with practical masers which are an order of magnitude smaller and lighter than the conventional large maser. The present results and considerations provide an excellent basis for confidence that this goal may be attained.

References

1. Harry E. Peters, "Small, Very Small, and Extremely Small Hydrogen Masers," Proceedings, 32nd Annual Symposium on Frequency Control, USAERADCOM, Fort Monmouth, NJ, 1978.

2. This work is supported by the U.S. Air Force, RADC(ET), Deputy for Electronic Technology, Hanscom AFB, MA.

3. H.E. Peters, "New Hydrogen Maser Designs," Proceedings, 34th Annual Symposium on Frequency Control, USAERADCOM, Fort Monmouth, NJ, 1980.

4. C. Adoin, M. Desaintfuscien, and J.P. Schermann, "Application of The Transient Behavior to The Hydrogen Maser," Proceedings, 22nd Annual Symposium of Frequency Control, U.S. Army Electronics Command, Fort Monmouth, NJ, 1968.

5. H.T.M. Wang, "An Oscillating Compact Hydrogen Maser," Proceedings, 34th Annual Symposium on Frequency Control, USAERADCOM, Fort Monmouth, NJ, 1980.

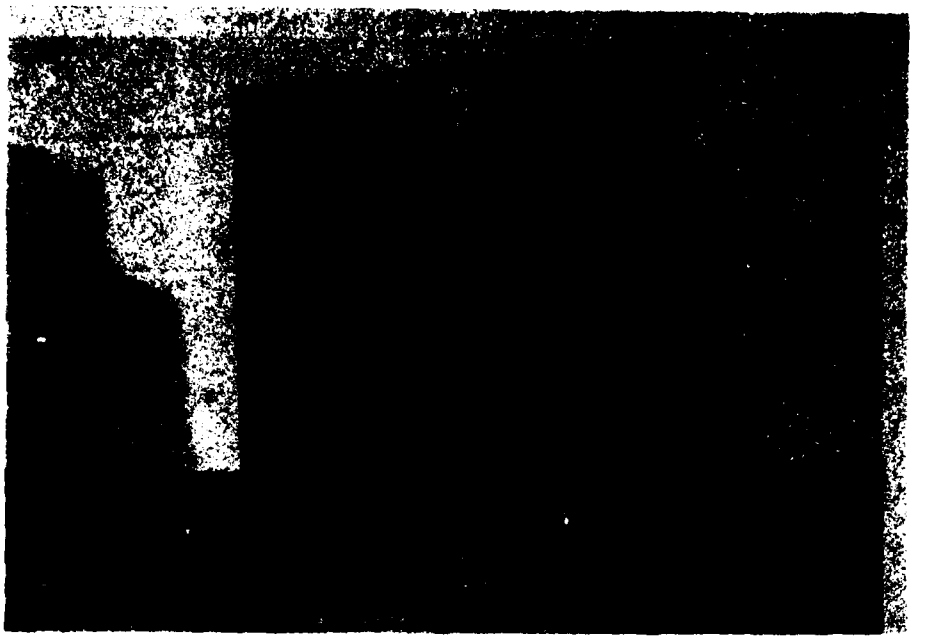


Figure 1. Small Hydrogen Maser.

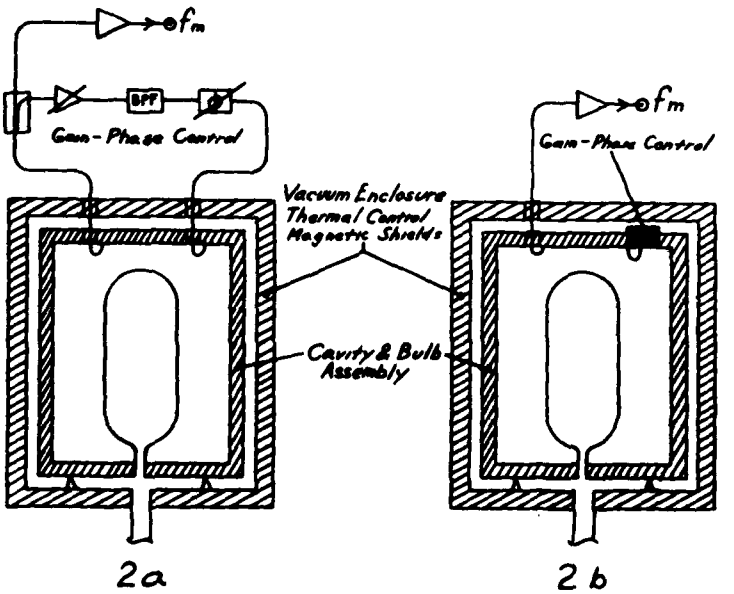


Figure 2. Cavity Active Q Enhancement Methods.

EXTREMELY SMALL HYDROGEN MASER

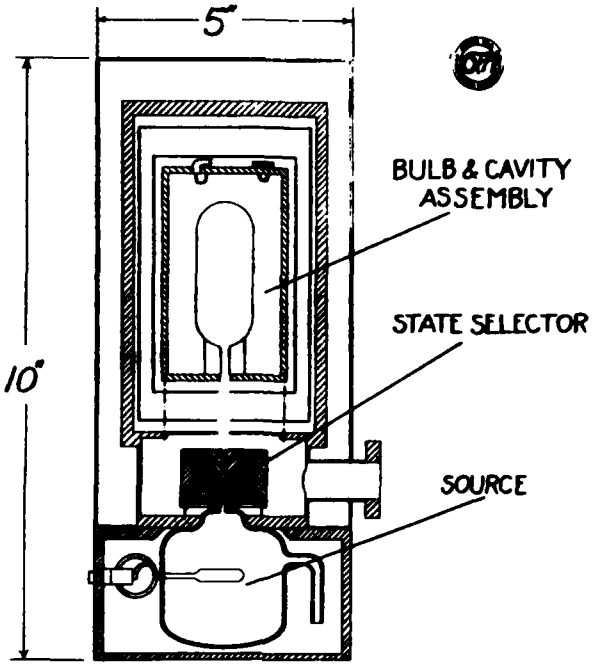


Figure 3. Extremely Small Hydrogen Maser.

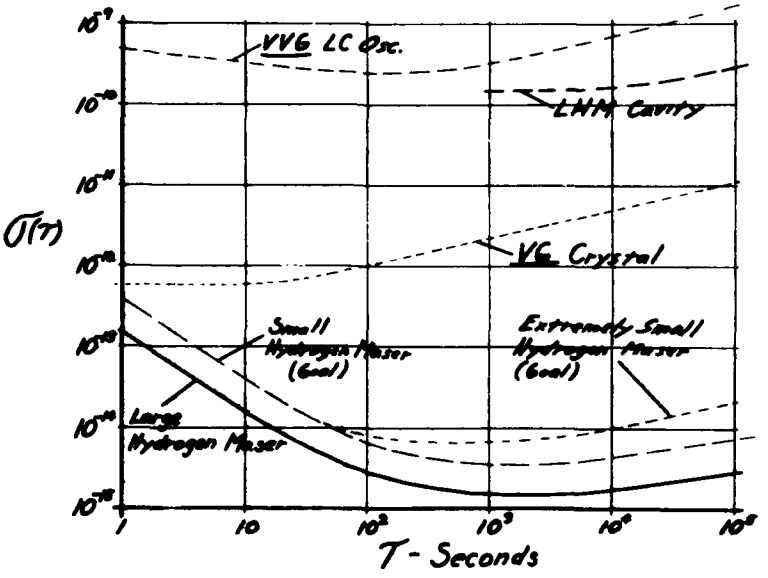


Figure 4. Stability Goals.

PROPERTIES OF SIGNAL SOURCES AND MEASUREMENT METHODS

D. A. Howe, D. W. Allan, and J. A. Barnes

TABLE OF CONTENTS

Summary A1

Introduction. A1

I. The Sine Wave and Stability A1
1.1 Common Methods of Measuring Frequency Stability A1
A. Beat Frequency Method A4
B. Dual Mixer Time Difference (DMTD) System. A4
C. Loose Phase Lock Loop Method. A5
D. Tight Phase Lock Loop Method. A6
E. Time Difference Method. A7

II. Measurement Methods Comparison. A7

III. Characterization. A9
3.1 Non-random Fluctuations A9
3.2 Random Fluctuations A10

IV. Analysis of Time-Domain Data. A10

V. Confidence of the Estimate and Overlapping Samples. A13

VI. Maximal Use of the Data and Determination of the Degrees of Freedom. A15
6.1 Use of Data A15
6.2 Determining the Degrees of Freedom. A16

VII. Example of Time-Domain Signal Processing and Analysis A17

VIII. Spectrum Analysis A19
8.1 The Loose Phase Locked Loop A21
8.2 Equipment for Frequency Domain Stability Measurements. A22
8.3 Procedure and Example. A24

IX. Power-Law Noise Processes A25

X. Pitfalls in Digitizing the Data A26
10.1 Discrete-continuous Processes A26
10.2 Digitizing the Data A26
10.3 Aliasing. A27
10.4 Some History of Spectrum Analysis Leading to the Fast Fourier Transform A29
10.5 Leakage A31
10.6 Picket Fence Effect A33
10.7 Time Domain-Frequency Domain Transforms A34
A. Integral Transform. A34
B. Fourier Series. A34
C. Discrete Fourier Transform. A35

XI. Translation from Frequency Domain Stability Measurement to Time Domain Stability Measurement and Vice-Versa A35
11.1 Procedure A35

XII. Causes of Noise Properties in a Signal Source A37
12.1 Power-law Noise Processes A37
12.2 Other Types of Noise. A38

XIII. Conclusion. A42

References. A42

Bibliography. A42

Appendix. A44

Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, Ft. Monmouth, NJ 07703. May 1981

PROPERTIES OF SIGNAL SOURCES AND MEASUREMENT METHODS

D. A. Howe, D. W. Allan, and J. A. Barnes

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

Summary

This paper is a review of frequency stability measurement techniques and of noise properties of frequency sources.

First, a historical development of the usefulness of spectrum analysis and time domain measurements will be presented. Then the rationale will be stated for the use of the two-sample (Allan) variance rather than the classical variance. Next, a range of measurement procedures will be outlined with the trade-offs given for the various techniques employed. Methods of interpreting the measurement results will be given. In particular, the five commonly used noise models (white PM, flicker PM, white FM, flicker FM, and random walk FM) and their causes will be discussed. Methods of characterizing systematics will also be given. Confidence intervals on the various measures will be discussed. In addition, we will point out methods of improving this confidence interval for a fixed number of data points.

Topics will be treated in conceptual detail. Only light (fundamental) mathematical treatment will be given.

Although traditional concepts will be detailed, two new topics will be introduced in this paper: (1) accuracy limitations of digital and computer-based analysis and (2) optimizing the results from a fixed set of input data.

The final section will be devoted to fundamental (physical) causes of noise in commonly used frequency standards. Also transforms from time to frequency domain and vice-versa will be given.

Key Words. Frequency stability; Oscillator noise modeling; Power law spectrum; Time-domain stability; Frequency-domain stability; White noise; Flicker noise.

Introduction

Precision oscillators play an important role in high speed communications, navigation, space tracking, deep space probes and in numerous other important applications. In this paper, we will review some precision methods of measuring the frequency and frequency stability of precision oscillators. Development of topics does not rely heavily on mathematics. The equipment and set-up for stability measurements are outlined. Examples and typical results are presented. Physical interpretations of common noise processes are discussed. A table is provided by which typical frequency domain stability characteristics may be translated to time domain stability characteristics and vice-versa.

I. THE SINE WAVE AND STABILITY

A sine wave signal generator produces a voltage that changes in time in a sinusoidal way as shown in figure 1.1. The signal is an oscillating signal because the sine wave repeats itself. A cycle of the oscillation is produced in one period "T". The phase is the angle "Φ" within a cycle corresonding to a particular time "t".

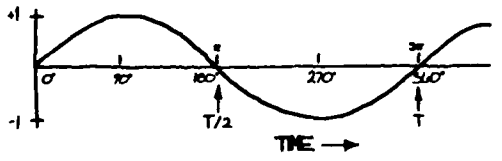


FIGURE 1.1

It is convenient for us to express angles in radians rather than in units of degrees, and positive zero-crossings will occur at even multiples of π-radians. The frequency "ν" is the number of cycles in one second, which is the reciprocal of period (seconds per cycle). The expression describing the voltage "V" out of a sine wave signal generator is given by $V(t) = V_p \sin[\Phi(t)]$ where V_p is the peak voltage amplitude. Equivalent expressions are

$$V(t) = V_p \sin\left(2\pi \frac{t}{T}\right)$$

and

$$V(t) = V_p \sin (2\pi\nu t).$$

Consider figure 1.2. Let's assume that the maximum value of "V" equals 1, hence "V_p" = 1. We say that the voltage "V(t)" is normalized to unity. If we know the frequency of a signal and if the signal is a sine wave, then we can determine the incremental change in the period "T" (denoted by Δt) at a particular angle of phase.

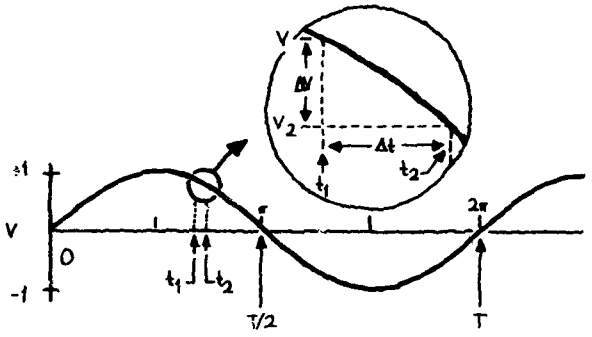


FIGURE 1.2

Note that no matter how big or small Δt may be, we can determine ΔV. Let us look at this from another point of view. Suppose we can measure ΔV and Δt. From this, there is a sine wave at a unique minimum frequency corresponding to the given ΔV and Δt. For infinitesimally small Δt, this frequency is called the instantaneous frequency at this t. The smaller the interval Δt, the better the approximation of instantaneous frequency at t.

When we speak of oscillators and the signals they produce, we recognize that an oscillator has some nominal frequency at which it operates. The "frequency stability" of an oscillator is a term used to characterize the frequency fluctuations of the oscillator signal. There is no formal definition for "frequency stability". However, one usually refers to frequency stability when comparing one oscillator with another. As we shall see later, we can define particular aspects of an oscillator's output then draw conclusions about its relative frequency stability. In general terms,

"Frequency stability is the degree to which an oscillating signal produces the same value of frequency for any interval, Δt, throughout a specified period of time".

Let's examine the two waveforms shown in figure 1.3. Frequency stability depends on the amount of time involved in a measurement. Of the two oscillating signals, it is evident that "2" is more stable than "1" from time t_1 to t_3 assuming the horizontal scales are linear in time.

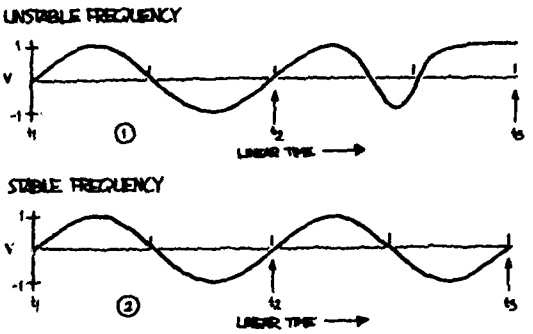


FIGURE 1.3

From time t_1 to t_2, there may be some question as to which of the two signals is more stable, but it's clear that from time t_2 to t_3, signal "1" is at a different frequency from that in interval t_1 to t_2.

If we want an oscillator to produce a particular frequency ν_0, then we're correct in stating that if the oscillator signal frequency deviates from ν_0 over any interval, this is a result of something which is undesirable. In the design of an oscillator, it is important to consider the sources of mechanisms which degrade the oscillator's frequency stability. All undesirable mechanisms cause random (noise) or systematic processes to exist along with the sine wave signal of the oscillator. To account for the noise components at the output of a sine wave signal generator, we can express the output as

$$V(t) = [V_0 + \varepsilon(t)] \sin [2\pi\nu_0 t + \phi(t)]. \quad (1.1)$$

where V_0 ≡ nominal peak voltage amplitude,
$\varepsilon(t)$ ≡ deviation of amplitude from nominal,
ν_0 ≡ nominal fundamental frequency,
$\phi(t)$ ≡ deviation of phase from nominal.

Ideally "ε" and "ϕ" should equal zero for all time. However, in the real world there are no perfect oscillators. To determine the extent of the noise components "ε" and "ϕ", we shall turn our attention to measurement techniques.

The typical precision oscillator, of course, has a very stable sinusoidal voltage output with a frequency ν and a period of oscillation T, which is the reciprocal of the frequency ($\nu = 1/T$). One goal is to measure the frequency and/or the frequency stability of the sinusoid. Instability is actually measured, but with little confusion it is often called stability in the literature. Naturally, fluctuations in frequency correspond to fluctuations in the period. Almost all frequency measurements, with very few exceptions, are measurements of phase or of the period fluctuations in an oscillator, not of frequency, even though the frequency may be the readout. As an example, most frequency counters sense the zero (or near zero) crossing of the sinusoidal voltage, which is the point at which the voltage is the most sensitive to phase fluctuations.

One must also realize that any frequency measurement involves two oscillators. In some instances, one oscillator is in the counter. It is impossible to purely measure only one oscillator. In some instances one oscillator may be enough better than the other that the fluctuations measured may be considered essentially those of the latter. However, in general because frequency measurements are always dual, it is useful to define:

$$y(t) = \frac{\nu_1 - \nu_0}{\nu_0} \qquad (1.2)$$

as the fractional frequency difference or deviation of oscillator one, ν_1, with respect to a reference oscillator ν_0 divided by the nominal frequency ν_0. Now, y(t) is a dimensionless quantity and useful in describing oscillator and clock performance; e.g., the time deviation, x(t), of an oscillator over a period of time t, is simply given by:

$$x(t) = \int_0^t y(t')dt' \qquad (1.3)$$

Since it is impossible to measure instantaneous frequency, any frequency or fractional frequency measurement always involves some sample time, Δt or "τ"--some time window through which the oscillators are observed; whether it's a picosecond, a second, or a day, there is always some sample time. So when determining a fractional frequency, y(t), in fact what is happening is that the time deviation is being measured say starting at some time t and again at a later time, $t + \tau$. The difference in these two time deviations, divided by τ gives the average fractional frequency over that period τ:

$$\bar{y}(t) = \frac{x(t + \tau) - x(t)}{\tau} \quad . \qquad (1.4)$$

Tau, τ, may be called the sample time or averaging time; e.g., it may be determined by the gate time of a counter.

What happens in many cases is that one samples a number of cycles of an oscillation during the preset gate time of a counter; after the gate time has elapsed, the counter latches the value of the number of cycles so that it can be read out, printed, or stored in some other way. Then there is a delay time for such processing of the data before the counter arms and starts again on the next cycle of the oscillation. During the delay time (or process time), information is lost. We have chosen to call it dead time and in some instances it becomes a problem. Unfortunately for data processing in typical oscillators the effects of dead time often hurt most when it is the hardest to avoid. In other words, for times that are short compared to a second when it is very difficult to avoid dead time, that is usually where dead time can make a significant difference in the data analysis. Typically for many oscillators, if

the sample time is long compared to a second, the dead time makes little difference in the data analysis, unless it is excessive.[1] New equipment or techniques are now available which contribute zero or negligible dead time.[2]

In reality, of course, the sinusoidal output of an oscillator is not pure, but it contains noise fluctuations as well. This section deals with the measurement of these fluctuations to determine the quality of a precision signal source.

We will describe five different methods of measuring the frequency fluctuations in precision oscillators.

1.1 Common Methods of Measuring Frequency Stability

A. Beat frequency method

The first system is called a heterodyne frequency measuring method or beat frequency method. The signal from two independent oscillators are fed into the two ports of a double balanced mixer as illustrated in figure 1.4.

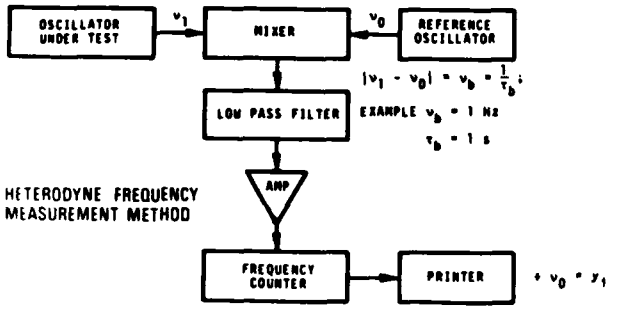


FIGURE 1.4

The difference frequency or the beat frequency, ν_b, is obtained as the output of a low pass filter which follows the mixer. This beat frequency is then amplified and fed to a frequency counter and printer or to some recording device. The fractional frequency is obtained by dividing ν_b by the nominal carrier frequency ν_0. This system has excellent precision; one can measure essentially all state-of-the-art oscillators.

B. Dual mixer time difference (DTMD) system

This system shows some significant promise and has just begun to be exploited. A block diagram is shown is figure 1.5.

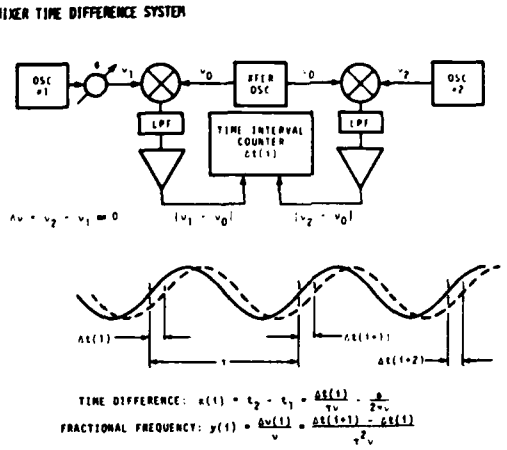


FIGURE 1.5

To preface the remarks on the DMTD, it should be mentioned that if the time or the time fluctuations can be measured directly, an advantage is obtained over just measuring the frequency. The reason is that one can calculate the frequency from the time without dead time as well as know the time behavior. The reason, in the past, that frequency has not been inferred from the time (for sample times of the order of several seconds and shorter) is that the time difference between a pair of oscillators operating as clocks could not be measured with sufficient precision (commercially the best that is available is 10^{-11} seconds). The system described in this section demonstrates a precision of 10^{-13} seconds. Such precision opens the door to making time measurements as well as frequency and frequency stability measuements for sample times as short as a few milliseconds and longer, all without dead time.

In figure 1.5, oscillator 1 could be considered under test and oscillator 2 could be considered the reference oscillator. These signals go to the ports of a pair of double balanced mixers. Another oscillator with separate symmetric buffered outputs is fed to the remaining other two

ports of the pair of double balanced mixers. This common oscillator's frequency is offset by a desired amount from the other two oscillators. Then two different beat frequencies come out of the two mixers as shown. These two beat frequencies will be out of phase by an amount proportional to the time difference between oscillator 1 and 2—excluding the differential phase shift that may be inserted. Further, the beat frequencies differ in frequency by an amount equal to the frequency difference between oscillators 1 and 2.

This measurement technique is very useful where one has oscillator 1 and oscillator 2 on the same frequency. This is typical for atomic standards (cesium, rubidium, and hydrogen frequency standards).

Illustrated at the bottom of figure 1.5 is what might represent the beat frequencies out of the two mixers. A phase shifter may be inserted as illustrated to adjust the phase so that the two beat rates are nominally in phase; this adjustment sets up the nice condition that the noise of the common oscillator tends to cancel (for certain types of noise) when the time difference is determined. After amplifying these beat signals, the start port of a time interval counter is triggered with the positive zero crossing of one beat and the stop port with the positive zero crossing of the other beat. Taking the time difference between the zero crossings of these beat frequencies, one measures the time difference between oscillator 1 and oscillator 2, but with a precision which has been amplified by the ratio of the carrier frequency to the beat frequency (over that normally achievable with this same time interval counter). The time difference x(i) for the i^{th} measurement between oscillators 1 and 2 is given by eq (1.5).

$$x(i) = \frac{\Delta t(i)}{\tau_b \nu_o} - \frac{\Phi}{2\pi\nu_o} + \frac{k}{\nu_o} \qquad (1.5)$$

where $\Delta t(i)$ is the i^{th} time difference as read on the counter, τ_b is the beat period, ν_o is the nominal carrier frequency, Φ is the phase delay in radians added to the signal of oscillator 1, and k is an integer to be determined in order to remove the cycle ambiguity. It is only important to know k if the absolute time difference is desired; for frequency and frequency stability measurements and for time fluctuation measurements, k may be assumed zero unless one goes through a cycle during a set of measurements. The fractional frequency can be derived in the normal way from the time fluctuations.

$$y_{1,2}(i, \tau) = \begin{cases} \dfrac{\nu_1(i, \tau) - \nu_2(i, \tau)}{\nu_o} \\ \dfrac{x(i+1) - x(i)}{\tau} \\ \dfrac{\Delta t(i+1) - \Delta t(i)}{\tau_b^2 \nu_o} \end{cases} \qquad (1.6)$$

In eqs (1.5) and (1.6), assumptions are made that the transfer (or common) oscillator is set at a lower frequency than oscillators 1 and 2, and that the voltage zero crossing of the beat $\nu_1 - \nu_c$ starts and that $\nu_2 - \nu_c$ stops the time interval counter. The fractional frequency difference may be averaged over any integer multiple of τ_b:

$$y_{1,2}(i, m\tau_b) = \frac{x(i+m) - x(i)}{m\tau_b} \qquad (1.7)$$

where m is any positive integer. If needed, τ_b can be made to be very small by having very high beat frequencies. The transfer (or common) oscillator may be replaced with a low phase-noise frequency synthesizer, which derives its basic reference frequency from oscillator 2. In this set-up the nominal beat frequencies are simply given by the amount that the output frequency of the synthesizer is offset from ν_2. Sample times as short as a few milliseconds are easiliy obtained. Logging the data at such a rate can be a problem without special equipment. The latest NBS time scale measurement system is based on the DMTD and is yielding an excellent cost benefit ratio.

C. Loose phase lock loop method

This first type of phase lock loop method is illustrated in figure 1.6. The signal from an oscillator under test is fed into one port of a

mixer. The signal from a reference oscillator is fed into the other port of this mixer. The signals are in quadrature, that is, they are 90 degrees out

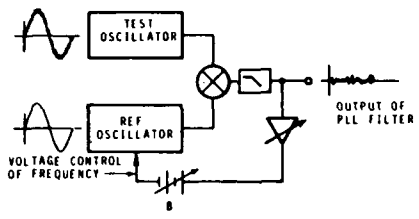


FIGURE 1.6

of phase so that the average voltage out of the mixer is nominally zero, and the instantaneous voltage fluctuations correspond to phase fluctuations rather than to amplitude fluctuations between the two signals. The mixer is a key element in the system. The advent of the Schottky barrier diode was a significant breakthrough in making low noise precision stability measurements. The output of this mixer is fed through a low pass filter and then amplified in a feedback loop, causing the voltage controlled oscillator (reference) to be phase locked to the test oscillator. The attack time of the loop is adjusted such that a very loose phase lock (long time constant) condition exists. This is discussed later in section VIII.

The attack time is the time it takes the servo system to make 70% of its ultimate correction after being slightly disturbed. The attack time is equal to $1/\pi w_h$, where w_h is the servo bandwidth. If the attack time of the loop is about a second then the voltage fluctuations will be proportional to the phase fluctuations for sample times shorter than the attack time. Depending on the coefficient of the tuning capacitor and the quality of the oscillators involved, the amplification used may vary significantly but may typically range from 40 to 80 dB via a good low noise amplifier. In turn this signal can be fed to a specturm analyzer to measure the Fourier components of the phase fluctuations. This system of frequency-domain analysis is discussed in sections VIII to X. It is of particular use for sample times shorter than one second (for Fourier frequencies greater than 1 Hz) in analyzing the characteristics of an oscillator. It is specifically very useful if one has discrete side bands such as 60Hz or detailed structure in the spectrum. How to characterize precision oscillators using this technique will be treated in detail later in section IX and XI.

One may also take the output voltage from the above amplifier and feed it to an A/D converter. This digital output becomes an extremely sensitive measure of the short term time or phase fluctuations between the two oscillators. Precisions of the order of a picosecond are easily achievable.

D. Tight phase lock loop method

The second type of phase lock loop method (shown in figure 1.7) is essentially the same as the first in figure 1.6 except that in this case the loop is in a tight phase lock condition; i.e., the attack time of the loop should be of the order of a few milliseconds. In such a case, the phase fluctuations are being integrated so that the voltage output is proportional to the frequency

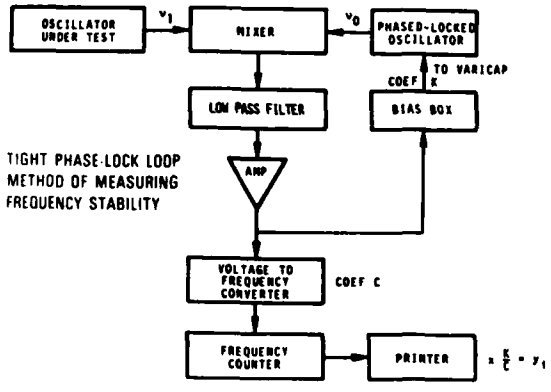


FIGURE 1.7

fluctuations between the two oscillators and is no longer proportional to the phase fluctuations for sample times longer than the attack time of the loop. A bias box is used to adjust the voltage on the varicap to a tuning point that is fairly linear and of a reasonable value. The voltage

fluctuations prior to the bias box (biased slightly away from zero) may be fed to a voltage to frequency converter which in turn is fed to a frequency counter where one may read out the frequency fluctuations with great amplification of the instabilities between this pair of oscillators. The frequency counter data are logged with a data logging device. The coefficient of the varicap and the coefficient of the voltage to frequency converter are used to determine the fractional frequency fluctuations, y_i, between the oscillators, where i denotes the i^{th} measurement as shown in figure 1.7. It is not difficult to achieve a sensitivity of a part in 10^{14} per Hz resolution of the frequency counter, so one has excellent precision capabilities with this system.

E. Time difference method

The last measurement method we will illustrate is very commonly used, but typically does not have the measurement precision more readily available in the first four methods illustrated above. This method is called the time difference method, and is shown in figure 1.8. Because of the wide

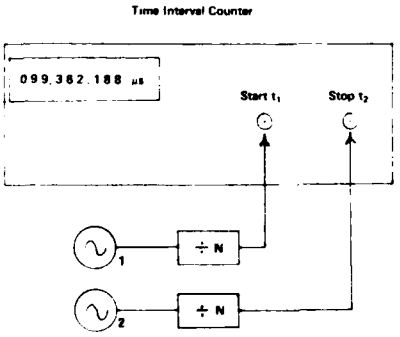


FIGURE 1.8

bandwidth needed to measure fast rise-time pulses, this method is limited in signal-to-noise ratio. However, some counters are commercially available allowing one to do signal averaging or to do precision rise-time comparison (precision of time difference measurements in the range of 10 ns to 10 ps are now available). Such a method yields a direct measurement of x(t) without any translation, conversion, or multiplication factors. Caution should be exercised in using this technique even if adequate measurement precision is available because it is not uncommon to have significant instabilities in the frequency dividers shown in figure 1.8--of the order of several nanoseconds. The technology exists to build better frequency dividers than are commonly available, but manufacturers have not yet availed themselves of state-of-the-art techniques in a cost beneficial manner. A trick to by-pass divider problems is to feed the oscillator signals directly into the time interval counter and observe the zero voltage crossing into a well matched impedance. (In fact, in all of the above methods one needs to pay attention to impedance matching, cable lengths and types, and connectors). The divided signal can be used to resolve cycle ambiguity of the carrier, otherwise the carrier phase at zero volts may be used as the time reference. The slope of the signal at zero volts is $2\pi V_p/\tau_1$, where $\tau_1 = 1/\nu_1$ (the period of oscillation). For V_p = 1 volt and a 5 MHz signal, this slope is 3m volts/ns, which is a very good sensitivity.

II. MEASUREMENT METHODS COMPARISON

When making measurements between a pair of frequency standards or clocks, it is desirable to have less noise in the measurement system than the composite noise in the pair of standards being measured. This places stringent requirements on measurement systems as the state-of-the-art of precision frequency and time standards has advanced to its current level. As will be shown, perhaps one of the greatest areas of disparity between measurement system noise and the noise in current standards is in the area of time difference measurements. Commercial equipment can measure time differences to at best 10^{-11}s, but the time fluctuations second to second of state-of-the-art standards is as good as 10^{-13}s.

The disparity is unfortunate because if time differences between two standards could be measured with adequate precision then one may also know the time fluctuations, the frequency differences, and the frequency fluctuations. In fact, one can set

Hierachy Status	Example of Measurement Method	Data deducible from the measurement			
		Time	δTime	Freq.	δFreq.
1	Dual Mixer Time Diff. or Time Interval Counter	$x(t) = \frac{\Delta t}{\tau_0}$	$\delta x(t,\tau) = x(t+\tau)-x(t)$	$y(t,\tau) = \frac{\delta x(t,\tau)}{\tau}$	$\delta y(t,\tau) = y(t+\tau,\tau)-y(t,\tau)$
2	Loose Phase-locked reference oscillator	can't measure	$\delta x(t,\tau) = \frac{\delta \phi(t,\tau)}{2\pi\nu}$	"	"
3	Heterodyne or beat frequency	can't measure	can't measure	$\lvert y(t,\tau)\rvert = \frac{\nu_{beat}}{\nu_0}$	"
4	Tight phase-locked reference oscillator	can't measure	can't measure	can't measure	$\delta y(t,\tau) = \frac{}{c} \cdot \delta V$

TABLE 2.1

Measurement Method	Time accuracy	Time stability	Frequency accuracy (1 day)	Frequency stability	Advantages	Disadvantages	Approximate Cost (k$)
Dual Mixer Time Difference	~100ps	0.1ps	$\sim 10^{-10}$	$\sim 1.10^{-13} \times \tau^{-1}$	No dead time; many choose sample time (1ms to as large as desired); oscillators may be at zero beat or different; measurement bandwidth easily changed; measurement noise typically below oscillators noise for tau=1s and longer; measures times, time stability, frequency, and frequency, stability.	More complex than other methods, and hence is more suceptable to extranious signal pickup, e.g., ground loops; the time difference is modulo the beat period, e.g., 200ns @ 5MHz.	~5 to 30
Time Interval Counter	~100ps	10ps	$\sim 10^{-10}$	$\sim 1.10^{-11} \times \tau^{-1}$	Wide bandwidth input allows a variety of signals; simple to use; cycle ambiguity almost never a problem; measures time, time stability, frequency, and frequency stability	Measurement noise typicaly in excess of oscillator instabilities for tau values up to and of the order of several thousand seconds, hence, typically not useful for short term stability analysis.	1 to 7
Loose Phase-lock reference* oscillator	---	~0.1ps	depends on calibration of varicap curve	$\sim 1.10^{-11} \times \tau^{-1}$	Useful for short term time stability analysis as well as spectrum analysis, excellent phase sensitivity--typically better than most oscillators.	Calibration depends on appropriate reference oscillator varicap curve characteristics.	<1
Heterodyne or Beat Frequency	---	---	$\sim 10^{-10}$	$\sim 1.10^{-13} \times \tau^{-1}$	Measurement noise can typically be made less than oscillator instabilities for tau =1s and longer.	Minimum tau determined by period of beat frequency--typically not adjustable; can not compare oscillators near zero beat; cannot tell which oscillator is high or low in frequency; dead time often associated with these measurements.	0.2 to 1
Tight phase-locked reference* oscillator	---	---	---	$\sim 1.10^{-13} \times \tau^{-1}$	Measurement noise typically less than oscillator instabilities for tau=1s and longer; good measurement system band-width control; dead time can be made small or negligible.	Need voltage controlled reference oscillator; frequency sensitivity is a function of varicap tuning curve, hence not conducive to measuring absolute frequency differences.	<1

* Assumes reference oscillator is available to the metrologist.

TABLE 2.2

up an interesting hierarchy of kinds of measurement systems: 1) those that can measure time, x(t); 2) those that can measure changes in time or time fluctuations $\delta x(t)$; 3) those that can measure frequency, $\nu(y \equiv (\nu-\nu_0)/\nu_0)$; and 4) those that can measure changes in frequency or frequency fluctuations, $\delta\nu$ ($\delta y \equiv \delta\nu/\nu_0$). As depicted in table 2.1, if a measurement system is of status 1 in this hierarchy, i.e., it can measure time, then time fluctuations, frequency and frequency fluctuations can be deduced. However, if a measurement system is only capable of measuring time fluctuations (status 2 - table 2.1), then time cannot be deduced, but frequency and frequency fluctuations can. If frequency is being measured (status 3 - table 2.1), then neither time nor time fluctuations may be deduced with fidelity because essentially all commercial frequency measuring devices have "dead time" (technology is at a point where that is changing with fast data processing speeds that are now available). Dead time in a frequency measurement destroys the opportunity of integrating the fractional frequency to get to "true" time fluctuations. Of course, if frequency can be measured, then trivially one may deduce the frequency fluctuations. Finally, if a system can only measure frequency fluctuations (status 4 - table 2.1), then neither time, nor time fluctuations, nor frequency can be deduced from the data. If the frequency stability is the primary concern then one may be perfectly happy to employ such a measurement system, and similarly for the other statuses in this measurement hierarchy. Obviously, if a measurement method of Status 1 could be employed with state-of-the-art precision, this would provide the greatest flexibility in data processing. From section 1, the dual mixer time difference system is purported to be such a method.

Table 2.2 is a comparison of these different measurement methods. The values entered are nominal; there may be unique situations where significant departures are observed. The time and frequency stabilities listed are the nominal second to second rms values. The accuracies listed are taken in an absolute sense. The costs listed are nominal estimates in 1981 dollars.

Figure 2.1 is a diagram indicating the sample time regions over which the various methods

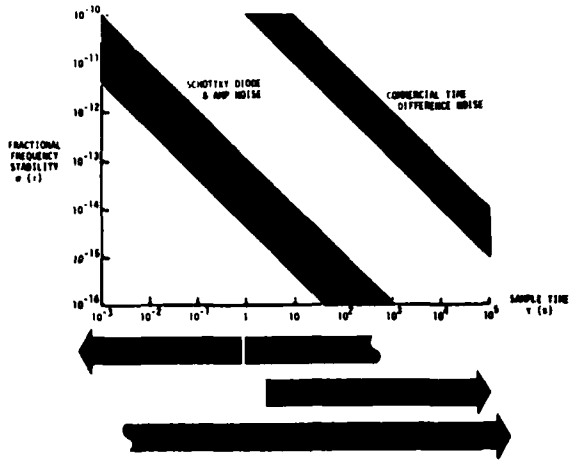


FIGURE 2.1

are most appropriately applied. The large diagonally oriented area indicates the typical noise limits of the measurement technique (at particular values of sample time indicated on the horizontal scale).

III. CHARACTERIZATION

Given a set of data of the fractional frequency or time fluctuations between a pair of oscillators, it is useful to characterize these fluctuations with reasonable and tractable models of performance. In so doing for many kinds of oscillators, it is useful to consider the flucuations as those that are random (may only be predicted statistically) and those that are non-random (e.g., systematics— those that are environmentally induced or those that have a causal effect that can be determined and in many cases can be predicted).

3.1 Non-random Fluctuations

Non-random fluctuations are usually the main cause of departure from "true" time or "true" frequency.

If, for example, one has the values of the frequency over a period of time and a frequency offset from nominal is observed, one may calculate directly that the phase error will accumulate as a ramp. If the frequency values show some linear drift then the time fluctuations will depart as a quadratic. In almost all oscillators, the above systematics, as they are sometimes called, are the primary cause of time and/or frequency departure. A useful approach to determine the value of the frequency offset is to calculate the simple mean of the set, or for determining the value of the frequency drift by calculating a linear least squares fit to the frequency. A least squares quadratic fit to the phase or to the time derivative is typically not as efficient an estimator of the frequency drift for most oscillators.

3.2 Random Fluctuations

After calculating or estimating the systematic or non-random effects of a data set, these may be subtracted from the data leaving the residual random fluctuations. These can usually be best characterized statistically. It is often the case for precision oscillators that these random fluctuations may be well modeled with power law spectral densities. This topic is discussed later in sections VIII to X. We have

$$S_y(f) = h_\alpha f^\alpha, \qquad (3.1)$$

where $S_y(f)$ is the one-sided spectral density of the fractional frequency fluctuations, f is the Fourier frequency at which the density is taken, h_α is the intensity coefficient, and α is a number modeling the most appropriate power law for the data. It has been shown[1,3] that in the time domain one can nicely represent a power law spectral density process using a well defined time-domain stability measure, $\sigma_y(\tau)$, to be explained in the next section. For example, if one observes from a log $\sigma_y^2(\tau)$ versus τ diagram a particular slope (call it μ) over certain regions of sample time, τ, this slope has a correspondence to a power law spectral density or a set of the same with some amplitude coefficient h_α. In particular, $\mu = -\alpha - 1$ for $-3 < \alpha < 1$ and $\mu \simeq -2$ for $1 < \alpha$. Further a correspondence exists between h_α and the coefficient for $\sigma_y(\tau)$. These coefficients have been calculated and appear in section XI. The transformations for some of the more common power law spectral densities have been tabulated making it quite easy to transform the frequency stability modeled in the time-domain over to the frequency domain and vice versa. Examples of some power-law spectra that have been simulated by computer are shown in figure 3.1. In descending order these

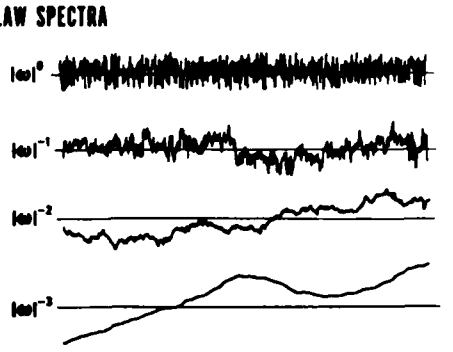


FIGURE 3.1

have been named white noise, flicker noise, random walk, and flicker walk (the ω in fig. 3.1 is angular Fourier frequency, $\omega = 2\pi f$).

Once the noise characteristics have been determined, one is often able to deduce whether the oscillators are performing properly or not and whether they are meeting either the design specifications or the manufacturers specifications. For example a cesium beam frequency standard or a rubidium gas cell frequency standard when working properly should exhibit white frequency noise, which is the same as random walk phase (or time) for tau values of the order of a few seconds to several thousand seconds (see also sec. XI).

IV. ANALYSIS OF TIME DOMAIN DATA

Suppose now that one is given the time or frequency fluctuations between a pair of precision oscillators measured, for example, by one of the techniques outlined in section I, and a stability analysis is desired. Let this comparison be depicted by figure 4.1. The minimum sample time

is determined by the measurement system. If the time difference or the time fluctuations are available then the frequency or the fractional frequency fluctuations may be calculated from one period of sampling to the next over the data length as indiciated in figure 4.1. Suppose further there are M values of the fractional frequency y_i. Now there are many ways to analyze these data. Historically, people have typically used the standard deviation equation shown in figure 4.1, $\sigma_{\text{std. dev.}}(\tau)$, where $\bar{y}$ is the average fractional frequency over the data set and is subtracted from each value of y_i before squaring, summing and dividing by the number of values minus one, (M-1), and taking the square root to get the standard deviation. At NBS, we have studied what happens to the standard deviation when the data set may be characterized by power law spectra which are more dispersive than classical white noise frequency fluctuations. In other words, if the fluctuations are characterized by flicker noise or any other non-white-noise frequency deviations, what happens to the standard deviation for that data set? One can show that the standard deviation is a function of the number of data points in the set; it is also a function of the dead time and of the measurement system bandwidth. For example, using flicker noise frequency modulation as a model, as the number of data points increases, the standard deviation monotonically increases without limit. Some statistical measures have been developed which do not depend upon the data length and which are readily usable for characterizing the random fluctuations in precision oscillators. An IEEE subcommittee on frequency stability has recommended what has come to be known as the "Allan variance" taken from the set of useful variances developed, and an experimental estimation of the square root of the Allan variance is shown as the bottom right equation in figure 4.1. This equation is very easy to implement experimentally as one simply need add up the squares of the differences between adjacent values of y_i, divide by the number of them and by two, and take the square root. One then has the quantity which the IEEE subcommittee has recommended for specification of stability in the time domain—denoted by $\sigma_y(\tau)$.

$$\sigma_y(\tau) = \left\langle \frac{1}{2} \left(\bar{y}(t+\tau) - \bar{y}(t)\right)^2 \right\rangle^{\frac{1}{2}}, \qquad (4.1)$$

where the brackets "<>" denote infinite time average. In practice this is easily estimated from a finite data set as follows:

$$\sigma_y(\tau) \approx \left[\frac{1}{2(M-1)} \sum_{i=1}^{M-1} \left(y_{i+1} - y_i\right)^2 \right]^{\frac{1}{2}}, \qquad (4.2)$$

where the y_i are the discrete frequency averages as illustrated in figure 4.1.

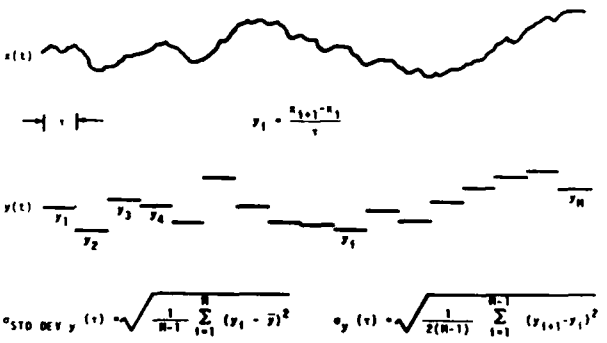


FIGURE 4.1

A simulated plot of the time fluctuations, x(t) between a pair of oscillators and of the corresponding fractional frequencies calculated from the time fluctuations each averaged over a sample time τ. At the bottom are the equations for the standard deviation (left) and for the time-domain measure of frequency stability as recommended by the IEEE subcommittee on frequency stability (right).

One would like to know how $\sigma_y(\tau)$ varies with the sample time, τ. A simple trick that one can use that is very useful if there is no dead time, is to average the previous values for y_1 and y_2 and call that a new y_1 averaged over 2τ, similarly average the previous values for y_3 and y_4 and call that a new y_2 averaged over 2τ etc., and finally apply the same equation as before to get $\sigma_y(2\tau)$. One can repeat this process for other desired integer multiples of τ and from the same data set

be able to generate values for $\sigma_y(m\tau)$ as a function of $m\tau$ from which one may be able to infer a model for the process that is characteristic of this pair of oscillators. If one has dead time in the measurements adjacent pairs cannot be averaged in an unambiguous way to simply increase the sample time. One has to retake the data for each new sample time--often a very time consuming task. This is another instance where dead time can be a problem.

How the classical variance (standard deviation squared) depends on the number of samples is shown in figure 4.2. Plotted is the ratio of the standard deviation squared for N samples to the standard deviation squared for 2 samples; $\langle\sigma^2(2,\tau)\rangle$ is the same as the Allan variance, $\sigma_y^2(\tau)$. One can see the dependence of the standard deviation upon the number of samples for various kinds of power

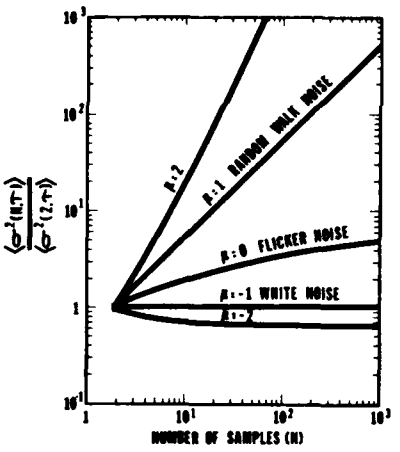


FIGURE 4.2

The ratio of the time average of the standard deviation squared for N samples over the time average of a two sample standard deviation squared as a function of the number of smaples, N. The ratio is plotted for various power law spectral densities that commonly occur in precision oscillators. The figure illustrates one reason why the standard deviation is not a convenient measure of frequency stability; i.e. it may be very important to specify how many data points are in a data set if you use the standard deviation.

law spectral densities commonly encountered as reasonable models for many important precision oscillators. Note, $\sigma_y^2(\tau)$ has the same value as the classical variance for the classical noise case (white noise FM). One main point of figure 4.2 is simply to show that with the increasing data length the standard deviation of the common classical variance is not well behaved for the kinds of noise processes that are very often encountered in most of the precision oscillators of interest.

One may combine eq (1.4) and eq (4.1), which yields an equation for $\sigma_y(\tau)$ in terms of the time difference or time deviation measurements.

$$\sigma_y(\tau) = \left\langle \frac{1}{2\tau^2} \left(x(t+2\tau) - 2x(t+\tau) + x(t)\right)^2 \right\rangle^{\frac{1}{2}}, \tag{4.3}$$

which for N discrete time readings may be estimated as,

$$\sigma_y(\tau) \cong \left[\frac{1}{2(N-2)\tau^2} \sum_{i=1}^{N-2} (x_{i+2} - 2x_{i+1} + x_i)^2 \right]^{\frac{1}{2}}, \tag{4.4}$$

where the i denotes the number of the reading in the set of N and the nominal spacing between readings is τ. If there is no dead time in the data and the original data were taken with the x's spaced by τ_o, we can pick τ in eq (4.4) to be any integer multiple of τ_o, i.e., $\tau = m\tau_o$:

$$\sigma_y(M\tau_o) \cong \left[\frac{1}{2(N-2m)m^2\tau_o^2} \sum_{i=1}^{N-2m} (x_{i+2m} - 2x_{i+m} + x_i)^2 \right]^{\frac{1}{2}}, \tag{4.5}$$

Equation (4.5) has some interesting consequences because of the efficient data usage in terms of the confidence of the estimate as will be explained in the next section.

EXAMPLE: Find the Allan variance, $\sigma_y^2(\tau)$, of the following sequence of fractional frequency fluctuation values y_k, each value averaged over one second.

$y_1 = 4.36 \times 10^{-5}$ $\quad$ $y_5 = 4.47 \times 10^{-5}$
$y_2 = 4.61 \times 10^{-5}$ $\quad$ $y_6 = 3.96 \times 10^{-5}$
$y_3 = 3.19 \times 10^{-5}$ $\quad$ $y_7 = 4.10 \times 10^{-5}$
$y_4 = 4.21 \times 10^{-5}$ $\quad$ $y_8 = 3.08 \times 10^{-5}$

(assume no dead-time in measurement of averages)

Since each average of the fractional frequency fluctuation values is for one second, then the first variance calculation will be at τ = 1s. We are given M = 8 (eight values); therefore, the number of pairs in sequence is M-1 = 7. We have:

Data values y_k (x 10^{-6})	First differences $(y_{k+1} - y_k)$ (x 10^{-5})	First difference squared $(y_{k+1} - y_k)^2$ (x 10^{-10})
4.36	----	----
4.61	0.25	0.06
3.19	-1.42	2.02
4.21	1.02	1.04
4.47	0.26	0.07
3.96	-0.51	0.26
4.10	0.14	0.02
3.08	-1.02	1.04
		4.51

$$\sum_{k=1}^{M-1} (y_{k+1} - y_k)^2 = 4.51 \times 10^{-10}$$

Therefore,

$$\sigma_y^2(1s) = \frac{4.51 \times 10^{-10}}{2(7)} = 3.2 \times 10^{-11}$$

and

$$\sigma_y(\tau) = [\sigma_y^2(1s)]^{\frac{1}{2}} = [3.2 \times 10^{-11}]^{\frac{1}{2}} = 5.6 \times 10^{-6}$$

Using the same data, one can calculate the variance for τ = 2s by averaging pairs of adjacent values and using these new averages as data values for the same procedure as above. For three second averages (τ = 3s) take adjacent threesomes and find their averages and proceed in a similiar manner. More data must be acquired for longer averaging times.

One sees that with large numbers of data values, it is helpful to use a computer or programmable calculator. The confidence of the estimate on $\sigma_y(\tau)$ improves nominally as the square root of the number of data values used. In this example, M=8 and the confidence can be expressed as being no better the $1/\sqrt{8} \times 100\% = 35\%$. This then is the allowable error in our estimate for the τ = 1s average. The next section shows methods of computing and improving the confidence interval.

V. CONFIDENCE OF THE ESTIMATE AND OVERLAPPING SAMPLES[4]

One can imagine taking three phase or time measurements of one oscillator relative to another at equally spaced intervals of time. From this phase data one can obtain two, adjacent values of average frequency. From these two frequency measurements, one can calculate a single sample Allan (or two-sample) variance (see fig. 5.1). Of course this variance does not have high precision or confidence since it is based on only one frequency difference.

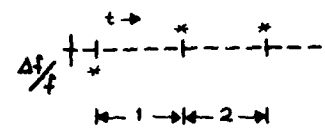


FIGURE 5.1

Statisticians have considered this problem of quantifying the variability of quantities like the Allan Variance. Conceptually, one could imagine repeating the above experiment (of taking the three phase points and calculating the Allan Variance), many times and even calculating the distribution of the values.

For the above cited experiment we know that the results are distributed like the statistician's chi-square distribution with one degree of freedom. That is, we know that for most common oscillators the first difference of the frequency is a normally distributed variable with the typical bell-shaped curve and zero mean. However, the square of a normally distributed variable is NOT normally distributed. That is easy to see since the square is always positive and the normal curve is completely symmetric and negative values are as likely as positive. The resulting distribution is called a chi-square distribution, and it has ONE "degree of freedom" since the distribution was obtained by considering the squares of individual (i.e., one independent sample), normally distributed variables.

In contrast, if we took five phase values, then we could calculate four consecutive frequency values, as in figure 5.2. We could then take the

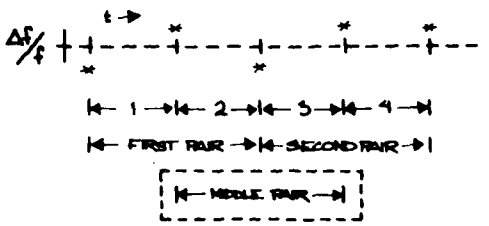


FIGURE 5.2

first pair and calculate a sample Allan Variance, and we could calculate a second sample Allan Variance from the second pair (i.e., the third and fourth frequency measurements). The average of these two sample Allan Variances provides an improved estimate of the "true" Allan Variance, and we would expect it to have a tighter confidence interval than in the previous example. This could be expressed with the aid of the chi-square distribution with TWO degrees of freedom.

However, there is another option. We could also consider the sample Allan Variance obtained from the second and third frequency measurements. That is the middle sample variance. Now, however, we're *in trouble because clearly* this last sample Allan Variance is NOT independent of the other two. Indeed, it is made up of parts of each of the other two. This does NOT mean that we can't use it for improving our estimate of the "true" Allan Variance, but it does mean that we can't just assume that the new average of three sample Allan Variances is distributed as chi-square with three degrees of freedom. Indeed, we will encounter chi-square distributions with fractional degrees of freedom. And as one might expect, the number of degrees of freedom will depend upon the underlying noise type, that is, white FM, flicker FM, or whatever.

Before going on with this, it is of value to review some concepts of the chi-square distribution. Sample variances (like sample Allan Variances) are distributed according to the equation:

$$\chi^2 = \frac{(d.f.)\cdot s^2}{\sigma^2} \qquad (5.1)$$

where S^2 is the sample Allan Variance, χ^2 is chi-square, d.f. is the number of degrees of freedom (possibly not an integer), and σ^2 is the "true" Allan Variance we're all interested in knowing--but can only estimate imperfectly. Chi-square is a random variable and its distribution has been studied extensively. For some reason, chi-square is defined so that d.f., the number of degrees of freedom, appears explicitly in eq (5.1). Still, χ^2 is a (implicit) function of d.f., also.

The probability density for the chi-square distribution is given by the relation

$$\rho(\chi^2) = \frac{1}{2^{d.f.}\,\Gamma\left(\frac{d.f.}{2}\right)} (\chi^2)^{\frac{d.f.}{2}-1} e^{\frac{-\chi^2}{2}} \qquad (5.2)$$

where $\Gamma\left(\frac{d.f.}{2}\right)$ is the gamma function, defined by the integral

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\, dx \qquad (5.3)$$

Chi-square distributions are useful in determining specified confidence intervals for variances and standard deviations. Here is an example. Suppose we have a sample variance $s^2 = 3.0$ and we know that this variance has 10 degrees of freedom. (Just how we can know the degrees of freedom will be discussed shortly.) Suppose also that we want to know a range around our sample value of $s^2 = 3.0$ which "probably" contains the true value, σ^2. The desired confidence is, say, 90%. That is, 10% of the time the true value will actually fall outside of the stated bounds. The usual way to proceed is to allocate 5% to the low end and 5% to the high end for errors, leaving our 90% in the middle. This is arbitrary and a specific problem might dictate a different allocation. We now resort to tables of the chi-square distribution and find that for 10 degrees of freedom the 5% and 95% points correspond to:

$$\chi^2(.05) = 3.94 \quad \text{for d.f.} = 10 \quad \chi^2(.95) = 18.3 \qquad (5.4)$$

Thus, with 90% probability the calculated sample variance, s^2, satisfies the inequality:

$$3.94 \leq \frac{(d.f.)\cdot s^2}{\sigma^2} \leq 18.3 \qquad (5.5)$$

and this inequality can be rearranged in the form

$$1.64 \leq \sigma^2 \leq 7.61 \qquad (5.6)$$

or, taking square roots:

$$1.28 \leq \sigma \leq 2.76 \qquad (5.7)$$

Now someone might object to the form of eq (5.7) since it seems to be saying that the true sigma falls within two limits with 90% probability. Of course, this is either true or not and is not subject to a probabilistic interpretation. Actually eq (5.7) is based on the idea that the true sigma is not known and we estimate it with the square root of a sample variance, s^2. This sample variance is a random variable and is properly the subject of probability, and its value (which happened to be 3.0 in the example) will conform to eq (5.7) nine times out of ten.

Typically, the sample variance is calculated from a data sample using the relation:

$$s^2 = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \bar{x})^2 \qquad (5.8)$$

where it is implicitly assumed that the x_n's are random and uncorrelated (i.e., white) and where $\bar{x}$ is the sample mean calculated from the same data set. If all of this is true, then s^2 is chi-square distributed and has N-1 degrees of freedom.

Thus, for the case of white x_n, and a conventional sample variance (i.e., eq (5.8)), the number of degrees of freedom are given by the equation:

$$d.f. = N-1 \qquad (5.9)$$

The problem of interest here is to obtain the corresponding equations for Allan Variances using overlapping estimates on various types of noise (i.e., white FM, flicker FM, etc.).

Other authors (Lesage and Audoin, and Yoshimura) have considered the question of the variance of the Allan Variances without regard to the distributions. This is, of course, a closely related problem and use will be made of their results. These authors considered a more restrictive set of overlapping estimates than will be considered here, however.

VI. MAXIMAL USE OF THE DATA AND DETERMINATION OF THE DEGREES OF FREEDOM.

6.1 Use of Data

Consider the case of two oscillators being compared in phase and exactly N values of the phase difference are obtained. Assume that the data are taken at equally spaced intervals, τ_0. From these N phase values, one can obtain N-1 consecutive values of average frequency and from these one can compute N-2 individual, sample Allan Variances (not all independent) for $\tau = \tau_0$. These N-2 values can be averaged to obtain an estimate of the Allan Variance at $\tau = \tau_0$. The variance of this variance has been calculated by the above cited authors.

Using the same set of data, it is also possible to estimate the Allan Variances for integer multiples of the base sampling interval, $\tau = n\tau_0$. Now the possibilities for overlapping sample Allan Variances are even greater. For a phase data set of N points one can obtain exactly N-2n sample Allan Variances for $\tau = n\tau_0$. Of course only a fraction of these are generally independent. Still the use of ALL of the data is well justified (see fig. 6.1).

Consider the case of an experiment extending for several weeks in duration with the aim of getting estimates of the Allan Variance for tau values equal to a week or more. As always the purpose is to estimate reliably the "true" Allan Variance as well as possible--that is, with as tight an uncertainty as possible. Thus one wants

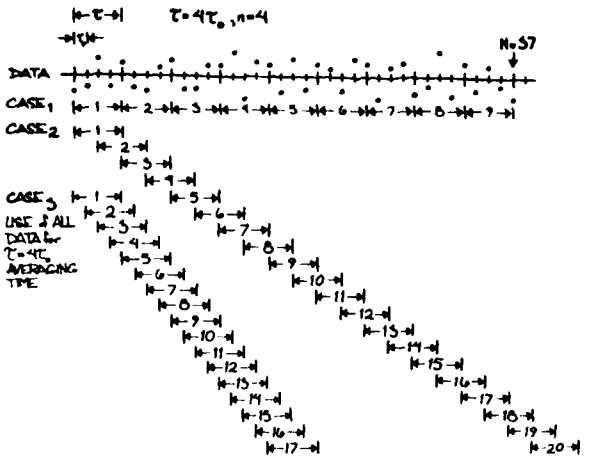


FIGURE 6.1

to use the data as efficiently as possible since obtaining more data can be very expensive. The most efficient use is to average all possible sample Allan Variances of a given tau value that one can compute from the data.

The problem comes in estimating how tight the confidence intervals really are—that is, in estimating the number of degrees of freedom. Clearly, if one estimates the confidence intervals pessimistically, then more data is needed to reach a specified tolerance, and that can be expensive. The other error of over-confidence in a questionable value can be even more expensive. Ideally, one has realistic confidence estimates for the most efficient use of the data, which is the intent of this writing.

6.2 Determining the Degrees of Freedom

In principle, it should be possible to determine analytically the equations corresponding to eq (5.9) for all cases of interest. Unfortunately the analysis becomes quite complicated. Exact computer algorithms were devised for the cases of white phase noise, white frequency modulation and random walk FM. For the two flicker cases (i.e., flicker FM and PM) a completely empirical approach was used. Due to the complexity of the computer programs, empirical fits were devised for all five noise types.

The approach used is based on three equations relating to the chi-square distribution:

$$\frac{\chi^2}{(d.f.)} = \frac{S^2}{\sigma^2} \tag{6.1}$$

$$E[\chi^2] = d.f. \tag{6.2}$$

$$Var[\chi^2] = 2(d.f.) \tag{6.3}$$

where the expression $E[\chi^2]$ means the "expectation," or average value of χ, $Var[\chi^2]$ is the variance of χ^2, and d.f. is the number of degrees of freedom.

A computer was used to simulate phase data sets of some length, N, and then Allan Variances with $\tau = n\tau_0$ were calculated for all possible samples. This "experiment" was repeated at least 1000 times using new simulated data sets of the same spectral type, and always of the same length, N. Since the data were simulated on a computer, the "true" Allan Variance, σ^2, was known for many of the noise models and could be substituted into eq (6.1). From the 1000 values of s^2/σ^2, distributions and sample variances were obtained. The "experimental" distributions were compared with theoretical distributions to verify that the observed distributions truely conformed to the chi-square distribution.

The actual calculation of the degrees of freedom were made using the relation:

$$d.f. = \frac{2(\sigma^2)^2}{Var(s^2)} \tag{6.4}$$

which can be deduced from eqs (6.1), (6.2), and (6.3). The $Var(s^2)$ was estimated by the sample variance of the 1000 values of the average Allan Variances, each obtained from a phase data set of length N.

Of course this had to be repeated for various values of N and n, as well as for each of the five common noise types: white PM, flicker PM, white FM, flicker FM, and random walk FM. Fortunately, certain limiting values are known and these can be used as checks on the method. For example, when (N-1)/2=n, only one Allan Variance is obtained from each data set and one should get about one degree of freedom for eq (6.4), which was observed

in fact. Also for n=1 the "experimental" conditions correspond to those used by Lesage and Audoin, and by Yoshimura. Indeed, the method also was tested by verifying that it gave results consistent with eq (5.9) when applied to the conventional sample variance. Thus, combining eq (6.4) with the equations for the variance of the Allan Variances from Lesage and Audoin and Yoshimura, one obtains:

$$\begin{aligned}
&\text{White PM} && \text{d.f.} = \frac{18(N-2)^2}{35N-88}, \quad N \geq 4 \\
&\text{Flicker PM} && \text{d.f.} = ? \\
&\text{White FM} && \text{d.f.} = \frac{2(N-2)^2}{3N-7} \\
&\text{Flicker FM} && \text{d.f.} = \frac{2(N-2)^2}{2.3N-4.9} \\
&\text{Random Walk FM} && \text{d.f.} = N-2
\end{aligned} \tag{6.5}$$

for n=1. Unfortunately, their results are not totally consistent with each other. Where inconsistency arose the value in best agreement with the "experimental" results was chosen.

The empirical equations which were fit to the "experimental" data and the known values are summarized below:

$$\begin{aligned}
&\text{White PM} && \text{d.f.} \cong \frac{(N+1)(N-2n)}{2(N-n)} \\
&\text{Flicker PM} && \text{d.f.} \cong \exp\left(\ln\frac{N-1}{2n} \ \ln\frac{(2n+1)(N-1)}{4}\right) \\
&\text{White FM} && \text{d.f.} \cong \left[\frac{3(N-1)}{2n} - \frac{2(N-2)}{N}\right] \cdot \frac{4n^2}{4n^2+5} \\
&\text{Flicker FM} && \text{d.f.} \cong \begin{cases} \dfrac{2(N-2)}{2.3N-4.9}, & \text{for } n=1 \\ \dfrac{5N^2}{4n(N+3n)}, & \text{for } n \geq 2 \end{cases} \\
&\text{Random Walk FM} && \text{d.f.} \cong \frac{N-2}{n} \ \frac{(N-1)^2-3n(N-1)+4n^2}{(N-3)^2}
\end{aligned} \tag{6.6}$$

The figures in Appendix I demonstrate the fit to the "experimental" data.

It is appropriate to give some estimate of just how well these empirical equations approach the "true" values. The equations have approximately (a few percent) the correct assymptotic behavior at n=1 and n=(N-1)/2. In between, the values were tested (using the simulation results) over the range of N=5 to N=1025 for n=1 to n=(N-1)/2 changing by octaves. In general, the fit was good to within a few percent. We must acknowledge that distributional problems with the random number generators can cause problems, although there were several known values which should have revealed these problems if they are present. Also for three of the noise types the exact number of degrees of freedom were calculated for many values of N and n and compared with the "Monte Carlo" calculations. The results were all very good.

Appendix I presents the data in graphical form. All values are thought to be accurate to within one percent or better for the cases of white PM, white FM, and random walk FM. A larger tolerence should be allowed for the flicker cases.

VII. EXAMPLE OF TIME-DOMAIN SIGNAL PROCESSING AND ANALYSIS

We will analyze in some detail a commercial portable clock, Serial No. 102. This cesium was measured against another commercial cesium whose stability was well documented and verified to be better than the one under test. Plotted in figure 7.1 are the residual time deviations after removing

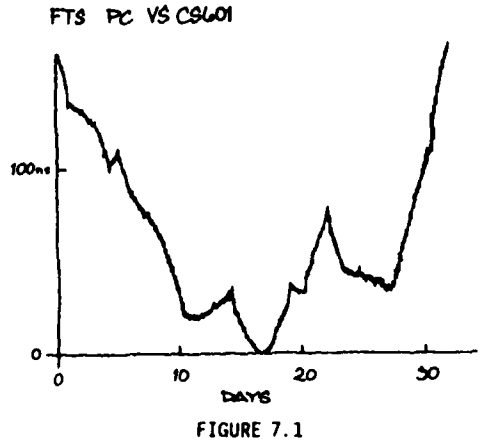


FIGURE 7.1

a mean frequency of 4.01 parts in 10^{13}. Applying the methods described in section IV and section V, we generated the $\sigma_y(\tau)$ diagram shown in figure 7.2.

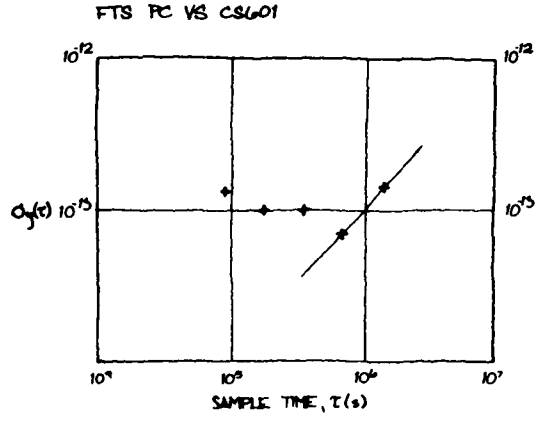


FIGURE 7.2

One observes that the last two points are proportional to τ^{+1} and one is suspicious of a significant frequency drift.

If one calculates the drift knowing that $\sigma_y(\tau)$ is equal to the drift times $\frac{\tau}{\sqrt{2}}$ a drift of 1.22×10^{-14} per day is obtained. A linear least squares to the frequency was removed and sections IV and V were applied again. The linear least squares fit showed a drift of 1.23×10^{-14} per day, which is in excellent agreement with the previous calculated value obtained from $\sigma_y(\tau)$. Typically, the linear least squares will give a much better estimate of the linear frequency drift than will the estimate from $\sigma_y(\tau)$ being proportional to τ^{+1}.

Figure 7.3 gives the plot of the time residuals after removing the linear least squares and figure 7.4 is the corresponding $\sigma_y(\tau)$ vs. τ diagram. From the 33 days of data, we have used the 90% confidence interval to bracket the stability estimates and one sees a reasonable fit corresponding to white noise frequency modulation at a level of $4.4 \times 10^{-11}\ \tau^{-\frac{1}{2}}$. This seemed excessive

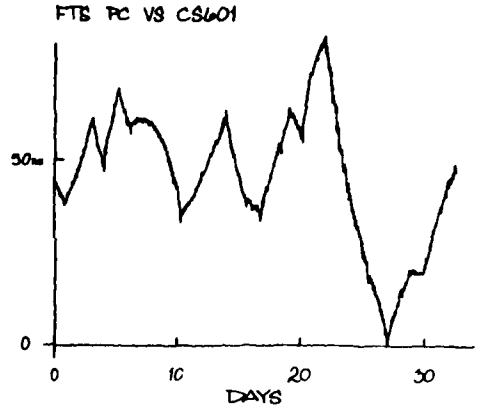


FIGURE 7.3

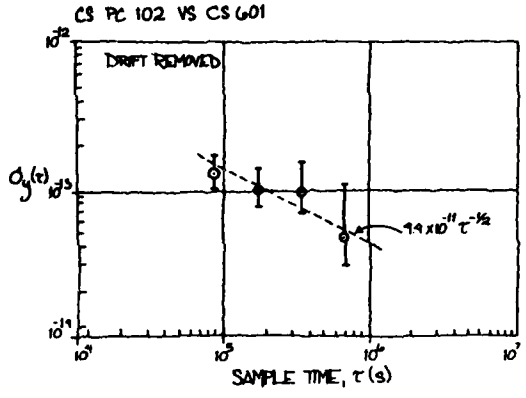


FIGURE 7.4

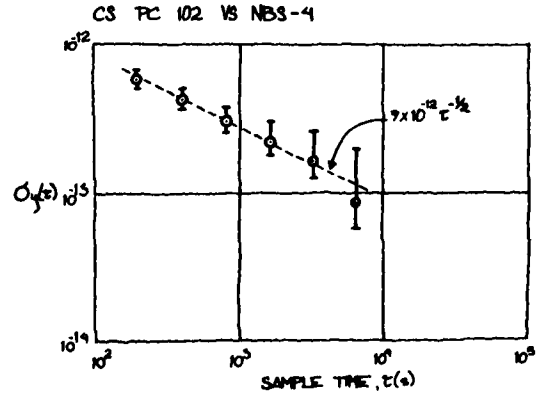


FIGURE 7.5

in terms of the typical performance of this particular cesium and in as much as we were doing some other testing within the environment, such as working on power supplies and charging and discharging batteries, we did some later tests. Figure 7.5 is a plot of $\sigma_y(\tau)$ after the standard had been left alone in a quiet environment and had been allowed to age for about a week. One observes that the white noise frequency modulation level is more than a factor of 4 improved over the previous data. This led us to do some studies on the effects of the power supply on the cesium frequency as one is charging and discharging batteries, which proved to be significant. One notices in figure 7.4 that the $\sigma_y(\tau)$ values plotted are consistent within the error bars with flicker noise frequency modulation. This is more typical of the kind of noise one would expect due to such environmental perturbations as discussed above.

Careful time- and/or frequency-domain analyses can lead to significant insights into problems and their solutions and is highly recommended by the authors. The frequency-domain techniques will be next approached.

VIII. SPECTRUM ANALYSIS

Another method of characterizing the noise in a signal source is by means of spectrum analysis. To understand this approach, let's examine the waveform shown in figure 8.1.

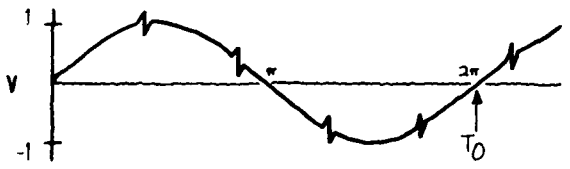


FIGURE 8.1

Here we have a sine wave which is perturbed for short instances by noise. Some loosely refer to these types of noises as "glitches". The waveform has a nominal frequency over one cycle which we'll call "ν_0" ($\nu_0 = \frac{1}{T}$). At times, noise causes the instantaneous frequency to differ markedly from the nominal frequency. If a pure sine wave signal of frequency ν_0 is subtracted from this waveform, the remainder is the sum of the noise components. These components are of a variety of frequencies and the sum of their amplitudes is nearly zero except for the intervals during each glitch when their amplitudes momentarily reinforce each other. This is shown graphically in figure 8.2.

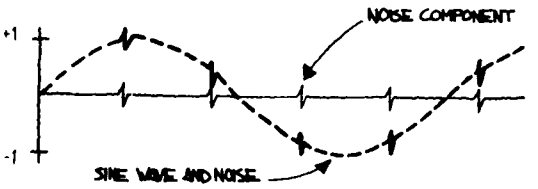


FIGURE 8.2

One can plot a graph showing rms power vs. frequency for a given signal. This kind of plot is called the power spectrum. For the waveform of figure 8.1 the power spectrum will have a high value at ν_0 and will have lower values for the signals produced by the glitches. Closer analysis reveals that there is a recognizable, somewhat constant repetition rate associated with the glitches. In fact, we can deduce that there is a significant amount of power in another signal whose period is the period of the glitches as shown in figure 8.2. Let's call the frequency of the glitches ν_s. Since this is the case, we will observe a noticeable amount of power in the spectrum at ν_s with an amplitude which is related to the characteristics of the glitches. The power spectrum shown in figure 8.3 has this feature. A predominant ν_s component has been depicted, but other harmonics also exist.

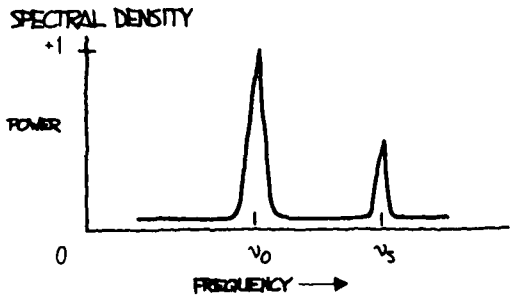


FIGURE 8.3

Some noise will cause the instantaneous frequency to "jitter" around ν_0, with probability of being higher or lower than ν_0. We thus usually find a "pedestal" associated with ν_0 as shown in figure 8.4.

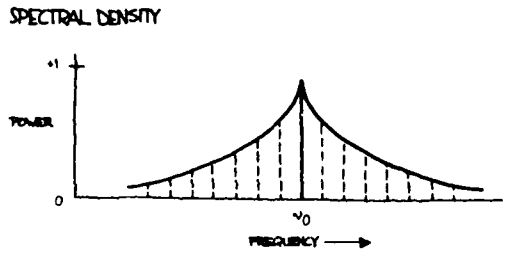


FIGURE 8.4

The process of breaking down a signal into all of its various components of frequency is called Fourier expansion (see sec. X). In other words, the addition of all the frequency components, called Fourier frequency components, produces the original signal. The value of a Fourier frequency is the difference between the frequency component and the fundamental frequency. The power spectrum can be normalized to unity such that the total area under the curve equals one. The power spectrum normalized in this way is the power spectral density.

The power spectrum, often called the RF spectrum, of V(t) is very useful in many applications. Unfortunately, if one is given the RF spectrum, it is impossible to determine whether the power at different Fourier frequencies is a result of amplitude fluctuations "$\varepsilon(t)$" or phase fluctuations "$\phi(t)$. The RF spectrum can be separated into two independent spectra, one being the spectral density of "$\varepsilon(t)$" often called the AM power spectral density and the other being the spectral density of "$\phi(t)$".

For the purposes here, the phase-fluctuation components are the ones of interest. The spectral density of phase fluctuations is denoted by $S_\phi(f)$ where "f" is Fourier frequency. For the frequently encountered case where the AM power spectral density is negligibly small and the total modulation of the phase fluctuations is small (mean-square value is much less than one rad^2), the RF spectrum has approximately the same shape as the phase spectral density. However, a main difference in the representation is that the RF spectrum includes the fundamental signal (carrier), and the phase spectral density does not. Another major difference is that the RF spectrum is a power spectral density and is measured in units of watts/hertz. The phase spectral density involves no "power" measurement of the electrical signal. The units are radians^2/hertz. It is tempting to think of $S_\phi(f)$ as a "power" spectral density because in practice it is measured by passing V(t) through a phase detector and measuring the detector's output power spectrum. The measurement technique makes use of the relation that for small deviations ($\delta\phi \ll 1$ radian),

$$S_\phi(f) = \left(\frac{V_{rms}(f)}{V_s} \right)^2 \qquad (8.3)$$

where $V_{rms}(f)$ is the root-mean-square noise voltage per $\sqrt{Hz}$ at a Fourier frequency "f", and V_s is the sensitivity (volts per radian) at the phase quadrature output of a phase detector which is comparing the two oscillators. In the next section, we will look at a scheme for directly measuring $S_\phi(f)$.

One question we might ask is, "How do frequency changes relate to phase fluctuations?" After all it's the frequency stability of an oscillator that is a major consideration in many applications. The frequency is equal to a rate of change in the phase of a sine wave. This tells us that fluctuations in an oscillator's output frequency are related to phase fluctuations since we must change the rate of "$\phi(t)$" to accomplish a shift in "$\nu(t)$", the frequency at time t. A rate of change of total "$\Phi_T(t)$" is denoted by "$\dot{\Phi}_T(t)$". We have then

$$2\pi\nu(t) = \dot{\Phi}_T(t) \qquad (8.4)$$

The dot denotes the mathematical operation of differentiation on the function Φ_T with respect to its independent variable t.* From eq (8.4)

* As an analogy, the same operation relates the position of an object with its velocity.

and eq (1.1) we get

$$2\pi\nu(t) = \dot{\phi}_T(t) = 2\pi\nu_0 + \dot{\phi}(t)$$

Rearranging, we have

$$2\pi\nu(t) - 2\pi\nu_0 = \dot{\phi}(t)$$

or

$$\nu(t) - \nu_0 = \frac{\dot{\phi}(t)}{2\pi} \qquad (8.5)$$

The quantity $\nu(t) - \nu_0$ can be more conveniently denoted as $\delta\nu(t)$, a change in frequency at time t. Equation (8.5) tells us that if we differentiate the phase fluctuations $\phi(t)$ and divide by 2π, we will have calculated the frequency fluctuation $\delta\nu(t)$. Rather than specifying a frequency fluctuation in terms of shift in frequency, it is useful to denote $\delta\nu(t)$ with respect to the nominal frequency ν_0. The quantity $\frac{\delta\nu(t)}{\nu_0}$ is called the fractional frequency fluctuation** at time t and is signified by the variable y(t). We have

$$y(t) = \frac{\delta\nu(t)}{\nu_0} = \frac{\dot{\phi}(t)}{2\pi\nu_0} \qquad (8.6)$$

The fractional frequency fluctuation y(t) is a dimensionless quantity. When talking about frequency stability, its appropriateness becomes clearer if we consider the following example. Suppose in two oscillators $\delta\nu(t)$ is consistently equal to + 1 Hz and we have sampled this value for many times t. Are the two oscillators equal in their ability to produce their desired output frequencies? Not if one oscillator is operating at 10 Hz and the other at 10 MHz. In one case, the average value of the fractional frequency fluctuation is 1/10, and in the second case is 1/10,000,000 or 1×10^{-7}. The 10 MHz oscillator is then more precise. If frequencies are multiplied or divided using ideal electronics, the fractional stability is not changed.

In the frequency domain, we can measure the spectrum of frequency fluctuations y(t). The spectral density of frequency fluctuations is denoted by $S_y(f)$ and is obtained by passing the signal from an oscillator through an ideal FM detector and performing spectral analysis on the resultant output voltage. $S_y(f)$ has dimensions of (fractional frequency)^2/Hz or Hz^{-1}. Differentiation of $\phi(t)$ corresponds to multiplication by $\frac{f}{\nu_0}$ in terms of spectral densities. With further calculation, one can derive that

$$S_y(f) = \left(\frac{f}{\nu_0}\right)^2 S_\phi(f) \qquad (8.7)$$

We will address ourselves primarily to $S_\phi(f)$, that is, the spectral density of phase fluctuations. For noise-measurement purposes, $S_\phi(f)$ can be measured with a straightforward, easily duplicated equipment set-up. Whether one measures phase or frequency spectral densities is of minor importance since they bear a direct relationship. It is important, however, to make the distinction and to use eq (8.7) if necessary.

8.1 The Loose Phase-Locked Loop

Section I, 1.1, C described a method of measuring phase fluctuations between two phase-locked oscillators. Now we will detail the procedure for measuring $S_\phi(f)$.

Suppose we have a noisy oscillator. We wish to measure the oscillator's phase fluctuations relative to nominal phase. One can do this by phaselocking another oscillator (called the reference oscillator) to the test oscillator and mixing the two oscillator signals 90° out of phase (phase quadrature). This is shown schematically in figure 8.9. The two oscillators are at the same frequency in long term as guaranteed by the phase-lock loop (PLL). A low-pass filter (to filter the R.F. sum component) is used after the mixer since the difference (baseband) signal is the one of interest. By holding the two signals at a relative phase difference of 90°, short-term phase fluctuations between the test and reference oscillators will appear as voltage fluctuations out of the mixer.

** Some international recommendations replace "fractional" by "normalized".

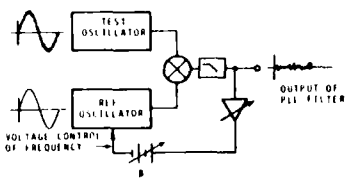


FIGURE 8.9

With a PLL, if we can make the servo time constant very long, then the PLL bandwidth as a filter will be small. This may be done by lowering the gain A_v of the loop amplifier. We want to translate the phase modulation spectrum to baseband spectrum so that it is easily measured on a low frequency spectrum analyzer. With a PLL filter, we must keep in mind that the reference oscillator should be as good or better than the test oscillator. This is because the output of the PLL represents the noise from both oscillators, and if not properly chosen, the reference can have noise masking the noise from the test oscillator. Often, the reference and test oscillators are of the same type and have, therefore, approximately the same noise. We can acquire a meaningful measurement by noting that the noise we measure is from two oscillators. Many times a good approximation is to assume that the noise power is twice that which is associated with one oscillator. $S_\phi(f)$ is general notation depicting spectral density on a per hertz basis. A PLL filter output necessarily yields noise from two oscillators.

The output of the PLL filter at Fourier frequencies above the loop bandwidth is a voltage representing phase fluctuations between reference and test oscillator. It is necessary to make the time-constant of the loop long compared with the inverse of the lowest Fourier frequency we wish to measure. That is, $\tau_c > \frac{1}{2\pi\ f(\text{lowest})}$. This means that if we want to measure $S_\phi(f)$ down to 1 Hz, the loop time-constant must be greater than $\frac{1}{2\pi}$ seconds. One can measure the time-constant by perturbing the loop (momentarily disconnecting the battery is convenient) and noting the time it takes for the control voltage to reach 70% of its final value. The signal from the mixer can then be inserted into a spectrum analyzer. A preamp may be necessary before the spectrum analyzer.

The analyzer determines the mean square volts that pass through the analyzer's bandwidth centered around a pre-chosen Fourier frequency f. It is desireable to normalize results to a 1 Hz bandwidth. Assuming white phase noise (white PM), this can be done by dividing the mean square voltage by the analyzer bandwidth in Hz. One may have to approximate for other noise processes. (The phase noise sideband levels will usually be indicated in rms volts-per-root-Hertz on most analyzers.)

8.2 Equipment for Frequency Domain Stability Measurements

(1) Low-noise mixer

This should be a high quality, double-balanced type, but single-ended types may be used. The oscillators should have well-buffered outputs to be able to isolate the coupling between the two input RF ports of the mixer. Results that are too good may be obtained if the two oscillators couple tightly via signal injection through the input ports. We want the PLL to control locking. One should read the specifications in order to prevent exceeding the maximum allowable input power to the mixer. It is best to operate near the maximum for best signal-to-noise out of the IF port of the mixer and, in some cases, it is possible to drive the mixer into saturation without burning out the device.

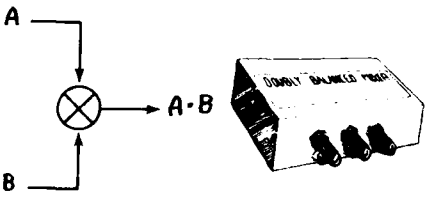


FIGURE 8.10

(2) Low-noise DC amplifier

The amount of gain A_v needed in the loop amplifier will depend on the amplitude of the mixer output and the degree of

varactor control in the reference oscillator. We may need only a small amount of gain to acquire lock. On the other hand, it may be necessary to add as much as 80 dB of gain. Good low-noise DC amplifiers are available from a number of sources, and with cascading stages of amplification, each contributing noise, it will be the noise of the first stage which will add most significantly to the noise being measured. If a suitable low-noise first-stage amplifier is not readily available, a schematic of an amplifier with 40 dB of gain is shown in figure 8.11 which will serve nicely for the first stage. Amplifiers with very low equivalent input noise performance are also available from many manufacturers. The response of the amplifier should be flat from DC to the highest Fourier frequency one wishes to measure. The loop time-constant is inversely related to the gain A_v and the determination of A_v is best made by experimentation knowing that $\tau_c < \frac{1}{2\pi f\ (\text{lowest})}$.

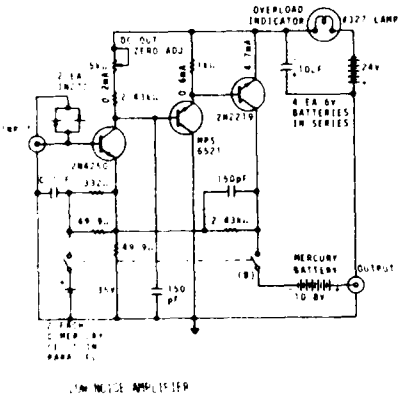


FIGURE 8.11

(3) Voltage-controlled reference quartz oscillator

This oscillator should be a good one with specifications available on its frequency domain stability. The reference should be no worse than the test oscillator. The varactor control should be sufficient to maintain phase-lock of the reference. In general, low quality test oscillators may have varactor control of as much as 1×10^{-6} fractional frequency change per volt. Some provision should be available on the reference oscillator for tuning the mean frequency over a frequency range that will enable phase-lock. Many factors enter into the choice of the reference oscillator, and often it is convenient to simply use two test oscillators phase-locked together. In this way, one can assume that the noise out of the PLL filter is no worse than 3 dB greater than the noise from each oscillator. If it is uncertain that both oscillators are contributing approximately equal noise, then one should perform measurements on three oscillators taking two at a time. The noisier-than-average oscillator will reveal itself.

(4) Spectrum analyzer

The signal analyzer typically should be capable of measuring the noise in rms volts in a narrow bandwidth from near 1 Hz to the highest Fourier frequency of interest. This may be 50 kHz for carrier frequencies of 10 MHz or lower. For voltage measuring analyzers, it is typical to use units of "volts per $\sqrt{\text{Hz}}$". The spectrum analyzer and any associated input amplifier will exhibit high-frequency rolloff. The Fourier frequency at which the voltage has dropped by 3 dB is the measurement system bandwidth f_h, or $\omega_h = 2\pi f_h$. This can be measured directly with a variable signal generator.

Section X describes how analysis can be performed using a discrete fourier transform analyser. Expanding digital technology has made the use of fast-fourier transform analysis affordable and compact.

Rather than measure the spectral density of phase fluctuations between two oscillators, it is possible to measure the phase fluctuations introduced by a device such as an active filter or amplifier. Only a slight modification of the existing PLL filter equipment set up is needed. The scheme is shown in figure 8.12.

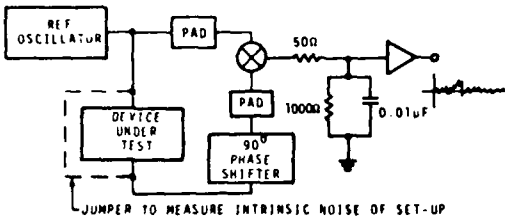


FIGURE 8.12

Figure 8.12 is a differential phase noise measurement set-up. The output of the reference oscillator is split so that part of the signal passes through the device under test. We want the two signals going to the mixer to be 90° out of phase, thus, phase fluctuations between the two input ports cause voltage fluctuations at the output. The voltage fluctuations then can be measured at various Fourier frequencies on a spectrum analyzer.

To estimate the noise inherent in the test set-up, one can in principle bypass the device under test and compensate for any change in amplitude and phase at the mixer. The PLL filter technique must be converted to a differential phase noise technique in order to measure inherent test equipment noise. It is a good practice to measure the system noise before proceeding to measurement of device noise.

A frequency domain measurement set-up is shown schematically in figure 8.13. The component

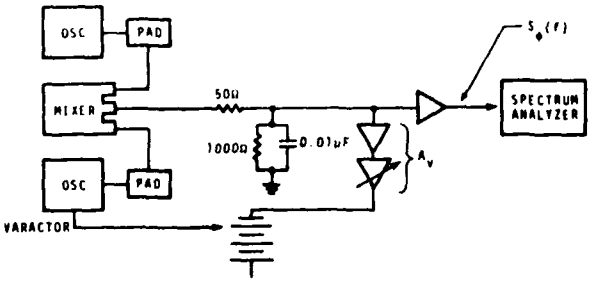


FIGURE 8.13

values for the low-pass filter out of the mixer are suitable for oscillators operating at around 5 MHz.

The active gain element (a_v) of the loop is a DC amplifier with flat frequency response. One may replace this element by an integrator to achieve high gain near DC and hence, maintain better lock of the reference oscillator in long term. Otherwise long-term drift between the reference and test oscillators might require manual re-adjustment of the frequency of one or the other oscillator.[5]

8.3 Procedure and Example

At the input to the spectrum analyzer, the voltage varies as the phase fluctuations in short-term

$$S_\phi(f) = \left(\frac{V_{rms}\ (f)}{V_s}\right)^2$$

V_s is the phase sensitivity of the mixer in volts per radian. Using the previously described equipment set-up, V_s can be measured by disconnecting the feedback loop to the varactor of the reference oscillator. The peak voltage swing is equal to V_s in units of volts/rad if the resultant beat note is a sine wave. This may not be the case for state-of-the-art $S_\phi(f)$ measurements where one must drive the mixer very hard to achieve low mixer noise levels. Hence, the output will not be a sine wave, and the volts/rad sensitivity must be estimated by the slew-rate (through zero volts) of the resultant square-wave out of the mixer/amplifier.

The value for the measured $S_\phi(f)$ in decibels is given by:

$$S_\phi(f) = 20 \log \frac{V_{rms}\text{Voltage at f}}{V_s\text{full-scale }\Phi\text{-detector voltage}}$$

EXAMPLE: Given a PLL with two oscillators such that, at the mixer output:

V_s = 1 volt/rad

V_{rms}(45 Hz) = 100 nV per root hertz
solve for S_ϕ(45 Hz).

$$S_\phi(45\ \text{Hz}) = \left(\frac{100\ \text{nV}\ (\text{Hz}^{-\frac{1}{2}})}{1\ \text{V/rad}}\right)^2 = \left(\frac{10^{-7}}{1}\right)^2\ \text{rad}^2\ \text{Hz}^{-1}$$

$$= 10^{-14}\ \frac{\text{rad}^2}{\text{Hz}}$$

In decibels,

$$S_\phi(45\ \text{Hz}) = 20\ \log\ \frac{100\ \text{nv}}{1\ \text{V}} = 20\ \log\ \frac{10^{-7}}{10^{0}}$$

$$= 20\ (-7) = -140\ \text{dB at 45 Hz}$$

In the example, note that the mean frequency of the oscillators in the PLL was not essential to computing $S_\phi(f)$. However, in the application of $S_\phi(f)$, the mean frequency ν_o is necessary information. Along with an $S_\phi(f)$, one should always attach ν_o. In the example above ν_o = 5 MHz, so we have

$$S_\phi\ (45\ \text{Hz}) = 10^{-14}\ \frac{\text{rad}^2}{\text{Hz}},\ \nu_o = 5\ \text{MHz}.$$

From eq (8.7), $S_y(f)$ can be computed as

$$S_y\ (45\ \text{Hz}) = \left(\frac{45}{5 \times 10^{6}}\right)^2\ 10^{-14}\ \frac{\text{rad}^2}{\text{Hz}}$$

$$S_y\ (45\ \text{Hz}) = 8.1 \times 10^{-25}\ \text{Hz}^{-1},\ \nu_o = 5\ \text{MHz}.$$

IX. POWER-LAW NOISE PROCESSES

Power-law noise processes are models of precision oscillator noise that produce a particular slope on a spectral density plot. We often classify these noise processes into one of five categories. For plots of $S_\phi(f)$, they are:

1. Random walk FM (random walk of frequency), S_ϕ plot goes down as $1/f^4$.
2. Flicker FM (flicker of frequency), S_ϕ plot goes down as $1/f^3$.
3. White FM (white of frequency), S_ϕ plot goes down as $1/f^2$.
4. Flicker PM (flicker of phase), S_ϕ plot goes down as $1/f$.
5. White PM (white of phase), S_ϕ plot is flat.

Power law noise processes are characterized by their functional dependence on Fourier frequency. Equation 8.7 relates $S_\phi(f)$ to $S_y(f)$, the spectral density of frequency fluctuations. Translation of $S_y(f)$ to time-domain data $\sigma_y(\tau)$ for the five model noise processes is covered later in section XI.

The spectral density plot of a typical oscillator's output usually is a combination of different power-law noise processes. It is very useful and meaningful to categorize the noise processes. The first job in evaluating a spectral density plot is to determine which type of noise exists for a particular range of Fourier frequencies. It is possible to have all five noise processes being generated from a single oscillator, but, in general, only two or three noise processes are dominant. Figure 9.1 is a graph of $S_\phi(f)$ showing the five noise processes on a log-log scale. Figure 9.2 shows the spectral density of phase fluctuations for a typical high-quality oscillator.

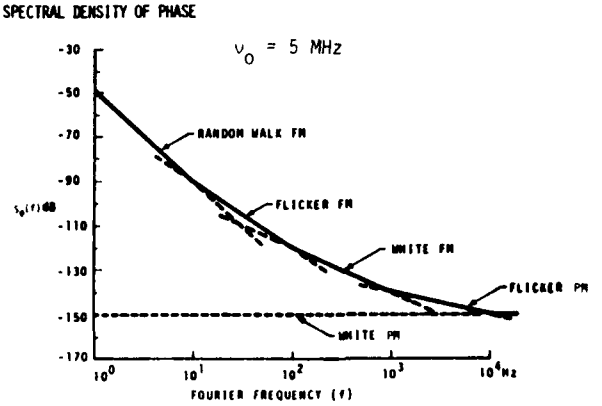


FIGURE 9.1

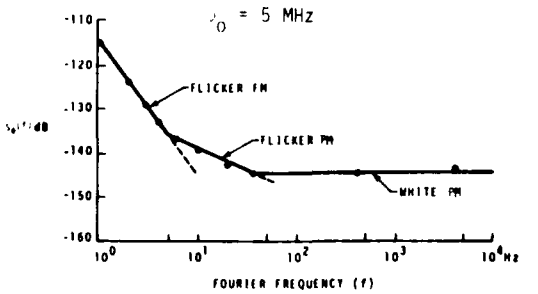


FIGURE 9.2

X. PITFALLS IN DIGITIZING THE DATA

The advent and prolific use of digital computers has changed the manner in which processing of analog signals takes place if a computer is used. This section addresses the most common problems in such analyses.

10.1 Discrete-Continuous Processes

Digital processing implies that data must be presented to a computer or other processor as an array of numbers whether in a batch or in a time series. If the data are not already in this form (it usually is not when considering frequency stability measurements), then it is necessary to transform to this format by digitizing. Usually, the signal available for analysis is a voltage which varies with frequency or phase difference between two oscillators.

10.2 Digitizing the Data

Digitizing the data is the process of converting a continuous waveform into discrete numbers. The process is completed in real time using an analog-to-digital converter (ADC). Three considerations in the ADC are of importance here:

1. Conversion time
2. Resolution (quantization uncertainty)
3. Linearity

An ADC "looks at" an incoming waveform at equispaced intervals of time T. Ideally, the output of the ADC is the waveform (denoted by y(t)) multiplied by a series of infinitely narrow sampling intervals of unit height as in figure 10.1. We have at t = T

$$y_1(t) = y(t)\delta(t-T) = y(T)\delta(t-T) \qquad (10.1)$$

where $\delta(t-T)$ is a delta function. If y(t) is continuous at t = nT and n = 0, ±1, ±2,..., then

$$y_\ell(t) = \sum_{n=-\infty}^{\infty} y(nT)\ \delta\ (t-nT) \qquad (10.2)$$

ℓ = integer

The delta function respresentation of a sampled waveform eq (10.2) is useful when a subsequent continuous integration is performed using it.[6]

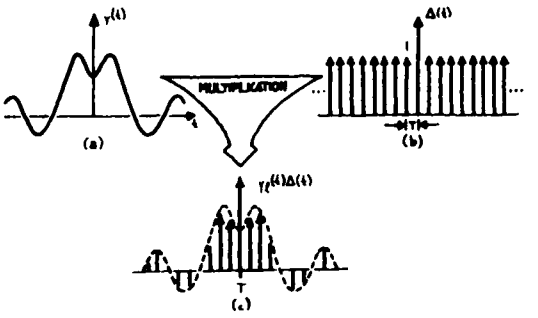


FIGURE 10.1

In ADC's, the input signal is sampled during an aperture time and held for conversion to a digital number, usually binary. Sampling and processing takes time which is specified as the conversion time. This is the total time required for a complete measurement at one sample to achieve a given level of accuracy. If $y_\ell(t)$ is the ideal discrete-time representation of continuous process y(t), then the ADC output denoted by $y'_\ell(t)$ is:

$$y'_\ell = y_{\ell-d} \pm \varepsilon\left(\frac{dy}{dt}\right) \qquad (10.3)$$

where "d" is the conversion time and $\varepsilon \cdot \frac{dy}{dt}$ the accuracy tolerance at "d" as a function of rate-of-change in y(t). In general a trade-off exists between d and ε. For example, for a commonly available, high-quality 10-bit ADC, a conversion time of d = 10 μs yields a maximum error of 3%. Whereas given a 30 μs conversion time, we can obtain 0.1% maximum error.

The error due to conversion time "d" is many times negligible since processing in digital filters and spectrum analysis takes place after the converter. Conversion time delay can be of critical concern, however, where real-time processing at speeds of the order of "d" become important

such as in digital servo loops where corrections are needed for fast changing errors.

A portion of the conversion-time error is a function of the rate of change $\frac{dy}{dt}$ of the process if the sample-and-hold portion of the ADC relies on the charging of a capacitor during an aperture time. This is true because the charge cycle will have a finite time-constant and because of aperture time uncertainty. For example, if the time-constant is 0.1 ns (given by say a 0.1 Ω source resistance charging a 0.001 μfd capacitor), then a 0.1% nominal error will exist for slope $\frac{\Delta y}{\Delta t}$ = 1V/μs due to charging. With good design, this error can be reduced. The sampling circuit (before charge) is usually the dominant source of error and logic gate-delay jitter creates an aperture time uncertainty. The jitter typically is between 2-5 ns which means an applied signal slewing at, say, 1 V/μs produces an uncertainty of 2-5 mV. Since $\varepsilon \cdot \frac{dy}{dt}$ is directly proportional to signal slewing rate, it can be anticipated that high-level, high-frequency components of y(t) will have the greatest error in conversion. For typical ADC's, less than 0.1% error can be achieved by holding $\frac{\Delta y}{\Delta t}$ to less than 0.2 V/μs.

The continuous process y(t) is partitioned into 2^n discrete ranges for n-bit conversion. All analog values within a given range are represented by the same digital code, usually assigned to the nominal midrange value. There is, therefore, an inherent quantization uncertainty of ± ½ least-significant bit (LSB), in addition to other conversion errors. For example, a 10-bit ADC has a total of 1024 discrete ranges with a lowest order bit then representing about 0.1% of full scale and quantization uncertainty of ± 0.05%.

We define the dynamic range of a digital system as the ratio between the maximum allowable value of the process (prior to any overflow condition) and the minimum discernable value. The dynamic range when digitizing the data is set by the quantizing uncertainty, or resolution, and is the ratio of 2^n to ½ LSB. (If additive noise makes coding ambiguous to the ½ LSB level, then the dynamic range is the ratio of 2^n to the noise uncertainty, but this is usually not the case.) For example, the dynamic range of a 10-bit system is 2^n = 1024 to ½, or 2048 to 1. Expressed in dB's, this is

$$20 \log 2048 = 66.2 \text{ dB}$$

if referring to a voltage-to-code converter.

The converter linearity specifies the degree to which the voltage-to-code transfer approximates a straight line. The nonlinearity is the deviation from a straight line drawn between the end points (all zeros to all ones code). It is usually not acceptable to have nonlinearity greater than ½ LSB which means that the sum of the positive errors or the sum of the negative errors of the individual bits must not exceed ½ LSB (or ± ½ LSB). The linearity specification used in this context includes all effects such as temperature errors under expected operating temperature extremes and power supply sensitivity errors under expected operating supply variations.

10.3 Aliasing

Figure 10.1 illustrates equispaced sampling of continuous process y(t). It is important to have a sufficient number of samples/second to properly describe information in the high frequencies. On the other hand, sampling at too high a rate may unnecessarily increase the processing labor. As we reduce the rate, we see that sample values could represent low or high frequencies in y(t). This property is called aliasing and constitutes a source of error similiar to "imaging" which occurs in analog frequency mixing schemes (i.e., in the multiplication of two different signals).

If the time between samples (k) is T seconds, then the sampling rate is $\frac{1}{T}$ samples per second. Then useful data in y(t) will be from 0 to $\frac{1}{2T}$ Hz and frequencies higher than $\frac{1}{2T}$ Hz will be folded into the lower range from 0 to $\frac{1}{2T}$ Hz and confused with data in this lower range. The cutoff frequency is then given by

$$f_s = \frac{1}{2T} \qquad (10.4)$$

and is sometimes called the "Nyquist frequency."

We can use the convolution theorem to simply illustrate the existence of aliases. This theorem states that multiplication in the time domain corresponds to convolution in the frequency domain, and the time domain and frequency domain representations are Fourier transform pairs. The Fourier transform of y(t) in figure 10.1(a) is denoted by Y(f), thus:

$$Y(f) = \int_{-\infty}^{\infty} y(t)e^{-j2\pi ft}dt \qquad (10.5)$$

and

$$y(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(f)e^{j2\pi ft}df \qquad (10.6)$$

The function Y(f) is depicted in figure 10.2(a). The Fourier transform of $\Delta(t)$ is shown in figure 10.2(b) and is given by $\Delta(f)$ where applying the discrete transform yields

$$\Delta(f) = \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T}), \qquad (10.7)$$

recalling that

$$\Delta(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT), \qquad (10.8)$$

from eq (10.2)

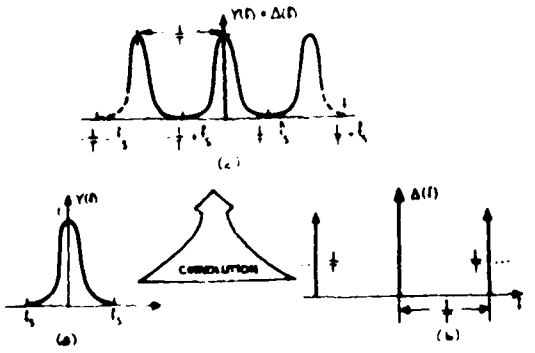


FIGURE 10.2

Y(f) convolved with Δf is denoted by Y(f)* $\Delta(f)$ and is shown in figure 10.2(c). We see that the transform Y(f) is repeated with origins at $f = \frac{n}{T}$. Conversely, high frequency data with information around $f = \frac{n}{T}$ will fold into the data around the origin between $-f_s$ and $+f_s$. In the computation of power spectra, we encounter errors as shown in figure (10.3).

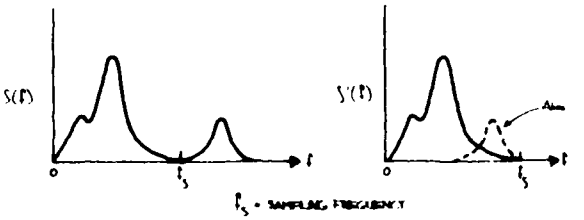


FIGURE 10.3

Aliased power spectra due to folding. (a) True Spectra, (b) Aliased Spectra.

Two pioneers in information theory, Harold Nyquist and Claude Shannon, developed design criteria for discrete-continuous processing systems. Given a specified accuracy, we can convey time-domain process y(t) through a finite bandwidth whose upper limit f_N is the highest significant spectral component of y(t). For discrete-continuous process $y_k(t)$, ideally the input signal spectrum should not extend beyond f_s, or

$$f_N \leq f_s \qquad (10.9)$$

where f_s is given by eq (10.4). Equation (10.9) is refered to as the "Shannon limit."

In practice, there is never a case in which there is absolutely no signal or noise component above f_N. Filters are used before the ADC in order to suppress components above f_N which fold into the lower bandwidth of interest. This so-called anti-aliasing filter usually must be quite sophisticated in order to have low ripple in the passband, constant phase delay in the passband, and steep rolloff characteristics. In examining the rolloff requirements of the anti-aliasing filter, we can apply a fundamental filter property that the output spectrum is equal to the input

spectrum multiplied by the square of the frequency response function; that is,

$$S(f)\ [H(f)]^2 = \underset{\text{out}}{S(f)} \qquad (10.10)$$

The filter response must be flat to f_N and attenuate aliased noise components at $\frac{n}{T} \pm f = 2nf_s \pm f$. In digitizing the data, the observed spectra will be the sum of the baseband spectrum (to f_N) and all spectra which are folded into the baseband spectrum

$$\begin{aligned} \underset{\text{observed}}{S(f)} &= S_0(f) + S_{-1}(2f_s - f) + S_{+1}(2f_s + f) \\ &\quad + S_{-2}(4f_s - f) \ldots + S_i\left(2i\left(f_s + \tfrac{1}{T} f\right)\right) \\ &= S_0(f) + \sum_{i=-M}^{M} S_i\left(2i\left(f_s + f\right)\right) \end{aligned} \qquad (10.11)$$

where M is an appropriate finite limit.

For a given rejection at an upper frequency, clearly the cutoff frequency f_c for the anti-aliasing filter should be as low as possible to relax the rolloff requirements. Recall that an nth order low-pass filter has frequency response function

$$H(f) = \frac{1}{1 + j\left(\frac{f}{f_c}\right)^n} \qquad (10.12)$$

and output spectrum

$$\underset{\text{out}}{S(f)} = \frac{S(f)}{1 + \left(\frac{f}{f_c}\right)^{2n}} \qquad (10.13)$$

and after sampling, we have (applying eq (10.11))

$$\underset{\text{observed}}{S(f)} = \frac{S_0(f)}{1 + \left(\frac{f}{f_c}\right)^{2n}} + \sum_{i=-M}^{M} \frac{S_i\left(2i\left(f_s + \frac{1}{T} f\right)\right)}{1 + \frac{2i\left(f_s + \frac{1}{T} f\right)^{2n}}{f_c}} \qquad (10.14)$$

If f_c is chosen to be higher than f_N, then the first term (baseband spectrum) is negligibly affected by the filter, which is our hope. It is the second term (the sum of the folded in spectra) which causes an error.

As an example of the rolloff requirement, consider the measurement of noise process n(t) at f = 400 Hz in a 1 Hz bandwidth on a digital spectrum analyzer. Suppose n(t) is white; that is,

$$S_n(f) = k_o \qquad (10.15)$$

k_o = constant

Suppose further that we wish to only measure the noise from 10 Hz to 1 kHz; thus f_N = 1 kHz. Let us assume a sampling frequency of $f_s = 2f_N$ or 2 kHz. If we impose a 1 dB error limit in S_{observed} and have 60 dB of dynamic range, then we can tolerate an error limit of 10^{-6} due to aliasing effects in this measurement, and the second term in eq (10.14) must be reduced to this level. We can choose f_c = 1.5kHz and obtain

$$\underset{\text{observed}}{S(f)} = k_o + \sum_{i=-M}^{M} \frac{k_o}{1 + \frac{2i\left(f_s + \frac{1}{T} f\right)^{2n}}{f_c}} \qquad (10.16)$$

The term in the series which contributes most is at i = -1, the nearest fold-in. The denominator must be 10^6 or more to realize the allowable error limit and at $n \geq 8$ this condition is met. The next most contributing term is i = +1 at which the error is $< 10^{-7}$ for n = 8, a negligible contribution. The error increases as f increases for a fixed n because the nearest fold-in (i = -1) is coming down in frequency (note fig. 10.2(c)) and power there is filtered less by the anti-aliasing filter. Let us look at the worst case (f = 1kHz) to determine a design criteria for this example. At f = 1 kHz, we must have $n \geq 10$.

Thus the requirement in this example is for a 10-pole low-pass filter (60 dB/octave rolloff).

10.4 Some History of Spectrum Analysis Leading to the Fast Fourier Transform

Newton in his Principia (1687) documented the first mathematical treatment of wave motion al-

though the concept of harmonics in nature was pointed out by Pythagoras, Kepler, and Galileo. However, it was the work of Joseph Fourier in 1807 which showed that almost any function of a real variable could be represented as the sum of sines and cosines. The theory was rigorously treated in a document in 1822.

In using Fourier's technique, the periodic nature of a process or signal is analyzed. Fourier analysis assumes we can apply fixed amplitudes, frequencies, and phases to the signal.

In the early 1900's two relatively independent developments took place: (1) radio electronics and electric power hardware were fast growing technologies; and (2) statistical analysis of events or processes which were not periodic became increasingly understood. The radio engineer explored signal and noise properties of a voltage or current into a load by means of the spectrum analyzer and measurement of the power spectrum. On the other hand, statisticians explored deterministic and stochastic properties of a process by means of the variance and self-correlation properties of the process at different times. Wiener (1930) showed that the variance spectrum (i.e., the breakdown of the variance with Fourier frequency) was the Fourier transform of the autocorrelation function of the process. He also theorized that the variance spectrum was the same as the power spectrum normalized to unit area. Tukey (1949) advocated the use of the variance spectrum in the statistical treatment of all processes because (1) it is more easily interpreted than correlation-type functions; and (2) it fortuitously is readily measureable by the radio engineer.

The 1950's saw rigorous application of statistics to communication theory. Parallel to this was the rapid advancement of digital computer hardware. Blackman and Tukey (1959) and Welch (1961) elaborated on other useful methods of deriving an estimate for the variance spectrum by taking the ensemble time-average sampled, discrete line spectra. The approach assumes the random process is ergodic. Some digital approaches estimate the variance spectrum using Wiener's theorem if correlation-type functions are useful in the analysis, but in general the time-averaged, sample spectrum is the approach taken since its implementation is direct and straightforward. Most always, ergodicity can be assumed.[8,9]

The variance of process y(t) is related to the total power spectrum by

$$\sigma^2[y(t)] = \int_{-\infty}^{\infty} S_y(f)\, df. \qquad (10.17)$$

Since

$$\sigma^2[y(t)] = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} y^2(t)\, dt \qquad (10.18)$$

we see that if y(t) is a voltage or current into a 1-ohm load, then the mean power of y(t) is the integral of $S_y(f)$ with respect to frequency over the entire range of frequencies $(-\infty,\infty)$. $S_y(f)$ is, therefore, the power spectrum of process y(t). The power spectrum curve shows how the variance is distributed with frequency and should be expressed in units of watts per unit of frequency, or volts squared per unit of frequency when the load is not considered.

Direct estimation of power spectra has been carried out for many years through the use of analog instruments. These have variously been referred to as sweep spectrum analyzers, harmonic analyzers, filter banks, and wave analyzers. These devices make use of the fact that the spectrum of the output of a linear system (analog filter) is the spectrum of the input multiplied by the square of the system's frequency response function (real part of the transfer characteristic). Note eq (10.10). If y(t) has spectrum $S_y(f)$ feeding a filter with frequency response function H(f), then its output is

$$S(f)_{\text{filtered}} = [H(f)]^2\, S_y(f) \qquad (10.19)$$

If H(f) is rectangular in shape with width Δf, then we can measure the contribution to the total power spectrum due to $S_y(f \pm \frac{\Delta f}{2})$.

The development of the fast Fourier transform (FFT) in 1965 made digital methods of spectrum

estimation increasingly attractive. Today the choice between digital or analog methods depends more on the objectives of the analysis rather than on technical limitations. However, many aspects of digital spectrum analysis are not well known by the casual user in the laboratory while the analog analysis methods and their limitations are understood to a greater extent.

Digital spectrum analysis is realized using the discrete Fourier transform (DFT), a modified version of the continuous transform depicted in eqs (10.5) and (10.6). By sampling the input waveform y(t) at discrete intervals of time $t_\ell = \Delta t$ representing the sampled waveform by eq (10.2) and integrating eq (10.5) yields

$$Y(f) = \sum_{\ell=-\infty}^{\infty} y(\ell t)e^{-j2\pi f\ell T} \qquad (10.20)$$

Equation (10.20) is a Fourier series expansion. Because f(t) is specified as being bandlimited, the Fourier transform as calculated by eq (10.20) is as accurate as eq (10.5); however, it cannot extend beyond the Nyquist frequency, eq (10.4).

In practice we cannot compute the Fourier transform to an infinite extent, and we are restricted to some observation time T consisting of $n\Delta t$ intervals. This produces a spectrum which is not continuous in f but rather is computed with resolution Δf where

$$\Delta f = \frac{1}{n\Delta t} = \frac{1}{T} \qquad (10.21)$$

With this change, we get the discrete finite transform

$$Y(m\Delta f) = \sum_{n=0}^{N-1} y_\ell(t)e^{-j2\pi m\Delta fnt} \qquad (10.22)$$

The DFT computes a sampled Fourier series, and eq (10.22) assumes that the function y(t) repeats itself with period T. $Y(m\Delta f)$ is called the "line spectrum." A comparison of the DFT with the continuous Fourier transform is shown later in part 10.7.

The fast Fourier transform (FFT) is an algorithm which efficiently computes the line spectrum by reducing the number of adds and multiplies involved in eq (10.22). If we choose $T/\Delta t$ to equal a rational power of 2, then a symmetric matrix can be derived through which $y_\ell(t)$ passes and quickly yields $Y(m\Delta f)$. An N-point transformation by the direct method requires a processing time proportional to N^2 whereas the FFT requires a time proportional to $N \log_2 N$. The approximate ratio of FFT to direct computing time is given by

$$\frac{N \log_2 N}{N^2} = \frac{\log_2 N}{N} = \frac{Y}{N} \qquad (10.23)$$

where $N = 2^Y$. For example, if $N = 2^{10}$, the FFT requires less than 1/100 of the normal processing time.

We must calculate both the magnitude and phase of a frequency in the line spectrum, i.e., the real and imaginary part at the given frequency. N points in the time domain allow N/2 complex quantities in the frequency domain.

The power spectrum of y(t) is computed by squaring the real and imaginary components, adding the two together and dividing by the total time T. We have

$$S_y(m\Delta f) = \frac{R[Y(m\Delta f)]^2 + I[Y(m\Delta f)]^2}{T} \qquad (10.24)$$

This quantity is the sampled power spectrum and again assumes periodicity in process y(t) with total period T.[10]

10.5 Leakage

Sampled digital spectrum analysis always involves transforming a finite block of data. Continuous process y(t) is "looked at" for T time through a data window which can functionally be described by

$$y'(t) = w(t)\cdot y(t) \qquad (10.25)$$

where w(t) is the time domain window. The time-discrete counterpart to eq (10.25) is

$$y'_\ell(t) = w_\ell(t)\cdot y_\ell(t) \qquad (10.26)$$

and $w_\ell(t)$ is now the sampled version of w(t) derived similarly to eq (10.2). Equation (10.26) is equivalent to convolution in the frequency domain, or

$$Y'(m\Delta f) = W(m\Delta f)*Y(m\Delta f) \qquad (10.27)$$

Y'(mΔf) is called the "modified" line spectrum due to convolution of the original line spectrum with the Fourier transform of the time-domain window function.

Suppose the window function is rectangular, and

$$w_\ell(t) = 1, \quad \frac{-T}{2} \leq t \leq \frac{T}{2}$$
$$= 0, \quad t > \left|\frac{T}{2}\right| \qquad (10.28)$$

This window is shown in figure 10.4(a). The Fourier transform of this window is

$$W(m\Delta f) = T\,\frac{\sin\pi\, m\Delta f\, NT}{\pi m\Delta f NT} \qquad (10.29)$$

and is shown in figure 10.4(b). If y(t) is a sine wave, we convolve the spectrum of the sinusoid, a delta function, with W(mΔf).

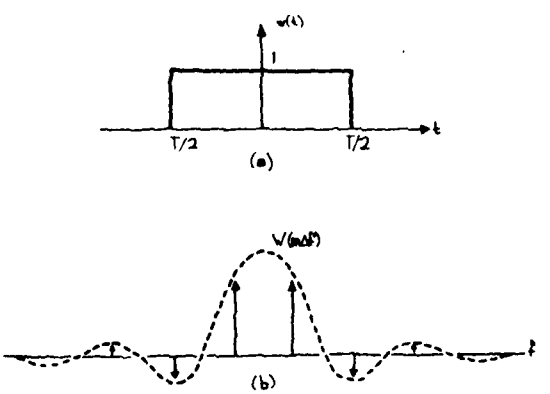


FIGURE 10.4

The transform process (eq 10.22) treats the sample signal as if it were periodically extended. Discontinuities usually occur at the ends of the window function in the extended version of the sampled waveform as in figure 10.5(c). Sample spectra thus represent a periodically extended sampled waveform, complete with discontinuites at its ends, rather than the original waveform.

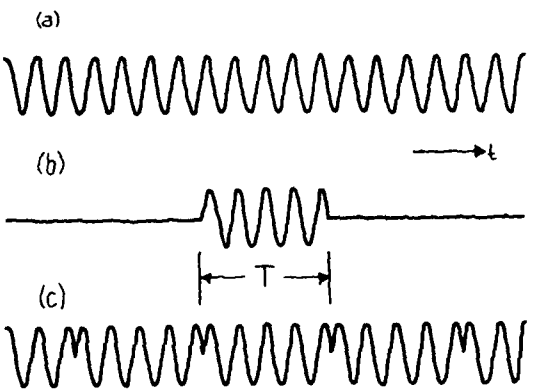


FIGURE 10.5

Spurious components appear near the sinusoid spectrum and this is referred to as "leakage." Leakage results from discontinuites in the periodically extended sample waveform.

Leakage cannot be eliminated entirely, but one can choose an appropriate window function w(t) in order to minimize its effect. This is usually done at the expense of resolution in the frequency domain. An optimum window for most cases is the Hanning window given by:

$$w(t) = [\tfrac{1}{2} - \tfrac{1}{2}\cos(\tfrac{2\pi t}{T})]^a \qquad (10.30)$$

for $0 \leq t \leq T$ and "a" designates the number of times the window is implemented. Figure 10.6(a) shows the window function and 10.6(b) shows the Hanning line shape in the frequency domain for various numbers of "Hanns." Note that this window

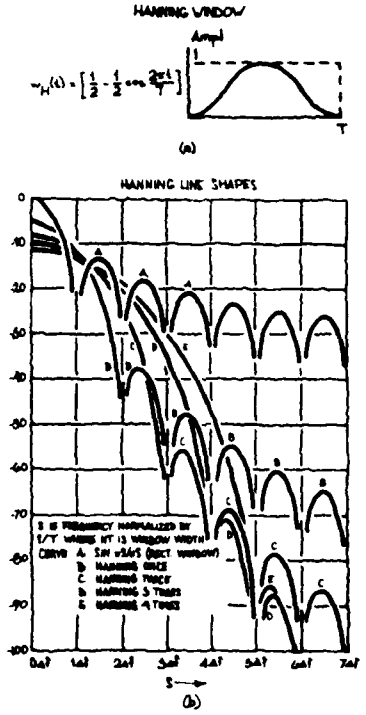


FIGURE 10.6

eliminates discontinuities due to the ends of sample length T.

Each time the Hanning window is applied, the sidelobes in the transform are attenuated by 12 dB/octave, and the main lobe is widened by $2\Delta f$. The amplitude uncertainty of an arbitrary sine wave input is reduced as we increase the number of Hanns; however, we trade off resolution in frequency.

The effective noise bandwidth indicates the departure of the filter response away from a true rectangularly shaped filtered response (frequency domain). Table 10-I lists equivalent noise bandwidth corrections for up to three applications of the Hanning window.[11]

Number of Hanns	Equivalent Noise Bandwidth
1	1.5 Δf
2	1.92 Δf
3	2.31 Δf

TABLE 10.I

10.6 Picket-Fence Effect

The effect of leakage discussed in the previous section gives rise to a sidelobe type response that can be tailored according to the time-window function through which the analyzed signal passes as a block to be transformed to the frequency domain. Using the Hanning window diminishes the amplitudes of the sidelobes, however, it increases the effective bandwidth of the passband around the center frequency. This is because the effective time-domain window length is shorter than a perfect rectangular window. Directly related to the leakage (or sidelobe) effect is one called the "picket-fence" effect. This is because the sidelobes themselves resemble a frequency response which has geometry much like a picket fence.

The existence of both sidelobe leakage and the resultant picket-fence effect are an artifact of the way in which the FFT analysis is performed. Frequency-domain analysis using analog filters involves a continuous signal in and a continuous signal out. On the other hand, FFT analysis involves a continuous signal in, but the transform to the frequency domain is performed on blocks of data. In order to get discrete frequency information from a block, the assumption is made that the block represents one period of a periodic signal. The picket-fence effect is a direct consequence of this assumption. For example, consider a sinewave signal which is transformed from a time-varying voltage to a frequency-domain representation through an FFT. The block of data to be transformed will be length, T, in time. Let's say that the block, T, represents only $4\frac{1}{2}$ cycles of the input sinewave as in figure 10.5. Artificial sidebands will be created in the transform to the frequency domain, whose frequency spacing equals $\frac{1}{T}$, or the reciprocal of the block length. This represents a worst-case condition for sidelobe generation and creates a large number of spurious discrete frequency components as shown in figure 10.7(b).

If, on the other hand, one changes the block time, T, so that the representation is an integral number of cycles of the input sinewave, then the transform will not contain sidelobe leakage compo-

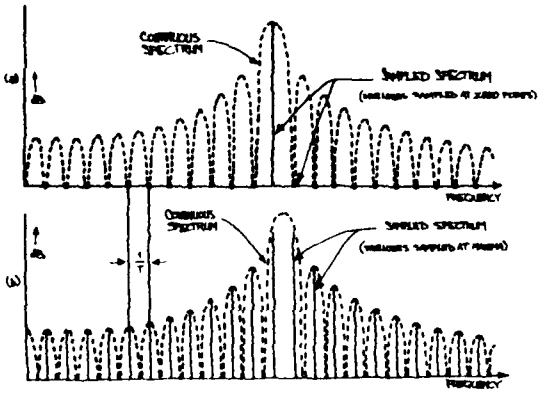


FIGURE 10.7

nents and the artificial sideband frequency components disappear. In practice, when looking at complex input signals, the block time, T, is not synchronous with any component of the transformed part of the signal. As a result, discrete frequency components in the frequency domain have associated with them sidebands which come and go depending on the phase of the time window, T, relative to the sine components of the incoming signal. The effect is much like looking through a picket fence at the sidebands.[12]

An analogy to the sidelobe leakage and picket-fence effect is to record the incoming time-varying signal on a tape loop, which has a length of time, T. The loop of tape then repeats itself with a period of T. This repeating signal is then coupled to a scanning or filter-type spectrum analyzer. The phase discontinuity between the end of one passage of the loop and the beginning of the loop on itself again represents a phase-modulation component, which gives rise to artificial sidebands in the spectrum analysis. A word of caution — this is not what actually happens in a FFT analyzer (i.e., there is no recirculating memory). However, the Fourier transform treats the incoming block as if this were happening.

10.7 Time Domain-Frequency Domain Transforms

A. Integral transform

Figure 10.8 shows the well-known integral transform, which transforms a continuous time-domain signal extending over all time into a

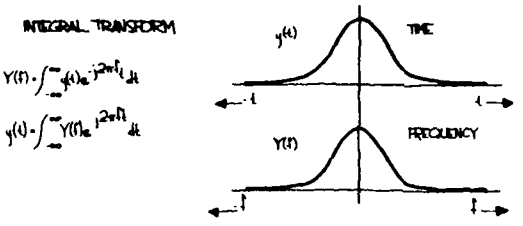


FIGURE 10.8

continuous frequency-domain signal extending over all frequency. This is the ideal transform. In practice, however, one deals with finite times and bandwidths. The integral transform then, at best, is an estimate of the transform and is so for only short, well-behaved signals. That is, the signal goes to zero at infinite time and at infinitely high frequency.

B. Fourier series

The Fourier-series transform assumes periodicity in the time-domain signal for all time. Only one period of the signal (for time T) is required for this kind of transform. The Fourier series treats the incoming signal as periodic with period, T, and continuous. The transformed spectrum is then discrete with infinite harmonic components with frequency spacing of $\frac{1}{T}$. This is shown in figure 10.9.

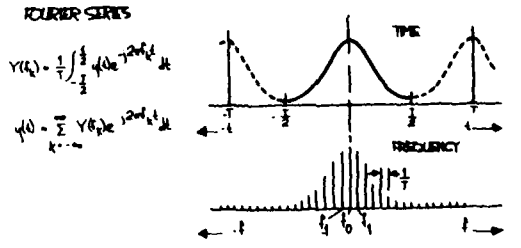


FIGURE 10.9

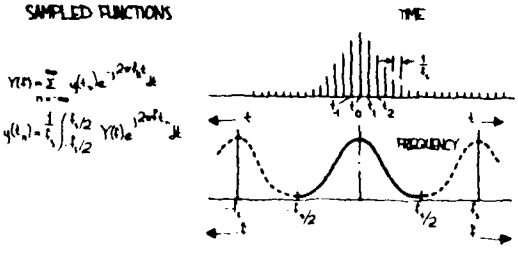


FIGURE 10.10

Figure 10.10 shows the transform from a sampled time-domain signal to the frequency domain. Note that in the frequency domain, the result is repetitive in frequency. This effect, commonly called aliasing, is discussed earlier in section 10.3. Figure 10.9 and figure 10.10 show the symmetry between the time- and frequency-domain transforms of discrete lines.

C. Discrete fourier transform

Finally, we have the sampled, periodic (assumed) time-domain signal which is transformed to a discrete and repetitive (aliased) frequency-domain representation. This is shown in figure 10.11

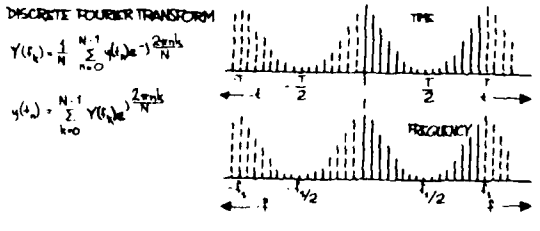


FIGURE 10.11

XI. TRANSLATION FROM FREQUENCY DOMAIN STABILITY MEASUREMENT TO TIME DOMAIN STABILITY MEASUREMENT AND VICE-VERSA.

11.1 Procedure

Knowing how to measure $S_\phi(f)$ or $S_y(f)$ for a pair of oscillators, let us see how to translate the power-law noise process to a plot of $\sigma_y^2(\tau)$. First, convert the spectrum data to $S_y(f)$, the spectral density of frequency fluctuations (see sections III and VIII). There are two quantities which completely specify $S_y(f)$ for a particular power law noise process: (1) the slope on a log-log plot for a given range of f and (2) the amplitude. The slope we shall denote by "α"; therefore f^α is the straight line (on log-log scale) which relates $S_y(f)$ to f. The amplitude will be denoted "h_α"; it is simply the coefficient of f for a range of f. When we examine a plot of spectral density of frequency fluctuations, we are looking at a representation of the addition of all the power-law processes (see sec. IX). We have

$$S_y(f) = \sum_{\alpha=-\infty}^{\infty} h_\alpha f^\alpha \qquad (11.1)$$

In section IX, five power-law noise processes were outlined with respect to $S_\phi(f)$. These five are the common ones encountered with precision oscillators. Equation (8.7) relates these noise processes to $S_y(f)$. One obtains

		with respect to $S_y(f)$	slope on log-log paper
1.	Random Walk FM	(f^{-2}) . . .	$\alpha = -2$
2.	Flicker FM	(f^{-1}) . . .	$\alpha = -1$
3.	White FM	(1) . . .	$\alpha = 0$
4.	Flicker ϕM	(f) . . .	$\alpha = 1$
5.	White ϕM	(f^2) . . .	$\alpha = 2$

Table 11.1 is a list of coefficients for translation from $\sigma_y^2(\tau)$ to $S_y(f)$ and from $S_\phi(f)$ to

$\sigma_y^2(\tau)$. In the table, the left column is the designator for the power-law process. Using the middle column, we can solve for the value of $S_y(f)$ by computing the coefficient "a" and using the measured time domain data $\sigma_y^2(\tau)$. The rightmost column yields a solution for $\sigma_y^2(\tau)$ given frequency domain data $S_\phi(f)$ and a calculation of the appropriate "b" coefficient.

$S_y(f) = H_\alpha f^\alpha$ $\alpha =$	$S_y(f) = a\,\sigma_y^2(\tau)$ $a =$	$\sigma_y^2(\tau) = b\,S_\phi(f)$ $b =$
2 (white phase)	$\frac{(2\pi)^2 \tau^2 f^2}{3 f_h}$	$\frac{3 f_h}{(2\pi)^2 \tau^2 \nu_0^2}$
1 (flicker noise)	$\frac{(2\pi)^2 \tau^2 f}{1.038 + 3 \ln(\omega_h \tau)}$	$\frac{[1.038 + 3 \ln(\omega_h \tau)]f}{(2\pi)^2 \tau^2 \nu_0^2}$
0 (white frequency)	2τ	$\frac{f^2}{2 \tau \nu_0^2}$
-1 (flicker frequency)	$\frac{1}{2 \ln(2) \cdot f}$	$\frac{2 \ln(2) \cdot f^3}{\nu_0^2}$
-2 (random walk frequency)	$\frac{6}{(2\pi)^2 \tau f^2}$	$\frac{(2\pi)^2 \tau f^4}{6 \nu_0^2}$

TABLE 11.1

Conversion table from time domain to frequency domain and from frequency domain to time domain for common kinds of interger power law spectral densities; $f_h (= \omega_h/2\pi)$ is the measurement system bandwidth. Measurement reponse should be within 3 dB from D.C. to f_h (3 dB down high-frequency cutoff is at f_h).

$$S_\phi(f) = \frac{\nu_0^2}{f^2} S_y(f)$$

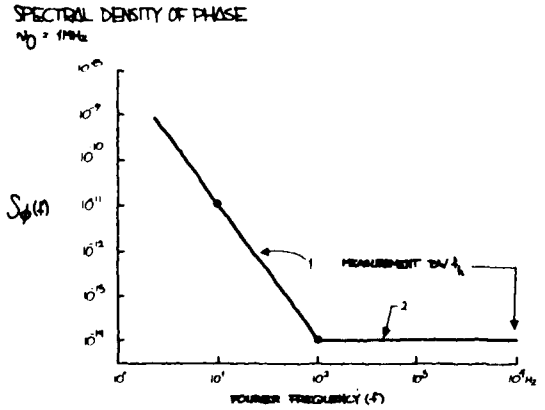


FIGURE 11.1

EXAMPLE:

In the phase spectral density plot of figure 11.1, there are two power-law noise processes for oscillators being compared at 1 MHz. For region 1, we see that when f increases by one decade (that is, from 10 Hz to 100 Hz), $S_\phi(f)$ goes down by three decades (that is, from 10^{-11} to 10^{-14}). Thus, $S_\phi(f)$ goes as $1/f^3 = f^{-3}$. For region 1, we identify this noise process as flicker FM. The rightmost column of table 11.1 relates $\sigma_y^2(\tau)$ to $S_\phi(f)$. The row designating flicker frequency noise yields:

$$\sigma_y^2(\tau) = \frac{2\ell n(2) \cdot f^3}{\nu_0^2} S_\phi(f)$$

One can pick (arbitrarily) a convenient Fourier frequency f and determine the corresponding values of $S_\phi(f)$ given by the plot of figure 11.1. Say, f = 10, thus $S_\phi(10) = 10^{-11}$. Solving for $\sigma_y^2(\tau)$, given ν_0 = 1 MHz, we obtain:

$$\sigma_y^2(\tau) = 1.39 \times 10^{-20}$$

therefore, $\sigma_y(\tau) = 1.18 \times 10^{-10}$. For region 2, we have white PM. The relationship between $\sigma_y^2(\tau)$ and

$S_\phi(f)$ for white PM is:

$$\sigma_y^2(\tau) = \frac{3f_h}{(2\pi)^2 \tau^2 \nu_0^2} S_\phi(f)$$

Again, we choose a Fourier frequency, say f = 100, and see that $S_\phi(100) = 10^{-14}$. Assuming $f_h = 10^4$ Hz, we thus obtain:

$$\sigma_y^2(\tau) = 7.59 \times 10^{-24} \cdot \frac{1}{\tau^2}$$

therefore,

$$\sigma_y(\tau) = 2.76 \times 10^{-12} \cdot \frac{1}{\tau}.$$

The resultant time domain characterization is shown in figure 11.2.

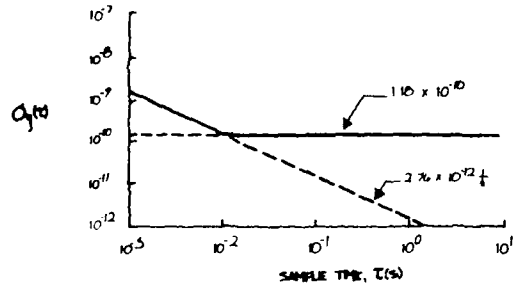


FIGURE 11.2

The translation of $S_\phi(f)$ of figure 11.1 yields this $\sigma_y(\tau)$ plot.

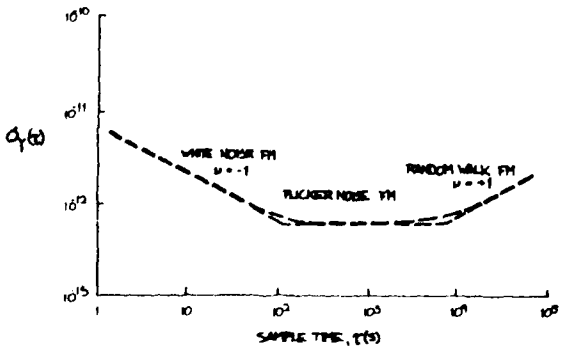


FIGURE 11.3

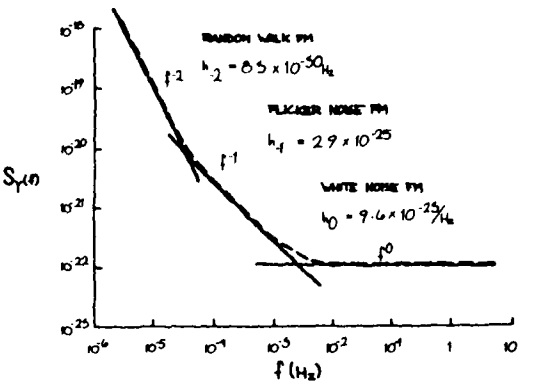


FIGURE 11.4

Figures 11.3 and 11.4 show plots of time-domain stability and a translation to frequency domain. Since table 11.1 has the coefficients which connect both the frequency and time domains, it may be used for translation to and from either domain.

XII. CAUSES OF NOISE PROPERTIES IN A SIGNAL SOURCE

12.1 Power-law Noise Processes

Section IX pointed out the five commonly used power-law models of noise. With respect to $S_\phi(f)$, one can estimate a staight line slope (on a log-log scale) which corresponds to a particular noise type. This is shown in figure 12.1 (also fig.9.1).

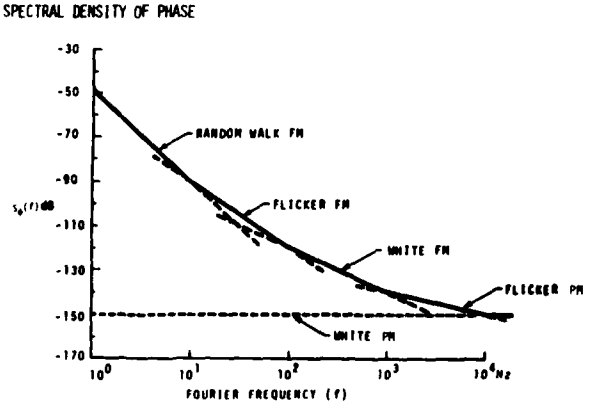


FIGURE 12.1

We can make the following general remarks about power-law noise processes:

1. Random walk FM ($1/f^4$) noise is difficult to measure since it is usually very close to the carrier. Random walk FM usually relates to the OSCILLATOR'S PHYSICAL ENVIRONMENT. If random walk FM is a predominant feature of the spectral density plot then MECHANICAL SHOCK, VIBRATION, TEMPERATURE, or other environmental effects may be causing "random" shifts in the carrier frequency.

2. Flicker FM ($1/f^3$) is a noise whose physical cause is usually not fully understood but may typically be related to the PHYSICAL RESONANCE MECHANISM OF AN ACTIVE OSCILLATOR or the DESIGN OR CHOICE OF PARTS USED FOR THE ELECTRONICS, or ENVIRONMENTAL PROPERTIES. Flicker FM is common in high-quality oscillators, but may be masked by white FM ($1/f^2$) or flicker PM ($1/f$) in lower-quality oscillators.

3. White FM ($1/f^2$) noise is a common type found in PASSIVE-RESONATOR FREQUENCY STANDARDS. These contain a slave oscillator, often quartz, which is locked to a resonance feature of another device which behaves much like a high-Q filter. Cesium and rubidium standards have white FM noise characteristics.

4. Flicker PM ($1/f$) noise may relate to a physical resonance mechanism in an oscillator, but it usually is added by NOISY ELECTRONICS. This type of noise is common, even in the highest quality oscillators, because in order to bring the signal amplitude up to a usable level, amplifiers are used after the signal source. Flicker PM noise may be introduced in these stages. It may also be introduced in a frequency multiplier. Flicker PM can be reduced with good low-noise amplifier design (e.g., using rf negative feedback) and hand-selecting transistors and other electronic components.

5. White PM (f^0) noise is broadband phase noise and has little to do with the resonance mechanism. It is probably produced by similar phenomena as flicker PM ($1/f$) noise. STAGES OF AMPLIFICATION are usually responsible for white PM noise. This noise can be kept at a very low value with good amplifier design, hand-selected components, the addition of narrowband filtering at the output, or increasing, if feasible, the power of the primary frequency source.[13]

12.2 Other types of noise

A commonly encountered type of noise from a signal source or measurement apparatus is the presence of 60 Hz A.C. line noise. Shown in figure 12.2 is a constant white PM noise source with 60 Hz, 120 Hz and 180 Hz components added. This kind of noise is usually caused by AC power getting into the measurement system or the source under test. In the plot of $S_\phi(f)$, one observes discrete line spectra. Although $S_\phi(f)$ is a measure of spectral density, one can interpret the line spectra with no loss of generality, although one usually does not refer to spectral densities when characterizing discrete lines. Figure 12.3 is the time domain representation of the same white phase modulation level with 60 Hz noise. Note that the amplitude of $\sigma_y(t)$ varies up and down depending on sampling time. This is because in the time domain the sensitivity to a periodic wave varies directly as the sampling interval. This effect (which is an alias effect) is a very powerful tool for filtering out a periodic wave imposed on a signal source. By sampling in the time domain at integer periods, one is virtually insensitive to the periodic (discrete line) term.

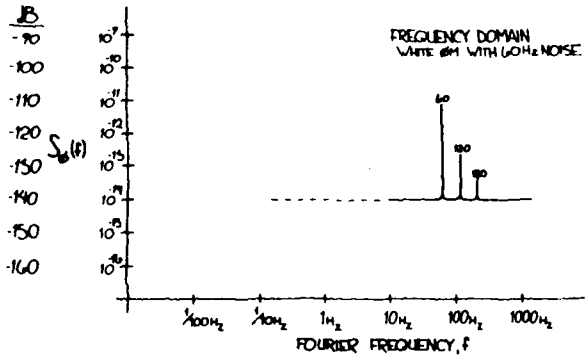


FIGURE 12.2

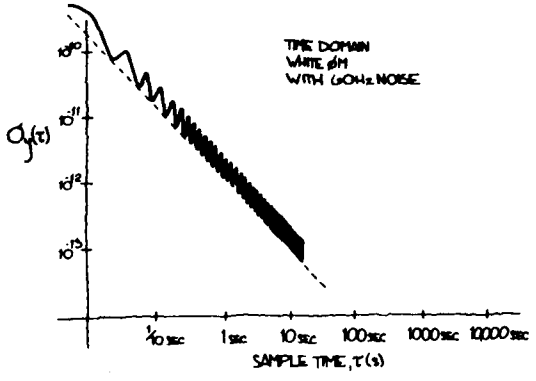


FIGURE 12.3

For example, diurnal variations in data due to day to day temperature, pressure, and other environmental effects can be eliminated by sampling the data once per day. This approach is useful for data with only one periodic term.

Figure 12.4 shows the kind of plot one might see of $S_\phi(f)$ with vibration and acoustic sensitivity in the signal source with the device under vibration. Figure 12.5 shows the translation to the time domain of this effect. Also noted in figure 12.4 is a (typical) flicker FM behavior in the low frequency region. In the translation to time domain (fig. 12.5), the flicker FM behavior masks the white PM (with the superimposed vibration characteristic) for long averaging times.

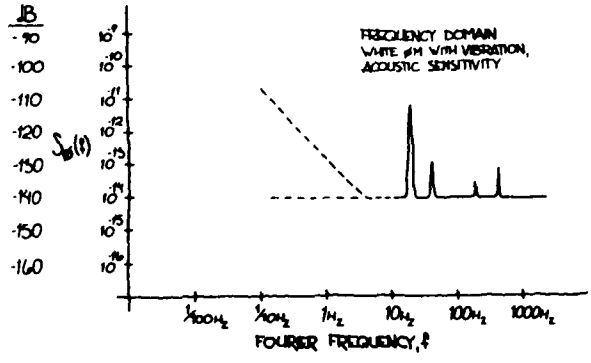


FIGURE 12.4

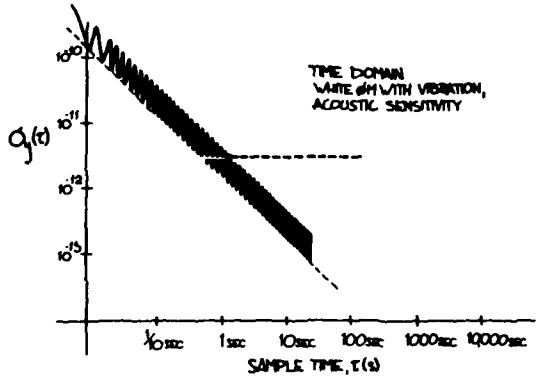


FIGURE 12.5

Figure 12.6 shows examples of plots of two power law processes ($S_\phi(f)$) with a change in the flicker FM level. (Example 1 is identical to the example given in sec. XI.) Figure 12.7 indicates the effect of a lower flicker FM level as translated to the time domain. Note again the existence of both power law noise processes. However for a given averaging time (or Fourier frequency) one noise process may dominate over the other.

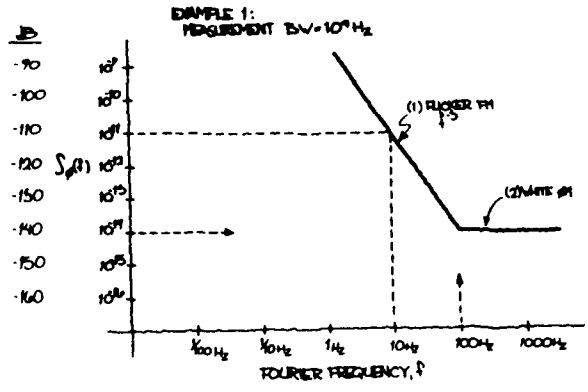


FIGURE 12.6a

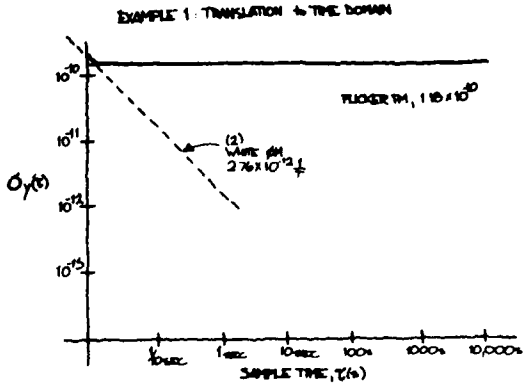


FIGURE 12.7a

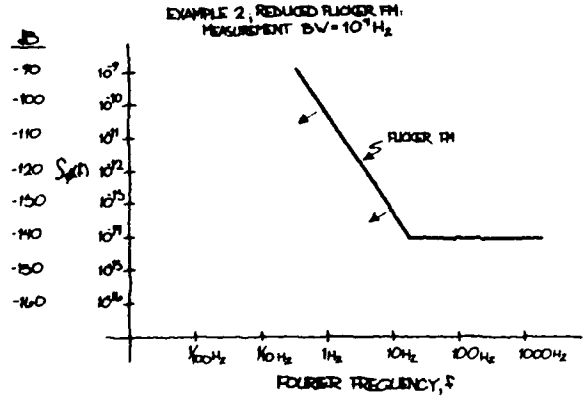


FIGURE 12.6b

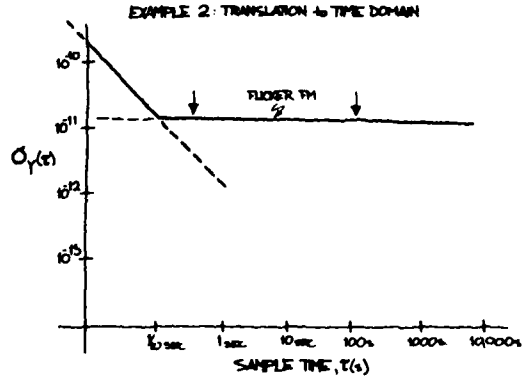


FIGURE 12.7b

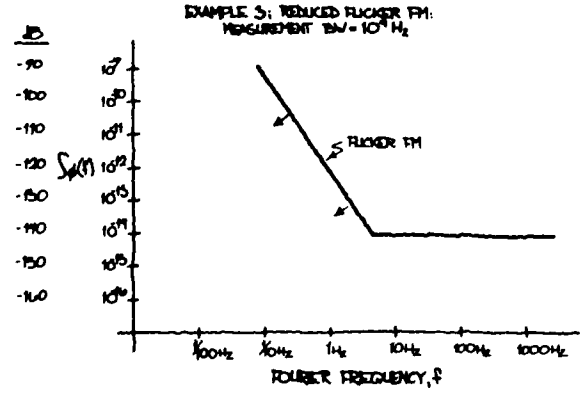


FIGURE 12.6c

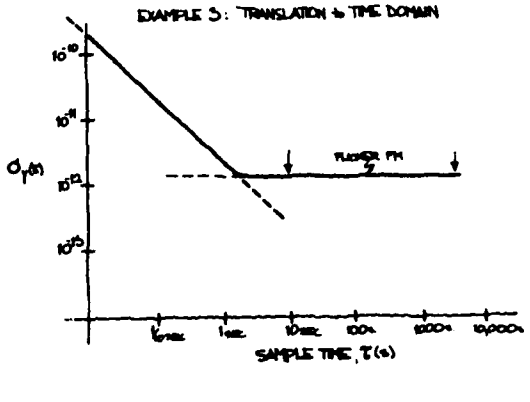


FIGURE 12.7c

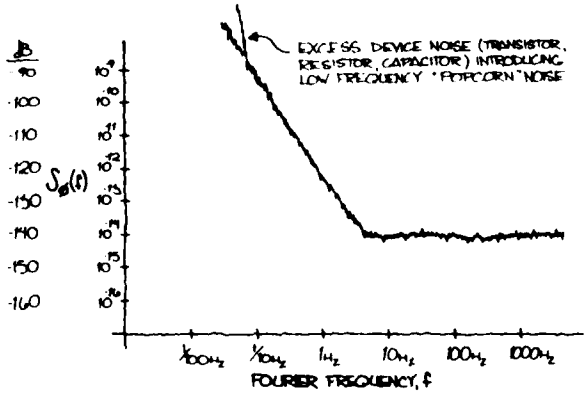


FIGURE 12.8

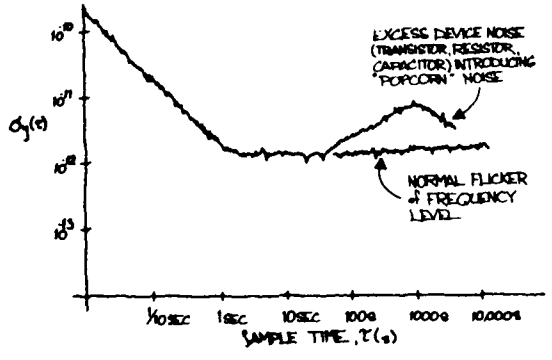


FIGURE 12.9

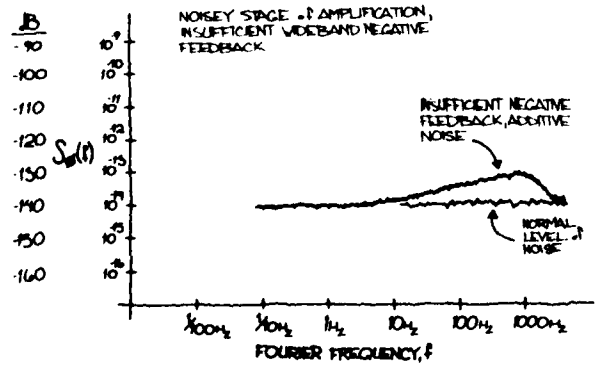


FIGURE 12.10

Excess device noise from transistors, capacitors, resistors, and the like can introduce a low frequency noise which has been referred to as "popcorn" noise because of its sonic qualities. Figure 12.8 shows a plot of $S_\phi(f)$ from a signal source having such excess low frequency noise. Figure 12.9 is the time domain representation. The rise in amplitude of σ_y for long averaging times is particularly aggravating. The solution to this kind of problem if it is introduced by devices is to carefully grade the devices in the assembly and testing process.

Stages of amplification following a signal source many times rely on local degenerate or overall negative feedback schemes in order to minimize the excess noise from active gain elements (such as transistors). This is the recommended design approach. However, phase shift in the negative feedback circuit or poor bandwidth in the gain elements can result in poor high frequency noise behavior. Figure 12.10 shows a kind of result one might see as a gradual rise in $S_\phi(f)$ because of insufficient negative feed back at high Fourier frequencies.

Section X discussed aliasing in the frequency domain. Figure 12.11 shows the resultant measurement anomaly due to digital sampling of a poorly bandlimited (anti-aliased) white noise source. Noise voltage above the sampling frequency f_s is folded into the analysis region of interest. Note also that the stopband ripple characteristics are folded into the high-frequency portion of the passband. For a given sampling frequency, a compromise exists between increasing the high-

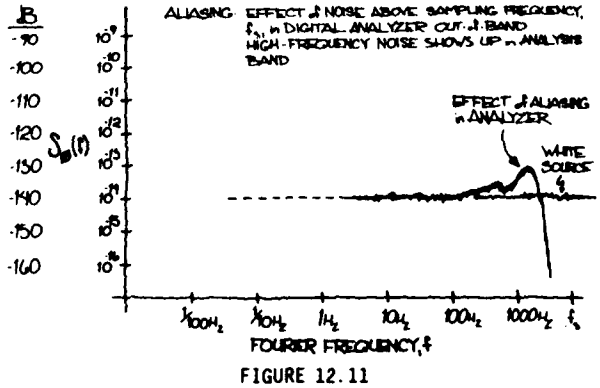


FIGURE 12.11

frequency extent of the analysis band and improving the anti-aliasing filter's stopband rejection. Section X has an example of the filter requirements for a particular case.

XIII. CONCLUSION

This writing highlights major aspects of time-domain and frequency-domain oscillator signal measurements. The contents are patterned after lectures presented by the authors. The authors have tried to be general in the treatment of topics, and bibliography is attached for readers who would like details about specific items.

References

1. J. A. Barnes, Andrew R. Chi, Leonard S. Cutler, Daniel J. Healey, David B. Leeson, Thomas E. McGunigal, James A. Mullen, Jr., Warren L. Smith, Richard L. Sydnor, Robert F. C. Vessot, and Gernot M. R. Winkler "Characterization of Frequency Stability," NBS Tech. Note 394 (1970); Proc. IEEE Trans. on Instrum. and Meas. IM-20, 105 (1971).

2. D. W. Allan, "The Measurement of Frequency and Frequency Stability of Precision Oscillators," NBS Tech. Note 669 (1975); Proc. 6th PTTI Planning Meeting, 109 (1975).

3. D. W. Allan, "Statistics of Atomic Frequency Standards," Proc. IEEE 54, 221 (1966).

4. Sections V and VI are from notes to be published by J. A. Barnes.

5 S. R. Stein, "An NBS Phase Noise Measurement System Built for Frequency Domain Measurements Associated with the Global Positioning System, NBSIR 76-846 (1976). See also F. L. Walls and S. R. Stein, "Servo Technique in Oscillators and Measurement Systems," NBS Tech. Note 692 (1976).

6. C. D. McGillen and G. R. Cooper, Continous and Discrete Signal and System Analysis, Holt, Rinehart, and Winston, Inc. (1974).

7. E. O. Brigham, The Fast Fourier Transform, Prentice-Hall, Inc., New Jersey (1974).

8. Private communication with Norris Nahman, June 1981.

9. G. M. Jenkins and D. G. Watts, Spectral Analysis and Its Application, Holden-Day, San Francisco, 1968.

10. P. F. Panter, Modulation, Noise, and Spectral Analysis, McGraw-Hill, New York, 1965, pp. 137-141.

11. D. Babitch and J. Oliverio, "Phase Noise of Various Oscillators at Very Low Frequencies, Proc. 28th Annual Symp. on Frequency Control, May 1974. U. S. Army Electronics Command, Ft. Monmouth, NJ.

12. N. Thrane, "The Discrete Fourier Transform and FFT Analyzers," Bruel and Kjaer Technical Review, No. 1 (1979).

13. Much of the material presented here is contained in NBS Tech. Note 679, "Frequency Domain Stability Measurements: A Tutorial Introduction," D. A. Howe (1976).

Bibliography

General References

1. November or December of even-numbered years, IEEE Transactions on Instrumentation and Measurement (Conference on Precision Electromagnetic Measurements, held every two years).

2. Proceedings of the IEEE. Special issue on frequency stability, vol. 54, February 1966.

3. Proceedings of the IEEE. Special issue on time and frequency, vol. 60, no. 5, May 1972.

4. Proceedings of the IEEE-NASA Symposium on the Definition and Measurement of Short-Term Frequency Stability at Goddard Space Flight Center, Greenbelt, MD, November 23-24, 1964. Prepared by Goddard Space Flight Center (Scientific and Technical Information Division, National Aeronautics and Space Administration, Washington, DC, 1965). Copies available for $1.75 from the US Government Printing Office, Washington, DC 20402.

5. Proceedings of the Annual Symposium on Frequency Control. Prepared by the US Army Electronics Command, Fort Monmouth, NJ. Copies from Electronics Industry Association, 2001 Eye Street, N. W., Washington, DC 20006, starting with 32nd Annual or call (201) 544-2773 Price per copy, $20.00.

6. J. A. Barnes, A. R. Chi, L. S. Cutler, et al., "Characterization of Frequency Stability," IEEE Trans. on Intsrum. and Meas. IM-20, no. 2, pp. 105-120 (May 1971). Prepared by the Subcommittee on Frequency Stability of the Institute of Electrical and Electronics Engineers.

7. Time and Frequency: Theory and Fundamentals, Byron E. Blair, Editor, NBS Monograph 140, May 1974.

8. P. Kartaschoff, Frequency and Time, London, England: Academic Press, 1978.

9. G. M. R. Winkler, Timekeeping and Its Application, Advances in Electonics and Electron Physics, Vol. 44, New York: Academic Press, 1977.

10. F. M. Gardner, Phaselock Techniques, New York: Wiley & Sons, 1966.

11. E. A. Gerber, "Precision Frequency Control and Selection, A Bibliography," Proc. 33rd Annual Symposium on Frequency Control, pp. 569-728, 1979. Available from Electronics Industries Association, 2001 Eye St. NW, Washington, DC 20006.

12. J. Rutman, "Characterization of Phase and Frequency Instabilities in Precision Frequency Sources: Fifteen Years of Progress," Proc. IEEE, vol. 66, no. 9, September 1978, pp. 1048-1075.

Additional Specific References (selected)

1. D. W. Allan, "Statistics of Atomic Frequency Standards," Proc. IEEE, vol. 54, pp. 221-230, February 1966.

2. D. W. Allan, "The Measurement of Frequency and Frequency Stability of Precision Oscillators," NBS Tech. Note 669, July 1975.

3. J. R. Ashley, C. B. Searles, and F. M. Palka, "The Measurement of Oscillator Noise at Microwave Frequencies," IEEE Trans. on Microwave Theory and Techniques MTT-16, pp. 753-760, September 1968.

4. E. J. Baghdady, R. D. Lincoln, and B. D. Nelin, "Short-Term Frequency Stability: Theory, Measurement, and Status," Proc. IEEE, vol. 53, pp. 704-722, 2110-2111, 1965.

5. J. A. Barnes, "Models for the Interpretation of Frequency Stability Measurements," NBS Technical Note 683, August 1976.

6. J. A. Barnes, " A Review of Methods of Analyzing Frequency Stability," Proc. 9th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Nov. 1977, pp. 61-84.

7. J. A. Barnes, "Atomic Timekeeping and the Statistics of Precision Signal Generators," Proc. IEEE, vol. 54, pp. 207-220, February 1966.

8. 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 Tech. Note 375, January 1969.

9. J. A. Barnes and D. W. Allan, "A Statistical Model of Flicker Noise," Proc. IEEE, vol. 43, pp. 176-178, February 1966.

10. J. A. Barnes and R. C. Mockler, "The Power Spectrum and its Importance in Precise Frequency Measurements," IRE Trans. on Instrumentation I-9, pp. 149-155, September 1960.

11. W. R. Bennett, "Methods of Solving Noise Problems," Proc. IRE, May 1956, pp. 609-637.

12. W. R. Bennett, Electrical Noise, New York: McGraw-Hill, 1960.

13. C. Bingham, M. D. Godfrey, and J. W. Tukey, "Modern Techniques of Power Spectrum Estimation," IEEE Trans. AU, vol. 15, pp. 56-66, June 1967.

14. R. B. Blackman and J. W. Tukey, The Measurement of Power Spectra, New York: Dover, 1958.

15. E. O. Brigham and R. E. Morrow, "The Fast Fourier Transform," IEEE Spectrum, vol. 4, no. 12, pp. 63-70, December 1967.

16. J. C. Burgess, "On Digital Spectrum Analysis of Periodic Signals," J. Acoust. Soc. Am., vol. 58, no. 3, September 1975, pp. 556-557.

17. R. A. Campbell, "Stability Measurement Techniques in the Frequency Domain," Proc. IEEE-NASA Symposium on the Definition and Measurement of Short-Term Frequency Stability, NASA SP-80, pp. 231-235, 1965.

18. L. S. Cutler and C. L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards," Proc. IEEE, vol. 54, pp. 136-154, February 1966.

19. W. A. Edson, "Noise in Oscillators," Proc. IRE, vol. 48, pp. 1454-1466, August 1960.

20. C. H. Grauling, Jr. and D. J. Healey, III, "Instrumentation for Measurement of the Short-Term Frequency Stability of Microwave Sources," Proc. IEEE, vol. 54, pp. 249-257, February 1966.

21. D. Halford, "A General Mechanical Model for $|f|^{\alpha}$ Spectral Density Random Noise with Special Reference to Flicker Noise $1/|f|$," Proc. of the IEEE, vol. 56, no. 3, March 1968, pp. 251-258.

22. H. Hellwig, "Atomic Frequency Standards: A Survey," Proc. IEEE, vol. 63, pp. 212-229, February 1975.

23. H. Hellwig, "A Review of Precision Oscillators," NBS Tech. Note 662, February 1975.

24. H. Hellwig, "Frequency Standards and Clocks: A Tutorial Introduction," NBS Tech. Note 616-R, March 1974.

25. Hewlett-Packard, "Precision Frequency Measurements," Application Note 116, July 1969.

26. G. M. Jenkins and D. G. Watts, Spectral Analysis and its Applications, San Francisco: Holden-Day, 1968.

27. S. L. Johnson, B. H. Smith, and D. A. Calder, "Noise Spectrum Characteristics of Low-Noise Microwave Tubes and Solid-State Devices," Proc. IEEE, vol. 54, pp. 258-265, February 1966.

28. P. Kartaschoff and J. Barnes, "Standard Time and Frequency Generation," Proc. IEEE, vol. 60, pp. 493-501, May 1972.

29. A. L. Lance, W. D. Seal, F. G. Mendoza, and N. W. Hudson, "Automating Phase Noise Measurements in the Frequency Domain," Proc. 31st Annual Symposium on Frequency Control, 1977, pp. 347-358.

30. D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," Proc. IEEE L., vol. 54, pp. 329-330, February 1966.

31. P. Lesage and C. Audoin, "A Time Domain Method for Measurement of the Spectral Density of Frequency Fluctuations at Low Fourier Frequencies," Proc. 29th Annual Symposium on Frequency Control, 1975, pp. 394-403.

32. P. Lesage and C. Audoin, "Estimation of the Two-Sample Variance with Limited Number of Data," Proc. 31st Annual Symposium on Frequency Control, 1977, pp. 311-318.

33. P. Lesage and C. Audoin, "Characterization of Frequency Stability: Uncertainty due to Finite Number of Measurements," IEEE Trans on Instrum. and Meas. IM-22, pp. 157-161, June 1973.

34. W. C. Lindsey and C. M. Chie, "Specification and Measurement of Oscillator Phase Noise Instability," Proc. 31st Annual Symposium on Frequency Control, 1977, pp. 302-310.

35. D. G. Meyer, "A Test Set for Measurement of Phase Noise on High Quality Signal Sources," Proc. IEEE Trans on Instrum. and Meas. IM-19, No. 4, pp. 215-227, November 1970.

36. C. D. Motchenbacher and F. C. Fitchen, Low-Noise Electronic Design, New York: John Wiley & Sons, 1973.

37. D. Percival, "Prediction Error Analysis of Atomic Frequency Standards," Proc. 31st Annual Symposium on Frequency Control, 1977, pp. 319-326.

38. D. Percival, "A Heuristic Model of Long-Term Clock Behavior," Proc. 30th Annual Symposium on Frequency Control, 1976, pp. 414-419.

39. J. Shoaf, "Specification and Measurement of Frequency Stability," NBS Internal Report 74-396, November 1974.

40. J. L. Stewart, "Frequency Modulation Noise in Oscillators," Proc. IRE, vol. 44, pp. 372-376, March 1956.

41. R. F. C. Vessot, L. Mueller, and J. Vanier, "The Specification of Oscillator Characteristics from Measurements in the Frequency Domain," Proc. IEEE, pp. 199-206, February 1966.

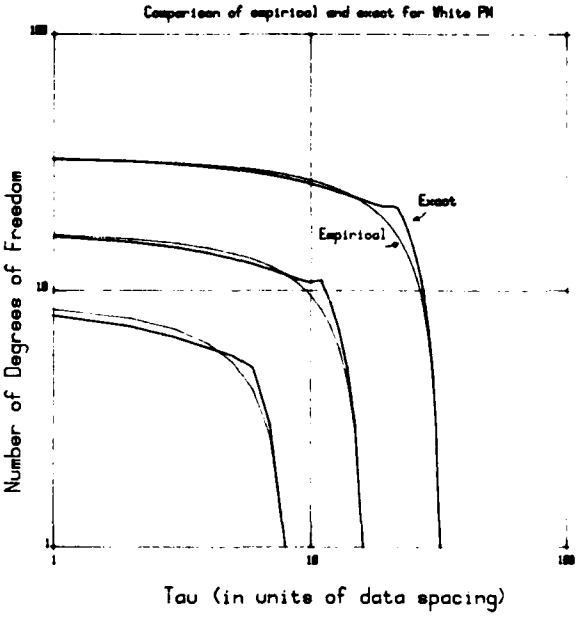
42. F. L. Walls, S. R. Stein, J. E. Gray, and D. J. Glaze, "Design Considerations in State-of-the-Art Signal Processing and Phase Noise Measurement Systems," Proc. 30th Annual Symposium on Frequency Control, 1976, pp. 269-274.

43. F. Walls and A. Wainwright, "Measurement of the Short-Term Stability of Quartz Crystal Resonators and the Implications for Quartz Oscillator Design and Applications," IEEE Trans. on Instrum. and Meas., pp. 15-20, March 1975.

44. G. M. R. Winkler, "A Brief Review of Frequency Stability Measures," Proc. 8th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Nov. 1976, pp. 489-527.

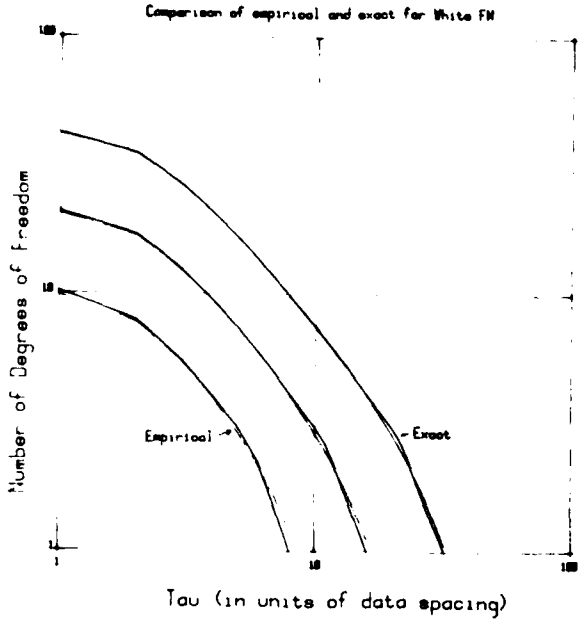
Appendix

DEGREES OF FREEDOM FOR ALLAN VARIANCE



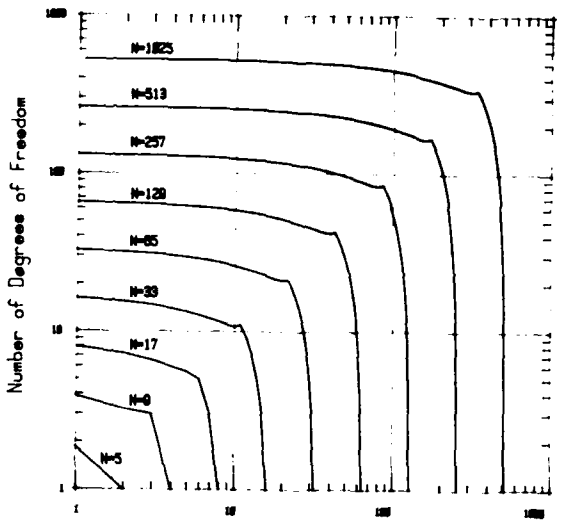
DEGREES OF FREEDOM FOR ALLAN VARIANCE

Comparison of empirical and exact for White FM



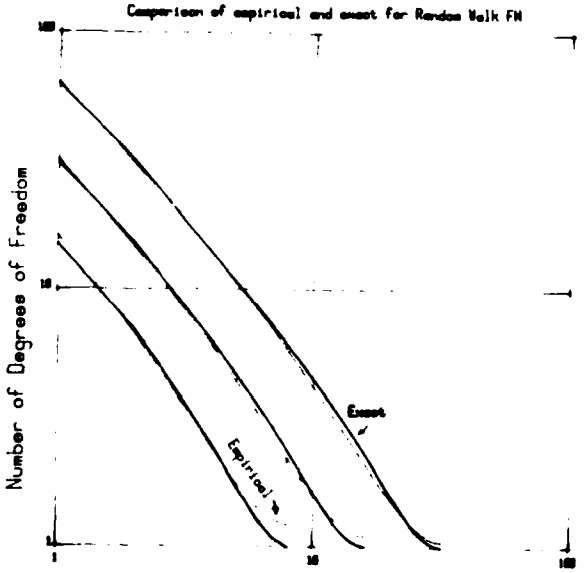
DEGREES OF FREEDOM FOR ALLAN VARIANCES

White Phase Noise



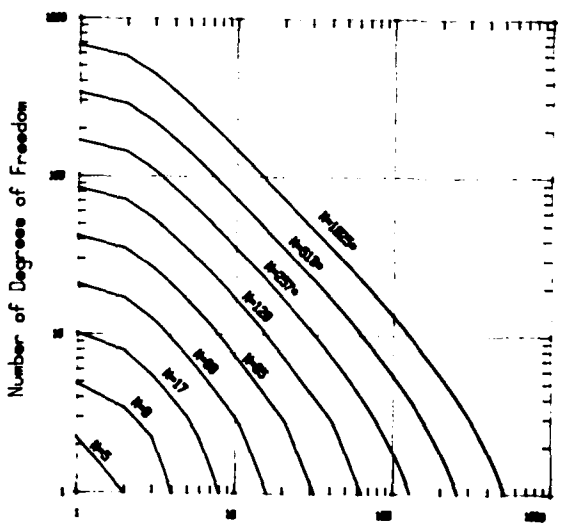
DEGREES OF FREEDOM FOR ALLAN VARIANCE

Comparison of empirical and exact for Random Walk FM



DEGREES OF FREEDOM FOR ALLAN VARIANCES

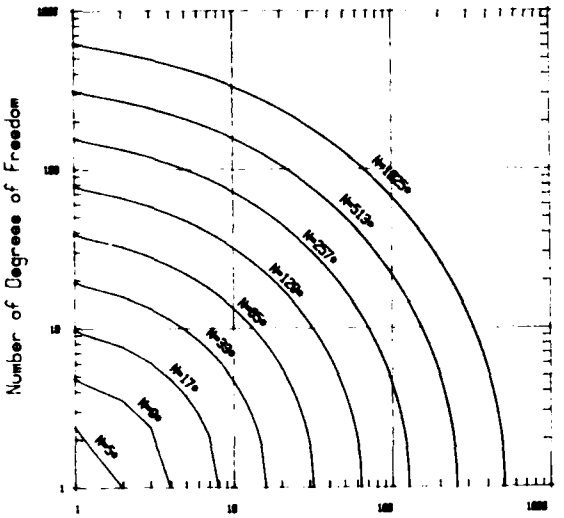
White Frequency Noise



(a) -Empirical Formula used

DEGREES OF FREEDOM FOR ALLAN VARIANCES

Flicker Phase Modulation

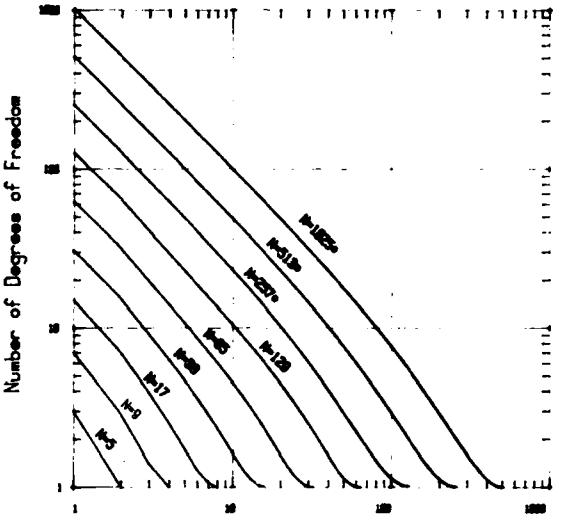


Tau (in units of data spacing)

(* -Empirical Formula used)

DEGREES OF FREEDOM FOR ALLAN VARIANCES

Random Walk FM

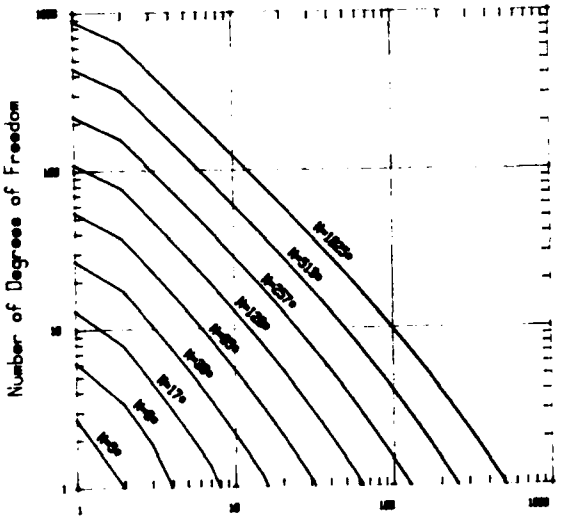


Tau (in units of data spacing)

(* -Empirical Formula used)

DEGREES OF FREEDOM FOR ALLAN VARIANCES

Flicker Frequency Modulation

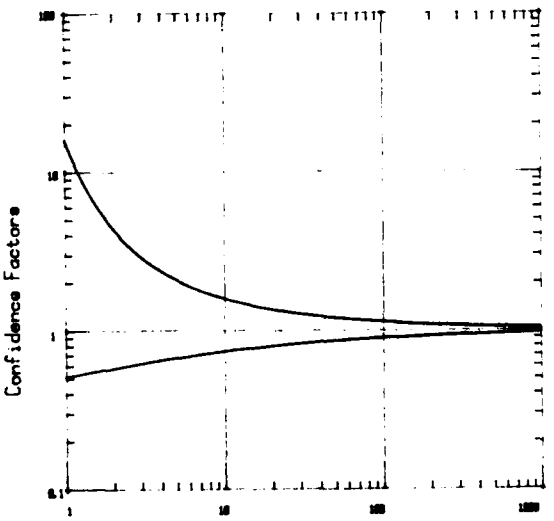


Tau (in units of data spacing)

(* -Empirical Formula used)

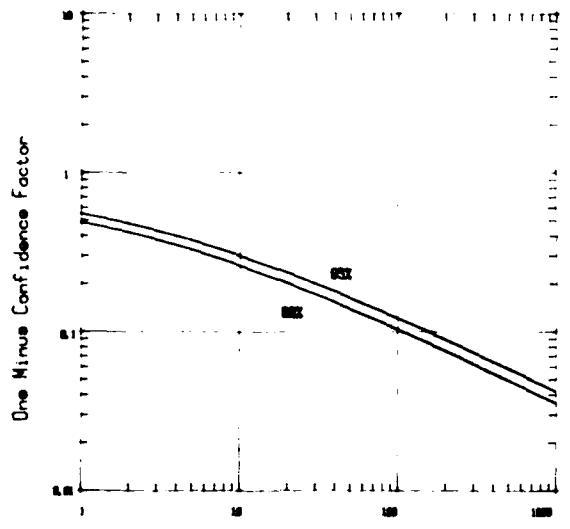
CONFIDENCE FACTORS FROM CHI-SQUARE

90% Intervals

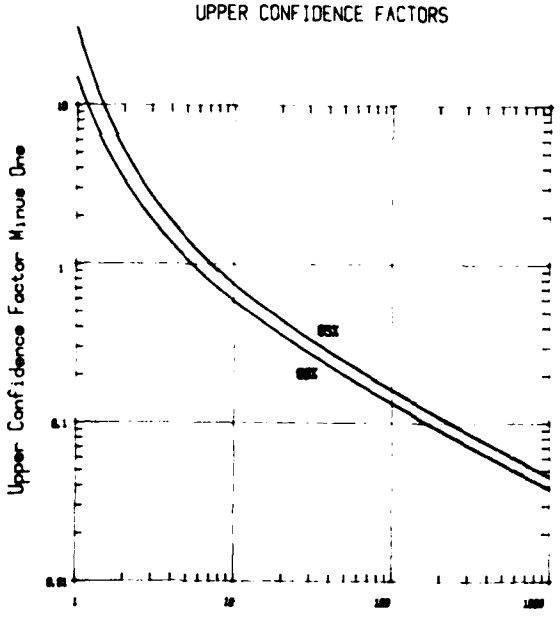


Number of Degrees of Freedom

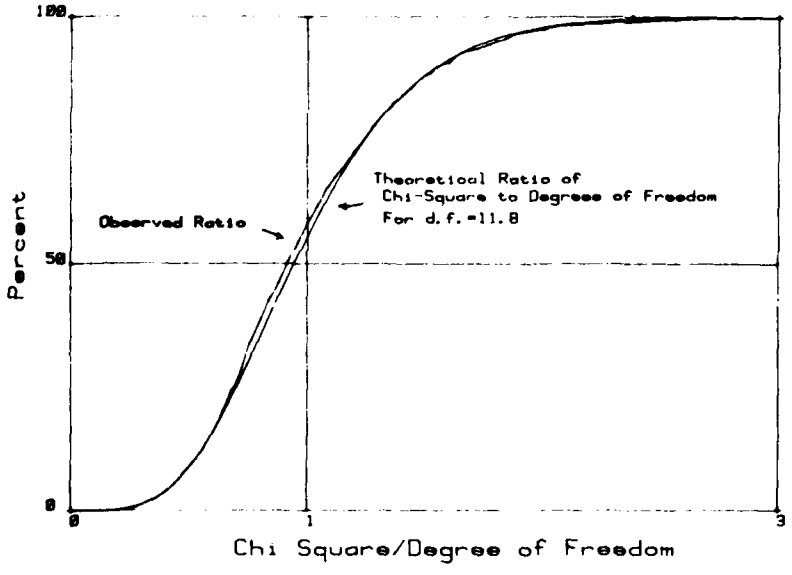
LOWER CONFIDENCE FACTORS



UPPER CONFIDENCE FACTORS



Distribution of Ratio of Observed Variance to Actual



AUTHOR INDEX

AUTHOR	PAGE
T. Adachi	149
D. Allan	470, 546
A. F. Armington	297
J. Asahara	304
J. H. Balbi	187
A. Ballato	71,376,576
J. Barnes	470
K. Barnhill	455
D. E. Beetley	263
J. C. Bergquist	602,625
R. J. Besson	122,340
L. Bidart	365
R. W. Birrell	60
B. R. Blitch	263
M. Bloch	415,553
R. B. Bossoli	317
V. E. Bottom	3
A. L. Bramble	406
J. C. Brice	312
R. N. Brown	329
M. D. Brunsman	436
M. Carel	349
R. L. Carter	358
B. H. T. Chai	401
J. L.Chambers	60
J. Chauvin	365
K. Chiba	646
F. Y. Cho	376
S. S. Chuang	130
A. Clements	546
D. D. Davis	546
H. Dehmelt	596
R. E. Dietterle	345
R. J. Dinger	144
D. J. Dodson	436
J. Dowsett	312
M. I. Dulmet	187
L. Dworsky	237
R. A. Eichinger	565
D. T. Elliott	174
T. C. English	637
F. K. Euler	329,492
M. Feldman	612,625
M. Feldmann	349
D. Fest	458
R. L. Filler	31,104,110,117
R. Fischer	595
E. D. Fletcher	157,312
J. M. Frank	40
J. J. Gagnepain	14,340,458,476
M. Gilden	395
G. Goldfrank	92
G. Goujon	476
J. Groslambert	458
T. W. Grudkowski	395
C. P. Guerin	122
D. Gunn	501
L. E. Halliburton	317
P. Handel	476
T. Hashi	646
D. J. Healey, III	440
J. Henaff	349
E. H. Hirt	537
J. Ho	92
W. H. Horton	271,592
J. P. Hou	222
J. A. Horrigan	297
M. Hozake	304
D. A. Howe	669
M. Y. Huang	436
W. M. Itano	602
D. Janiaud	340
E. Jechart	637
R. S. Kagiwada	401
A. Kahan	329
G. Kennedy	237
R. Kinsman	501
E. Knolmayer	56
D. R. Koehler	322
E. D. Kolb	291
B. Komiyama	335
H. Kotake	166
T. Kudama	166
S. Y. Kwan	440
K. M. Lakin	257
C. S. Lam	193
B. Y. Lao	345
J. J. Larkin	297
K. F. Lau	401
R. A. Laudise	291
P. C. Y. Lee	193,222
R. Lefevre	244
M. W. Levine	651
L. Lewis	612,625
H. G. Lipson	329
A. T. Lowe	48
T. Lukaszek	71,376
D. B. MacDonald	352
G. J. Malinowski	286
J. J. Martin	317
K. McDonald	657
D. N. McQuiddy, Jr.	516
R. F. Milsom	174
T. Musha	484
K. Nagai	304
M. Nakazawa	71,222
Y. Noguchi	484
C. Nyholm	286
J. J. O'Connor	297
S. Okano	166
M. Olivier	458
J. M. Owens	358
T. S. Payne	271
H. E. Peters	662
D. E. Phillips	428
B. R. Potter	352
M. A. Pugh	60
M. Redwood	174
A. M. Renard	455
V. R. Rosati	117
D. A. Rowe	345
J. J. Royer	250
N. J. Schneier	345
J. S. Schoenwald	345,383
B. K. Sinha	213
C. V. Smith, Jr.	358
R. C. Smythe	230,271,280
G. L. Snider	286
T. M. Snowden	263
J. J. Snyder	464
T. Stakelon	525
E. J. Staples	345,383
J. H. Staudte	583
D. S. Stevens	205
R. B. Stokes	401
D. A. Symonds	271
C. Takeuchi	149
T. Takeuchi	595,646
S. Taki	304
W. J. Tanski	388,415
D. M. Tennant	532
Y. Teramachi	484
H. F. Tiersten	205,230
A. Tirkel	511
Y. Tsuzuki	149
J. Uebersfeld	476
M. Valdois	340
J. P. Valentine	122
R. J. Valihura	60
D. C. L. Vangheluwe	157
J. R. Vig	104,110
A. Vulcan	415,553
F. L. Walls	602,625
R. W. Ward	99
A. Warner	92
W. Washington	104
M. Weiss	546
F. Welsh	525
C. E. Wheatley	428
J. White	657
D. F. Williams	376
R. C. Williamson	13
D. J. Wineland	602
J. Wise	345,383
S. T. Workman	60
M. D. Yakos	537
K. Yamazaki	166
N. Yannoni	492
K. H. Yen	401

SPECIFICATIONS AND STANDARDS GERMANE TO FREQUENCY CONTROL

Institute of Electrical and Electronic Engineers

Order through: IEEE Service Center
445 Hoes Lane
Piscataway, NJ 08854
(201/981-0060)

176-1978 Piezoelectricity $ 9.00

177-1966 Piezoelectric Vibrators, Definitions and Methods of Measurements for (ANSI C83.17-1970) $ 4.00

180-1962 Ferroelectric Crystal Terms, Definitions of $ 3.00

319-1971 Piezomagnetic Nomenclature $ 4.00

Electronic Industries Association

Order through: Electronic Industries Assn.
2001 Eye Street, N.W.
Washington, DC 20006
(202/457-4900)

(a) Holders and Sockets

RS-192-A, Holder Outlines and Pin Connections for Quartz Crystal Units. (Standard Dimensions for older types.) $ 6.80

RS-367, Dimensional and Electrical Characteristics Defining Receiver Type Sockets. (Including crystal sockets.) $20.20

RS-417, Crystal Outlines (Standard dimensions and pin connections for current quartz crystal units - 1974.) $ 7.80

(b) Production Tests

RS-186-E,(All Sections),Standard Test Methods for Electronic Component Parts $42.00

(c) Application Information

Components Bulletin No.6, Guide for the Use of Quartz Crystals for Frequency Control $ 4.90

(d) RS-477, Cultured Quartz (Apr. 81) $ 5.50

International Electrotechnical Commission

Order through: American National Standards Institute
1430 Broadway
New York, New York 10018

*ANSI can quote prices on specific IEC publications on a day to day basis only. All IEC and ISO standards have been removed from its Standards Catalog. Call ANSI, NYC (212/354-3300) for prices.

IEC Publication 122-1 (1976)

Quartz crystal units for frequency control and selection. Part 1: Standard values and test conditions. (Second edition)

IEC Publication 122-2 (1962) Section 3: Guide to the Use of Quartz Oscillator Crystals, including Amendment 1 (1969).

IEC Publication 122-3 (1977) Part 3: Standard Outlines and Pin Connections. (Second edition)

IEC Publication 283 (1968) Methods for the Measurement of Frequency and Equivalent Resistance of Unwanted Resonances of Filter Crystal Units

IEC Publication 302 (1969) Standard Definitions and Methods of Measurement for Piezoelectric Vibrators Operating Over the Frequency Range up to 30 MHz

IEC Publication 314 (1970) Temperature Control Devices for Quartz Crystal Units, including Supplement 314A Contents: General Characteristics & Standards; Test Conditions;Pin Connections

IEC Publication 314A (1971) First Supplement to Publication 314 (1970) Contents: Guide to the Use of Temperature Control Devices for Quartz Crystal Units

IEC Publication 368 (1971) Piezoelectric Filters including Amendment 1, Amendment 2, and Supplement 368A and 368B Contents: General Information & Standard Values; Test Conditions

IEC Publication 368A (1973) First Supplement to Publication 368 (1971) Contents: Guide to the Use of Piezoelectric Filters

IEC Publication 368B (1975) Second Supplement to Publication 368 (1971) Contents: Piezoelectric Ceramic Filters

IEC Publication 444 (1973) Basic Method for the Measurement of Resonance Frequency and Equivalent Series Resistance of Quartz Crystal Units by Zero Phase Technique in a π - Network

IEC Publication 483 (1976) Guide to Dynamic Measurements of Piezoelectric Ceramics with High Electromechanical Coupling

Department of Defense

Order through: Naval Publication & Form Center
5801 Tabor Avenue
Philadelphia, PA 19120

MIL-C-3098 Crystal Unit, Quartz, General Specification For

MIL-H-10056 Holders (Enclosures), Crystal, General Specifications For

MIL-STD-683 Crystal Units, Quartz/ And Holders, Crystal

MIL-F-28734 Frequency Standards, Cesium Beam, General Specifications For

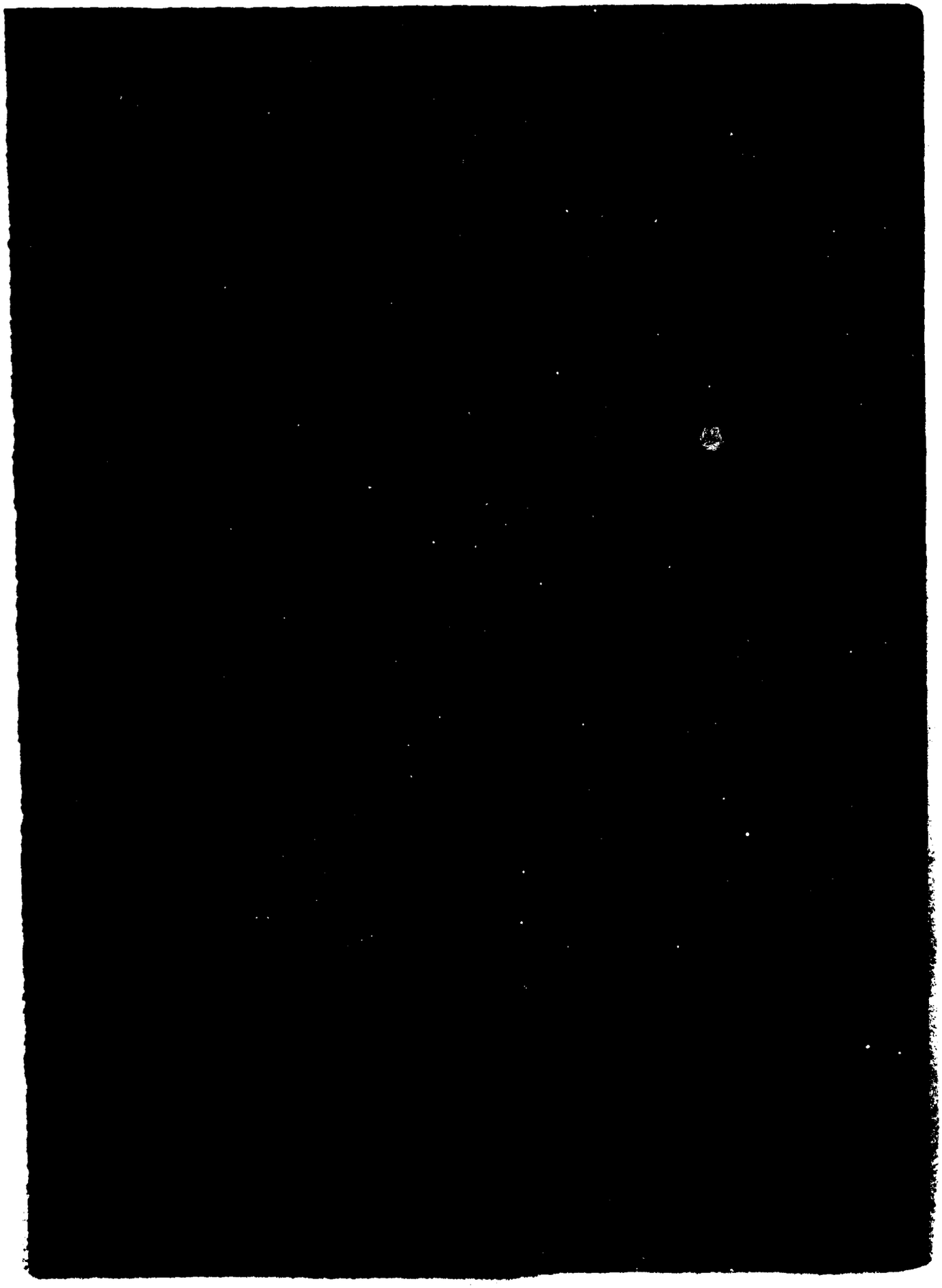
MIL-O-55310 Oscillators, Crystal, General Specification For

MIL-F-18327 Filters, High Pass, Low Pass, Band Pass Suppression and Dual Functioning, General Specification For

MIL-O-39021 Oven,Crystal,General Specification For

MIL-O-55240 Oscillators, Audio Frequency

MIL-F-28811 Frequency Standard, Cesium Beam Tube



END

DATE
FILMED

3-82

DTIC